Optimal Urban Transportation Policy: Evidence from Chicago[†]

Milena Almagro[*], Juan Camilo Castillo[†], Nathaniel Hickok[‡], Felipe Kup[§], and Tobias Salz[¶]

August 2, 2022

Preliminary and incomplete. Please do not cite or circulate.

Abstract

Despite high public transit subsidies, American commuters still overwhelmingly use socially inefficient cars. In this project, we quantify optimal urban transportation policies in the presence of congestion, network, and environmental externalities. Given dire municipal budgets for public transit, we account for budget considerations. We show theoretically that a budget constrained social planner introduces both quantity and quality distortions similar to those of a monopolist. We then move to an empirical analysis of the transportation system in Chicago based on a equilibrium model. To estimate supply and demand, we combine data from public and private sources that include the universe of public transit trips, ride hailing, taxi trips as well as car trips based on cell-phone location records. As a theoretical benchmark, we characterize optimal prices and service levels across all modes. Our main counterfactual determines optimal prices and service levels for existing public transit modes while allowing for profit maximizing responses of private market participants.

JEL classification: L00, L5, L13, L62, G41, G51.

Keywords: Transportation, Subsidy Design, Ramsey Pricing

[†]For helpful comments we thank seminar audiences at Booth, Harvard, MIT, Rice, Texas A&M, and Yale. We acknowledge financial support from the NSF (Award No. 1559013, Supplement), through the NBER's "Transportation Economics in the 21st Century" initiative. We thank Steve Berry, Jing Li, and Gabriel Kreindler for helpful comments. All remaining errors are our own.

[*] University of Chicago Booth, email: milena.almagro@chicagobooth.edu

[†] University of Pennsylvania Economics, email: jccast@upenn.edu

[‡] MIT Economics, email: nhickok@mit.edu

[§] University of Pennsylvania Economics, email: kupf@sas.upenn.edu

[¶] MIT Economics and NBER, email: tsalz@mit.edu

1 Introduction

Since the 1950s American urban transit has been generally characterized by the overwhelming use of personal cars. Just as recently as 2009, Americans undertook 320 billion trips using personal transportation and 7.5 billion using public transit. Only in very few urban areas more than five percent of people commute daily by public transit.[1] Moreover, public transit is generally associated with low utilization rates, as on average less than a third of bus seats are utilized, according to estimates from the department of transportation.[2] On top of the above, customers only pay a quarter of the trip cost directly through fares.[3] These facts suggest that American commuters largely prefer private alternatives over public transit, despite the latter being highly subsidized. However, compared to a public transit trip, a private trip emits twice the amount of CO_2 per passenger-mile.[4] One immediate question is whether subsidies should be even higher or do the low utilization rates indicate that higher subsidies are wasteful. More generally, are current public transit policies efficient?

The goal of this project is to quantify optimal urban transit policies. Given prevailing difficulties to build new transit infrastructure in America, which are well documented (Brooks and Liscow, 2021), we focus on interventions that are immediately implementable, such as taxes, subsidies, and adjustments in service levels. Determining the optimal level of these interventions requires accounting for several ingredients. First, one must understand how people substitute across different alternatives given mode prices and travel time. Moreover, public transit generally suffers from various externalities, such as road congestion as well as network and environmental externalities. A laissez-faire allocation is therefore socially inefficient, which calls for government intervention. In addition, any mitigation of these externalities has to be weighted against distributional concerns of public transit provision, such as to provide affordable transportation access to low-income households. In such case, the social planner may apply different welfare weights to different types of households.

In order to quantify these efficiency and distributional considerations, we first formulate a framework of a social planner who sets prices and service levels for road-based modes of transportation. Moreover, as transportation agencies usually have limited resources, we consider a budget-constrained planner. Concretely, she needs to satisfy a city budget given the operational costs and the collected revenue from the passenger's collected fares. Our framework reveals that absent any distortions, an unconstrained social planner would set price minus marginal cost equal to the marginal externality, which includes environmental externalities as well as the cost of congestion. This conclusion mimics the canonical result of optimal taxes in the pres-

[1] See https://www.bts.dot.gov.

[2] DOT: Public Transportation's Role in Responding to Climate Change

[3] newgeograghy.com

[4] A private car emits 0.96 pounds of CO2 per passenger-mile whereas public transit emits 0.45.

ence of social externalities. However, introducing budget considerations makes our solution depart from such answers for several reasons. First, as the planner collects revenues to cover costs and to meet the budget constraint, she has incentives to behave as a monopolist, which downward-distorts quantities. Second, given that our planner has the ability to choose service levels, she is directly affecting the quality that travellers experience in her transit system. These margin of adjustment distorts quality towards the marginal consumer, similar to a monopolist who controls both price and product quality (Spence, 1975).

Next, we apply this framework to the city of Chicago, the third most populous city in the US, and estimate a model of mode demand, mode supply, and the road congestion technology. Chicago is an ideal setting for our purposes for several reasons. First, it relies to a meaningful extent on both public and private transportation. This facilitates the identification of our demand model and means that transportation policy in this setting can be substantial. Second, Chicago is a city with large economic disparities, which means that it is important to account for the distributional effects of transportation policies. Lastly, Chicago is particularly good in terms of data availability.

We build a comprehensive dataset of all public and private trips from July 2019 to March 2020. We have access to the universe of public transit trips (excluding Metra, a suburban rail sysem), ride-share trips, and taxi trips. Unfortunately, there are no official records of trips done by private cars. To overcome this problem, we establish mobility patterns from individual cell-phone location records over this time horizon. Our data covers only forty percent of all active devices, but a battery of exercises reveal that this subsample is largely representative in terms of important sociodemographic factors such as income. Therefore, by the means of an appropriate re-scaling factor, we construct the total number of trips in the city using trips identified in our cell-phone data. Finally, given that we have the universe of public transit trips, we estimate the number of car trips as a residual, by subtracting public transit from the total number of trips constructed using our cell-phone data.

To estimate demand we follow Berry *et al.* (1995), while also imposing granular micro moments, as in Petrin (2002). This has the advantage that we can use standard inversion techniques to address potential endogeneity concerns, while micro moments allow us to capture how substitution depends on distance to the bus station, car ownership, and income. We define our main spatial unit of observation as a community area. Markets are defined at the hourly level from one community area to another. We impose that demand shocks for different modes are the same within a market. Finally, to construct our micro-moments, we zoom in and construct mode market share data at the census-tract level.[5]

[5]There is a total of 77 community areas and 900 census tracts in the city of Chicago.

Related Literature Our work relates to several strands of 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).

More recently, the literature has focused on short-run empirical estimates of such models. Many papers have focused congestion in models with a unique mode of transport. For example, Hall (2016) uses a bottleneck model to investigate the welfare effects of priced highway lanes, whereas Kreindler (2018) experimentally estimates the effects of peak congestion pricing. Within this strand of literature, Durrmeyer and Martínez (2022) is most closely related to ours. Like our study, they estimate an equilibrium model of mode substitution and investigate the welfare effects of urban transportation policies. They use it to investigate the welfare effects of traffic restrictions and road tolls. We, in contrast, use our model to construct a budget constrained social planner that chooses optimal prices and service levels to maximize welfare while keeping in mind distributional considerations. Moreover, the setting in Durrmeyer and Martínez (2022) is Paris, which differs substantially from Chicago in terms of the political economy of urban transportation systems. Therefore, we view these papers as complementary, given differences in the feasibility of the policies considered by policy-makers across the world.

There are several papers that investigate the long run effects of urban transportation policy, such as new rail infrastructure and residential location choices. Tsivanidis (2018) quantifies the equilibrium effects of new public transit infrastructure in Bogotá. Fajgelbaum and Schaal (2020) and Allen and Arkolakis (2020) characterize the optimal allocation of road infrastructure. However, these papers focus only on one mode of transportation and abstract away from incorporating rich heterogeneity in substitution patterns to preserve tractability and closed form solutions. We depart from their work by abstracting away from the residential choice model but allowing for rich demand substitution patterns across modes, which is a key margin of adjustment in the short-run. More recently, Barwick *et al.* (2021) explores the interaction between mode choices and residential location choices. We depart from their work by incorporating a planner's problem that takes location choices as given.

Our work is also related to equilibrium studies of transportation markets with search and matching frictions. Arnott (1996) shows that taxis should be subsidized because of increasing returns to scale. Lagos (2003) formulates a spatial search- and matching model and calibrates it to the New York City taxi market. Buchholz (2020) and Brancaccio *et al.* (2020a) estimate a structural dynamic model based on Lagos (2003) and apply it to the taxi and the dry bulk shipping industry respectively. Frechette *et al.* (2019) also estimate a dynamic model of the taxi market and micro-found a matching function based on the Manhattan street grid. Within this literature, our work is most closely related to Brancaccio *et al.* (2020b) who present a framework to compute the optimal prices in a dynamic search and matching market which

different types exhibits inefficiencies. Like Brancaccio *et al.* (2020b) we are able to characterize optimal policies but the setup of the models is quite different as ours applies to a model of mode choice whereas Brancaccio *et al.* (2020b) is formulates for a homogenous search and matching market. Unlike Brancaccio *et al.* (2020b) we allow for a matching function that exhibit non-constant returns to scale, and the social planner also sets qualities in addition to prices.

2 Background and Data

The section details necessary institutional context about the transportation landscape in Chicago as well as how we construct our data.

Background Our analysis focuses on Chicago for several reasons: it is a large city for which detailed data is available and it has a significant public transit infrastructure while also relying heavily on cars. Its transportation system, which is the second largest one in the United States, is operated by the Chicago Transit Authority (henceforth CTA). It includes a bus network of 152 routes with more than two thousand buses and a rapid transit system with eight routes and 144 stations, which is known as the "Chicago L". Prices are per ride and not dependent on distance. The full fare for bus and the L are $2.25 and $2.50 respectively.[6] The Chicago metropolitan area is also served by a commuter rail known as the Metra, which is operated by the Regional Transportation Authority (RTA). Metra prices are by zone and therefore depend on distance traveled unlike for buses and the L. Both the CTA and the RTA have a history of budget shortfalls, which suggests that it is important to account for budget considerations. 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.[7] Ride hail companies adjust prices dynamically according to demand and supply conditions.

Data Our demand analysis is based on aggregate moments and micro moments of mode market shares, prices, travel times, and demographics. We define a market as all trips from one community area to another within a given hour. Chicago is divided into 77 community areas, which are on average about the size of three square miles. To account for local demand conditions, we also construct additional trip characteristics across modes at the census tract area. We construct a dataset of mode prices, travel times, market shares, and demographics. To construct those, we rely on a variety of data sources.

[6] Reduced fares exist for students and seniors and a monthly pass for unlimited rides can be obtained at $75. See https://www.transitchicago.com/fares/ for additional details.

[7] See Chicago Taxi Fare Regulation.

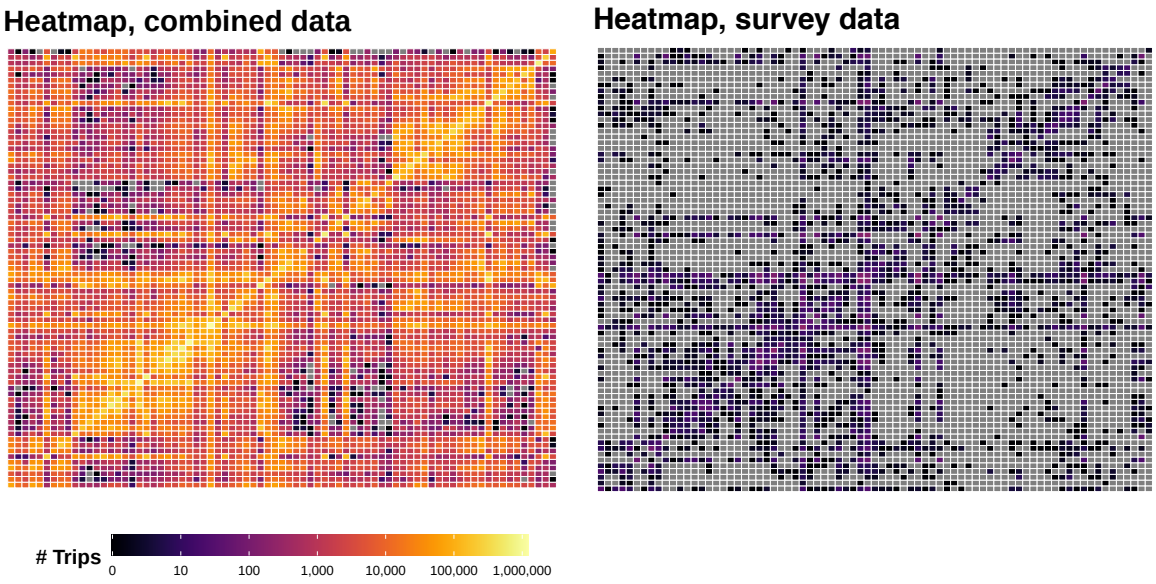


Figure 1: Combined vs. Survey Data: Flows Across Community Areas

Note: These figures show the number of trips from every origin community area to every destination community area in our combined data and in survey data. Gray points represent empty cells.

The first data source is administrative data from the CTA. The CTA publishes de-identified taxi and ride hail data at the trip level online. Those data include prices, locations, trip length and for pooled rides information on how many people were on the trip. We also have access to the universe of individual public transit trips for both buses and the L. For these trips we can observe the station or bus stop of origin, the time that the passenger tapped in, and an estimate of the drop-off L station or bus stop.

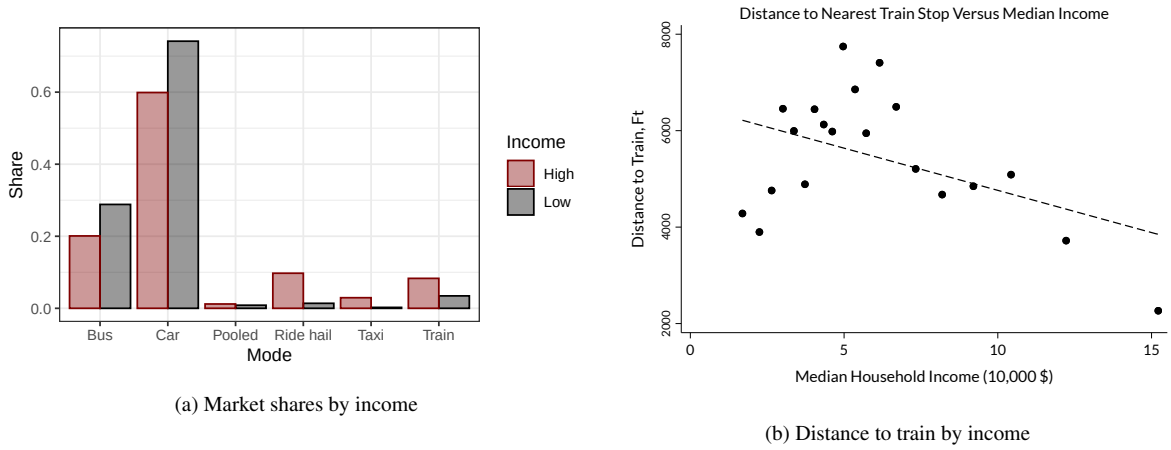
The second data source is based on cell-phone location records from Veraset, which records people's movements over time.[8] The dataset covers about forty percent of active cell-phone devices in the U.S. The data records the device ID and a sequence of GPS coordinates for each device. The frequency with which GPS coordinates are generated depends on which applications the user has installed. For our analysis we restrict ourselves 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 (see appendix A). We infer a device users' home location based on the modal GPS location at night time hours. We infer income and car ownership from census data. Since the cell-phone data only covers forty percent of devices and we further restrict ourselves to a subset of those, one may be concerned that that some demographic groups are over or underrepresented. We test for this based on census covariates that we can assign to devices with our inferred home location. The results from this test show that income groups are equally represented in the data. For additional details on this exercise see appendix B.

[8]https://www.veraset.com

The final output from this process is a dataset with a fraction of all trips that took place in Chicago. We scale up the number of trips by a factor such that the total number of trips is consistent with the numbers reported by the CTA ridership survey. One alternative would have been to use trips from the survey directly. Figure 1, however, shows that survey data is very sparse. Once it is disaggregated at the community area level, most of the origin-destination pairs have zero trips. This 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.

Our scaled-up cell phone data gives us the total number of trips for all modes. The CTA data gives us the number of trips for buses, subways, taxis, and ride-hailing apps. We construct the number of private car trips as the difference between the total number of trips minus the number of trips that we observe in the CTA data.

2.1 Summary Statistics and Descriptive Results



(a) Market shares by income

(b) Distance to train by income

Figure 2: Market shares and distance to nearest subway stop by income

Note: Subfigure (a) presents the market shares of each mode for low and high income travelers. Subfigure (b) plots the distance to the nearest subway station by income.

Figure 2a presents the market share of each mode for low and high income travelers. Cars and buses account for the largest shares. Low income people are more likely to use these two modes, which is explained in part because they do not use ride hailing or taxis to a large extend. Part of it is also likely driven by the fact that they tend to live far away from subway stations, as shown in figure 2b.

Figure 3 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

	mean	p10	p25	p50	p75	p90
7am to 9am	22%	8%	14%	22%	28%	36%
9am to 7am	14%	2%	6%	13%	20%	27%

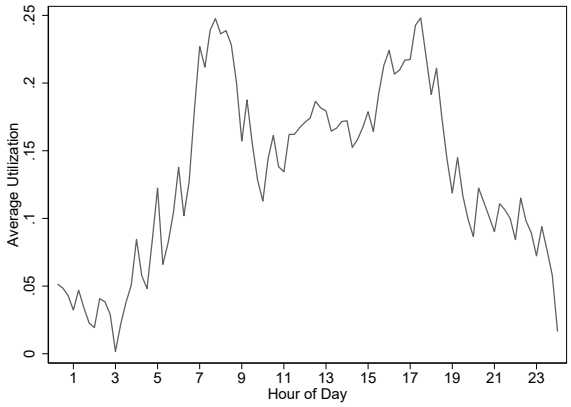


Figure 3: Bus utilization rates across different hours of the day.

Note: This figure shows the average utilization rate across bus routes for each hour of the day. To compute this average 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 utilization rate for a given bus route in a certain hour is then computed as the average utilization rate across each bus on the route over that hour.

within hour, we see that 90% of buses are at a utilization below 36% during the morning rush hour.

3 Theoretical framework

We now present a model that serves as a conceptual framework for our analysis. For simplicity, in this section we abstract away from two dimensions that our empirical model accounts for: the hour of the week and the origin and destination of a trip. Our theoretical results generalize once those are accounted for.

The main market participants are travelers, who decide which mode of transportation they want to use. There is also a central planner—the city government—who chooses prices and the capacity of each mode of transportation, which determines travel and waiting times.

Demand Travelers are characterized by a type $\theta \in \mathbb{R}^n$ with density $f(\cdot)$. A traveler of type θ gets utility $u_j(t_j, \theta) - p_j$ if she takes transportation mode $j \in \mathcal{J}$, 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). Each rider chooses the mode that maximizes her utility among $\mathcal{J} \cup \{0\}$, where 0 is the outside option of not taking a trip, which has zero utility (i.e., utility is measured relative to not taking a trip). Given vectors of prices $\mathbf{p}$ and total times $\mathbf{t}$ for all modes, demand for mode j is given by

$$q_j = q_j(\mathbf{p}, \mathbf{t}) = \int_{\Theta_j(\mathbf{p},\mathbf{t})} f(\theta)\, d\theta \tag{1}$$

where $\Theta_j(\mathbf{p}, \mathbf{t}) = \{\theta : j = \arg\max_{k \in \mathcal{J}}[u_k(t_k, \theta) - p_k], u_j(t_j, \theta) - p_j > 0\}$ is the set of travelers who choose mode j. We refer to the vector $\mathbf{q}$ as trips.

Gross utility and consumer surplus are given by

$$U(\mathbf{p}, \mathbf{t}) = \sum_j \int_{\Theta_j(\mathbf{p},\mathbf{t})} u(t_j, \theta) f(\theta)\, d\theta \qquad \text{and} \qquad CS(\mathbf{p}, \mathbf{t}) = \sum_j \int_{\Theta_j(\mathbf{p},\mathbf{t})} (u(t_j, \theta) - p_j) f(\theta)\, d\theta. \tag{2}$$

We assume that the demand system $q(\mathbf{p}, \mathbf{t})$ is invertible in both $\mathbf{p}$ and $\mathbf{t}$, a property that is satisfied by standard discrete choice models like our empirical demand model. When invertability is satisfied, we can directly work with quantities and travel times instead of prices, which simplifies the derivations. We thus define inverse functions $p(\mathbf{p}, \mathbf{t})$ and $t(\mathbf{q}, \mathbf{p})$, which tell us prices and travel times for a given number of travelers, holding the respective other endogenous variable fixed. Moreover, we can define $\Theta_j(\mathbf{q}, \mathbf{t}) = \Theta_j(p(\mathbf{q}, \mathbf{t}), \mathbf{t})$, which is the set of passengers choosing mode j when there are $\mathbf{q}$ travelers and travel times are $\mathbf{t}$. Lastly, we define the indirect gross utility function also in terms of quantities and travel times: $U(\mathbf{q}, \mathbf{t}) = U(p(\mathbf{q}, \mathbf{t}), \mathbf{t})$.

Transportation technology Total times are determined by a transportation technology T, which maps capacities and total number of travelers to travel times:

$$\mathbf{t} = T(\mathbf{q}, \mathbf{k}).^{9} \tag{3}$$

[9] We treat roads and railroads as exogenous, although they could be part of investment. This longer-run question is outside the scope of this work.

This function captures travel times, accounting for the effect of both traffic congestion and waiting times. The former arise from a bottleneck model, which restricts travel flows per unit of time and the latter from search and matching inefficiencies which lead to wait times before passengers can find a livery vehicle or board a bus or train. For taxis and ride hailing, the number of idle drivers decreases as more people use those modes, increasing the time riders need to wait before being picked up. Conversely, more drivers reduce waiting times. For public transit, the more buses and trains the higher the frequency with which they arrive, reducing waiting times. For private cars, streets get congested as more people drive their car, increasing travel times. This transportation technology accounts for spillovers across modes. Importantly, congestion affects all modes that use streets: as more people drive to their destination, for instance, streets get more congested, increasing the travel time of cars, taxis, buses, and ride hailing.

The government incurs a cost $C(\mathbf{q}, \mathbf{k})$, which 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 arguments.

Equilibrium Given a set of prices $\mathbf{p}$ and capacities $\mathbf{k}$, a market equilibrium is a set of trips $\mathbf{q}$ and times $\mathbf{t}$ such that (1) and (3) are satisfied.

There might exist multiple equilibria given $(\mathbf{p}, \mathbf{k})$.[10] However, 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. Thus, although $\mathbf{q}$ is an endogenous quantity, it will be more convenient to work with quantities instead of prices.

Social planner The city government chooses prices $\mathbf{p}$ and capacities $\mathbf{k}$. Its goal is to maximize welfare subject to a budget constraint.

Given the possible multiplicity of equilibria, a choice of $(\mathbf{p}, \mathbf{k})$ does not fully characterize the market outcome. We therefore think of the government as choosing an allocation $(\mathbf{q}, \mathbf{k})$ to maximize some objective function. It then sets $(\mathbf{p}, \mathbf{k}) = (p(\mathbf{q}, T(\mathbf{q}, \mathbf{k})), \mathbf{k})$.[11]

The government's revenue is $\Pi(\mathbf{q}, \mathbf{k}) = \sum_j p_j(\mathbf{q}, T(\mathbf{q}, \mathbf{k}))q_j - C(\mathbf{q}, \mathbf{k})$, and its budget is B. Welfare is equal to consumer surplus plus the government's revenue minus externalities, which can be written as $W(\mathbf{q}, \mathbf{k}) = U(\mathbf{q}, T(\mathbf{q}, \mathbf{k})) - C(\mathbf{q}, \mathbf{k}) - E(\mathbf{q}, \mathbf{k})$.[12] The social planner's objective

[10] The source of the problem is that travelers are playing a game, since their optimal choice depends on what they expect other agents' choices to be, and there might be multiple Nash equilibria in that game.

[11] The only concern is that the market could end up in a wrong equilibrium $(\mathbf{q}', \mathbf{t}') \neq (\mathbf{q}, \mathbf{t})$. However, the government can ensure the right equilibrium plays out in the short run: if agents expect an equilibrium with total times $\mathbf{t}'$, the government can set prices $p(\mathbf{q}, \mathbf{t}')$ 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).

[12] This expression can be derived by noting that $W = CS + \Pi - E = U - \sum_j p_j q_j + \sum_j p_j q_j - C - E$.

function is thus

$$\max_{\mathbf{q},\mathbf{k}} U(\mathbf{q}, T(\mathbf{q},\mathbf{k})) - C(\mathbf{q},\mathbf{k}) - E(\mathbf{q},\mathbf{k}) \qquad \text{s.t.} \qquad \sum_j p_j(\mathbf{q}, T(\mathbf{q},\mathbf{k}))q_j - C(\mathbf{q},\mathbf{k}) \geq -B. \quad (4)$$

In order to derive optimality conditions for this problem, we first define some useful notation. Let $\tilde{C}_j^q = \frac{\partial C}{\partial q_j} + \frac{\partial C}{\partial k_j}\frac{k_j}{q_j}$ and $\tilde{E}_j^q = \frac{\partial E}{\partial q_j} + \frac{\partial E}{\partial k_j}\frac{k_j}{q_j}$ represent the marginal cost and externality of a trip when holding the capacity per trip fixed—i.e., if the number of trips using mode j increase by 1%, so does capacity k_j. Similarly, let $\tilde{T}_{jk} = \frac{\partial T_j}{\partial q_k} + \frac{\partial T_j}{\partial k_k}\frac{k_k}{q_k}$ be the derivative of mode-j time with respect of the number of trips using mode k, holding the capacity per trip fixed. Also let $\bar{u}_j^T = \frac{\partial U}{\partial T_j}$ be the derivative of gross utility per traveler with respect to the total time of mode j. Finally, Ω_{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* (4) *can be written as*

$$p_j = \underbrace{\tilde{C}_j}_{\substack{\text{Marginal}\\ \text{cost}}} + \underbrace{\tilde{E}_j}_{\substack{\text{Marginal env.}\\ \text{externality}}} - \sum_k \underbrace{\bar{u}_k^T}_{\substack{\text{Mg. utility}\\ \text{of time}}} \underbrace{\tilde{T}_{kj}}_{\substack{\text{Network}\\ \text{effects}}} + \frac{\lambda}{1+\lambda}\Bigg\{ \underbrace{\sum_{k\in J} q_k \Omega_{kj}}_{\substack{\text{“Monopolist”}\\ \text{markup}}} - \tilde{E}_j - \underbrace{\sum_k (\tilde{u}_k^T - \bar{u}_k^T)\tilde{T}_{kj}}_{\substack{\text{Spence}\\ \text{distortion}}} \Bigg\} \quad (5)$$

$$\underbrace{\frac{\partial C}{\partial k_j}}_{\substack{\text{Mg. cost}\\ \text{of capacity}}} = - \underbrace{\frac{\partial E}{\partial k_j}}_{\substack{\text{Mg. env. externality}\\ \text{of capacity}}} + \underbrace{\sum_k \bar{u}_k^T \frac{\partial T_k}{\partial k_j}}_{\substack{\text{Mg. gross utility}\\ \text{from capacity}}} + \frac{\lambda}{1+\lambda}\Bigg\{ \frac{\partial E}{\partial k_j} + \underbrace{\sum_k (\tilde{u}_k^T - \bar{u}_k^T)\frac{\partial T_k}{\partial k_j}}_{\substack{\text{Spence}\\ \text{distortion}}} \Bigg\} \quad (6)$$

where λ 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 C.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, $p_j = \tilde{C}_j + \tilde{E}_j - \sum_k \bar{u}_k^T \tilde{T}_{kj}$, is equal to the marginal cost—accounting also for the marginal cost of additional capacity—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 capacity 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 (6) also takes a Pigouvian form with $\lambda = 0$. The marginal cost is the price of capacity 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 capacity is minus the marginal environmental externality plus the change in gross utility that a higher capacity induces due to times.

Once it accounts for the budget constraint, the social planner behaves somewhat like a monopolist. It not only cares about welfare but also about staying on budget, so it tries to increase its revenue to relax its constraint. This introduces three distortions that are captured by the final terms in both expressions. First, there is a "monopolist" markup in equation (5). Second, the government underweighs 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 $\lambda \to \infty$, 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 $\frac{\lambda}{1+\lambda} \to 1$.

4 Main model

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 c belongs to. Time is divided into hours $h \in \mathcal{H}$.

Recall that our model has several parts. First, we have travellers with fixed origin and destination who choose across modes of transport. Second, a supply side given by public transit service levels, taxi and ride-hailing drivers, Finally, a traffic congestion technology that aggregates the decisions of both sides of the market and maps into travel times.

4.1 Demand Model

Here we describe the empirical specification of our demand model that we use to estimate travelers' disutility for price and travel times, as well as unobserved determinants of mode convenience.

Our model has two levels of aggregation across traveller's choices. First, we define a market $M \in \mathcal{M}$ as the collection of people who make travel decisions from community area a to community area a' at a particular hour h. Second, within every market M, there are sub-markets $m = (c, c', h)$, the collection of people who make travel decisions from census tract $c \in a$ to census tract $c' \in a'$ during hour h. We denote by $M(m)$ the market m belongs to.

For each sub-market m, people $i \in \mathcal{I}_m$ arrive at exogenous arrival rates λ_m. They decide which mode $j \in \mathcal{J}_m \cup \{0\}$ to use, where $j = 0$ denotes staying put. They solve the following problem:

$$V_m^i = \max_{j \in \mathcal{J}_m \cup \{0\}} u_{mj}^i + \epsilon_j^i \quad = \max_{j \in \mathcal{J}_m \cup \{0\}} \quad \xi_{mj} - \alpha_T^i \cdot T_{mj} - \alpha_p^i \cdot p_{mj} + \epsilon_j^i, \tag{7}$$

where T_{mj} denotes the sum of the waiting and travel times for mode j, and p_{mj} is the price for mode j. Both are fixed within a sub-market. α_T^i and α_p^i are person i-specific time and price coefficients, respectively. Finally, ϵ_j^i is an idiosyncratic unobserved taste shock for a specific mode.

The joint distribution of α_T^i and α_p^i varies across sub-markets m:

$$(\alpha_p^i, \alpha_t^i) \sim F_m.$$

The random variation in coefficients across markets allows us to capture differences in mode availability, for instance due to differences in car ownership, or other economic factors such as income.

We assume that unobserved mode-specific demand shocks can be decomposed into a market fixed effect $\mu_{M(m)}$, a mode fixed effect μ_j, and an interaction term $v_{M(m),j}$:

$$\xi_{m,j} = \mu_{M(m)} + \mu_j + v_{M(m),j}.$$

Importantly, this term is constant within market. In other words, we assume that demand shocks are constant within community areas.

We assume that ϵ_j^i follows a Type I Extreme Value distribution and that the indirect utility of not taking a trip $j = 0$ is zero. Under this assumption, it follows that person i chooses mode j with probability given by:

$$\mathbb{P}_{mj}^i = \frac{\exp(u_{mj}^i)}{1 + \sum_{k \in \mathcal{J}_m} \exp(u_{mk})}.$$

Integrating over (α_p^i, α_t^i), we obtain that market shares and trips for mode j in sub-market m are given by:

$$\mathbb{P}_{mj} = \int \mathbb{P}_{mj}^i \, dF_m(\alpha_p^i, \alpha_t^i, T_{mj}^i) \qquad q_{mj} = \lambda_m \mathbb{P}_{mj}.$$

Finally, we obtain market shares and trips in market M by summing over sub-markets:

$$\mathbb{P}_{Mj} = \sum_{m \in M} f_m \mathbb{P}_{mj} \qquad q_{Mj} = \sum_{m \in M} q_{mj},$$

where $f_m = \frac{\lambda_m}{\lambda_{M(m)}}$ is the fraction of travelers in market $M(m)$ that arrive to sub-market m.

4.2 Public transit

We are still in the process of settin up a model of public transit. As a summary, there are two main factors that we will account for:

- *Travelers in different locations of the city have different walking times depending on the distance between their origin/destination and the nearest bus stop of subway station.*
- *The waiting time is inversely proportional to the frequency of routes.*

4.3 Taxis and ride hailing

Taxi and ride-hailing trips are provided by drivers that are hired by the city government at a wage w. The capacity k_{hj} of mode j during hour h is the size of the pool of drivers working for that mode. The government chooses the overall number of drivers in the city, but drivers themselves decide where they work.

We now describe a model that characterizes the waiting and travel times of riders as well as the costs and externalities, given the capacity k_{hj} and the number of trips $q_{aa'hj}$. It is composed of two main parts. First, a model of matching between travelers and drivers, which determines how long travelers and drivers need to wait before trips start. Second, a model of drivers' movements that determines the distribution of drivers across the city capturing two main forces: there are more drivers in areas with a net inflow of trips, and there are more drivers in places where earnings are high.

Matching At time h, let $\tilde{q}_{ahj}$ be the number of mode-j trips with origin a, and let L_{ahj} be the number of idle drivers in location a. We assume that there is a matching technology such that the expected waiting time for drivers before being matched T^D_{ahj}, the expected waiting time for riders before a trip starts T^W_{ahj}, and the expected pickup time T^P_{ahj} are given by

$$T^D_{ahj} = \frac{L_{ahj}}{q_{ahj}}, \qquad T^W_{ahj} = A^W_{aj} L^{-\phi_j}_{ahj}, \quad \text{and} \quad T^P_{ahj} = A^P_{aj} L^{-\chi_j}_{ahj}. \tag{8}$$

The expression for T^D_{ahj} is a simple application of Little's law.[13] In the other two expressions, A^W_{aj} and A^P_{aj} are scale factors that measure the overall matching inefficiency for mode j in location a. The parameters ϕ_j and χ_j are elasticities that determine how matching inefficiencies depend on the number of idle drivers. Note that this flexible specification nests simple models of matching in taxi and ride hailing markets.[14]

[13] The law states that the number of people waiting is equal to the matching rate times the average waiting time.

[14] In the taxi model in Lagos (2003), for instance, $\phi_j = 1$ and $A^P_{aj} = 0$. In the simplest ride-hailing model described by Castillo *et al.* (2018), $\phi_j = \psi_j = \frac{1}{n}$ in n-dimensional space.

Driver movements We model drivers' behavior as a continuous-time Markov chain. Drivers can be in two types of states: they can be idle in location a, or they can be driving a passenger from a to a'. We denote by $\mathbf{L}_{hj}$ the vector of the number of drivers in each state, whose components must add to k_{hj}.

We now describe how drivers transition between states. Every transition is characterized by a rate and by a payoff. Suppose drivers expect net earnings to be π_{ahj} for the next hour if they are idle in location a. Idle drivers transition from location a to location a' at a rate $\exp(\mu_{aa'} + \omega(\pi_{a'hj} - \pi_{ahj}))$. The intercept term $\mu_{aa'}$ measures, overall, how likely drivers are to move from a to a', which depends on how close both locations are and on how good the roads connecting them are. The coefficient ω measures to what extent drivers are more likely to go to high-earning locations.[15] This movement from a to a' entails a per-mile physical cost of driving.

Drivers that are idle in location a are matched to a rider that is going to location a' at a rate $\frac{q_{aa'hj}}{L_{ahj}}$. In that case, they obtain earnings that are equal to the trip fare minus the per-mile cost of picking up the driver and taking him to the destination. Once they are on a trip, they transition to being idle at a' with rate $\frac{1}{T^P_{ahj}+T^T_{aa'hj}}$, so that the average time it takes them to get to the destination is $T^P_{ahj} + T^T_{aa'hj}$, the sum of the pickup time and the travel time. Completing a trip has a payoff of zero.

Driver equilibrium These state transitions describe an irreducible matrix of transition rates $\mathbf{\Lambda} = \Lambda(\mathbf{q}_{hj}, \mathbf{L}_{hj}, \vec{\pi}_{hj})$, which depends on the number of trips and the number of drivers in each state (which affect matching rates and pickup times) and on the vector of expected earnings $\vec{\pi}_{hj}$ in all locations (which affect drivers' movements). Based on $\mathbf{\Lambda}$, it is straightforward to compute the steady state distribution of drivers $L(k_{hj}, \mathbf{\Lambda})$ as the unique vector $\mathbf{L}$ such that $\mathbf{\Lambda L} = 0$ and $1 \cdot \mathbf{L} = k_{hj}$. Furthermore, one can compute the expected net earnings $\vec{\pi}_{hj} = \Pi(\mathbf{\Lambda})$ by computing the evolution of the Markov chain (see appendix D).

A driver equilibrium for mode j during hour h is characterized by a driver distribution $\mathbf{L}_{hj}$, a transition matrix $\mathbf{\Lambda}_{hj}$, and a vector of payoffs $\vec{\pi}_{hj}$ such that

$$\mathbf{L}_{hj} = L(k_{hj}, \mathbf{\Lambda}_{hj}), \qquad \mathbf{\Lambda}_{hj} = \Lambda(\mathbf{q}_{hj}, \mathbf{L}_{hj}, \vec{\pi}_{hj}), \quad \text{and} \quad \vec{\pi}_{hj} = \Pi(\mathbf{\Lambda}_{hj}) \tag{9}$$

Appendix D shows that such an equilibrium exists and is unique. It also presents an algorithm to compute a driver equilibrium.

Having computed a driver equilibrium, total times $T(\mathbf{q}, \mathbf{k})$ can be computed as the sum of the travel time $T^T_{aa'hj}$ from the congestion model and the waiting time $T^W_{aa'hj}$. Costs $C(\mathbf{q}, \mathbf{k})$ are equal

[15] This specification results in state transition probabilities that take a multinomial logit form. Conditional on moving to an adjacent location, a driver moves from a to a' with probability $\frac{\exp(\mu_{aa'} + \omega(\pi_{a'hj} - \pi_{ahj}))}{\sum_{b \in N(a)} \exp(\mu_{ab} + \omega(\pi_{bhj} - \pi_{ahj}))}$, where $N(a)$ is the set of locations that are adjacent to a.

to the sum of per-mile costs and the wages of drivers, and externalities $E(\mathbf{q}, \mathbf{k})$ are computed as a sum of per-mile externalities.

4.4 Transportation technology

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

4.4.1 Traffic congestion

To account for traffic congestion, we model the city as a directed graph in which nodes are community areas $\mathcal{A}$, and edges (a, a') connect community areas that are spatially adjacent. Let $\mathcal{E}$ be the set of all edges. We assume that routes are exogenous. If a traveler uses mode j in market $M = (a, a', h)$, she follows a directed path $r_{Mj} = (e^1_{Mj}, e^2_{Mj}, \ldots)$ over edges that connects a with a'.

During hour h, the congestion on edge e is

$$Q_{eh} = \sum_j b_j \sum_{M \in \mathcal{M}^e_{hj}} q_{Mj},$$

where $\mathcal{M}^e_{hj}$ is the set of markets during hour h such that mode j goes through edge e. The weight b_j captures the fact that some modes congest more than others. Every marginal car trip congests streets, whereas a marginal bus or train traveler does not affect congestion.

We assume that the travel time for mode j over edge e at time h is given by

$$\tau_{jeh} = A_{je} Q^{\alpha_j}_{eh}$$

where A_{je} is an edge-specific term that measures the distance between the nodes e connects and the quality of infrastructure, and α_j is the elasticity of travel time with respect to congestion.

The travel time for a mode-j trip in market $M = (a, a', h)$ is given by

$$T^T_{Mj} = \sum_{e \in r_{Mj}} \tau_{jeh},$$

the sum of the travel times over all edges in the path r_{Mj}.

5 Estimation

5.1 Demand model

We estimate the demand model specified in section 4.1 by GMM using two sets of moments. The first set are the orthogonality between errors $v_{M,j}$ and market-specific instruments $z_{M,j}$:

$$E[vz] = 0.$$

We compute errors $v_{M,j}$ by finding values such that the predicted market share $\mathbb{P}_{Mj}$ and observed market shares s_{Mj} are equal, market by market.

The second set are micro moments within market M. We define $\sigma_{Mj} = \sum_{m \in M} \lambda_m (s_{mj} - \mathbb{P}_{mj}) z_m$, the within-market covariance between the deviation in shares $s_{mj} - \mathbb{P}_{mj}$ and a set of sub-market level instruments z_m. Our second set of moments says that this covariance is zero:

$$E[\sigma] = 0.$$

This moment exploits within-market variation. As one example, one would expect the market share of trains to be higher for trips departing from or arriving to the census tract where a train station is located. Trips to or from that census tract would have a lower time, so if we use the average time as an instrument z_m, this moment allows us to find time coefficients that rationalize these patterns. The identifying assumption is that these instruments are exogenous within community area.

Results [*We are currently working on the estimation of the demand side, and we expect to have results soon.*]

5.2 Other parts of the model

[*We are also working on the estimation of the rest of the model.*]

6 Counterfactuals

This is a list of the main counterfactuals results we expect to be able to present by November:

- *Computing the optimal prices and investments across all modes.*
- *Computing the optimal prices and investments across all modes, subject to a budget constraint.*
- *Computing the optimal prices and investment of public transit (subject to a budget constraint) accounting for the private response of taxi and ride-hailing markets.*

In all counterfactuals, we will present welfare results in terms of aggregate consumer surplus as well as hetereogeneity in travelers' utility by income and by neighborhood.

References

Allen, T. and Arkolakis, C. (2020). *The welfare effects of transportation infrastructure improvements*. Tech. rep., National Bureau of Economic Research.

Arnott, R. (1996). Taxi travel should be subsidized. *Journal of Urban Economics*, **40** (3), 316–333.

---, De Palma, A. and Lindsey, R. (1990). Economics of a bottleneck. *Journal of urban economics*, **27** (1), 111–130.

---, --- and --- (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, pp. 161–179.

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.

Berry, S., Levinsohn, J. and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, **63** (4), 841–890.

Brancaccio, G., Kalouptsidi, M. and Papageorgiou, T. (2020a). Geography, transportation, and endogenous trade costs. *Econometrica*, **88** (2), 657–691.

---, ---, --- and Rosaia, N. (2020b). *Search frictions and efficiency in decentralized transportation markets*. Tech. rep., National Bureau of Economic Research.

Brooks, L. and Liscow, Z. D. (2021). Infrastructure costs. *Available at SSRN 3428675*.

Buchholz, N. (2020). Spatial equilibrium, search frictions and efficient regulation in the taxi industry. *Working paper*.

Castillo, J. C., Knoepfle, D. and Weyl, G. (2018). Surge pricing solves the wild goose chase. *Working Paper*.

Durrmeyer, I. and Mart´nez, N. (2022). The welfare consequences of urban traffic regulations.

Fajgelbaum, P. D. and Schaal, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, **88** (4), 1411–1452.

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.

Hall, J. D. (2016). Pareto improvements from lexus lanes: The effects of pricing a portion of the lanes on congested highways, http://individual.utoronto.ca/jhall/documents/PI_from_LL.pdf.

KREINDLER, G. E. (2018). The welfare effect of road congestion pricing: Experimental evidence and equilibrium implications. *Working paper*.

LAGOS, R. (2003). An analysis of the market for taxicab rides in new york city. *International Economic Review*, **44** (2), 423–434.

PETRIN, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, **110** (4), 705–729.

SMALL, K. A. (1982). The scheduling of consumer activities: work trips. *The American Economic Review*, **72** (3), 467–479.

SPENCE, A. M. (1975). Monopoly, quality, and regulation. *The Bell Journal of Economics*, pp. 417–429.

TSIVANIDIS, N. (2018). The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá's transmilenio. *Unpublished manuscript*.

WEYL, E. G. (2010). A Price Theory of Multi-Sided Platforms. *American Economic Review*, **100** (4), 1642–1672.

Appendix

Appendix A Trip construction

Appendix B Representativeness of data

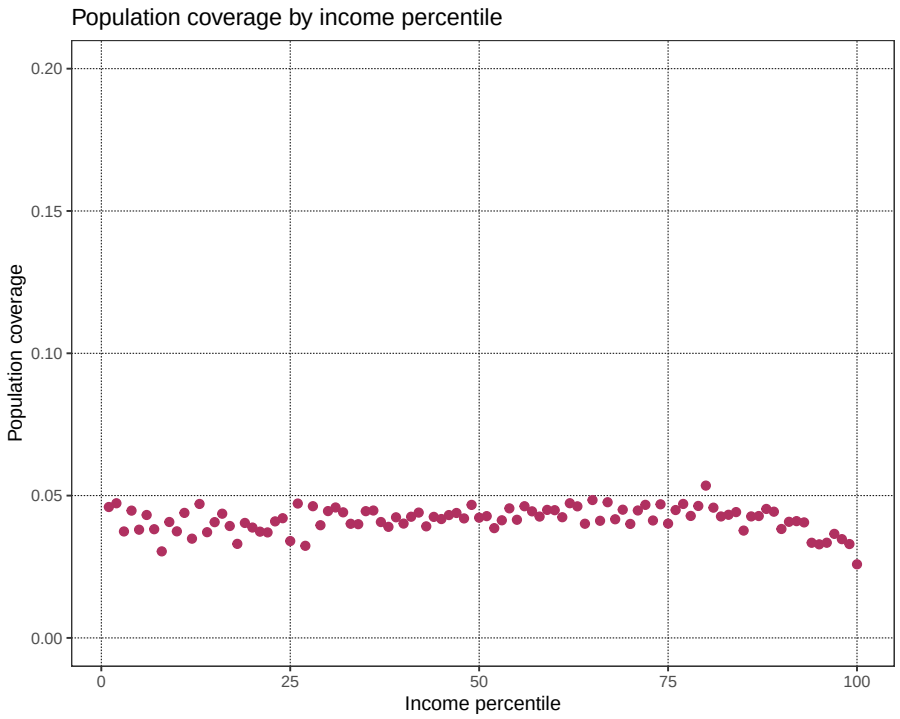


Figure 4: Representativeness across income groups

Note: This figure plots the fraction of the population in each income percentile that shows up in our data.

Appendix C Proofs

C.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 $\frac{\partial U}{\partial q_j} + \sum_k \frac{\partial U}{\partial t_k}\frac{\partial T_k}{\partial q_j} - \frac{\partial C}{\partial q_j} - \frac{\partial E}{\partial q_j} + \lambda\left(p_j + \sum_k q_k \frac{dp_k}{dq_j} - \frac{\partial C}{\partial q_j}\right) = 0$. 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 $\frac{\partial U}{\partial q_j} = p_j$. 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 $U(q,t) = \int_{\Theta_j(q,t)} u(t_j,\theta) f(\theta)\, d\theta$. 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 θ as q_j increases. This can also be written as $\sum_k \int_{\partial\Theta_k(q,t)} u_k(t_k,\theta) e_k(\theta) f(\theta)\, d\theta + \sum_{k,l>k} \int_{\partial\Theta_{kl}(q,t)} (u_k(t_k,\theta) - u_{lk}(t_l,\theta)) e_l(\theta) f(\theta)\, d\theta$. 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 $\frac{\partial q_k}{\partial q_j}$. It is thus equal to 1 for j and 0 for $k \neq j$. We can thus conclude that $\frac{\partial}{\partial q_j} U(q,t) = p_j$.

Substituting $\frac{\partial U}{\partial q_j} = p_j$, $\frac{dp_k}{dq_j} = \frac{\partial p_k}{\partial q_j} + \frac{\partial p_k}{\partial t_j}\frac{\partial T_k}{\partial q_j}$, and $\frac{dp_k}{dk_j} = \frac{\partial p_k}{\partial t_j}\frac{\partial T_k}{\partial k_j}$ on the first order conditions and a few steps of algebra yield

$$p_j = \frac{\partial C}{\partial q_j} - \frac{1}{1+\lambda} \sum_k \frac{\partial U}{\partial T_k}\frac{\partial T_k}{\partial q_j} + \frac{1}{1+\lambda}\frac{\partial E}{\partial q_j} - \frac{\lambda}{1+\lambda} \sum_k \left(q_k \frac{\partial p_k}{\partial q_j} + q_k \sum_{k'} \frac{\partial p_k}{\partial T_{k'}}\frac{\partial T_{k'}}{\partial q_j} \right), \tag{10}$$

$$\frac{\partial C}{\partial k_j} = \frac{1}{1+\lambda} \sum_k \frac{\partial U}{\partial T_k}\frac{\partial T_k}{\partial k_j} - \frac{1}{1+\lambda}\frac{\partial E}{\partial k_j} + \frac{\lambda}{1+\lambda} \sum_k q_k \sum_{k'} \frac{\partial p_k}{\partial T_{k'}}\frac{\partial T_{k'}}{\partial k_j}. \tag{11}$$

The final term in equation (10) 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 $\sum_{k'} q_{k'} \frac{\partial p_{k'}}{\partial T_k}$ 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 $\frac{\partial u_j}{\partial\theta} - \frac{\partial u_k}{\partial\theta}$ (and the inverse of $\frac{\partial u_j}{\partial\theta}$ for $v_j(\theta)$), and $\hat{n}_x(p,t,\theta)$, the unit normal component of the boundary $\partial\Theta_x(p,t)$ at θ. 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 $W_{jk}(p,t)$. 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 $\frac{\partial p_j}{\partial t_k}$ 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 $\frac{\partial u_j(t,\theta)}{\partial t}$ 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 $\tilde{u}_j^T \equiv \sum_k q_k \frac{\partial p_k}{\partial t_j}$ and the definitions of $\epsilon_{jk}^{T,q}$, $\epsilon_{jk}^{T,k}$, $\bar{u}_j^T$, Ω_{kj}, $\tilde{C}_j^q$, and $\tilde{E}_j^q$ into expres-

sions (10) and (11) yields expression (6) and

$$p_j = \underbrace{\frac{\partial C}{\partial q_j}}_{\substack{\text{Mg. cost}\\ \text{of a trip}}} + \underbrace{\frac{\partial E}{\partial q_j}}_{\substack{\text{Mg. env. ext.}\\ \text{of a trip}}} - \sum_k \underbrace{\bar{u}_k^T}_{\substack{\text{Mg. utility}\\ \text{of time}}} \underbrace{\frac{\partial T_k}{\partial q_j}}_{\substack{\text{Congestion}\\ \text{effect}}} + \frac{\lambda}{1+\lambda}\Bigg\{ \underbrace{\sum_{k\in J} q_k \Omega_{kj}}_{\substack{\text{“Monopolist”}\\ \text{markup}}} - \frac{\partial E}{\partial q_j} + \underbrace{\sum_k (\tilde{u}_k^T - \bar{u}_k^T)\frac{\partial T_k}{\partial q_j}}_{\substack{\text{Spence}\\ \text{distortion}}} \Bigg\}. \tag{12}$$

To obtain expression (5), multiply (6) by $\frac{k_j}{q_j}$ and subtract it from (12). □

Appendix D Driver equilibrium

Computing expected earnings Suppose that the state transition matrix is $\mathbf{\Lambda}$. Each state transition $s \to s'$ is associated with a certain payoff $y_{ss'}$, as described in the main text. Since payoffs only take place during state transitions, each state s is associated with a flow payoff that is equal to $y_s = \sum_{s'} \Lambda_{ss'} y_{ss'}$, a sum of the payoffs for each transition weighted by the rates at which those transitions occur.

Consider a driver that is in state s at time 0. His state at time t is a mixture over states that can be computed as $x(t) = P(t)e_s$, where e_s is the vector with a 1 in coordinate s and zero in every other coordinate and $P(t) = \exp(\Lambda t)$ is the evolution matrix of the Markov chain. His expected flow payoff at time t is $y \cdot x(t)$, so his expected payoffs between time zero until time Δ are equal to $\pi_s = \int_0^\Delta y \cdot x(t)\, dt = y \cdot \int_0^\Delta P(t)\, dt \cdot e_s$, and the vector of all such payoffs can be written as $\pi = \int_0^\Delta y \cdot x(t)\, dt = y \cdot \int_0^\Delta P(t)\, dt$.

Equilibrium computation Conditional on $\vec{\pi}_{hj}$, the joint solution of $\mathbf{L}_{hj} = L(k_{hj}, \mathbf{\Lambda}_{hj})$ and $\mathbf{\Lambda}_{hj} = \Lambda(\mathbf{q}_{hj}, \mathbf{L}_{hj}, \vec{\pi}_{hj})$ can be written as

$$\sum_{a'} L_{a'} m_{a'a} + \sum_{a'} L_{aa'} r_{aa'} = \sum_{a'} q_{aa'} + L_a \sum_{a'} m_{aa'} \quad \forall a \tag{13}$$

$$q_{aa'} = L_{aa'} r_{aa'} \quad \forall a, a' \tag{14}$$

wehere $m_{aa'} = \exp(\mu_{aa'} + \omega(\pi_{a'hj} - \pi_{ahj}))$ and $r_{aa'} = \frac{1}{T^P_{ahj} + T^T_{aa'hj}}$. From equation (14), we can solve for the values of $L_{aa'}$, and if we substitute (14) into (13), we can write it as a single matrix equation

$$\tilde{\Lambda}\tilde{L} = \tilde{q} \tag{15}$$

where $\tilde{L}$ is the vector of the number of idle drivers in every location, $\tilde{\Lambda}_{aa'} = m_{aa'} - r_{a'a}$ and $\tilde{q}_a = \sum_{a'} q_{a'a} - q_{aa'}$ are the net inflows to a due to trips. This system is linearly dependent, but replacing the last equation by $1 \cdot \tilde{L} = k_{hj} - \sum_{aa'} L_{aa'}$ yields a linear system with a unique solution. Call this solution $\check{L}(k_{hj}, \mathbf{q}_{hj}, \vec{\pi}_{hj})$, and let $\check{\Pi}(k_{hj}, \mathbf{q}_{hj}, \vec{\pi}_{hj}) = \Pi(\Lambda(\mathbf{q}_{hj}, \check{L}(k_{hj}, \mathbf{q}_{hj}, \vec{\pi}_{hj}), \vec{\pi}_{hj}))$.

Fixing k_{hj} and $\mathbf{q}_{hj}$, let $f(\pi) = \check{\Pi}(k_{hj}, \mathbf{q}_{hj}, \pi)$ (ommiting the vector notation). A driver equlibrium is a fixed point of f:

$$\pi = f(\pi). \tag{16}$$

Iterated application of f may diverge. However, a result from Castillo (2020) is useful. Let $\tilde{f}_\gamma(\pi) = (1-\gamma)\pi + \gamma f(\pi)$, where $\gamma \in \mathbb{R}$. This function has the same set of fixed points as f. And, as long as $(f(\pi) - f(\pi')) \cdot (\pi - \pi) < 0$ for all π, π', then there exists some sufficiently small $\gamma > 0$ such that $\tilde{f}_\gamma$ is a contraction mapping. This implies that there exists a unique fixed point of f. Furthermore, iteratively applying $\tilde{f}_\gamma$ converges to that fixed point.

Public Transit Potential

Milena Almagro, Juan Camilo Castillo, Nathaniel Hickok, Felipe Kup, and Tobias Salz

Motivation

Some Facts about Public Transit in the US

The current state of public transit:

- Customers only pay about 24% of the trip cost directly through fares.
  Source: newgeography.com
- The average bus utilization rate is very low.
  Source: DOT

Motivation

Some Facts about Public Transit in the US

The current state of public transit:

- Customers only pay about 24% of the trip cost directly through fares.
  Source: newgeography.com
- The average bus utilization rate is very low.
  Source: DOT
- Only 5% of commuting trips use public transit.
  Source: Census.gov

Motivation

Some Facts about Public Transit in the US

The current state of public transit:

- Customers only pay about 24% of the trip cost directly through fares.
  Source: newgeography.com
- The average bus utilization rate is very low.
  Source: DOT
- Only 5% of commuting trips use public transit.
  Source: Census.gov
- A private car emits about twice the amount of CO_2 per passenger mile as public transit.
  Source: DOT

Motivation

Some Facts about Public Transit in the US

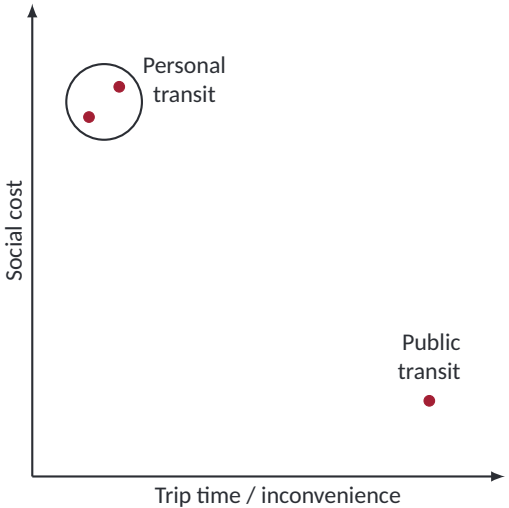
The current state of public transit:

- Customers only pay about 24% of the trip cost directly through fares.
  Source: newgeography.com
- The average bus utilization rate is very low.
  Source: DOT
- Only 5% of commuting trips use public transit.
  Source: Census.gov
- A private car emits about twice the amount of CO_2 per passenger mile as public transit.
  Source: DOT

→ Is current urban transit running at an **efficient** level?

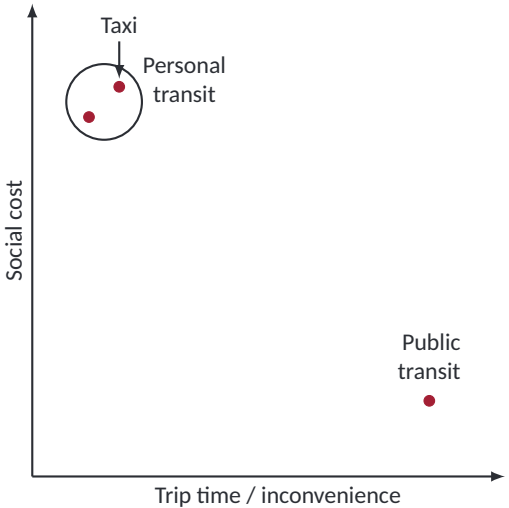
Urban Transportation

Toward a city planner's problem



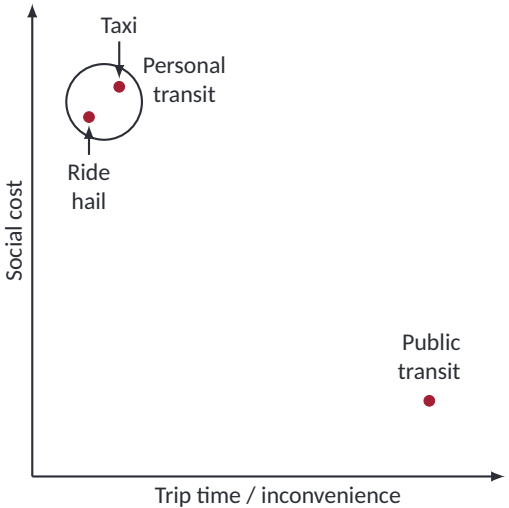
Urban Transportation

Toward a city planner's problem



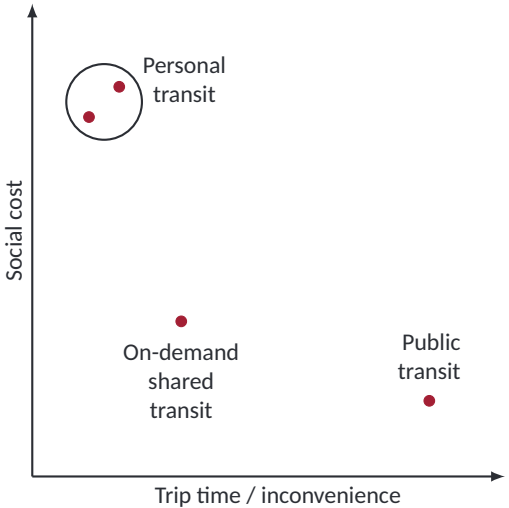
Urban Transportation

Toward a city planner's problem



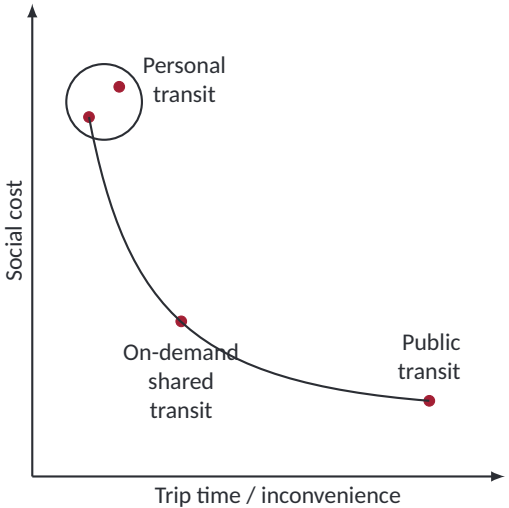
Urban Transportation

Toward a city planner's problem



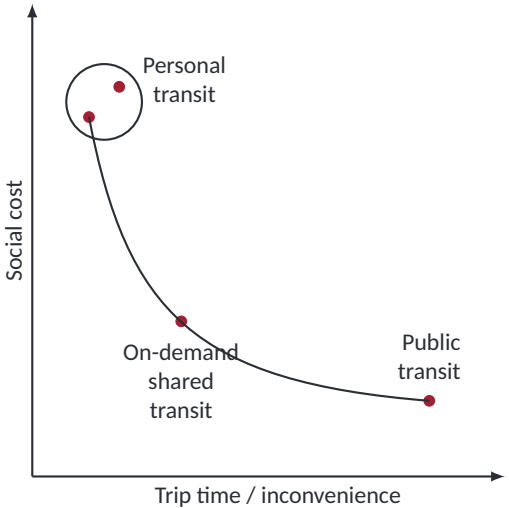
Urban Transportation

Toward a city planner's problem



Urban Transportation

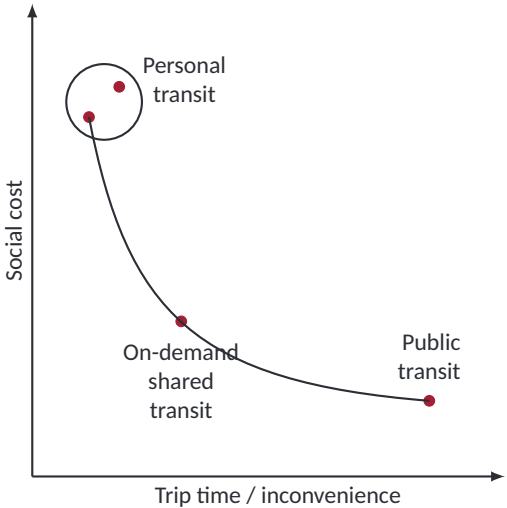
Toward a city planner's problem



City planner's problem:

Urban Transportation

Toward a city planner's problem

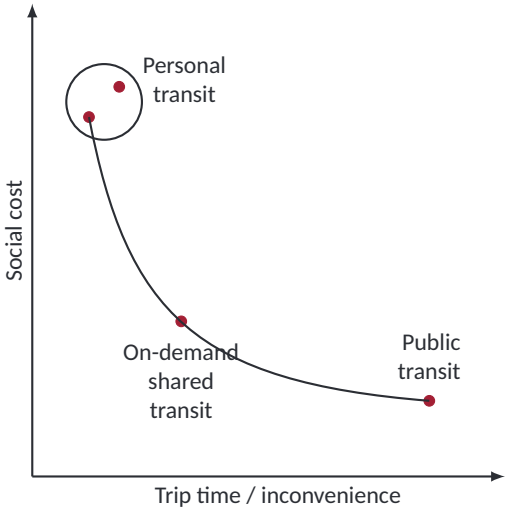


City planner's problem:

- Objective function:
  Consumer Surplus, Operating Costs, Externalities

Urban Transportation

Toward a city planner's problem

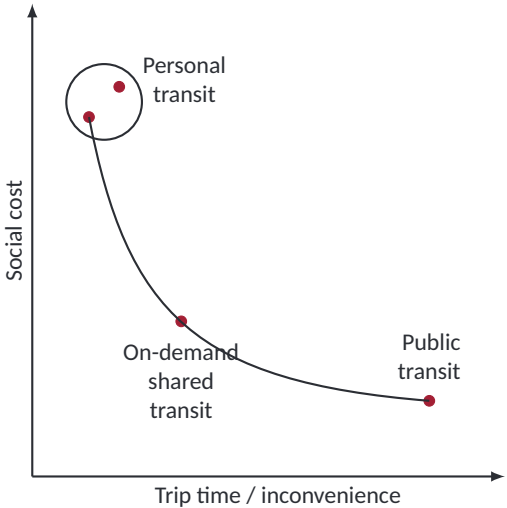


City planner's problem:

- Objective function:
  Consumer Surplus, Operating Costs, Externalities
- Set prices and capacities across modes:
  Map to de-centralized equilibrium quantities

Urban Transportation

Toward a city planner's problem

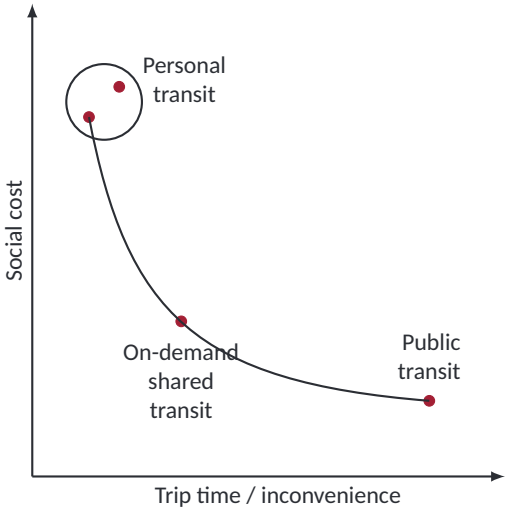


City planner's problem:

- Objective function:
  Consumer Surplus, Operating Costs, Externalities
- Set prices and capacities across modes:
  Map to de-centralized equilibrium quantities
- There is a budget constraint:
  Revenue - Costs $\leq$ City Budget

Urban Transportation

Toward a city planner's problem



City planner's problem:

- Objective function:
  Consumer Surplus, Operating Costs, Externalities
- Set prices and capacities across modes:
  Map to de-centralized equilibrium quantities
- There is a budget constraint:
  Revenue - Costs $\leq$ City Budget
- There is heterogeneity:
  In consumers and locations

Room for government intervention

Why regulate prices?

- Limited long-run infrastructure investments

Room for government intervention

Why regulate prices?

- Limited long-run infrastructure investments
- Externalities: Congestion, Environmental, Network

Room for government intervention

Why regulate prices?

- Limited long-run infrastructure investments
- Externalities: Congestion, Environmental, Network
- Budget constraint introduces monopoly-like distortions

Room for government intervention

Why regulate prices?

- Limited long-run infrastructure investments
- Externalities: Congestion, Environmental, Network
- Budget constraint introduces monopoly-like distortions
- Distributive role (not today)

This Project

Research Questions:

1. What are the optimal **prices** and **capacities** of the current urban transit system?

This Project

Research Questions:

1. What are the optimal **prices** and **capacities** of the current urban transit system?
2. What are the welfare and distributional gains from short-run investments such as:
   - improvements on **on-demand** technologies?
   - congestion pricing, congestion zones… (not today)?

This Project

Research Questions:

1. What are the optimal **prices** and **capacities** of the current urban transit system?
2. What are the welfare and distributional gains from short-run investments such as:
   - improvements on **on-demand** technologies?
   - congestion pricing, congestion zones... (not today)?

IO approach to urban transportation:

- Allow for rich substitution patterns that depend on observed and unobserved sources of variation.

Roadmap

1. Construct a dataset for all relevant modes for **Chicago**
2. Model of a City Planner
3. Discrete choice **mode demand** (McFadden 1974, Berry et al. 1995).
4. Transportation system: transit **technology** determines cost, convenience, and network externalities
5. Determine welfare effects of optimal prices and capacities
   - Accounting for environmental and congestion externalities + network effects

Counterfactual Results

Budget Constraint City Planner (Ramsey)

Preliminary!

Public transit shares increase by 8.4%:

Counterfactual Results

Budget Constraint City Planner (Ramsey)

Preliminary!

Public transit shares increase by 8.4%:

- Bus market share: 20% → 27%.

Counterfactual Results

Budget Constraint City Planner (Ramsey)

Preliminary!

Public transit shares increase by 8.4%:

- Bus market share: 20% → 27%.
- On-demand transit: 1.9% → 3.3%

Counterfactual Results

Budget Constraint City Planner (Ramsey)

Preliminary!

Public transit shares increase by 8.4%:

- Bus market share: 20% → 27%.
- On-demand transit: 1.9% → 3.3%
- Car market share: 70% → 65%.

Counterfactual Results

Budget Constraint City Planner (Ramsey)

Preliminary!

Public transit shares increase by 8.4%:

- Bus market share: 20% → 27%.
- On-demand transit: 1.9% → 3.3%
- Car market share: 70% → 65%.
- Taxi + ride hail: 7.7% → 4.1%.

Counterfactual Results

Budget Constraint City Planner (Ramsey)

Preliminary!

Public transit shares increase by 8.4%:

- Bus market share: 20% → 27%.
- On-demand transit: 1.9% → 3.3%
- Car market share: 70% → 65%.
- Taxi + ride hail: 7.7% → 4.1%.
- The cost of externalities is reduced by 1/3 or $2M ($CO_2$, congestion) per day.

Literature Review

1. **Transportation:** McFadden (1974), Small (1982), Small (2005), Parry and Small (2009)
2. **Spatial Equilibrium Models:**
   - Theory: Arnott (1996), Lagos (2003)
   - Empirics: Frechètte et al. (2019), Buchholz (2021), Brancaccio et al. (2020), Arora et al. (2020), Castillo (2020), Buchholz et al. (2020), Cairncross et al. (2021), Rosaia (2021), Leccese (2021), Brancaccio el at. (2021), Kreindler (2022), Durrmeyer and Martínez (2022)
3. **Transportation in the long run:** Tsivanidis (2018), Allen and Arkolakis (2020), Barwick et al. (2021)
4. **Geo-location Data and Mobility:** Miyahuchi et al. (2020), Glaeser et al. (2020), Couture et al. (2021)

Literature Review

1. **Transportation:** McFadden (1974), Small (1982), Small (2005), Parry and Small (2009)
2. **Spatial Equilibrium Models:**
   - Theory: Arnott (1996), Lagos (2003)
   - Empirics: Frechètte et al. (2019), Buchholz (2021), Brancaccio et al. (2020), Arora et al. (2020), Castillo (2020), Buchholz et al. (2020), Cairncross et al. (2021), Rosaia (2021), Leccese (2021), Brancaccio el at. (2021), Kreindler (2022), Durrmeyer and Martínez (2022)
3. **Transportation in the long run:** Tsivanidis (2018), Allen and Arkolakis (2020), Barwick et al. (2021)
4. **Geo-location Data and Mobility:** Miyahuchi et al. (2020), Glaeser et al. (2020), Couture et al. (2021)

$\Longrightarrow$ **A budget constrained city with heterogeneous agents setting optimal prices and capacities**

Roadmap

Roadmap

Data Sources

Chicago, June 2019-February 2020.

Data Sources

Chicago, June 2019-February 2020.

Raw data sets:

1. Individual cell phone location records: 40% of all devices.
2. Universe of public transit trips through MIT-CTA partnership.
3. Universe of taxi and ride hailing (pooled + single rides) trips from the city of Chicago.
4. Block level census data.
5. 2019 Chicago transit survey for validation and calibration.

– Car trips are identified as:

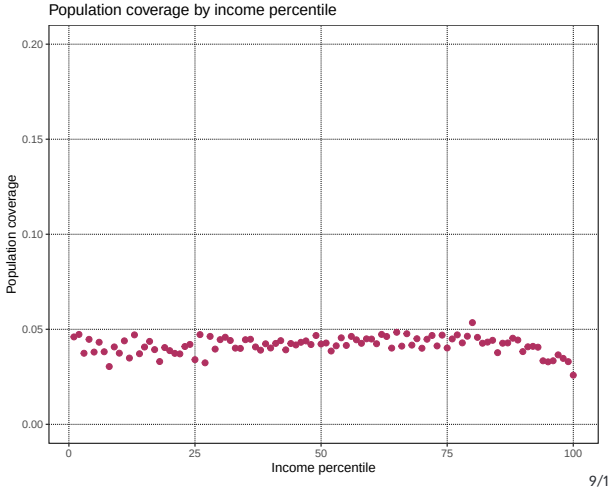
$$\text{Car Trips} = \text{Cell Phone Trips} - \text{Public Transit Trips} - \text{Ride-hailing Trips} - \text{Taxi Trips}$$

Goal: Hourly flows, prices, and travel times by mode across community areas + micro moments.

Cell Phone Data

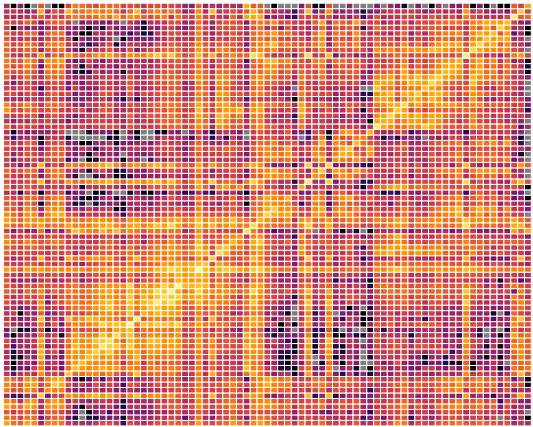
Representativeness Across Income Groups

1. Assign home census tract to cellphone id
2. Construct population coverage relative to census tract population in census
3. Order tracts by income percentiles

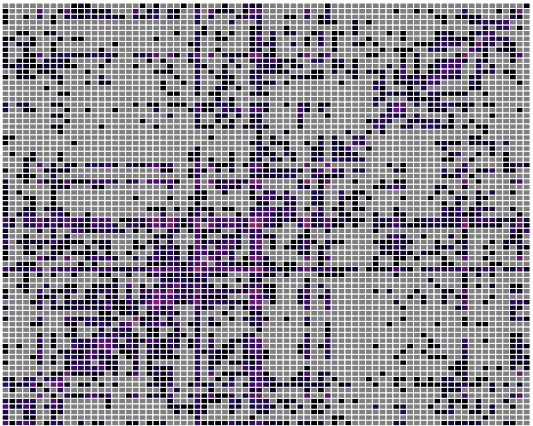


Combined vs. Survey Data: Flows Across Community Areas

Heatmap, combined data



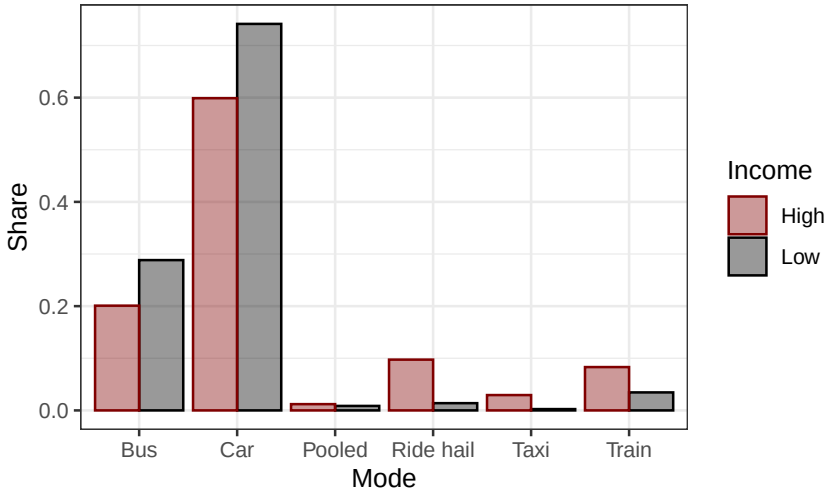
Heatmap, survey data



 # Trips

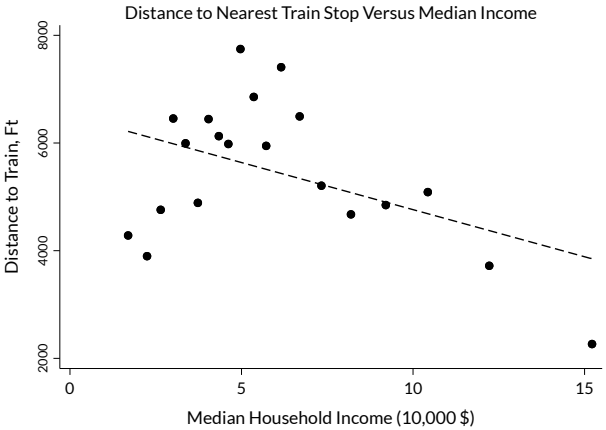
0 10 100 1,000 10,000 100,000 1,000,000

Mode Market Shares by Income



Heterogeneity across locations

Mode choice correlates with location's characteristics across space



	Car Ownership
Log Median HH Income	0.159***
	(0.006)
Log Distance to Train	0.051***
	(0.005)
Log Distance to Bus	0.101***
	(0.010)
Constant	-2.035***
	(0.086)
Observations	797
F	353.191

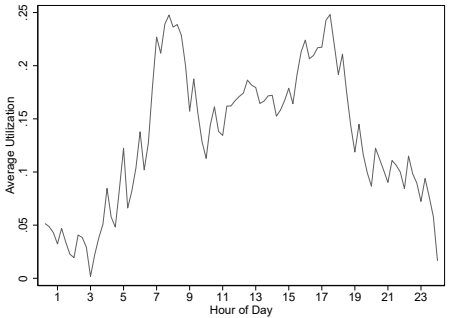
Summary Statistics

Utilization of Buses during Weekdays

The city runs 129 bus routes with 1,864 buses.

Utilization (across buses):

	mean	p10	p25	p50	p75	p90
7am to 9am	22%	8%	14%	22%	28%	36%
9am to 7am	14%	2%	6%	13%	20%	27%



Roadmap

Model
Outline

There are 3 parts:

1. Commuters make mode choices: (prices and times) $\mapsto$ (# of people choosing each mode)
2. Transportation technology: (# of people choosing each mode) $\mapsto$ (time and social cost)
3. City government: chooses prices and capacity of each mode, trades off welfare and budget (Ramsey)

Outside our model (for now):

- Supply side: government hires drivers at market wage

Next few slides: model with one O-D pair, one time period

- We estimate main model for different O-D pairs and times of the week

Model

Outline

Demand: $D_j(\mathbf{p}, \mathbf{t})$

- Prices p_j
- Travel time (wait + trip time) t_j

Model

Outline

Demand: $D_j(\mathbf{p}, \mathbf{t})$

- Prices p_j
- Travel time (wait + trip time) t_j

Transportation technology j is described by three functions of quantity and capacity $(\mathbf{q}, \boldsymbol{\kappa})$:

1. $\tau_j(q_j, \kappa_j)$: Total time (wait + travel)
2. $C(\mathbf{q}, \boldsymbol{\kappa})$: Cost (fuel, depreciation, labor)
3. $E(\mathbf{q}, \boldsymbol{\kappa})$: Externalities ($CO_2$, congestion),

Model

Outline

Demand: $D_j(\mathbf{p}, \mathbf{t})$

- Prices p_j
- Travel time (wait + trip time) t_j

Transportation technology j is described by three functions of quantity and capacity $(\mathbf{q}, \boldsymbol{\kappa})$:

1. $\tau_j(q_j, \kappa_j)$: Total time (wait + travel)
2. $C(\mathbf{q}, \boldsymbol{\kappa})$: Cost (fuel, depreciation, labor)
3. $E(\mathbf{q}, \boldsymbol{\kappa})$: Externalities ($CO_2$, congestion),

κ_j: Capacity of mode j, determines total times

- Taxis, ride-hailing, shared: Number of drivers working. More idle drivers → lower times.
- Buses, trains: Route frequency. More buses → lower times.

Model

City Government and Equilibrium

City government chooses prices and capacities $(\mathbf{p}, \boldsymbol{\kappa})$. Equilibrium $(\mathbf{q}, \mathbf{t})$ such that:

$$\underbrace{\mathbf{q} = \mathbf{D}(\mathbf{t}, \mathbf{p})}_{\text{demand}} \quad \text{and} \quad \underbrace{\mathbf{t} = \boldsymbol{\tau}(\mathbf{q}, \boldsymbol{\kappa})}_{\text{technology}}$$

Conditional on equilibrium, government's problem (Ramsey):

$$\max_{(\mathbf{q},\boldsymbol{\kappa})} \underbrace{\overbrace{U(\mathbf{t}, \mathbf{p}) - \mathbf{p} \cdot \mathbf{q}}^{\text{Consumer Surplus}} + \overbrace{\mathbf{p} \cdot \mathbf{q} - C(\mathbf{q}, \boldsymbol{\kappa})}^{\text{Gov. revenue \ Profits}} - E(\mathbf{q}, \boldsymbol{\kappa})}_{\text{Welfare}} \quad \text{s.t.} \quad \overbrace{\mathbf{p} \cdot \mathbf{q} - C(\mathbf{q}, \kappa) \geq B}^{\text{Budget constraint}}$$

$$\underbrace{\mathbf{q} = \mathbf{D}(\mathbf{t}, \mathbf{p}), \quad \mathbf{t} = \boldsymbol{\tau}(\mathbf{q}, \boldsymbol{\kappa})}_{\text{Equilibrium conditions}}$$

Optimal Pricing

Lagrangian for government's problem:

$$\max_{\mathbf{q},\boldsymbol{\kappa}} CS(\mathbf{q},\boldsymbol{\kappa}) - E(\mathbf{q},\boldsymbol{\kappa}) + (1+\lambda)\cdot\left\{\sum_j p_j(\mathbf{q},\boldsymbol{\kappa})\cdot q_j - C(\mathbf{q},\boldsymbol{\kappa}) - B\right\}$$

First-order conditions

1. Quantity:

$$p_j = \frac{\partial C}{\partial q_j} - \frac{1}{1+\lambda}\cdot\frac{\partial U}{\partial t_j}\cdot\frac{\partial t_j}{\partial q_j} + \frac{1}{1+\lambda}\cdot\frac{\partial E}{\partial q_j} - \frac{\lambda}{1+\lambda}\cdot\sum_k q_k\cdot\frac{\partial p_k}{\partial q_j}$$

2. Capacity:

$$\frac{\partial C}{\partial \kappa_j} = \frac{1}{1+\lambda}\cdot\frac{\partial U}{\partial t_j}\cdot\frac{\partial t_j}{\partial \kappa_j} - \frac{1}{1+\lambda}\cdot\frac{\partial E}{\partial \kappa_j} + \frac{\lambda}{1+\lambda}\cdot\sum_k q_k\cdot\frac{\partial p_k}{\partial \kappa_j}$$

Optimal Pricing

$$p_j - \underbrace{\tilde{C}_j}_{\text{mg. cost}} = \underbrace{\tilde{E}_j}_{\text{mg. externality}} - \underbrace{\bar{u}_j^T}_{\text{Utility of time}} \cdot \underbrace{(\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Returns to scale}} +$$

$$\frac{\lambda}{1+\lambda} \cdot \left\{ - \tilde{E}_j - \underbrace{\sum_{k \in J} q_k \cdot \Omega_{kj}}_{\text{Market power distortion}} - \underbrace{(\tilde{u}_j^T - \bar{u}_j^T) \cdot (\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Spence distortion}} \right\}$$

Optimal Pricing

$$p_j - \underbrace{\tilde{C}_j}_{\text{mg. cost}} = \underbrace{\tilde{E}_j}_{\text{mg. externality}} - \underbrace{\bar{u}_j^T}_{\text{Utility of time}} \cdot \underbrace{(\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Returns to scale}} +$$

$$\frac{\lambda}{1+\lambda} \cdot \left\{ - \tilde{E}_j - \underbrace{\sum_{k \in J} q_k \cdot \Omega_{kj}}_{\text{Market power distortion}} - \underbrace{(\tilde{u}_j^T - \bar{u}_j^T) \cdot (\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Spence distortion}} \right\}$$

Notation:

- $\tilde{C}_j = \frac{\partial C_j}{\partial q_j} + \frac{\partial C}{\partial \kappa_j} \frac{\kappa_j}{q_j}$: Mg. cost of a trip, keeping capacity per trip $\frac{\kappa_j}{q_j}$ constant
- $\tilde{E}_j = \frac{\partial E_j}{\partial q_j} + \frac{\partial E_j}{\partial \kappa_j} \frac{\kappa_j}{q_j}$: Mg. externality of a trip, keeping capacity per trip $\frac{\kappa_j}{q_j}$ constant

$\Longrightarrow$ *"Standard"* ***Pigouvian tax/subsidy***

Optimal Pricing

$$p_j - \underbrace{\tilde{C}_j}_{\text{mg. cost}} = \underbrace{\tilde{E}_j}_{\text{mg. externality}} - \underbrace{\bar{u}_j^T}_{\text{Utility of time}} \cdot \underbrace{(\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Returns to scale}} +$$

$$\frac{\lambda}{1+\lambda} \cdot \left\{ - \tilde{E}_j - \underbrace{\sum_{k \in J} q_k \cdot \Omega_{kj}}_{\text{Market power distortion}} - \underbrace{(\tilde{u}_j^T - \bar{u}_j^T) \cdot (\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Spence distortion}} \right\}$$

Notation:

- $\bar{u}_j^T = \frac{\partial CS}{\partial t_j} \frac{t_j}{q_j}$: If t_j increases by 1%, how does utility per commuter change?
- $\epsilon_j^{T,q} + \epsilon_j^{T,\kappa}$: If both # of trips and capacity increase by 1%, what is the change in total time?

$\Longrightarrow$ *Higher subsidy to modes with larger returns to scale*

Optimal Pricing

$$p_j - \underbrace{\tilde{C}_j}_{\text{mg. cost}} = \underbrace{\tilde{E}_j}_{\text{mg. externality}} - \underbrace{\bar{u}_j^T}_{\text{Utility of time}} \cdot \underbrace{(\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Returns to scale}} +$$

$$\frac{\lambda}{1+\lambda} \cdot \left\{ - \tilde{E}_j - \underbrace{\sum_{k \in J} q_k \cdot \Omega_{kj}}_{\text{Market power distortion}} - \underbrace{(\tilde{u}_j^T - \bar{u}_j^T) \cdot (\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Spence distortion}} \right\}$$

Notation:

- Ω: Inverse Jacobian of demand ($\sim$ inverse price elasticities)
- $\tilde{u}_j^T$: If t_j increases by 1%, how much does utility of *marginal* commuters change?

Optimal Pricing

$$p_j - \underbrace{\tilde{C}_j}_{\text{mg. cost}} = \underbrace{\tilde{E}_j}_{\text{mg. externality}} - \underbrace{\bar{u}_j^T}_{\text{Utility of time}} \cdot \underbrace{(\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Returns to scale}} +$$

$$\frac{\lambda}{1+\lambda} \cdot \left\{ - \tilde{E}_j - \underbrace{\sum_{k \in J} q_k \cdot \Omega_{kj}}_{\text{Market power distortion}} - \underbrace{(\tilde{u}_j^T - \bar{u}_j^T) \cdot (\epsilon_j^{T,q} + \epsilon_j^{T,\kappa})}_{\text{Spence distortion}} \right\}$$

Notation:

- Ω: Inverse Jacobian of demand ($\sim$ inverse price elasticities)
- $\tilde{u}_j^T$: If t_j increases by 1%, how much does utility of *marginal* commuters change on avg.?

Government cares about **budget**, so it behaves like a **monopolist** to some extent:

- Underweights externalities
- Market power markup. Larger for less elastic modes, accounts for cross elasticities
- Caters to marginal rather than average consumer

Roadmap

Demand

Market Definition:

- Community area o to d at hour h

Consumer covariates:

$x_i = (\text{car ownwership}_i, \text{precise location}_i, \text{income}_i)$

Choices depend on access:

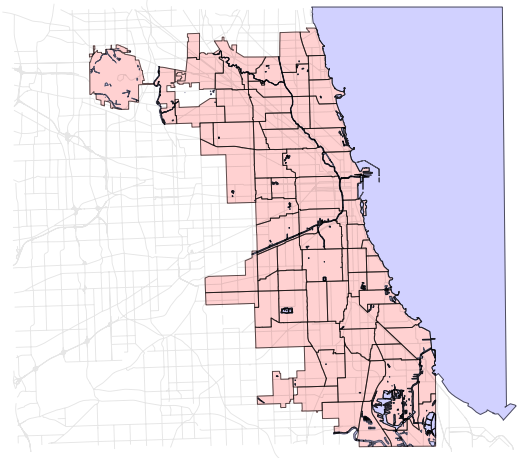
$$j^\star = \underset{j \in \mathcal{J}_{odh}(x_i)}{\text{argmax}}\ u^i_{odhj}$$

Utility of agent i:

$$u^i_{odhj} = -\alpha_t(x_i) \cdot t_{odhj}(x_i) - \alpha_p(x_i) \cdot p_{odhj} + \xi_{odhj} + \epsilon^i_{odhj},$$

where: $\xi_{odtj} = \mu_{h,j} + \mu_o + \mu_d + \nu_{odtj}$.

Chicago Community Areas

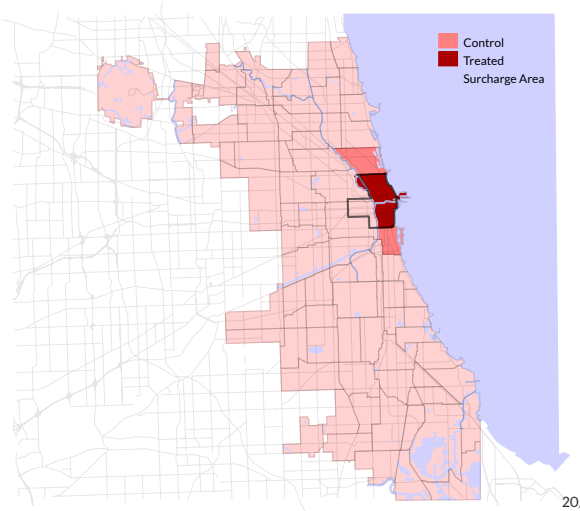


Identifying Variation for Price Elasticity: Ride-Hailing Trip Surcharge

We exploit a tax to ride-hailing that offers *temporal* and *spatial* price discontinuities:

- Introduced Jan 2020.
- Surcharge (higher tax) for trips to/from "Surcharge area".
- Active weekdays 6am-10pm.
- $1.75 for "standard", $0.65 for pool trips

Surcharge Areas (Treated) and Neighboring Areas (Control)
by Community Area



Identifying Variation for Price Elasticity: Ride-Hailing Trip Tax

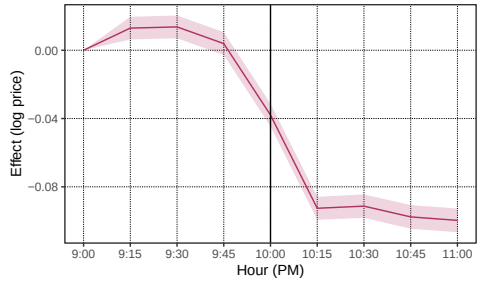
Specification:

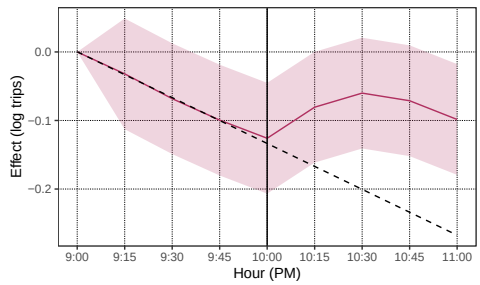
$$y_{o,d,t} = \mu_{o,d} + \alpha_t + \beta_t \cdot treat_{o,d} + \epsilon_{o,d,t}$$

where:

- $y_{o,d,t}$: Log price, log trips
- o, d: origin/destination community area
- t: Time of the day (15-min intervals)
- $treat_{o,d}$: Are $o \rightarrow d$ trips subject to surcharge?

Figures plot coefficients β_t 2020





Identifying Variation for Price Elasticity: Ride-Hailing Trip Tax

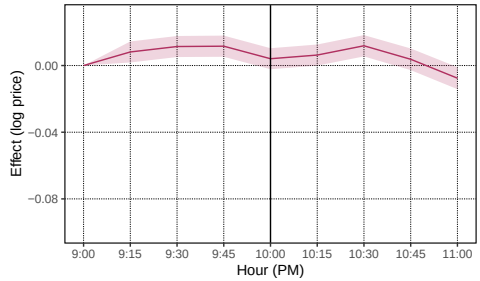
Specification:

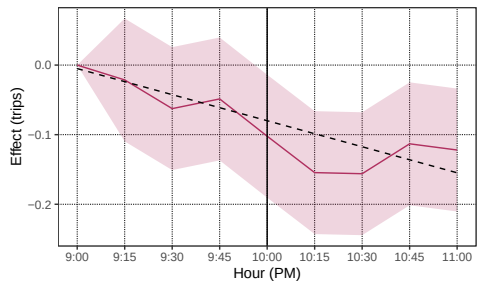
$$y_{o,d,t} = \mu_{o,d} + \alpha_t + \beta_t \cdot treat_{o,d} + \epsilon_{o,d,t}$$

where:

- $y_{o,d,t}$: Log price, log trips
- o, d: origin/destination community area
- t: Time of the day (15-min intervals)
- $treat_{o,d}$: Are $o \rightarrow d$ trips subject to surcharge?

Figures plot coefficients β_t 2019





Simple Demand Results

	(1)	(2)	(3)	(4)	(5)
Trip Time (in hours)	-4.051***	-1.632***	-2.002***	-1.717***	-1.722***
	(0.127)	(0.315)	(0.187)	(0.264)	(0.275)
Price	-0.219***	-0.128***	0.00829*		
	(0.00806)	(0.00938)	(0.00470)		
VOT	$18.49	$12.79	-241.57	$19.07	$19.14
Price Elasticity	-1.56	-.91	.06	-1.42	-1.42
Hour FE		✓	✓	✓	✓
Day of Week FE		✓	✓	✓	✓
O-D FE		✓	✓	✓	✓
Mode FE			✓	✓	✓
Elasticity from Policy				✓	✓
Mode-Hour-We FE					✓
R_a^2	0.653	0.398	0.630	0.552	0.583
Observations	743046	742803	742803	742803	742803

Simple Demand Results

	(1)	(2)	(3)	(4)	(5)
Trip Time (in hours)	-4.051***	-1.632***	-2.002***	-1.717***	-1.722***
	(0.127)	(0.315)	(0.187)	(0.264)	(0.275)
Price	-0.219***	-0.128***	0.00829*		
	(0.00806)	(0.00938)	(0.00470)		
VOT	$18.49	$12.79	-241.57	$19.07	$19.14
Price Elasticity	-1.56	-.91	.06	-1.42	-1.42
Hour FE		✓	✓	✓	✓
Day of Week FE		✓	✓	✓	✓
O-D FE		✓	✓	✓	✓
Mode FE			✓	✓	✓
Elasticity from Policy				✓	✓
Mode-Hour-We FE					✓
R_a^2	0.653	0.398	0.630	0.552	0.583
Observations	743046	742803	742803	742803	742803

Simple Demand Results

	(1)	(2)	(3)	(4)	(5)
Trip Time (in hours)	-4.051***	-1.632***	-2.002***	-1.717***	-1.722***
	(0.127)	(0.315)	(0.187)	(0.264)	(0.275)
Price	-0.219***	-0.128***	0.00829*		
	(0.00806)	(0.00938)	(0.00470)		
VOT	$18.49	$12.79	-241.57	$19.07	$19.14
Price Elasticity	-1.56	-.91	.06	-1.42	-1.42
Hour FE		✓	✓	✓	✓
Day of Week FE		✓	✓	✓	✓
O-D FE		✓	✓	✓	✓
Mode FE			✓	✓	✓
Elasticity from Policy				✓	✓
Mode-Hour-We FE					✓
R_a^2	0.653	0.398	0.630	0.552	0.583
Observations	743046	742803	742803	742803	742803

Simple Demand Results

	(1)	(2)	(3)	(4)	(5)
Trip Time (in hours)	-4.051***	-1.632***	-2.002***	-1.717***	-1.722***
	(0.127)	(0.315)	(0.187)	(0.264)	(0.275)
Price	-0.219***	-0.128***	0.00829*		
	(0.00806)	(0.00938)	(0.00470)		
VOT	$18.49	$12.79	-241.57	$19.07	$19.14
Price Elasticity	-1.56	-.91	.06	-1.42	-1.42
Hour FE		✓	✓	✓	✓
Day of Week FE		✓	✓	✓	✓
O-D FE		✓	✓	✓	✓
Mode FE			✓	✓	✓
Elasticity from Policy				✓	✓
Mode-Hour-We FE					✓
R_a^2	0.653	0.398	0.630	0.552	0.583
Observations	743046	742803	742803	742803	742803

Roadmap

Shared On-Demand Transit

Estimating Network Externalities

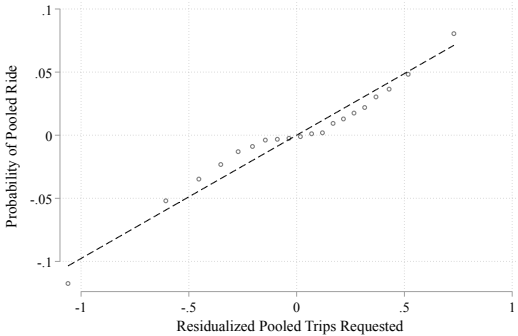
Estimate matching technology assumption by matching four moments:

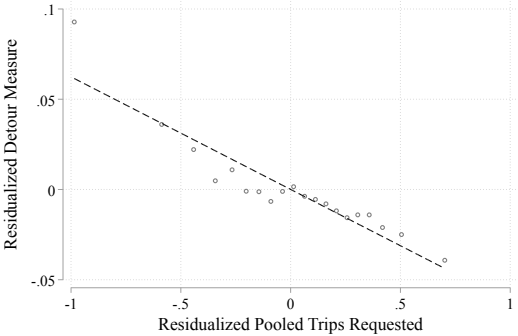
1. Match probability
2. Average detour
3. Elasticity of match probability wrt # of trips
4. Elasticity of average detour wrt # of trips

Microfoundation

Shared On-Demand Transit

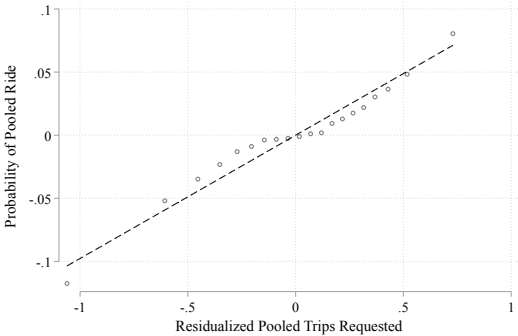
Estimating Network Externalities

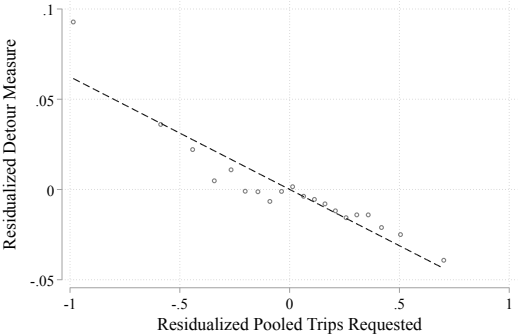




Shared On-Demand Transit

Estimating Network Externalities





Doubling the number of trip requests leads to:

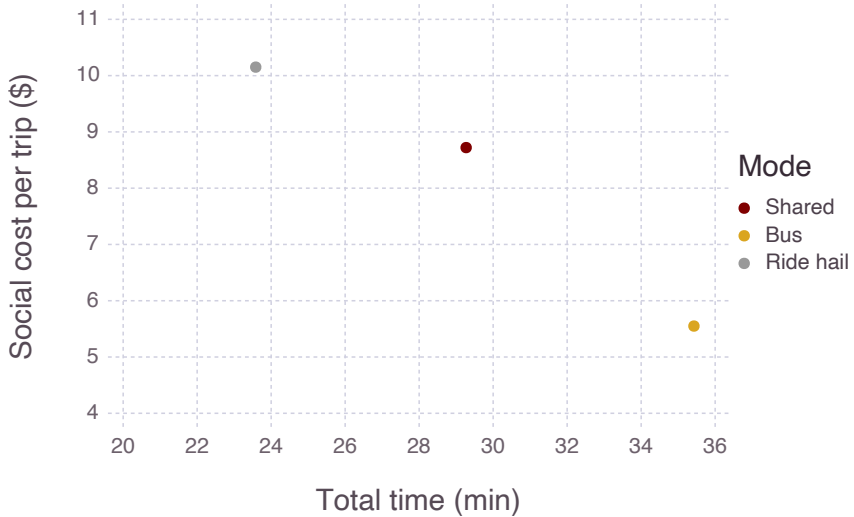
Extensive margin: 10% Increase in matches

Intensive margin: 8% reduction in detour

Roadmap

Status Quo

Mode Comparison



Counterfactual
Additional Assumptions

Externalities:

- CO_2 cost: $100 per ton; congestion externality: 10 cents per km.

City Budget:

- Pick budget constraint that corresponds to current transit spending.
- Capacity: city hires drivers at market wages.

Other (current) model restrictions:

- Bus technology.
  - Zero marginal cost, wait times inversely proportional to capital spending.
- Taxis and Ride Hail modeled as one technology.
- Random coefficient proportional to income.

Counterfactuals

Compare status quo to:

1. No shared / no ride hail
2. Ramsey (city planner with budget constraint)
3. Ramsey with more efficient sharing technology (50% detour reduction)

Counterfactual Results

	Status quo	No Shared / No Ride Hail	Ramsey	Improved Shared / Ramsey
Ride Hail Price ($)	24.5	∞	8.3	8.6
Bus Price ($)	2.5	2.5	0.2	0.7
On-demand Shared Price ($)	17.1	∞	0.8	3.1
Ride Hail T (mins)	20.0	∞	21.4	21.6
Bus T (mins)	36.5	36.5	32.7	32.8
On-demand Shared T (mins)	25.6	∞	24.7	22.6
Ride Hail %	7.7	0.0	4.1	3.2
Bus %	20.4	22.8	27.4	25.2
On-demand Shared %	1.9	0.0	3.3	7.7
Car %	70.1	77.2	65.2	63.9
Δ Welfare ($Million per day)	0.0	-14.4	22.32	25.1
Δ CS ($Million per day)	0.0	-8.1	20.1	22.5
Revenue / Profit ($Million per day)	-13.4	-19.6	-13.4	-13.4
Externalities ($Million per day)	7.6	7.7	5.4	5.5

Counterfactual Results

	Status quo	No Shared / No Ride Hail	Ramsey	Improved Shared / Ramsey
Ride Hail Price ($)	24.5	∞	8.3	8.6
Bus Price ($)	2.5	2.5	0.2	0.7
On-demand Shared Price ($)	17.1	∞	0.8	3.1
Ride Hail T (mins)	20.0	∞	21.4	21.6
Bus T (mins)	36.5	36.5	32.7	32.8
On-demand Shared T (mins)	25.6	∞	24.7	22.6
Ride Hail %	7.7	0.0	4.1	3.2
Bus %	20.4	22.8	27.4	25.2
On-demand Shared %	1.9	0.0	3.3	7.7
Car %	70.1	77.2	65.2	63.9
Δ Welfare ($Million per day)	0.0	-14.4	22.32	25.1
Δ CS ($Million per day)	0.0	-8.1	20.1	22.5
Revenue / Profit ($Million per day)	-13.4	-19.6	-13.4	-13.4
Externalities ($Million per day)	7.6	7.7	5.4	5.5

Counterfactual Results

	Status quo	No Shared / No Ride Hail	Ramsey	Improved Shared / Ramsey
Ride Hail Price ($)	24.5	∞	8.3	8.6
Bus Price ($)	2.5	2.5	0.2	0.7
On-demand Shared Price ($)	17.1	∞	0.8	3.1
Ride Hail T (mins)	20.0	∞	21.4	21.6
Bus T (mins)	36.5	36.5	32.7	32.8
On-demand Shared T (mins)	25.6	∞	24.7	22.6
Ride Hail %	7.7	0.0	4.1	3.2
Bus %	20.4	22.8	27.4	25.2
On-demand Shared %	1.9	0.0	3.3	7.7
Car %	70.1	77.2	65.2	63.9
Δ Welfare ($Million per day)	0.0	-14.4	22.32	25.1
Δ CS ($Million per day)	0.0	-8.1	20.1	22.5
Revenue / Profit ($Million per day)	-13.4	-19.6	-13.4	-13.4
Externalities ($Million per day)	7.6	7.7	5.4	5.5

Counterfactual Results

	Status quo	No Shared / No Ride Hail	Ramsey	Improved Shared / Ramsey
Ride Hail Price ($)	24.5	∞	8.3	8.6
Bus Price ($)	2.5	2.5	0.2	0.7
On-demand Shared Price ($)	17.1	∞	0.8	3.1
Ride Hail T (mins)	20.0	∞	21.4	21.6
Bus T (mins)	36.5	36.5	32.7	32.8
On-demand Shared T (mins)	25.6	∞	24.7	22.6
Ride Hail %	7.7	0.0	4.1	3.2
Bus %	20.4	22.8	27.4	25.2
On-demand Shared %	1.9	0.0	3.3	7.7
Car %	70.1	77.2	65.2	63.9
Δ Welfare ($Million per day)	0.0	-14.4	22.32	25.1
Δ CS ($Million per day)	0.0	-8.1	20.1	22.5
Revenue / Profit ($Million per day)	-13.4	-19.6	-13.4	-13.4
Externalities ($Million per day)	7.6	7.7	5.4	5.5

Graphs

Future Steps

Much Work Ahead!

- Add demand heterogeneity:
  - Income
  - Car ownership as choice set variation
  - Include heterogeneity in access
- Endogenize congestion (Couture, Duranton, and Turner (2018))
- Politically feasible counterfactuals
- Distributional effects

Conclusions

- Construct new dataset of all trips across all modes for a major urban area.
- Compute **optimal transit prices** and **capacities** with government budget constraints.
  $\Longrightarrow$ Application: can **on-demand** shared transit scale up?
- Preliminary results:
  - Public transit should be further subsidized
  - Potential of restricted on-demand sharing technology limited

Thank you!

Supply

Parametrizing Pooling Technology

Total trip time:

$$T = \underbrace{w(I)}_{\text{Waiting time}} + \underbrace{0.5 \cdot \tau}_{\text{Batch Length}} + \underbrace{T^0}_{\text{Direct time}} + \underbrace{P(q)}_{\substack{\text{Match} \\ \text{prob.}}} \cdot \underbrace{d(q)}_{\substack{\text{Expected} \\ \text{detour}}},$$

where I is the number of idle drivers, $w(I)$ and $d(q)$ are decreasing, $P(q)$ is increasing.

Supply

Parametrizing Pooling Technology

Total trip time:

$$T = \underbrace{w(I)}_{\text{Waiting time}} + \underbrace{0.5 \cdot \tau}_{\text{Batch Length}} + \underbrace{T^0}_{\text{Direct time}} + \underbrace{P(q)}_{\text{Match prob.}} \cdot \underbrace{d(q)}_{\text{Expected detour}},$$

where I is the number of idle drivers, $w(I)$ and $d(q)$ are decreasing, $P(q)$ is increasing.

Suppose that density of **trips with detour distance** x is proportional to

$$\overbrace{\alpha}^{\text{Efficiency}} \cdot \overbrace{q}^{\text{Requests}} \cdot \overbrace{\tau}^{\text{Batch length}} \cdot \frac{1}{\eta} \cdot x^{\overbrace{\eta - 1}^{\text{Skewness}}},$$

then $x \sim$ **Weibull** (also shortest path btn. vertices on random graph; Bauckhage et al., 2013).

Supply
Parametrizing Pooling Technology

Total trip time:

$$T = \underbrace{w(I)}_{\text{Waiting time}} + \underbrace{0.5 \cdot \tau}_{\text{Batch Length}} + \underbrace{T^0}_{\text{Direct time}} + \underbrace{P(q)}_{\substack{\text{Match} \\ \text{prob.}}} \cdot \underbrace{d(q)}_{\substack{\text{Expected} \\ \text{detour}}},$$

where I is the number of idle drivers, $w(I)$ and $d(q)$ are decreasing, $P(q)$ is increasing.

Suppose that density of **trips with detour distance** x is proportional to

$$\overbrace{\alpha}^{\text{Efficiency}} \cdot \overbrace{q}^{\text{Requests}} \cdot \overbrace{\tau}^{\substack{\text{Batch} \\ \text{length}}} \cdot \frac{1}{\eta} \cdot x^{\overbrace{\eta - 1}^{\text{Skewness}}},$$

then $x \sim$ **Weibull** (also shortest path btn. vertices on random graph; Bauckhage et al., 2013).

- Only match if detour $< \bar{x}$.
- $P(q)$ is the CDF of the shortest detour $1 - exp(-\alpha \cdot q \cdot \tau \cdot \bar{x}^{\eta})$.
- Expected detour $d(\bar{x}, \alpha, q, \tau)$: mean of detour, truncated at $\bar{x}$

Supply
Utilization

How much **driver time** does a trip take on average?

$$T^B = \underbrace{(1 - F(\bar{x}; \alpha, q, \tau))}_{\text{Prob. of no match}} \cdot \underbrace{T^0}_{\substack{\text{Busy time} \\ \text{if unmatched}}} + \underbrace{F(\bar{x}; \alpha, q, \tau)}_{\text{Match prob.}} \cdot \underbrace{\frac{1}{2} \cdot \left(T^0 + d(\bar{x}, \alpha, q, \tau) \right)}_{\text{Busy time if matched}}$$

The number of idle drivers is:

$$I = L - q \cdot T^B.$$

Waiting time (distance to the closest driver):

- Avg. waiting time is $w(I) = \frac{\rho}{\sqrt{I}}$ (Arnott 1996).

Additional **assumptions:**

- Optimal match radius depends on density: $\bar{x}(q) = \delta \cdot q^{-\beta}$
- Batch length τ fixed at one minute.
- Set ρ so that avg. ride-hailing time is 4 min

Supply

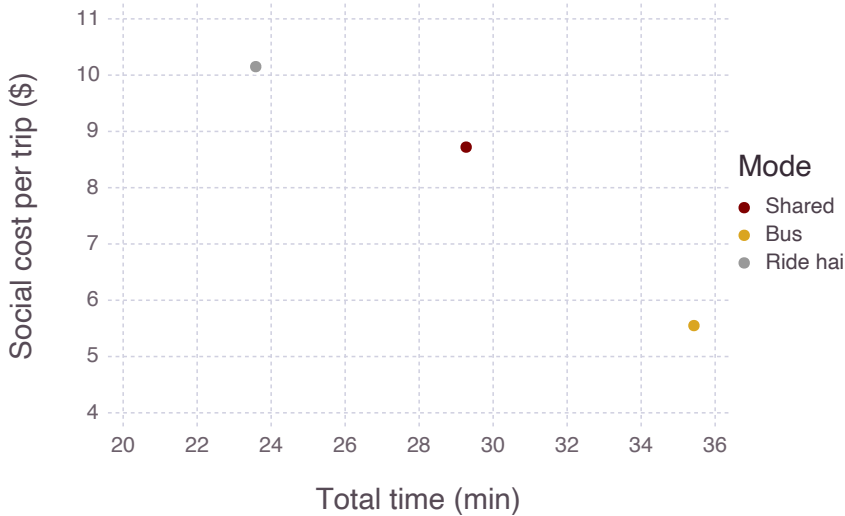
From model to the data

Given parameters $\alpha, \eta, \beta, \delta$, we can construct

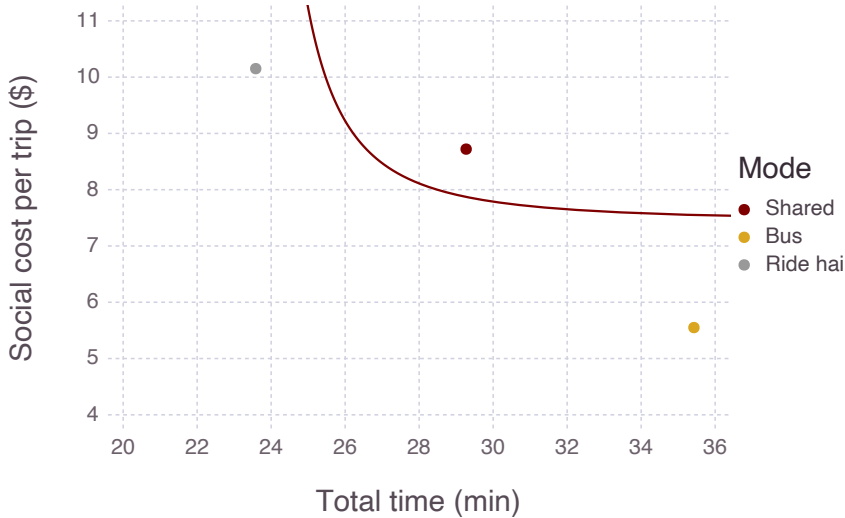
- $P(q)$ of being pooled $1 - exp(-\alpha \cdot q \cdot \tau \cdot \bar{x}^{\eta})$.
- $E(q)$ Expected detour $d(\bar{x}, \alpha, q, \tau)$

$\Longrightarrow$ Both functions $P(q)$ and $E(q)$ can be directly constructed from the data!

Improving current technologies

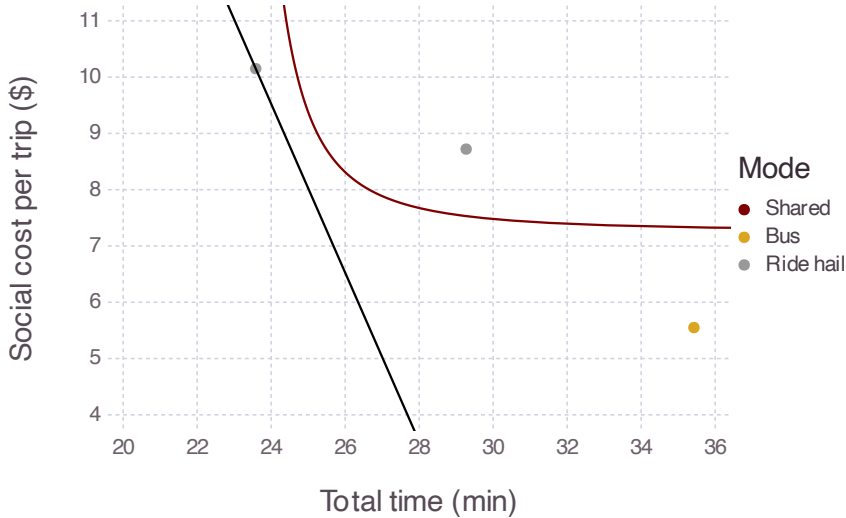


Improving current technologies



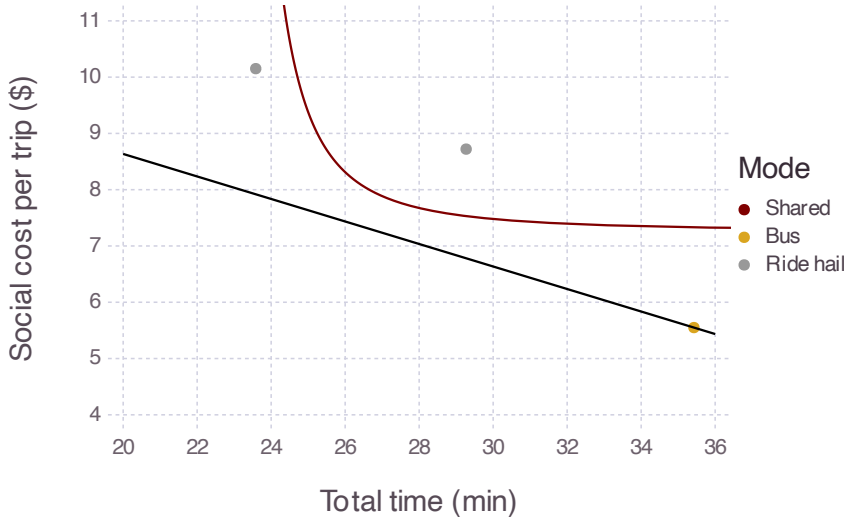
Back

Improving current technologies



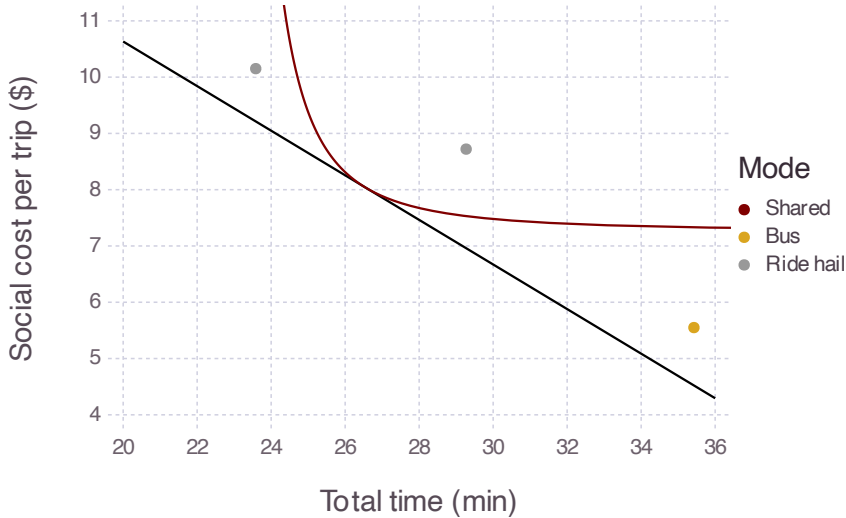
Back

Improving current technologies



Back

Improving current technologies



Back