Equilibrium Commuting Costs: The Role of Private and Public Transit

Jordán Mosqueda Juárez *
University of California San Diego

Click here for the latest version

October 20, 2025

Abstract

Developing cities rely on a mix of private minibuses and public transit, with many commutes being multimodal. This paper investigates how private providers' decisions shape commuting costs, considering complementarities with the public network, and the welfare and spatial consequences of policies that directly shift prices such as fare regulation and subsidies. I develop a quantitative spatial model in which commuters choose multimodal routes and private providers shape commuting costs through entry, pricing, and frequencies, affecting congestion and network-wide costs. The model is disciplined with newly-collected geographic and service data covering the near-universe of transit lines in the Mexico City metropolitan area. To identify key substitution and congestion elasticities, I exploit road-link-level speed changes induced by an exogenous subway-line collapse. Counterfactual analyses suggest that price-based policies can generate welfare gains comparable to infrastructure expansions. The mechanisms underscore that the endogenous response of the private sector and network-wide cost interactions are central to understand the effects of transit interventions.

*I am grateful to my co-chairs Fabian Eckert and Marc Muendler, and to my committee members Nir Jaimovich, Craig McIntosh, and Fabian Trottner for their guidance and feedback. I am especially indebted to Fabian Eckert and Fabian Trottner for their sustained mentorship and encouragement throughout this project. I am also grateful to Treb Allen, Lucas Conwell, Marco González-Navarro, Jim Hamilton, Gordon Hanson, Juan Herreño, Gaurav Khanna, Munseob Lee, Shaun McRae, Jorge Pérez Pérez, Esteban Rossi-Hansberg, Nick Tsivanidis, Johannes Wieland, Román Zárate, and numerous participants at seminars, for conversations that improved the paper. I am thankful to my mentors and colleagues that made my research stay at Banco de México very productive. I acknowledge financial support from the Department of Economics at UCSD and UC Alianza MX, and give special thanks to Natalia Volkow and her team for all the support at INEGI's microdata lab in Mexico City. All errors are on my own.

1 INTRODUCTION

Developing cities have long relied on private, decentralized, and often informal transit providers to satisfy the mobility needs of their populations. Despite major public investment, current political-economy and capacity constraints signal that many cities will continue to operate *mixed* systems with private providers at the core. In Mexico City—where transit accounts for roughly two-thirds of commuting—about 80% of trips are on privately owned minibuses and around 60% are multimodal.[1] In such settings, commuting costs are shaped by interactions across lines or markets: changing the cost of one leg shifts entry incentives and passenger flows on connected legs, with welfare and spatial consequences. At the same time, infrastructure expansions are fiscally costly and slow. To ease commuting costs, governments often rely on operational and price instruments such as fare regulation and subsidies.[2] Regulation could distort private entry and affect congestion when fares diverge from heterogeneous costs, and subsidies in a market could spill over to connected markets, affecting entry incentives and service provision. Yet the effects of these price-shifting policies have received limited attention.

Standard commuting models (Ahlfeldt et al., 2015; Allen and Arkolakis, 2014; Monte et al., 2018) are not fully suited to study these issues as they often treat commuting costs as iceberg terms exogenously parameterized by travel times. In mixed systems, both the time and price components of commuting costs are endogenous outcomes of private entry. Providers can directly influence prices and travel times (via frequencies and speeds), and cost changes can propagate to other markets due to multimodal cost complementarities and congestion. Ignoring these equilibrium feedbacks could misstate policy effects. Recent work recognizes some of these margins when studying private markets in isolation (Conwell, 2024) or localized private displacement from new public lines (Björkegren et al., 2025). What is missing is a general framework in which (i) private and public providers coexist at scale, and (ii) prices and times are determined in equilibrium. This would further allow us to study policies that shift prices or entry incentives rather rather than only time, e.g. when new infrastructure is built (Tsivanidis, 2019; Zárate, 2024).

This paper investigates how commuting costs are determined in equilibrium in such mixed systems and how price-shifting policies affect welfare and the spatial distribution of economic activity. I start by developing a general quantitative spatial framework that features private agents that make entry decisions and choose service characteristics in the presence of a broader mass-transit network, generating frequencies and trip times that interact through congestion on shared road links. Then, I calibrate the model to newly-collected data on the near-universe of private and public transit lines in the metropolitan area of Mexico City, and identify two key elasticities using quasi-experimental variation generated by the collapse of a subway line. Finally, I use the framework to evaluate two price-shifting policies. In the baseline, deregulating private fares increases welfare by about 0.9%, eliminating the metro fare increases welfare by roughly 0.5%, and applying both jointly yields a net gain of around 1.4%.

[1] Calculated using data from an origin-destination survey, INEGI *Encuesta Origen Destino 2017*, representative of more than 6 million trips.

[2] These policies are widespread across Latin America, Africa, and Southeast Asia.

Theory. The model has two components: demand of commuting and supply of transportation. In the demand side, commuters make location and route decisions, and generate route-level demand for trips (Allen and Arkolakis, 2022; Bordeu, 2023). Workers trade off time and income to choose (multimodal) routes from a given route choice set. For each location pair, there maybe one or several routes to travel; some may involve single rides, some other combinations. In this environment, a route is defined as a sequence of 'legs', each of which can be a different market, e.g. minibus line A and subway line B, or simply minibus line C. In each leg, commuters wait for a vehicle to come, pay a trip fare, and ride for some time—potentially encountering congestion on the road. The effective cost of a route, thus, comprises the time and pecuniary costs of all the markets that are used along the route. Under standard assumptions about the idiosyncratic taste of workers,[3] the commuting cost index between a given origin-destination pair aggregates the ratios of price and effective work-time net of commuting with some elasticity of substitution across routes.

The supply side—novel in this class of models—consists of many decentralized private markets (minibus lines) operating on the road network alongside a fixed public mass-transit network. In each market, a set of potential firms decides whether to enter by paying a fixed cost. Entrants maximize profits by choosing prices and the number of trips subject to vehicle capacity and a time budget per shift, given some individual residual demand. Markets are heterogenous: both the fixed cost of entry and variable cost are market-specific. The variable cost is a function of the time it takes to complete a trip, which increases with congestion caused by entrants across all markets that share road links. The strength of the increase in congestion to entrants is governed by an elasticity of congestion. Free entry and market clearing determine the equilibrium number of entrants, prices, and service frequency. These outcomes pin down travel times, waits, and pecuniary costs along routes—which are ultimately the components of the commuting cost index that commuters face.

There are two new key mechanisms in the model that govern how interactions across markets take place: a route-cost substitution/complementarity effect and a frequency-congestion trade-off. First, because routes are composed of potentially many markets and agents care about aggregate route costs, changes in the monetary or time cost of a single market directly affect connecting markets. In a first proposition, I show that to a first order, the elasticity of a route flow with respect to the cost of a given market depends on the elasticity of substitution across routes, the share of expenditure devoted to commuting, and the relative importance of a market on a given route. So, for example, a fare subsidy to one market effectively translates into subsidies for potentially many connected markets within the route. Second, more entry has an ambiguous effect on frequency. Entry directly increases frequency, but congestion caused by entry decreases it by increasing trip times for all markets that share a given road. In a second proposition, I show that the frequency gains in a market could be offset by congestion depending on the elasticity of congestion to additional entrants and the relative presence of a market in a given road link.

[3] Workers receive an idiosyncratic taste shock distributed nested Fréchet for location decisions in an upper nest, and route decisions in a lower nest.

These mechanisms are key to understand the effects of any transit intervention. If we consider these interactions across markets, it is not obvious that policies that lower times (e.g. infrastructure expansions) or prices (e.g. mass transit subsidies) will strictly improve commuting costs: the equilibrium adjustment of the private sector in response to changes in demand is crucial to assess the overall effects.

Quantification. I quantify the model using as setting one of the largest metro areas in the world: the Metropolitan Area of Mexico City, with 22 million residents. This is an attractive setting for several reasons. It features substantial variation in demographics and access to different types of transit, pairing a public mass transit network with an extensive privately operated network of minibuses. This megacity is one of the few cities in the world with existing granular data that covers the near-universe of *combi/micro/colectivo*[4] routes, making it possible to measure service and coverage at the city scale. Further, the network has experienced major transit disruptions, providing unique opportunities to leverage exogenous variation and learn about the key parameters that govern route preferences and congestion. Finally, the city has different ongoing policies such as private transit fare regulations and public transit subsidies, so it is an appealing laboratory to learn about mixed systems and the effects of policies.

On top of rich granular microdata on wages, rents, and commuting flows, I gathered novel data on the near-universe of private and public transit lines.[5] To do so, I simulated Google Maps transit trips across origin–destination pairs and recovered the geography and key service characteristics of roughly 80% of the transit routes in the system–around 2,000 in total. For each transit line, I recovered the name, length, trip time, and frequency. Crucially, I back out the implied observed number of entrants operating each line using the definition of frequency.[6] Inspection of the data reveals that the aggregate private network is 19 times larger relative to the public one, highlighting its core role in transportation. Further, private markets are longer, faster, and more frequent on average than public transit. To account for congestion on the road, I matched each line to the OpenStreetMap road network. Because Mexico City operates with ongoing fare subsidies and regulations, I calibrate the model under a version in which fares are set uniformly across firms within each market and across space. This regulated-prices version, along with the observed attributes, constitutes the observed environment to which all parameters are matched and serves as the baseline for counterfactual analyses.[7]

Identification of key elasticities. Next, I identify two key elasticities exploiting plausibly exogenous variation from a natural experiment. The first is the route-substitution elasticity, which governs how commuters trade-off income and time and reallocate across routes when relative costs change. The second is the elasticity of congestion to additional entrants, which determines how entry affects frequencies and travel times on shared roads. Together, these

[4]Sistema de Transporte Público Concesionado is the name that refers to private providers that offer transit services–commonly known as combi, micro or colectivo–, that operate on predetermined routes. They take several forms, such as minibuses of different sizes, but are mostly vans.

[5]The data is a census of all transit lines that was collected by a firm, WhereIsMyTransport, and was made available through Google Maps. More details on the data section.

[6]Frequency is the inverse of the headway. The headway is the time between two vehicles.

[7]In this calibration environment, price adjustments are shut down; the supply-side adjustment margins are entry, trip times, and waits.

parameters map demand shifts into equilibrium changes in prices, waits, and speeds, so their values are crucial to assess the effects of policies. In 2021, the elevated portion of a subway line collapsed.[8] The subway shock exogenously increased costs for some routes, forcing commuters to substitute toward alternatives, mostly constituted by private providers. Since route-level flows are very difficult to observe, I overcome this challenge by focusing instead on road speeds, which map directly to changes in flows. The changes in flows induced entry and relocation of minibuses, raising congestion on certain road links relatively more than others. The identification strategy then compares speed changes across two sets of links: those in *od* pairs where the shock leaves only a single viable route, which isolates congestion effects, and those in *od* pairs with multiple remaining alternatives, which capture substitution. The SMM exercise consists of choosing the elasticities that best replicate these observed patterns of speed changes. I find large values of these elasticities relative to related literature (Allen and Arkolakis, 2022; Bordeu, 2023; Mosquera, 2024), suggesting that congestion forces are substantial and commuters respond strongly to changes in route costs.

Quantitative exercises. With the calibrated framework in hand, I study two policies: a uniform fare regulation in the private sector and a subway fare subsidy. These policies are of both academic and policy interest because they act through understudied mechanisms and entail salient political and fiscal trade-offs. Further, they reveal relevant insights about service provision. Under regulation, price adjustments are effectively shut down, so the counterfactual reveals where transit would be more needed under market conditions. Similarly, removing the metro subsidy shows not only where private transit is more needed as a complement to public transit, but where it acts as a substitute (Björkegren et al., 2025).

Private fares in the metropolitan area are set by state authorities and are virtually uniform across space.[9] They are collected almost exclusively in cash, with infrequent, city-wide politically negotiated adjustments. Because costs vary widely across corridors (length, terrain, congestion) but fares are uniform, the fare may bind differentially across markets. In the counterfactual, I allow for the full general equilibrium adjustment of prices—holding subsidies constant. As a result, prices realign with heterogeneous route costs and capacity is reallocated toward high-demand, local peripheral markets. Overall commuting costs decrease as quicker trips due to eased congestion offset the frequency losses where entry falls, and prices fall on average due to improved costs of operation. Quantitatively, deregulation increases welfare by about 0.9%, decentralizing economic activity toward less-productive peripheral districts. Larger markets connecting towards central areas lose flows while more local, shorter markets gain flows.

Regarding transit subsidies, public systems—most notably the metro, which is mostly located in central areas—operate with a flat fare that is invariant of distance or number of (metro) legs. Government officials have claimed the subsidy to be as large as 72% of the true cost of a trip,

[8] The full *Línea 12 Tláhuac-Mixcoac* was closed for three years. It connects the South-East outskirts towards central areas and carries around 150,000 passengers a day.

[9] In practice, all fares across markets follow the same two-part tariff (a base fare plus a per-km increment after a distance threshold), but observed trip fares exhibit little cross-space variation, so the regulation pins down a near-uniform price.

with the implied gap to stated cost recovered through the city budget.[10] Given the metro's high ridership, this policy represents a large recurrent transfer that shapes the generalized cost of multimodal trips throughout the network. Removing the subsidy—holding constant privates' regulation—raises metro prices and this generates a reallocation of demand towards substitute private routes. Entry in substitute routes located in central areas increases congestion, decreasing trip times but improving frequency. On the contrarty, complement (feeder) markets lose demand and service frequency worsens, increasing wait times. The net effect in commuting costs is negative. However, increases in workers' disposable income due to eliminating the need to fund the subsidy offset these negative effects, leading to an overall welfare gain of roughly 0.5%. Economic activity is decentralized towards peripheral districts, mostly driven by central congestion and increased overall costs to move within and to central areas.

Evaluating these two policies jointly yields a net welfare gain near 1.4% while saving fiscal resources. These magnitudes are roughly comparable to those found by Zárate (2024), positive gain of $\approx 0.6 - 0.8\%$ following the opening of a new subway line in Mexico City, or by Tsivanidis (2019), who found a positive gain of $\approx 0.6 - 2.3\%$ following the opening of a new BRT system in Bogotá.[11] The broader lesson is that price-shifting policies interact through endogenous supply and demand on a heterogeneous network, and can deliver substantial effects comparable to building new infrastructure. Accounting for these interactions is crucial for measuring the impacts of any transit intervention in a city with a mixed transit system.

Related literature. This paper relates to three strands of the literature: urban transportation, endogenous trade costs, and the broader quantitative spatial literature. In the urban transportation literature, prior work has examined the welfare effects of new publicly-provided mass transit infrastructure by focusing on single modes, e.g. BRT in Bogotá (Tsivanidis, 2019), cablecar in Medellín (Khanna et al., 2024) or a subway line in Mexico City (Zárate, 2024). Similarly, Conwell (2024) has explored the welfare effects of alleviating micro externalities such as boarding and queuing times in private markets in Cape Town. A closely related paper (Björkegren et al., 2025) has studied this relationship by studying how the private sector responds to the roll-out of BRT lines in Lagos. The authors find that, when the government rolls out 13 new BRT lines–along the same corridors as the private providers–, frequencies and prices fall in the private sector. These relevant findings suggest that there are strong substitutability forces at play. My findings supplement their findings by showing that we must also consider the private sector as a large complement to mass transit. I show that when we increase the cost of mass transit (the analogous exercise in their paper would be to decrease the cost of mass transit), entry increases in substitute markets and frequencies rise, but in connecting markets the opposite occurs: entry and frequencies decrease.

Another closely related paper (Almagro et al., 2024) studies how a government can adjust

[10] Actual cost is 5 pesos, but the former governor claimed that the cost would be 18 pesos without the subsidy. https://www.reforma.com/costaria-18-pesos-entrada-al-metro-sin-subsidio-batres/ar2833638

[11] These numbers are not fully directly comparable, though. Despite all models pertaining to the same class of quantitative spatial models, there are important differences in the mechanisms studied in each paper. For example, Zárate (2024) includes a labor reallocation margin from informal to formal jobs, which drives part of the gains. His model is calibrated using virtually the same battery of data, though.

prices and frequencies of public transit as well as road pricing in order to mitigate on-the-road congestion and environmental externalities in Chicago. This paper contributes to this literature by studying the endogenous determination of prices and frequencies via entry of private agents coexisting with public mass transit, in a city-wide general equilibrium framework.

Furthermore, I contribute to the endogenous trade costs literature (Brancaccio et al., 2020; Allen and Arkolakis, 2022; Allen et al., 2023) by endogenizing urban transport costs with time and budget constraints in the demand side, and private entry in the supply side. Standard commuting models (Ahlfeldt et al., 2015; Allen and Arkolakis, 2014; Monte et al., 2018) often assume iceberg commuting costs measured in utility terms, abstracting away from income effects. This is a reasonable assumption to study infrastructure improvements in richer cities, but insufficient to study developing-city settings due to binding budget constraints (Bryan et al., 2025). In my model, commuting times and prices arise from firms' entry, pricing decisions, and congestion feedbacks in general equilibrium. Also, to my best knowledge, this is the first paper that provides an estimate of the elasticity of substitution across (transit) routes exploiting quasi-experimental variation from large transit disruptions.

Finally, I propose an alternative way to analyze multimodal networks (Fuchs and Wong, 2024) that does not explicitly rely on edge-level analysis—and instead relies directly on readily available outputs (Google Maps)—but still preserves the tractability and efficiency to solve a large-scale spatial model with many markets. In particular, I depart from the multiplicative iceberg edge-level cost assumptions (Fuchs and Wong, 2024; Allen and Arkolakis, 2022) and instead assume additive costs at the route level to be able to express driver-level residual demand purely as a function of their own price and their own market's price index. This ensures tractability and facilitates computation.

2 Model

I build on the canonical commuting model (Ahlfeldt et al., 2015; Allen and Arkolakis, 2014, 2022; Bordeu, 2023; Tsivanidis, 2019; Zárate, 2024), and add two elements. On the demand side, I include commuting in the budget and time constraints, so that agents trade off income and work time when choosing potentially multi-leg routes. On the supply side, private agents offer commuting services and coexist with a predetermined mass-transit network. As a result, commuting monetary costs and time are determined in equilibrium.

2.1 Environment

Geography and transportation network. The geography consists of J locations indexed by o (origin) and d (destination). Locations are connected by an exogenous transportation network comprising the road and transit networks. The transit network includes public transit (e.g., subway) and private transit (minibuses) and may overlap with the road network—particularly for buses.

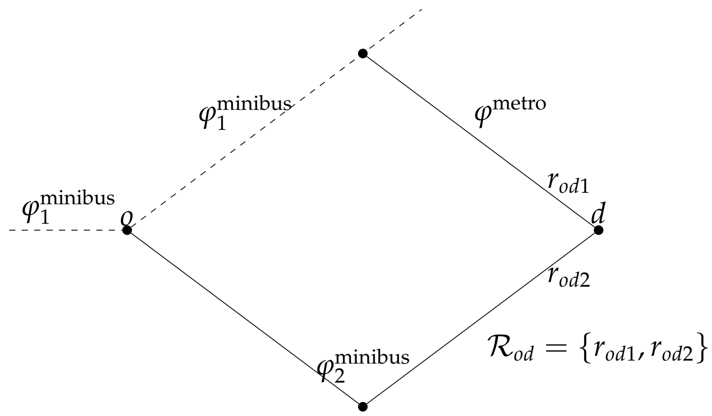
Transit markets. A market φ is a sequence of edges in the network. All public and private markets' geographies are predetermined and fixed (e.g., a specific bus or subway line). There are two types of markets: *public* and *private*. Markets have four characteristics: (i) price P_φ, (ii) trip time t_φ^{trip}, (iii) frequency $Freq_\varphi$, and (iv) wait time t_φ^{wait} (from the user's perspective). Private markets respond to demand conditions (details below). Publicly operated markets' attributes are treated as fixed and exogenous.

Routes. Each location pair od is connected by a set of routes $\mathcal{R}_{od}$. A *route* is a sequence of nodes

$$r = \{r_0 \xrightarrow{\varphi} r_1 \xrightarrow{\varphi} \cdots \xrightarrow{\varphi} r_K\},$$

where φ denotes the market used on a leg $r_{l-1} \xrightarrow{\varphi} r_l$. Under this definition, route choice implies mode choice (and its order). A route can traverse several markets; here a "market" corresponds to what is often called a "bus line" in practice.

FIGURE 1. ILLUSTRATION OF ALTERNATIVE ROUTES BETWEEN o AND d.



Note: The figure shows an example of a route choice set between an origin–destination pair with two routes. The top route uses a segment of minibus line 1 (dashed) and then the metro; the bottom route uses only minibus line 2.

Fundamentals. Each location has fundamental productivity A, residential land H, commercial land H^c, and amenities B_{od} associated with living in o and working in d.

2.2 Demand for commuting

Each worker ω chooses residence, workplace, and route to maximize

$$u_{odr}(\omega) = B_{od} \left(\frac{X(\omega)}{\alpha_x} \right)^{\alpha_x} \left(\frac{H_o(\omega)}{\alpha_h} \right)^{\alpha_h} \left(\frac{C_{odr}(\omega)}{\alpha_c} \right)^{\alpha_c} \varepsilon_{odr}(\omega), \tag{1}$$

subject to

$$X(\omega) + Q_o H_o(\omega) + P_{odr} C_{odr}(\omega) = w_d \, n_{odr} \, (1 + \Omega), \tag{2}$$

$$\bar{T} = n_{odr} + t_{odr}, \tag{3}$$

where X is the numeraire good, H is housing, C_{odr} is a CES bundle of commuting trips via route r, w_d is the wage, Q_o is rent, P_{odr} and t_{odr} are the monetary and time costs of commuting via r, and n_{odr} is work time. B_{od} are amenities that capture the fundamental value that workers receive from living in o and working in d.[12] The term Ω captures proportional income adjustments from land income and taxes.[13]

Indirect utility for a worker choosing (o, d, r) is

$$V_{odr}(\omega) = B_{od} \, \frac{\tilde{w}_d}{Q_o^{\alpha_h} \, \tau_{odr}} \, \varepsilon_{odr}(\omega), \quad \tilde{w}_d \equiv w_d(1 + \Omega), \tag{4}$$

where the generalized commuting cost is

$$\tau_{odr} \equiv \frac{P_{odr}^{\alpha_c}}{\bar{T} - t_{odr}}. \tag{5}$$

This expression for the commuting cost combines both time and price and has an intuitive interpretation: it is the monetary cost per unit of effective work time.

Each worker receives a one-time shock and makes two decisions, one for each nest: 1) residence and workplace, and 2) the route (and modes) to commute. Let $\varepsilon_{odr}(\omega)$ be distributed extreme-value type II distribution (nested Fréchet):

$$F(\vec{\varepsilon}) = \exp\left[- \sum_{o,d} \left(\sum_{r \in \mathcal{R}_{od}} \varepsilon_{odr}^{-\rho} \right)^{\theta/\rho} \right], \tag{6}$$

The shape parameters in this distribution measure (inverse) substitutability across residence-workplace pairs (θ) and routes (ρ), and $\theta < \rho$, reflecting the fact that it is easier to substitute across routes than locations.

In the upper nest, the share of workers choosing (o, d) is

$$\lambda_{od} = \frac{\left(\frac{B_{od} \, \tilde{w}_d}{Q_o^{\alpha_h} \, \tau_{od}} \right)^{\theta}}{\sum_{o'} \sum_{d'} \left(\frac{B_{o'd'} \, \tilde{w}_{d'}}{Q_{o'}^{\alpha_h} \, \tau_{o'd'}} \right)^{\theta}}, \tag{7}$$

[12] In terms of quantification, this amenity shifter serves as a structural residual that allows me to match rich data on commuting flows.

[13] This enters multiplicatively so that it does not distort spatial allocations to a first order. See Sections 2.6 and 2.7.

where the od CES commuting index with elasticity ρ is[14]

$$\tau_{od} \equiv \left(\sum_{r \in \mathcal{R}_{od}} \tau_{odr}^{-\rho} \right)^{-1/\rho}. \tag{8}$$

We can further decompose the full probability that a worker will choose odr into the probability that the od pair is chosen and the (lower-nest) probability that route r is chosen conditional on living/working in od:

$$\lambda_{odr} = \underbrace{\frac{\frac{B_{od} w_d^{\theta}}{Q_o^{\alpha_h \theta} \tau_{od}^{\theta}}}{\sum_{od} \frac{B_{od} w_d^{\theta}}{Q_o^{\alpha_h \theta} \tau_{od}^{\theta}}}}_{\lambda_{od}} \times \underbrace{\frac{\tau_{odr}^{-\rho}}{\sum_r \tau_{odr}^{-\rho}}}_{\lambda_{r|od}}. \tag{9}$$

Commuting costs at the route level. Given that routes are comprised of potentially many markets due to multi-modal/multi-legged trips, the total monetary and time cost of traveling through a route odr is the sum of the individual market's costs:

$$P_{odr} \equiv \sum_{\varphi \in r} P_{\varphi}, \tag{10}$$

$$t_{odr} \equiv \sum_{\varphi \in r} \left(t_{\varphi}^{\text{wait}} + \gamma_{odr}^{\varphi} \, t_{\varphi}^{\text{trip}} \right), \tag{11}$$

where $\gamma_{odr}^{\varphi} \in [0,1]$ is the fraction of market φ's trip used by route r. For example, a route could use only half of a bus line, so $\gamma_{odr}^{\varphi} = 0.5$. Note that once n markets are within a route, they become perfect complements. Given that agents do not care about individual costs of markets but as the aggregate route cost, a cost change in one market will directly impact the flow through other markets within the route. I summarize the implications of complementarity in the following proposition.

Proposition 1 (First-order route-flow response and within-route complementarity). Fix an OD pair (o, d) and hold λ_{od} and t_{odr} fixed with respect to P_{φ}. Define the route-level cost share of market φ on route r by

$$s_{\varphi|r} \equiv \frac{\mathbf{1}\{\varphi \in r\} \, P_{\varphi}}{P_{odr}} \in [0,1], \quad \bar{s}_{\varphi} \equiv \sum_{k \in \mathcal{R}_{od}} \lambda_{k|od} \, s_{\varphi|k}.$$

Then, the elasticity of the conditional flow with respect to the price of one market is

$$\frac{\partial \ln \lambda_{r|od}}{\partial \ln P_{\varphi}} = \rho \, \alpha_c \left(\bar{s}_{\varphi} - s_{\varphi|r} \right),$$

and the elasticity of the aggregate probability across all routes that use such market with re-

[14] For detailed derivations, refer to Appendix **??**.

spect to the price is

$$\frac{\partial}{\partial \ln P_\varphi}\left(\sum_{r\in\mathcal{R}_\varphi}\lambda_{r|od}\right) = \rho\,\alpha_c\,(\Lambda_\varphi - 1)\,\bar{s}_\varphi \leq 0, \quad \Lambda_\varphi \equiv \sum_{r\in\mathcal{R}_\varphi}\lambda_{r|od}.$$

Proof. See Appendix A.2. □

Proposition 1 shows that the strength of the response of flows along a route to an increase in the price of a specific market depends on the elasticity of substitution across routes (ρ), the commuting expenditure share (α_c), and the market's relative importance in the cost of the route $s_{\varphi|r}$. In particular, it shows that the conditional route share decreases with P_φ if $\bar{s}_\varphi \leq s_{\varphi|r}$, and increases if $\bar{s}_\varphi \geq s_{\varphi|r}$. This means that if the market's expenditure share is larger than the average, then that market is relatively important and increasing its price will lead to a reduction of the conditional flow. Further, an increase in P_φ decreases the aggregate probability of choosing *any* route that uses φ, unless $\Lambda_\varphi = 1$. This means that if all of the routes in the *od* choice set use market φ, then there is no aggregate reduction in the probability but merely a reshuffling of the probability mass across routes.

Demand for trips in each market φ . Given the decisions of workers to travel through different routes, demand for each transit market φ is determined by aggregating across all workers that chose routes that contained such market. Recall that a route can be composed of many markets (legs), so a trip along a route generates one or more trips for the markets contained in the route. Demand for trips in a given market is[15]

$$D_\varphi = \sum_{od}\sum_{r|\varphi\in r}\frac{\alpha_c\tilde{w}_d(\bar{T} - t_{odr})}{P_{odr}}\lambda_{odr}\bar{L} \tag{12}$$

2.3 Supply of transportation

I make three main assumptions about the transportation market structure:

1. Markets (φ) are **exogenously defined** in terms of their geography.

2. Private markets are **segmented**: there is a potential mass of entrants in each market.

3. There is free entry into each market.

The first assumption implies that the actual lines or roads through which minibuses travel do not change, and are given. This paper examines the intensive margin of transit, i.e. how supply varies within each market, rather than the extensive margin of where do we add or suppress lines.[16] The second and third assumptions allow me to preserve tractability and facilitate

[15] Detailed derivations in Appendix A.3

[16] This is a complex optimal transit network problem that has been explored in developing settings (Kreindler et al., 2023), although for the design of BRT lines. An interesting line of future research is the optimal design of minibus corridors, given public infrastructure.

the computation of the equilibrium of the model while preserving a realistic characterization of these markets. Private markets' specific characteristics—market structure, regulation, and operation—vary slightly across the globe in the sense that political, administrative, and economic contexts define these attributes, but overall they share common attributes.[17]

Driver's demand in a market φ. A worker consumes a CES bundle of trips in individual buses within market φ with elasticity of substitution χ. Aggregating demand for a given driver across all workers that consume in that market, we get CES demand for driver i:[18]

$$q_{i,\varphi} = \left(\frac{p_{i,\varphi}}{P_\varphi}\right)^{-\chi} D_\varphi. \tag{13}$$

Firm's problem. Individual firms (one can think of a minibus/driver as a firm) decide whether to enter the market by paying a fixed cost of entry f^e_φ. Upon entry, firms maximize profits by chooseing the price and number of trips that maximize their profits, subject to a capacity constraint q^c_φ. Firms face an invidiualCES demand, given by equation (13). One can think of this capacity as the number of seats in a minibus. The problem of the firm is:

$$\begin{aligned}
\max_{p_{i,\varphi}, n_{i,\varphi}} \ \pi_{i,\varphi} &= \left(p_{i,\varphi} \frac{q_{i,\varphi}}{n_{i,\varphi}} - \delta_\varphi \, t^{\text{trip}}_\varphi\right) n_{i,\varphi} - f^e_\varphi, \\
\text{s.t.} \quad q_{i,\varphi} &= \left(\frac{p_{i,\varphi}}{P_\varphi}\right)^{-\chi} D_\varphi, \\
\frac{q_{i,\varphi}}{n_{i,\varphi}} &\le q^c, \\
n_{i,\varphi} \, t^{\text{trip}}_\varphi &= \bar{T}^d.
\end{aligned}$$

In equilibrium, the firm chooses the price p_i such that it operates at full capacity,[19] i.e. $\frac{q_{i,\varphi}}{n_{i,\varphi}} = q^c$, which yields

$$p_{i,\varphi} = \left(\frac{D_\varphi}{q^c n_{i,\varphi}}\right)^{\frac{1}{\chi}} P_\varphi. \tag{14}$$

The market price index is

$$P_\varphi \equiv \left(\sum_{i\in\varphi} p_{i,\varphi}^{1-\chi}\right)^{\frac{1}{1-\chi}} = P_\varphi \left(\frac{D_\varphi}{q^c n_{i,\varphi}}\right)^{\frac{1}{\chi}} M_\varphi^{\frac{1}{1-\chi}}. \tag{15}$$

[17]Minibuses are privately owned, drivers may or may not be the owners, corridors are often explicitly defined and markets are segmented either by the government or by an association. Entrants pay entry costs to either the government or to an association. For example, in the specific Mexican context, markets are often conformed by associations of drivers that collectively bargain with the government on corridor concessions and vehicle permits. Drivers generally do not own the units they drive, so they typically rent a bus from a minibus owner that belongs to the association, and they pay a fixed cost to the owner. Once they enter, they do not operate in many markets simultaneously.

[18]Detailed derivations in Appendix A.3

[19]This is analogous to (Allen et al., 2023), where they explore the endogenous determination of trade costs using the trucker industry in Colombia. Trucker decisions are microfounded in a two-stage game where they choose the capacity and then compete in prices. Given capacity, they choose the price that ensures that they use all their available capacity.

From equation (15), it follows that the following market-clearing relation must hold:

$$D_\varphi(\mathbf{M}) = M_\varphi^{\chi/(\chi-1)}\, q_\varphi^c\, n_{i,\varphi},, \quad \forall \varphi \in \Phi. \tag{16}$$

This equation states that the total aggregate supply of "seats" must be equal to all trips demanded. The number of entrants M_φ is pinned down by this relationship. Note that demand depends on the full vector of entrants denoted by $\mathbf{M}$ (because entry determines commuting costs, as described in detail below), so equation (16) describes a system of $|\Phi|$ equations, where Φ is the set of markets. If we define a fixed-point operator by solving for M_φ in equation (16), then we have an update map for the vector of entrants that facilitates the computation of the model by applying usual fixed-point algorithms.

Entrants and free entry. Because all firms are homogeneous, entrants enter each market until everyone's profits are driven to zero. Substituting the price equation (14) and the market-clearing relationship (16) into profits, we obtain:

$$\pi_{i,\varphi} = \left(M^{\frac{1}{\chi-1}} P_\varphi q^c - \delta_\varphi t_\varphi^{\text{trip}} \right) n_{i,\varphi} - f_\varphi^e. \tag{17}$$

Imposing zero profits $\pi_{i,\varphi} = 0$ implies that:

$$P_\varphi = M_\varphi^{-\frac{1}{\chi-1}} \frac{t_\varphi^{\text{trip}}}{\bar{T}^d} \left(\frac{\delta_\varphi \bar{T}^d + f_\varphi^e}{q_\varphi^c} \right). \tag{18}$$

This effectively pins down the market price index as a decreasing function of the entrants and capacity, and increasing in cost parameters.

Trip times, wait times, and frequency. Having pinned down the number of entrants, then we can determine the rest of the service characteristics. For quantification purposes, I assume that the time it takes a minibus to complete a lap along its line or market, i.e. the trip time, is a function of the total number of units from all markets that pass through a given road link ℓ and an elasticity of congestion to an additional entrant, ϕ. Let each road link ℓ used by any market carry the total flow

$$S_\ell(\mathbf{M}) \;\equiv \sum_{\varphi:\, \ell \in \varphi} M_\varphi.$$

Then, the trip time in a given market will be the aggregation of the congestion measure at the road link level, across all road links that comprise the market:

$$t_\varphi^{\text{trip}}(\mathbf{M}) = \bar{t}_\varphi \sum_\ell S_\ell(\mathbf{M})^\phi. \tag{19}$$

Frequency, by definition, is the inverse of the headway, or the time between two units:

$$Freq_\varphi(\mathbf{M}) \equiv \frac{M_\varphi}{t_\varphi^{\text{trip}}(\mathbf{M})}. \tag{20}$$

Assuming that buses arrive uniformly at each stop, wait time (from the perspective of a worker waiting for the bus) is

$$t_{\varphi}^{\text{wait}}(\mathbf{M}) = \frac{1}{2}\frac{1}{Freq_{\varphi}(\mathbf{M})}. \tag{21}$$

This functional form implies that a worker that arrives at a bus stop will wait on average half of the headway between buses.[20] Importantly, note that all of these market-specific outcomes depend not only on the market's own characteristics, i.e. entrants, but on the entry of all markets that are related to the market through congestion. This is why the elasticity of congestion—and congestion more generally as a negative externality for both customers and suppliers—plays a major role in determining service characteristics. Due to the presence of this externality, it is at first glance ambiguous whether it is desirable to have an additional bus on the road. On the one hand, the benefit of an additional bus is an increase in the market's frequency, at the cost of congestion for all markets that share the roads with the market, so trip time increases for everyone, decreasing frequencies. The following proposition summarizes this trade-off and characterizes the threshold for the elasticity of congestion at which benefits are not outweighed by costs.

Proposition 2 (Frequency–congestion trade-off). Fix a market φ with entrants M_φ and hold $M_{-\varphi}$ fixed. Define the link-level share of φ on link ℓ and the ψ-specific congestion weights

$$s_{\varphi|\ell} \;\equiv\; \frac{M_\varphi}{S_\ell(M)} \in [0,1], \qquad w_{\ell|\psi} \;\equiv\; \frac{S_\ell(M)^\phi}{\sum_{j\in\psi} S_j(M)^\phi}, \qquad \sum_{\ell\in\psi} w_{\ell|\psi} = 1.$$

Then:

$$\frac{\partial \ln Freq_\varphi}{\partial \ln M_\varphi} \;=\; 1 \;-\; \phi\,\beta_\varphi, \qquad \beta_\varphi \;\equiv\; \sum_{\ell\in\varphi} w_{\ell|\varphi}\, s_{\varphi|\ell} \in [0,1],$$

and, for any $\psi \neq \varphi$,

$$\frac{\partial \ln Freq_\psi}{\partial \ln M_\varphi} \;=\; -\,\phi\,\beta_{\psi\leftarrow\varphi}, \qquad \beta_{\psi\leftarrow\varphi} \;\equiv\; \sum_{\ell\in\psi} w_{\ell|\psi}\, s_{\varphi|\ell} \in [0,1].$$

Proof. See Appendix A.4. □

Proposition 2 essentially characterizes a threshold value of the elasticity of congestion to determine whether entry in a market will increase or decrease frequency in that market. This threshold depends on the relative presence of that market in the links that it uses and shares with other markets, β_φ. The second part of Proposition 2 states that entry in a market strictly decreases frequencies for other markets. Also, it is useful to think about two extreme cases: when a market exclusively occupies certain road links, and when a market is virtually absent

[20] To obtain this, we can assume some process of arrival of buses with a parameter $Freq(.)$ that controls its mean and variance. For example, in the case of a Poisson process with parameter $\lambda = Freq(.)$, because a Poisson random variable is memoryless, we would have that $t_{\varphi}^{\text{wait}}(M_\varphi) = \frac{1}{Freq_\varphi(M_\varphi)}$, or that people wait on average the full headway between buses. The particular process that I am assuming is a uniform arrival of buses, which is one the simplest processes to model bus arrivals. This seems plausible if we think that buses are coordinated to some degree in terms of schedule – in the Mexican context, many minibus associations impose these sort of controls.

in some links. The following corollary gives us benchmark values for the elasticity of congestion for which the benefit of entry outweighs the cost.

Corollary 1 (Two simple benchmarks).

1. If φ is the *only* user of each of its links ($s_{\varphi|\ell} = 1$), then $\beta_\varphi = 1$, and the threshold value such that congestion costs exactly outweigh frequency benefits is $\phi = 1$ since

$$\frac{\partial \ln Freq_\varphi}{\partial \ln M_\varphi} = 1 - \phi.$$

2. If φ is a *small* user on heavily shared links ($s_{\varphi|\ell} \approx 0$), then $\beta_\varphi \approx 0$ and

$$\frac{\partial \ln Freq_\varphi}{\partial \ln M_\varphi} \approx 1.$$

This corollary gives us a worst-case scenario: if the elasticity is unity or above, we will only get costs by entry in these exclusively-transited links. This is not common as many roads are shared, so in general $\beta_\varphi < 1$, suggesting that the elasticity needs to be more than unity to get positive net costs with entry.

2.4 Residential land markets

I assume that each location has an exogenous land supply devoted to housing H_o, and Q_o is the housing price that equates demand for housing in each location to the housing supply:

$$Q_o = \frac{\alpha_h Y_o R_o}{H_o}, \quad \forall o \in J. \tag{22}$$

where $Y_o \equiv \sum_d \sum_r \lambda_{odr|o} w_d(\bar{T} - t_{odr})$ is the average income of workers ω living in o and $R_o = \sum_d \sum_r \bar{L}\lambda_{odr}$ is the mass of residents in a location, so equation (22) equates aggregate expenditure for housing in each location to the income generated in the housing rental market.

2.5 Final good and input (labor and commercial land) markets

Each location has a representative firm that produces a costlessly traded numeraire final good using a Cobb-Douglas technology that combines fundamental productivity, labor units L_d, and commercial floorspace H_d^c. The production function is given by:

$$y_d = A_d \Big(\frac{L_d}{\beta}\Big)^\beta \Big(\frac{H_d^c}{1-\beta}\Big)^{1-\beta}. \tag{23}$$

I assume that labor markets and commercial land markets are perfectly competitive, so the wage and commercial rent are pinned down by the (inverse) demand functions that stem from cost minimization of the representative firm. Wages in each location are given by

$$w_d = A_d \Big(\frac{\beta}{1-\beta} \frac{H_d^c}{L_d} \Big)^{1-\beta}, \tag{24}$$

and commercial rents are given by

$$Q_d^c = A_d \left(\frac{1-\beta}{\beta} \frac{L_d}{H_d^c} \right)^{\beta}. \tag{25}$$

2.6 Worker's land portfolio

I assume that each worker in the economy owns a share of the aggregate residential and commercial land revenue, and that is distributed proportionally to a worker's labor income $y_{odr} \equiv w_d(\bar{T} - t_{odr})$. Denote aggregate residential land revenue as

$$I^H \equiv \sum_o Q_o H_o,$$

and aggregate commercial land revenue as

$$I^C \equiv \sum_d Q_d^c H_d^c,$$

so that aggregate land income is, thus: $I = I^H + I^C$. Therefore, the total income of a worker consists of labor income and land portfolio:

$$y_{odr} \left(1 + \nu\right),$$

where

$$\nu = \frac{I}{\sum_{odr} y_{odr} \lambda_{odr} \bar{L}}.$$

Note that I assumed that land income is proportionally distributed to labor income such that spatial allocations are not affected, to a first order.[21]

2.7 Government funding to provide fare subsidies in public transit

Total public subsidy is

$$\mathcal{S} \equiv \sum_{\varphi_{\text{public}}} \left(p^*_{\varphi_{\text{public}}} - p_{\varphi_{\text{public}}}\right) D_{\varphi_{\text{public}}},$$

where p^* denotes the price in absence of subsidies, and p denotes the actual price. I assume that the government runs a balanced budget and that it levies a tax η on workers' total income. That is, the government collects a share η of workers' aggregate income to fund the subsidy

[21] The multiplicative term cancels in the numerator and denominator of the choice probabilities equations. However, GE adjustments through price-levels shifting will effectively modify spatial allocations.

such that tax revenue equals the subsidy amount:

$$\eta \sum_{odr} y_{odr}(1+\nu)\lambda_{odr}\bar{L} = \mathcal{S}.$$

Similar to the land portfolio, I assume that the tax is proportional to workers' income, such that spatial allocations are not affected, to a first order. Total disposable income of a worker is thus

$$\tilde{y}_{odr} = y_{odr}\left(1+\nu\right)\left(1-\eta\right),$$

and therefore $\Omega = \nu - \eta(1+\nu)$, presented in the commuter's budget constraint in Section 2.2.

2.8 General equilibrium

An equilibrium in this economy is a vector of prices $\left(w, Q, Q^c\right)_j$ across the J locations, a vector of transportation-market prices and times $\left(P, t^{\text{trip}}, t^{\text{wait}}\right)_\varphi$, an allocation of entrants $\{M\}_\varphi$, an allocation of residents and effective labor across locations $\{R, L\}_j$, a welfare constant $\bar{W}$, a land revenue and tax constants (ν, η) that, given fundamental location productivities A, amenities B, and residential H and commercial land supply H^c:

1. Residents maximize utility and residents' welfare is ex-ante equalized to $\bar{W}$:

$$\bar{W} \equiv \mathbb{E}[\max V_{odr}(\omega)] \approx \left(\sum_o \sum_d \left(\frac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}\tau_{od}}\right)^\theta\right)^{\frac{1}{\theta}}. \tag{26}$$

2. Transportation firms maximize profits, zero-profits hold, and demand of commuting trips equals supply of seats:

$$D_\varphi(\mathbf{M}) = M_\varphi^{\frac{\chi}{\chi-1}} q_\varphi^c n_{i,\varphi}.$$

3. Commuting gravity equations for flows hold:

$$\lambda_{odr} = \underbrace{\frac{\frac{B_{od}\tilde{w}_d^\theta}{Q_o^{\alpha_h\theta}\tau_{od}^\theta}}{\sum_{od}\frac{B_{od}\tilde{w}_d^\theta}{Q_o^{\alpha_h\theta}\tau_{od}^\theta}}}_{\lambda_{od}} \times \underbrace{\frac{\tau_{odr}^{-\rho}}{\sum_r \tau_{odr}^{-\rho}}}_{\lambda_{r|od}},$$

the number of residents in each location satisfies

$$R_o = \sum_d \sum_r \bar{L}\lambda_{odr},$$

and the effective labor units in each location satisfy

$$L_d = \sum_o \sum_r \bar{L}\lambda_{odr}(\bar{T} - t_{odr}).$$

4. Residential land markets clear:
$$Q_o = \frac{\alpha_h Y_o R_o}{H_o}.$$

5. Commercial land markets clear:
$$Q_d^c = A_d \left(\frac{1-\beta}{\beta} \frac{L_d}{H_d^c} \right)^{\beta}.$$

6. Labor markets clear:
$$w_d = A_d \Big(\frac{\beta}{1-\beta} \frac{L_d}{H_d^c} \Big)^{1-\beta}.$$

7. Government runs a balanced budget:
$$\eta \sum_{odr} y_{odr} (1+\nu) \lambda_{odr} \bar{L} = \sum_{\varphi_{\text{public}}} \left(p^*_{\varphi_{\text{public}}} - p_{\varphi_{\text{public}}} \right) D_{\varphi_{\text{public}}}.$$

8. Final good market clears, by courtesy of Walras's Law.

The algorithm to solve the general equilibrium in this model is described in Appendix B.

3 Quantification

I quantify the model focusing on the Metropolitan Area of Mexico City. We need two main sources of data. First, we need the transit network and the attributes of transit lines such as the number of operating units, frequency, length, and trip times. These data are often very hard to get in developing countries, as private and informal lines are hard to survey due to lack of resources. Notable examples of papers that collected their own data of the informal transit sector are Conwell (2024) in Cape Town, and Björkegren et al. (2025) in Lagos. A big part of the data effort in this paper was collecting this information for one of the largest metropolitan areas in the world, Mexico City. Second, in addition to the transit network, we need the more standard data sources such as granular wages, rents, and population shares and commuting flows. In this section I will first describe the setting and institutional context, then describe the collection of the network and calibration of network-related parameters, and then I discuss the calibration of demand and supply parameters, along with their data sources.

3.1 Commuting in the Metropolitan Area of Mexico City

The Metropolitan Area of Mexico City, home to more than 22 million people, combines a large scale mixed transit system, with substantial heterogeneity across neighborhoods in terms of access to transit and demographics. Mexico City shares many features of the large urban conglomerates around the world, e.g. throughout South America, Africa, South-East Asia. The metro area includes Mexico City, which has 16 municipalities, and 60 other municipalities from the neighboring states, the State of Mexico (59) and Hidalgo (1). While the center of

Mexico City enjoys a high living standard close to many developed-world cities, the outer districts are characterized by less educated and lower income inhabitants. While Mexico City has a subway system and mass rapid transit (MRTS) systems, public transportation in the State of Mexico comprises only one line of BRT and one subway line—representing 11 out of 192 subway stations—, even though it houses more than half of the overall population.

Daily mobility relies not only on an extensive public system—Metro, Metrobus, Tren Ligero, Cablebus, RTP buses and trolleybus—but also on a massive network of privately-operated *micros/combis/colectivos*. These small, flexible minibuses provide coverage in areas where mass transit is less accessible. Using detailed trip-level data from INEGI's Encuesta Origen Destino 2017 I document some motivating facts that describe how commuting is carried out in Mexico City. Table 1 shows a more detailed breakdown of the commuting shares by system.

1. **Transit vs Car.** Transit represents 62% of all trips, almost doubling the use of private car that represents 35% of trips. Further, 43% of households in sample report having a car.[22]
2. **Public vs Private Transit** 83% of all transit trips are done using private transit in some leg, dwarfing the 17% of public transit.
3. **Complementarity Through Multimodal Trips** 60% of trips that involve using private transit are multimodal.

TABLE 1—COMMUTING TRIPS IN SAMPLE OF 6.4M TRIPS ON A TYPICAL WEEKDAY IN 2017

Trips	Number	Share (%)
Total	60,559	100.00
Walk only	8,435	13.90
Wo/ walk only	52,124	86.10
Transit	32,128	61.60
Private	26,687	83.10
Unimodal	10,735	40.20
Multimodal	15,952	59.80
Public	5,441	16.90
Unimodal	4,154	76.30
Multimodal	1,287	23.70
Private car	18,434	35.40
Both	1,562	3.00

Note: Data comes form INEGI's Encuesta Origen Destino 2017, which is a representative survey that describes trips made by commuters, with detail information about the legs of each trip. Share corresponds to the immediate upper category, e.g. transit and private car add to 100%, and correspond 86.10% of non-walking-only trips. Multimodal includes multilegged trips in the case of private transit, so for example, a trip involving two minibuses would count as multimodal.

Most of the commuting gravitates from the outskirts towards the center. About half of the trips (3.1 million) begin inside Mexico City and the other half (3.4 million) begin outside Mexico City. However, 3.8 million trips end inside Mexico City, that is, there is a substantial net inflow

[22]This suggests that car vs transit substitution might be low.

commuting towards Mexico City. Figure 2 shows the delimitation of Mexico City and the geography of the districts along the geography of all public and private transit lines.

3.2 Institutional context

Market structure of the private sector. In Mexico City's metropolitan area, concessioned transit—known locally as combis, micros, peseros or colectivos—operates under a distinctive market structure. Most routes are controlled by route associations, but the day-to-day operation is carried out by individual drivers who typically do not own the vehicles they operate. Instead, a driver rents a minibus from an owner or association, and is required to pay a fixed daily quota (the so-called *cuota*). The driver collects fares in cash throughout the day, and once the quota has been covered, any remaining revenue becomes their income. This arrangement creates strong incentives to maximize passenger loads and complete as many trips as possible, shaping both the economic logic and the driving practices of the sector.

Fare regulation. The fares that drivers are allowed to charge are set by the state's transportation authority– the MA consists of mainly two states: State of Mexico and Mexico City. These concessioned microbuses rely exclusively on cash payment, so fare integration with the rest of the network is lacking. The fare consists in a base fare plus an additional per-km increase after some distance threshold. These two fare components are slightly different across state borders.[23] In practice, all trips are virtually charged the same amount, will little distance variation, except for the very long commutes. For calibration purposes, I introduce a version of the model that accounts for this regulation, as described below.

Public transit subsidies. Mexico City's government subsidizes all its transit systems[24], most notably the metro. The Metro's flat fare has been kept at 5 pesos per trip since 2013, with no increase in the number of legs or distance traveled. Unlike private microbuses, public systems use a unified electronic fare medium—the *Tarjeta de Movilidad Integrada*, so payment is integrated across modes even when operators and costs differ. Local authorities have stated that the price of the metro without the subsidy would be 18 pesos, so almost four times larger than it currently is. This amounts to a large subsidy budget: in 2024, the metro reported 1.17 billon trips, so a subsidy of 13 pesos per trip amounts to 15.21 billion pesos, or around 830 million US dollars. This represents roughly half of the total transportation budget and around 5% of the total local government's expenditure.[25]

[23] In the State of Mexico, the base fare in 2018 was 10 pesos and a 25 cent increase for each km after 5 km. In Mexico City, the base fare for the first 5 km was 5.50 pesos, for trips between 5 and 12 km, and 6.50 pesos beyond 12 km.

[24] Metro, Metrobús (BRT), Tren Ligero, Trolleybus, Buses RTP.

[25] From *Proyecto de Presupuesto de Egresos de la CDMX 2025*, transportation expenses are 38.7 billion pesos and total net expenditures are 291.5 billion pesos.

3.3 Transit network attributes

Mexico City's MA is one of the few developing megacities in the world that has detailed data on the private transit network.[26] Using the Google Maps Directions API, I simulated transit trips across all origin–destination pairs. From these simulations, I was able to recover the geography and attributes of roughly 80% of the transit routes in the system–around 2,000 in total. Table 2 provides some descriptive statistics. An interesting fact is that private lines are on average longer, faster and more frequent than public lines. Furthermore, in terms of coverage, the private network is 19 times larger and denser relative to the public one, as can be seen in Table 3. A striking fact is that 29% of all census tracts have at least 10 private lines going through them, while virtually no census tract has a comparable amount of public lines. This simple statistic shows the reach of private networks.

Figure 2 shows the overall resulting network. As it can be seen, the outskirts of the city rely heavily and almost exclusively on private providers. To account for own-district commuting, I simulated additional within-district trips to better capture local networks.[27] For each transit line, I obtained transit line names and GIS data, as well as key service characteristics such as route length, trip times, and frequency.[28] Importantly, I backed out the number of units operating each line M_φ using the definition of frequency.

Route choice sets. For every origin–destination pair, I assume that the relevant choice set of routes $\mathcal{R}_{od}$ corresponds to the alternatives provided by Google Maps and that for the purpose of solving the equilibrium, it says fixed. With this assumption, I depart from optimal routing frameworks (Allen and Arkolakis, 2022; Fuchs and Wong, 2024; Bordeu, 2023) that either explicitly enumerate all (infinite) paths Allen and Arkolakis (2022), or find the optimal routes and modes at each node recursively Fuchs and Wong (2024). There are two main reasons of why I depart from these frameworks: data limitations and tractability. First, to build the network from scratch and apply novel multimodal routing tools (Fuchs and Wong, 2024), I would need the source GTFS network information that describes stop nodes, switch nodes, and transfer rules–which is unavailable. Google's algorithm already encodes stop locations, complex transfer penalties, timed transfers, and minimum walk times–elements that are unobservable for me.[29] Using Google routes plausibly proxies the feasible space better than a hand-built graph with guessed transfers and guessed values of the elements that I mentioned. Furthermore, to stay tractable, Fuchs and Wong (2024) assume multiplicative iceberg edge-level costs. This assumption would make my CES-driver-level demand intractable, which is ultimately what allows me to solve the model efficiently. My assumption of (non-iceberg) additive route-level costs allows me to express residual demand for a driver purely as a function

[26] A firm, WhereIsMyTransport, created a census of all transit lines, both public and private, and recovered a GTFS dataset with all the information required to use with routing tools. Then, Google Maps had access to this dataset and so one can simulate transit trips with the complete network using the Google Maps.

[27] In this model, the commuting costs to do own-district commuting are not zero (or one if one thinks of iceberg commuting costs). Commuting costs are positive, so even if you travel within the district, you spend some time and money. Importantly, own-district flows should be allocated as demand to some market.

[28] I direct the interested reader to Appendix C.1 for more details about the collection of data and processing.

[29] These elements are also the elements that make multimodal transit routing different from road/rail routing.

TABLE 2—DESCRIPTIVE STATISTICS BY TYPE OF TRANSIT LINES

Variable	N	Mean	Min	Max	Median	SD
Private						
Length (km)	1658	28.0	2.1	135.9	22.2	19.8
Trip Time (min)	1658	82.4	7.3	294.8	75.1	43.2
Speed (km/h)	1658	20.0	7.5	95.3	17.7	8.2
Headway (min)	1573	8.6	1.0	30.0	8.0	4.4
Number of Vehicles/Units	1573	11.4	0.7	62.2	9.4	7.8
Public						
Length (km)	92	25.3	4.6	68.6	22.1	14.0
Trip Time (min)	92	97.7	12.7	271.5	85.5	53.5
Speed (km/h)	92	17.0	9.3	44.1	12.2	8.5
Headway (min)	91	9.7	2.0	43.0	6.0	9.1
Number of Vehicles/Units	91	15.8	1.5	71.3	13.2	12.1

Note: Private refers to all transit lines that are operated by private operators with public concessions (transporte público concesionado), which mostly includes minibuses and minivans. Public refers to all transit lines that are operated by a central public agency (e.g., subway, trains, metrobus, cablebus, trolleybus, and RTP buses). Data was collected from simulating trips in Google Maps. The number of vehicles or units is indirectly inferred from the definition of frequency (or inverse of headway), i.e. number of units divided by the time it takes to complete a trip or lap. Speed is calculated implicitly from length and trip time.

TABLE 3—TRANSIT NETWORK COVERAGE BY TYPE OF TRANSIT IN MEXICO CITY

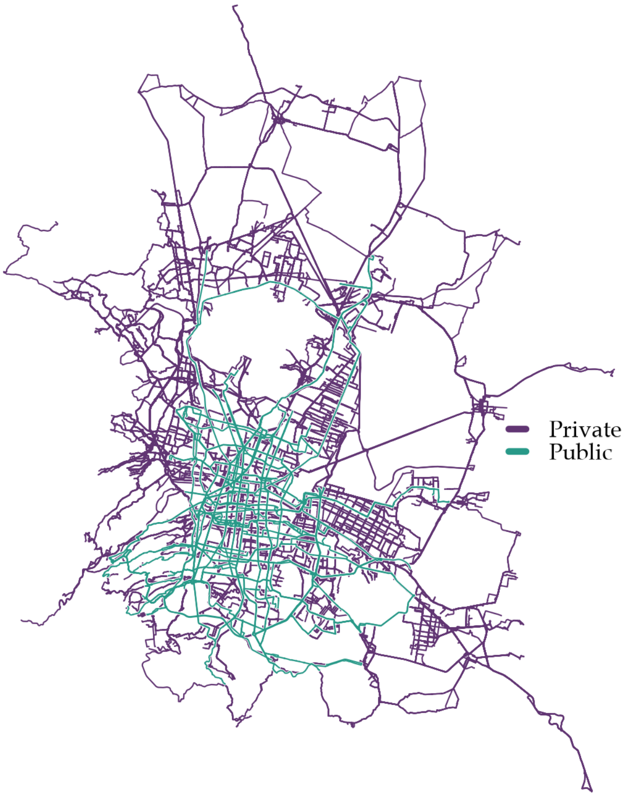
Metric	Private	Public
Total network length (km)	38307.45	2019.04
Network length density (total km length / total city area in km2)	16.07	0.85
Average number of lines passing through a census tract	9.05	0.59
Share of census tracts with at least 1 line	0.68	0.29
Share of census tracts with at least 5 lines	0.46	0.02
Share of census tracts with at least 10 lines	0.29	0.00

Note: Census tracts are INEGI's definition of AGEB. Total city area was calculated summing the area of the union of these census tracts. Private refers to all transit lines that are operated by private operators with concessions (transporte público concesionado), which mostly includes minibuses and minivans. Public refers to all transit lines that are operated by a central public agency (e.g., subway, trains, metrobus, cablebus, trolleybus, and RTP buses). Data was collected from Google Maps.

of its own price and the driver's own market price index, separable from other markets.[30] With multiplicative costs, I would not be able to collapse driver demand to a line-specific price contribution without carrying the entire path composition and prices of other legs/markets and therefore I would not be able to express entry and market-clearing in a system that allows fixed-point solutions. Although this assumption–that the choice set is fixed and corresponds to Google Maps alternatives–greatly simplifies the quantification, it has limitations. The main limitation is that substantial network changes (e.g. new infrastructure) could alter the route menu, and recomputing the full choice set or augmenting it with generated paths would be

[30]See Appendix A.3.

FIGURE 2. PUBLIC AND PRIVATE TRANSIT LINES IN MEXICO CITY M.A.



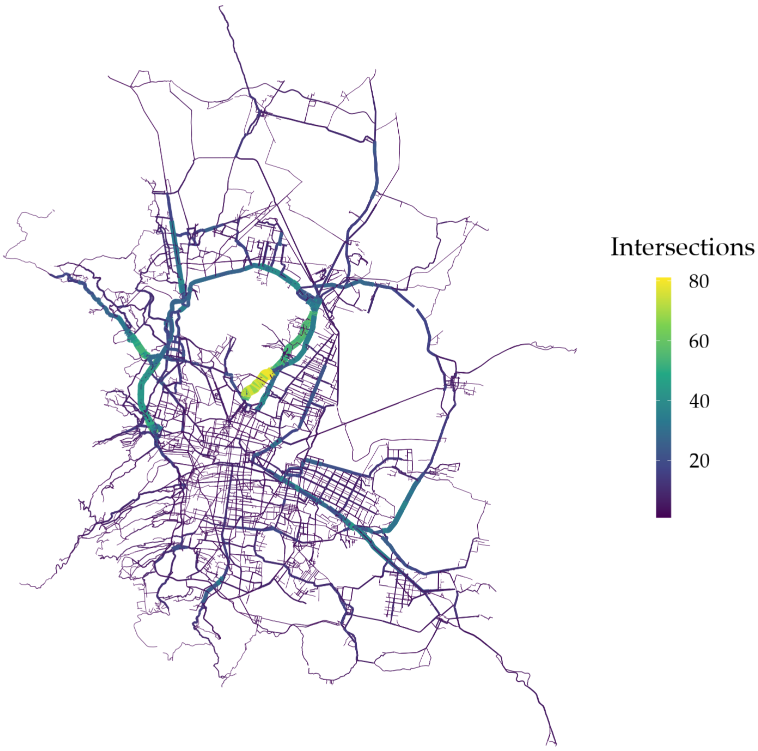
Note: Figure shows private and public transit lines in the metropolitan area of Mexico City. Transit line geographic data was collected from Google Maps.

necessary.

Road network. To account for congestion at the road segment level, I assume that the congestion in a given road segment is a function of the total number of buses that pass through that segment, with some elasticity. I spatially match all private transit lines to the OpenStreetMaps (OSM) road network, so I have a mapping of the transit lines that pass through each OSM road link. As Figure 3 shows, overlapping of private transit lines over a given OSM link can be quite substantial. For example, some OSM links carry over 80 different transit lines.

Trip time shifters and shares of market usage. From the information provided in the Google Maps API, for a given trip, I observe all the legs and lines used in the route, the travel time in each leg of the trip and the length of the leg. With this information I can compute, for each route, the share of market usage γ^{φ}_{odr}. So, for example, if a route alternative uses a market for only 25% of the full length of that market, then $\gamma^{\varphi}_{odr} = 0.25$. Then, having recovered the trip time ($t^{\text{trip}}_{\varphi}$) that each unit has to complete in a lap in a given market, and observing the number of entrants in each market, I can recover the trip time shifters $\bar{t}_{\varphi}$ that exactly match the observed time in that market, given the observed number of entrants.

FIGURE 3. INTERSECTIONS OF PRIVATE TRANSIT LINES AND OSM LINKS



Note: Figure shows all the OpenStreetMap road links in the economy, colored by the number of private bus lines that intersect each link. Line widths also represent number of intersections. The road network includes primary, secondary, and tertiary road links exclusively. Transit GIS data comes from Google Maps.

3.4 Data sources and calibration of demand and supply parameters

Geography, population, and commuting flows. I define a location to be the districts from the most recent origin-destination travel survey, Encuesta Origen Destino (EOD) 2017 collected by INEGI.[31] I set the size of the economy $\bar{L}$ to be the total number of workers that commute through transit. From EOD I also obtain the observed commuting flows at the *od* level, λ_{od}, that will serve to invert the *od*-specific amenities. Furthermore, from EOD, I obtain the total number of residents R and workers L in each location.

Wages, rents, and land shares. Observed wages come from the INEGI Economic Census 2019. I aggregate tract-level workplace wages to the EOD district level. Residential and commercial land shares come from the Urban Planning Ministries from Mexico City and Estado de México, SEIDUVI, and SEIDU. Residential rents come from the Housing Ministry, INFONAVIT, where I aggregate to the district level transaction-level home prices sold from 2018 to 2020.

Fundamentals inversion. To invert the model to obtain fundamental productivities A and

[31] Instituto Nacional de Estadística, Geografía e Información.

amenities B_{od}, I rely on the wage, rent, and population flows data described above. To obtain amenities, I rely on the commuting flow equation, and to obtain productivities, I rely on the inverse labor demand equation. A more detailed description can be found in Appendix C.2 and Appendix C.3.

Expenditure shares and time endowments. I calculate the share of expenditure devoted to housing $\alpha_h = 0.25$ and the share devoted to commuting $\alpha_c = 0.09$ via transit using data from INEGI's Encuesta Nacional de Gasto e Ingreso de los Hogares (ENIGH) 2018. Then, I set the time endowment of workers $\bar{T} = 14$ using sleep, work, commute, and other activities' times from Encuesta Nacional del Uso del Tiempo (ENUT 2019). Regarding the time endowment of firms/drivers, I do not have any data to discipline this parameter, so I set it to be $\bar{T}^d = 24$.

Elasticity of migration (location choice nest). I take the value of $\theta = 2$ from the literature, and follow Zárate (2024), who has a closely-related model for Mexico City. This value is similar to what Tsivanidis (2019) found in the Bogotá context.

Fixed costs of entry, capacities, and marginal costs per trip. For calibration of supply-side cost parameters, it will be important to have a version of the model where prices of individual drivers $p_{i,\varphi}$ cannot adjust, and instead are fixed exogenously to a uniform price $\bar{p}_\varphi$. This version of the model will correspond to the 'observed world', and cost parameters will be calibrated to replicate observed outcomes in this world – entry, frequency, trip time, and price. In Appendix A.5 I describe this model more in detail. The only difference that this version of model has relative to the model described before is that the price cannot adjust, so all adjustment happens through entry and trip time adjustment. To calibrate the fixed costs of entry f^e_φ, vehicle capacities q^c_φ, and the marginal cost per trip δ, I target the observed equilibrium outcomes–entry, frequency, trip times, and prices. I obtain firm-level demand from equilibrium route-level demand. Then, compute the number of trips a bus could perform in a day (based on trip times and drivers' time endowment) and compare it to the number of trips required to satisfy demand at a baseline capacity, 15 seats. This comparison identifies markets where demand exceeds market capacity at baseline capacity, in which case vehicle capacity is scaled up to ensure service can be delivered. Rescaling capacities, I obtain a vector that ranges from 15 to 446, with a mean of 29 seats. With demand, trip times, and adjusted capacities in hand, I then back out a value of $\delta = 36.4$ that ensures that no market with observed positive entry earns negative variable profits, and consequently, makes all entry costs non-negative. Finally, fixed costs of entry f^e_φ are calibrated as the residual profits that rationalize the zero-profit condition under free entry. This procedure jointly pins down entry costs, capacities, and marginal operating costs in a way that is consistent with observed prices, demand, and technological constraints on service provision.

(CES) Elasticity of substitution across firms. I model the demand of individual firms as CES with elasticity χ primarily for tractability and computation purposes. Having this smooth convexity in demand allows for the use of contraction mapping algorithms to find the large equilibrium ($\approx 1,400$) vector of entrants M. I do not have any data to discipline this parameter but, intuitively, we should expect it to be large if we believe that minibuses in a market are strongly substitutable. These minibuses are virtually homogeneous and there is little reason to assume that a commuter will wait more time for a *particular* minibus to come. We could argue that $\chi \to \infty$ as minibuses might split the market evenly in real life. For these reasons I set $\chi = 15$.[32]

Table 4 summarizes the parameter calibration.

TABLE 4—PARAMETER CALIBRATION AND DATA SOURCE/METHOD OF CALIBRATION

Parameter		Value	Source/Method of Calibration
J	Locations	192	EOD 2017 districts
$\bar{L}$	Workers using transit	3,608,702	EOD 2017
A	Productivities		Economic Census & Inversion
B	Amenities		INFONAVIT rents & Inversion
H	Residential land		SEIDUVI & SEIDU
H^c	Commercial land		SEIDUVI & SEIDU
$\bar{T}$	Time endowment	14 hours	Time use survey
α_h	Housing exp. share	0.245	ENIGH 2018
α_c	Commuting exp. share	0.085	ENIGH 2018
θ	Elasticity of migration	2	From literature
ρ	Elasticity of substitution across routes	7.4	Natural experiment
ϕ	Elasticity of congestion to M	0.77	Natural experiment
q^c_φ	Capacity constraint		Rescaled to match observed entrants and demand
Υ^φ_{odr}	Shares of market usage across routes		Google Maps
χ	Elasticity of substitution across firms (buses)	15	
δ	Marginal cost per time unit	36.4	Calibrated to get nonnegative fixed costs of entry
f^e_φ	Fixed costs of entry		To match observed p,M,t from Google Maps
$\bar{t}_\varphi$	Trip time shifters		To match observed travel times from Google Maps
$\bar{T}^d$	Time endowment of drivers	24 hours	

Note: The first block of parameters corresponds to the environment and location fundamentals. The second block shows preference (demand side) parameters. The third block shows transportation parameters (supply side).

3.5 Identification of elasticities with natural experiment

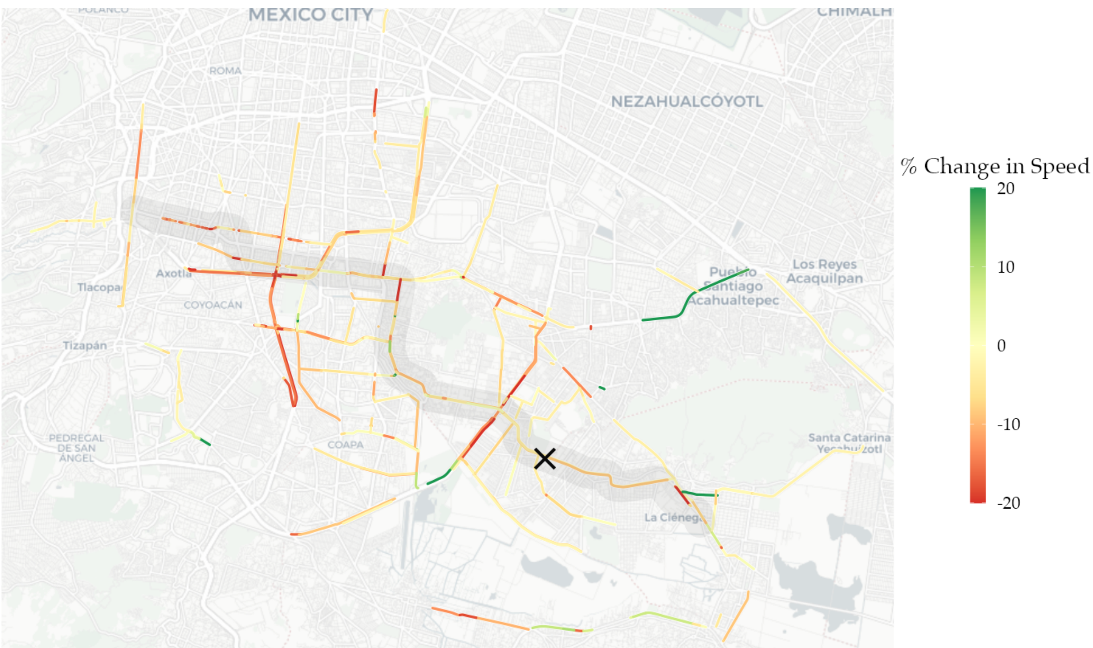
How can we learn about substitution and congestion elasticites without observing changes in flows? A main data challenge is to observe commute flows at the route level– we often observe flows exclusively at the *od* level. To overcome this challenge, instead of focusing directly on flows, we could focus on a mapping of flows. In particular, we can observe how traffic changes at the road level after a change in the cost of some routes. A second challenge, however, is to obtain plausibly exogenous variation in such costs. To overcome this second challenge, we can exploit exogenous variation generated by the collapse of a major subway line.

Identification argument using a subway shock. The objective is to identify the elasticity of substitution across routes ρ, and the elasticity of congestion to additional entrants ϕ, using only observable moments such as changes in speed in alternative routes. On May 3rd 2021, an elevated portion of subway line *Línea 12 Tláhuac-Mixcoac* collapsed.[33] The subway shock exogenously increased costs for some routes and it was plausibly a completely random event

[32]Results do not change quantitatively or qualitatively if we set, $\chi \in [10, 20]$.
[33]More details about this shock can be found in Appendix D.1

in terms of both timing and location. This shock forced affected users to divert towards alternative routes, spanning minibus lines, public buses lines, and an emergency BRT line that the government improvised. This shock potentially induced entry and relocation of minibuses to meet the demand, inducing congestion in some links relatively more than others. Therefore, we can learn about congestion from observing within-link speed variation, and learn about substitution from across-link speed variation.

FIGURE 4. TOMTOM ROAD-LEVEL SPEED DATA: CHANGES PRE/POST SUBWAY SHOCK



Note: Figure shows the spatial distribution of speed changes at the road-link level before and after the collapse of the subway line, for a sample of road links acquired from TomTom. The 'X' marked in black represents the site of the collapse. The subway line is depicted as the shaded grey buffer. Data was obtained from TomTom Traffic Stats API.

SMM exercise using traffic data. The goal is then, to simulate a subway shock in the model ($P_{\text{metro}} \to \infty$) and compute the equilibrium response in demand and, in turn, supply and trip times (speed). The demand response implies a relocation of commuters towards alternative routes, affecting service provision–entry and trip times–, and ultimately road-link-level speed. Then compare the speed moments generated by the model to the data,[34] and find the vector of parameters (ρ, ϕ) that best fits the data. Figure 4 shows the distribution of the changes in speed for a sample of road links near the subway line. Average speed decreased $\approx 6\%$.[35]

What moments are most informative about substitution of routes ρ vs congestion ϕ? To try to isolate variation that identifies congestion vs substitution, I focus on different sets of links. Consider an *od* pair with a set of routes $\mathcal{R}_{od}$. Suppose that some of those routes use the subway line in some leg. Now, shock the subway. All of those routes become impassable, so the set is reduced. Depending on the substitution strength ρ, the flow of commuters will distribute

[34] I obtained road-link-level speed data from TomTom Traffic Stats API in nearby street corridors to study the change in traffic patterns, pre and post shock. Full details on the data and processing, can be found in Appendix D.2

[35] The distribution of speed changes is shown in Appendix E figure 19.

among those remaining alternatives, and depending on the congestion strength ϕ, the speed in road links contained in those alternatives will respond accordingly. However, if we focus on links that are contained in *singleton* routes, i.e. $\{r \in \mathcal{R}'_{od} : |\mathcal{R}'_{od}| = 1\}$, where $\mathcal{R}'_{od}$ denotes the post-shock set, we can purge the variation coming from substitution of routes, given that by definition, there is no substitution in a single route. Therefore, I target the distribution of speed changes pre-post shock over the set of *singleton links* to identify ϕ, and the distribution of speed changes over the set of *non-singleton links* to identify ρ. In particular, I target the quantiles $(0.1, 0.25, 0.5, 0.75, 0.9)$ for both sets of links. Furthermore, I only consider links that are within a 3-kilometer buffer from the subway line, as the surrounding links are more likely to capture variation stemming exclusively from the collapse of the subway line. [36]

TABLE 5—IDENTIFIED ELASTICITIES AND MODEL FIT

	Target quantiles	Model	Target	Error (%)
(Singleton) Links to identify $\phi = 0.77$	0.10	-0.17	-0.15	0.10
	0.25	-0.08	-0.11	0.22
	0.50	-0.07	-0.07	0.01
	0.75	-0.04	-0.03	0.51
	0.90	-0.01	0.01	1.78
(Non-singleton) Links to identify $\rho = 7.40$	0.10	-0.12	-0.12	0.01
	0.25	-0.09	-0.08	0.22
	0.50	-0.02	-0.04	0.44
	0.75	0.00	-0.02	1.13
	0.90	0.05	0.00	73.16

Note: Table shows the fit of the model with the identified parameters to the target moments. These moments are quantiles of the distribution of speed changes across road links contained within a 3-km buffer of the collapsed subway line. To identify congestion, only links that belong to singleton routes after the shock were used. That is, routes that become singleton elements in OD route choice sets following the collapse of the subway. To identify substitution, links that belong to non-singleton routes were used.

Table 5 summarizes the identified elasticities and the model fit. I find values of 0.77 for the congestion parameter, and 7.4 for the substitution parameter. In terms of the magnitude of these parameters, Bordeu (2023) estimates the elasticity of congestion of cars in Chile to be 0.14, Allen and Arkolakis (2022) estimate 0.49 in the US roads context, Adler et al. (2020) estimate 0.16 for public transit buses in Rome, and Mosquera (2024) estimated 0.79 for medallion taxis in New York City. So, the value that I find is on the high end of what the literature has found, suggesting that bus-related congestion in Mexico City is substantial and similar to that of taxis in New York. In terms of the substitution across routes parameter, this is the first paper—to the best of my knowledge—to estimate it using quasi-experimental variation, in a transit setting. For context, Allen and Arkolakis (2022) estimate a value of 8 in the context of choosing routes when driving in US highways, so this suggests that Mexican commuters are about as sensitive.

Limitations of this identification approach. In this exercise, I attempted to identify the elas-

[36] Although I make robustness check for 1.5, 3, and 20km buffers, and parameters do not change substantially. Table 8 in the shows these results.

ticity of substitution and congestion in the transit context. To the extent that affected users switched to the use of private cars, I may be identifying a mix of bus-related and car congestion, and arguably substitution of transit towards private vehicles. If this case, the interpretation of the parameters would be different. While I cannot measure substitution towards private cars following the shock using existing data, Census 2020 (pre-shock data) reveals that car ownership in the Tláhuac area – which is where most of the demand of the line comes from – is low. The mean share of households that own at least one vehicle among census tracts in this area is 35%, which is lower to the western and south-western areas of the city.[37] This is suggestive evidence that car-related action is likely limited.

4 ENHANCING WELFARE THROUGH PRICE-SHIFTING POLICIES

What do we gain by making commuting costs τ an endogenous object? Recall that in standard models τ is exogenous, often enters directly in the utility function, and in virtually all cases is parametrized as a function of time between two locations. In frameworks that assume exogenous τ, we are unable to analyze policies that shift prices directly or that have general equilibrium effects through time and congestion by modifying entry incentives. In the framework proposed here, we can analyze how different channels and policies affect τ.

$$d\ln\tau_{odr} = \underbrace{\alpha_c \frac{\partial \ln P_{odr}}{\partial \ln M} d\ln M}_{\text{pricing / entry effect}} - \underbrace{\frac{\partial \ln(T - t_{odr})}{\partial \ln M} d\ln M}_{\text{congestion / frequency effect}} + \underbrace{\text{direct policy terms}}_{\text{fares, subsidies, etc.}}.$$

By inspecting the elements of τ, we can see that policies that shift prices can have direct impacts and indirect ones coming from entry or congestion effects. Therefore, in order to fully understand the effects of such policies on commuting costs and welfare, we need to take into account and understand these equilibrium adjustments.

Recall that welfare in this economy is primarily a function of commuting costs, wages, and rents:

$$\bar{W} = f(\tau, w, Q)$$

If the objective of a government is to enhance welfare by reducing τ, what policies achieve such objectives? Recent literature has studied the effects of policies that involve building new (public) infrastructure (Zárate, 2024; Tsivanidis, 2019; Björkegren et al., 2025), which improve market access by reducing commuting times. However, building infrastructure can be substantially fiscally costly, hard to plan, disruptive, and slow. In practice, many governments around the globe reach for easier-to-implement levers that directly affect prices, most notably price regulations to the private sector and fare subsidies.[38] These policies act on commuting costs τ through distinct margins and in different parts of the network.

[37] Figure 16 in Appendix D.3 shows the spatial heterogeneity.

[38] Bangladesh, Mexico, Philippines, Tanzania, South Africa, Argentina, Chile, Colombia, Brazil, Mexico, India, Indonesia. For a more detailed description of these examples, see Tables 9 and 10 in Appendix E.

Price regulation changes private operators' prices, alters entry incentives by changing profits, and as a result frequencies, wait times, and trip times change too. In the Mexico City setting in particular, price regulation is virtually uniform across space, and given substantial heterogeneity in transit lines costs and characteristics, this may create a suboptimal allocation of entrants across space. Subsidies to public transit reallocates riders across modes, indirectly affect entry incentives in public-transit-exposed routes, and can create or eliminate congestion in complementary or substitute private markets, respectively. Further, although easier to implement than building infrastructure, these policies still involve fiscal trade-offs: private fare regulation is virtually budget-free, while subsidies can take a significant amount of resources. How much do these policies matter quantitatively?

These policies cannot be studied using standard models, so I take advantage of the framework proposed here and analyze the effects of these policies taking as status quo a world where both regulation and subsidies are in place.

4.1 Counterfactual 1: let market forces determine transit prices

Consider the baseline environment in which the subway has a subsidized fare and private operators' prices are regulated. Then, allow private operators' prices to be determined in equilibrium. In such scenario, we can achieve an efficiency gain of $\approx$ 0.9%, driven mostly by a generalized decrease in commuting cost. To see this, let me decompose the change in the welfare expression given in equation 26 into the contribution of each of its elements as:

$$\Delta \ln \bar{W} \;\approx\; \underbrace{\sum_{o,d} \bar{\lambda}_{od}\, \Delta \ln \tilde{w}_d}_{\text{Disposable Income}} \;-\; \underbrace{\alpha_h \sum_{o,d} \bar{\lambda}_{od}\, \Delta \ln Q_o}_{\text{Rents}} \;-\; \underbrace{\sum_{o,d} \bar{\lambda}_{od}\, \Delta \ln \tau_{od}}_{\text{Commuting}} \,,$$

Table 6 shows the decomposition of welfare following price deregulation in the private sector. As we can see, mostly all of the change in welfare is coming from a change in commuting costs, which in turn are driven by around 80% by prices, and 20% by times. There is relatively little action coming from wage and rent adjustment, partly explained by the relatively low migration elasticity. This generalized improvement in commuting costs, however, masks a large heterogeneity across markets and space, as I explain next.

Effects across markets. Figure 5 shows such heterogeneity across markets, with a common pattern that explains the welfare gains. Prices mostly fall (though they rise in some markets), largely because costs decline as marginal trip costs—driven by trip times—fall. Trip times fall when congestion eases with fewer entrants. Despite fewer entrants, *capacity* (effective seats, i.e., entrants $\times$ trips per entrant) increases: faster trips allow each entrant to complete more trips per shift, meeting higher demand where it arises. Wait times mostly decline even when the number of entrants falls, because the "negative congestion" effect—shorter trip times raising trips per entrant—more than offsets the frequency loss from reduced entry.

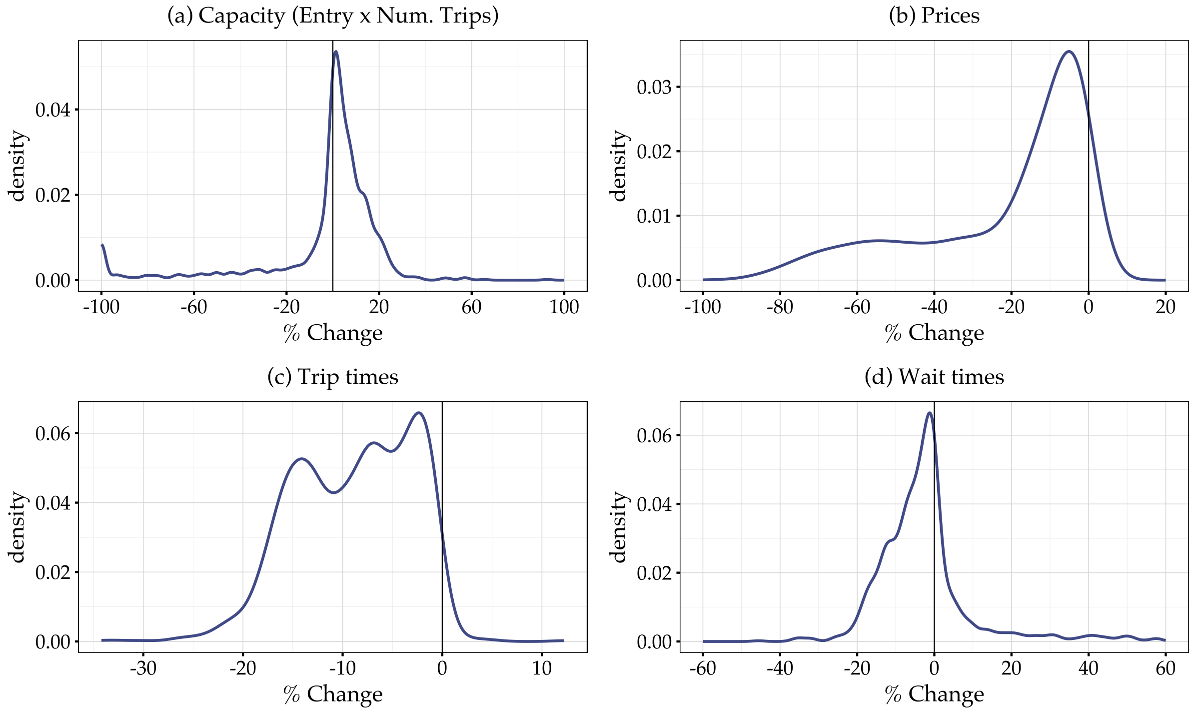
Which markets gain and lose customers? Figure 6 shows the change in flows across markets.

TABLE 6—WELFARE DECOMPOSITION: DEREGULATE PRICES

Component	Percent change
ΔW (%)	0.94
$\Delta\tau$ (%)	1.03
ΔP (%)	0.81
Δtrip time (%)	0.19
Δwait time (%)	0.02
$\Delta\tilde{w}$ (%)	-0.05
ΔQ (%)	-0.03

Note: Table shows the welfare decomposition across its main components, commuting costs (and sub-components), disposable income, and rents.

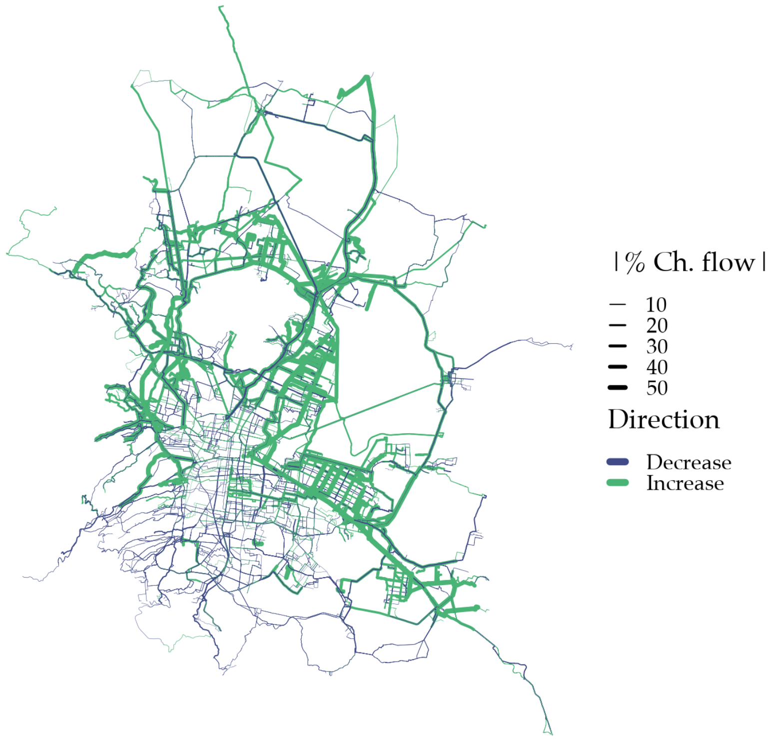
FIGURE 5. MARKET CHARACTERISTICS FOLLOWING PRICE DEREGULATION



Note: Figure shows the distribution of the changes of market characteristics following the policy counterfactual in the model. In panel (a), it shows the distribution of the change in overall 'seat' capacity in the market, which is composed of the total amount of entrants multiplied by the number of trips that each unit completes during their time endowment or shift. The mass at -100% signifies that there is exit, i.e. no entrants after the policy change. Panel (b) shows the changes in price indices, i.e. the CES aggregates of individual prices. Panel (c) shows the changes in the time it takes to complete a trip, which reflects mostly decreases due to less congestion. Panel (d) shows the distribution of changes in wait time. For visualization purposes, the range of these densities does not reflect the actual ranges of the changes, as I zoom in (and not trim) where most of the mass is located. Though there is virtually no mass on larger ranges. Full ranges are available upon request.

We can see that shorter and more local markets that connect outskirts districts are the ones that gain customers, particularly in the Eastern and South-Eastern part of the city. On the other hand, markets that are located in more central areas, and those that connect towards central areas lose customers. This is partly explained because their prices reflect relatively higher costs: there is a negative correlation between the change in flow (through the market) and length, and a positive correlation between the change in flow and distance to the CBD.[39]

FIGURE 6. CHANGE IN FLOWS ACROSS MARKETS FOLLOWING FARE DEREGULATION



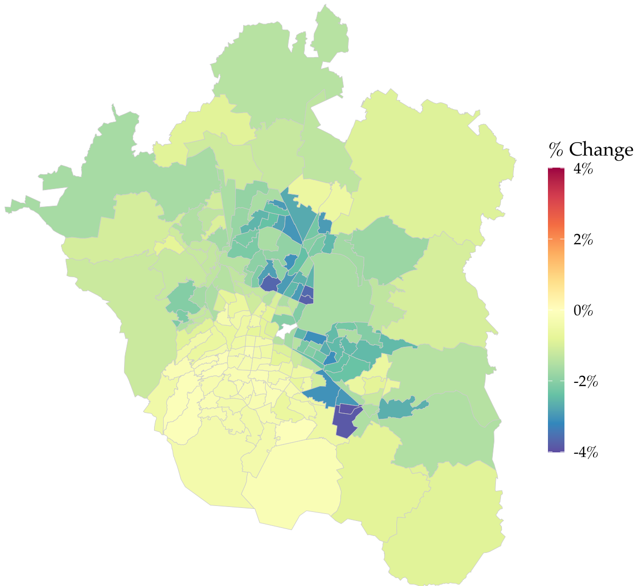
Note: Figure shows the change in flows at the market level (aggregating across all routes that use a given market), following the policy counterfactual in the model. The thickness of lines represents the absolute value in the percent change of the flow, and the color denotes whether it was a positive or negative change.

Commuting flows increase among outskirt districts and overall economic activity becomes more decentralized, as commuting becomes much more smooth for outer districts relative to central districts. Figure 7 shows the change in an origin-specific commuting cost index $\tau_o = \sum_d \lambda_{d|o} \tau_{od}$ that averages commuting cost indexes across destinations. South-Eastern areas gain up to four times more commuting access relative to central districts. These locations tend to be much less productive than central locations and with larger amenities, as reflected by the fundamental productivity A and a measure of local amenities $B_o = \sum_d B_{od}$. Figure 8 shows the correlation between these measures of productivity and amenities, and the corresponding change in wages and rents. There is a positive correlation between productivity and the change in wages and a negative correlation between amenities and the change in rents. This suggests on the one hand that less productive locations gained workers and wages decreased due to labor supply pressures, and on the other hand that rents increased as well in those locations due to housing demand pressure. Although the overall wage and rent effects on welfare are

[39]Scatter plot shown in additional figure 17 in Appendix E.

very limited.

FIGURE 7. COMMUTING COST INDEX CHANGE FOLLOWING DEREGULATION



Note: Figure shows the change in the commuting cost index by origin, weighted-averaged across destinations using conditional flow shares as weights.

FIGURE 8. FUNDAMENTALS AND PRICE CHANGES AFTER DEREGULATION

Note: Figure shows the correlation between location fundamentals and the change of local prices. Every dot is a location. Panel (a), shows that there is a positive relationship between local productivity and changes in wages. In panel (b), there is a negative relationship between local (average) amenities and changes in rents.

These results suggest that removing price regulations could improve efficiency in the economy mainly by improving service (wait and trip times), lowering congestion, and prices; and that the benefits of smoother commuting costs would be primarily realized by residents in the Eastern outskirts, characterized by lower productivity.

4.2 Counterfactual 2: remove the subsidy for subway fares

Consider the same baseline environment as before in which the subway has a subsidized fare and private operators' prices are regulated. Now, eliminate the 72% subway fare subsidy.[40] In terms of aggregate effects, there is an efficiency gain of $\approx 0.5\%$. Although this might seem a little counterintuitive at first, removing the subsidy frees up resources that increase disposable income of all residents in the economy.

Note that previous papers analyzing the effects of transit infrastructure do not consider a government budget constraint, so the funding of infrastructure does not come at any cost for residents, which may be relevant to consider a full GE response. The increase in commuting costs and rents—driven by an increase in housing demand—contribute negatively to welfare, but these effects are overturned by the increase in disposable income. This is shown in the first column of table 7, which shows the decomposition of welfare into its components. Also, we can further decompose the contribution of the change in commuting costs into price, wait and trip time. Most of the aggregate commuting cost adjustment comes from prices—which receive a direct impact of subsidy removal—, whereas times (trip and wait) account for around 6% of the total adjustment. These aggregate changes, however, hide substantial heterogeneous impacts across users and space.

TABLE 7—WELFARE DECOMPOSITION: SUBSIDY REMOVAL

Component	Percent change
ΔW (%)	0.49
$\Delta \tau$ (%)	-0.77
ΔP (%)	-0.72
Δtrip time (%)	-0.05
Δwait time (%)	0.00
$\Delta \tilde{w}$ (%)	1.67
ΔQ (%)	-0.40

Note: Table shows the welfare decomposition across its main components, commuting costs (and subcomponents), disposable income, and rents.

Effects across markets. The removal of the subway subsidy directly affects commuting costs via prices going up, and indirectly affects times through the endogenous adjustment of the private sector. The aggregate effects presented above mask substantial heterogeneity. Different private markets have heterogeneous exposure to the subway: some markets could act more like "last-mile" or "feeders", and some markets could be direct substitutes to the subway. For example, a market that frequently appears in routes that rely on the subway in some leg is exposed to it in a complementarity sense. On the other hand, a market that is frequently used in routes that serve as alternative to routes that use the subway is also exposed but in a substitutability sense. To better understand how different markets respond to these subway price changes, let me define the following exposure measures:

[40] I am focusing in the subsidy reported by authorities. They have stated that the true cost of the subway would be 18 pesos, while the actual price is 5 pesos. This yields an implied 72% subsidy.

- *Complementarity (metro-using) exposure* for market φ:

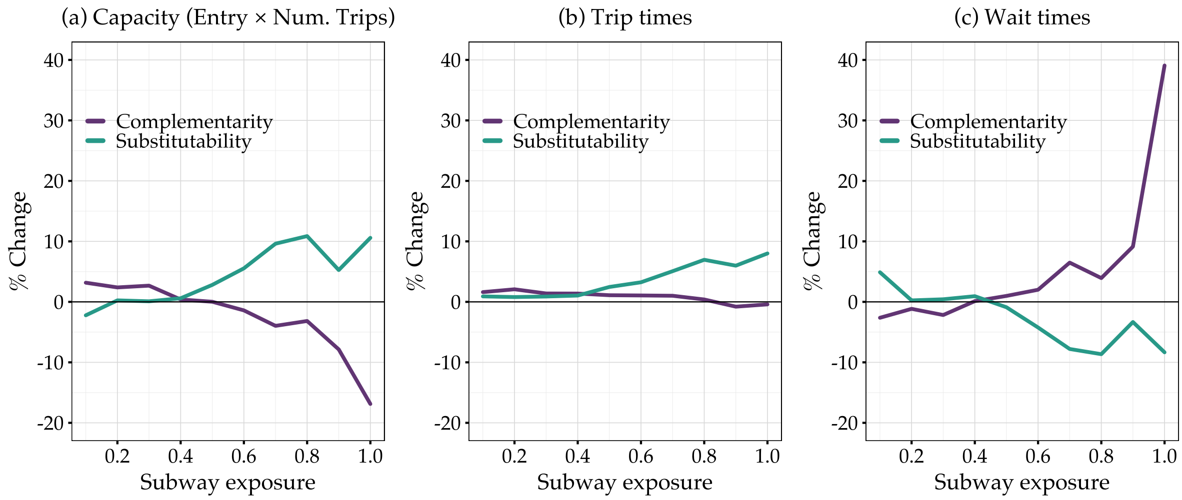
$$\text{Expo}_{\varphi}^{\text{comp}} = \frac{\sum_{(o,d)} \sum_{r \in \mathcal{R}_{od}} \mathbf{1}\{\varphi \in r\} \, \mathbf{1}\{M(r) = 1\}}{\sum_{(o,d)} \sum_{r \in \mathcal{R}_{od}} \mathbf{1}\{\varphi \in r\}}$$

- *Substitutability (alternative-to-metro) exposure* for market φ:

$$\text{Expo}_{\varphi}^{\text{subs}} = \frac{\sum_{(o,d)} \sum_{r \in \mathcal{R}_{od}} \mathbf{1}\{M_{od} = 1\} \, \mathbf{1}\{\varphi \in r\} \, \mathbf{1}\{M(r) = 0\}}{\sum_{(o,d)} \sum_{r \in \mathcal{R}_{od}} \mathbf{1}\{M_{od} = 1\} \, \mathbf{1}\{\varphi \in r\}}$$

Here $M(r) = 1$ if route r contains at least one metro segment (0 otherwise), and $M_{od} = 1$ if there exists at least one $r \in \mathcal{R}_{od}$ with $M(r) = 1$ (i.e., the od has a metro option). The complementarity measure is effectively a share: of all the routes that use some market φ, how many of these routes use the metro in some leg. For example, a market that is exclusively used in routes that connect to the subway would have a share of 1, and it would be a strong complement. Analogously, the substitutability measure is a share that captures how many routes that use a market φ are substitutes to metro-using routes, conditional on appearing as route choices within a given od-pair route choice set. For example, if the metro appears in half of all the routes in which some market φ is used, then the substitutability measure would be 0.5.

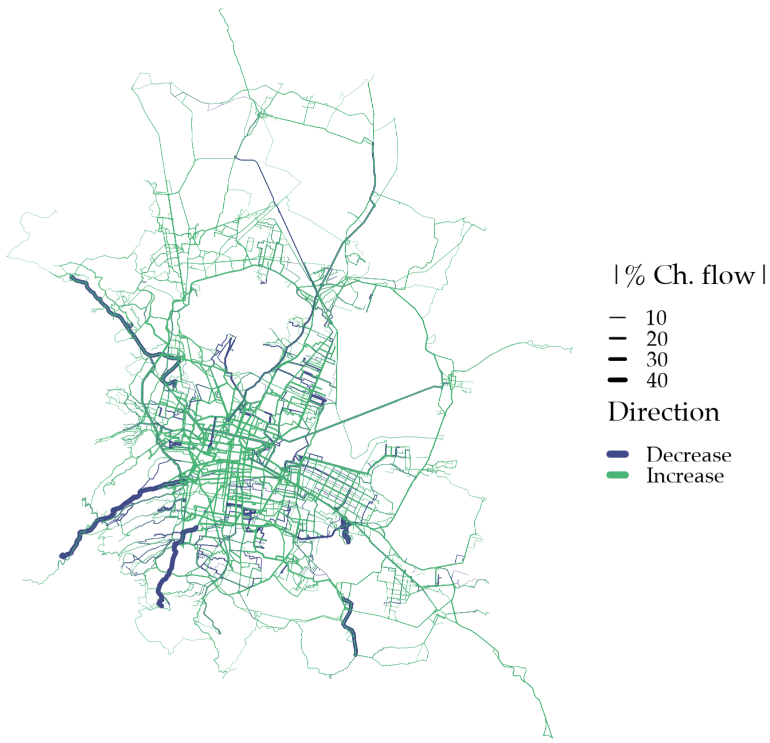
FIGURE 9. EFFECTS OF SUBSIDY REMOVAL BY METRO EXPOSURE



Note: Figure shows the average change in market characteristics across markets for each exposure decile, by type of exposure (complementarity or substitutability shares). The complementarity measure captures: of all the routes that use some market φ, how many of these routes use the metro in some leg. The substitutability measure captures: how many routes that use a market φ are substitutes to metro-using routes, conditional on appearing as route choices within a given od-pair route choice set.

Figure 9 shows the effects across private markets on capacity (entry multiplied by number of trips), trip times, and wait times, depending on their subway exposure. Removing the subsidy has a small effect on markets that are not exposed to the subway, i.e. with an exposure of around 0.4 or less. However, effects can be quite substantial for more exposed markets. Substitute markets experience a gain in customers due to relative prices changing and substitute towards alternative routes, so entry in those markets increases up to 10%, along with some congestion–an increase in trip times up to 7%. Frequencies improve as well in these markets, improving wait times by around 10% as well. Complementary markets, or subway 'feeders', experience a drop in demand and reduce capacity of up to around 15%. Trip times in these markets are virtually unchanged, most likely due to the increased presence of the other entrants that also share some road links. Most notably, wait times in very connected markets increase up to 35% as they become deserted and frequencies drop. Although in the aggregate the positive and negative effects of improved service in substitute and complementary markets are to some extent canceled out, these disaggregated findings suggest that users are very heterogeneously affected by this policy across space.

FIGURE 10. CHANGE IN FLOWS ACROSS MARKETS FOLLOWING SUBSIDY REMOVAL



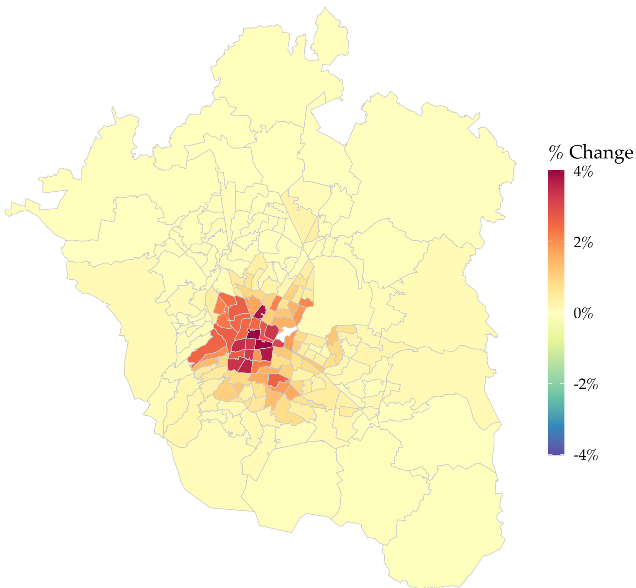
Note: Figure shows the change in flows at the market level (aggregating across all routes that use a given market), following the policy counterfactual in the model. The thickness of lines represents the absolute value in the percent change of the flow, and the color denotes whether it was a positive or negative change.

Given that the subway network is mostly located within central areas of Mexico City, all the substitution takes places in central districts. Figure 10 shows the change in flows across private lines. As it can be seen, most markets located in central areas are the ones that gain customers, reflecting route substitution away from routes that use the (now more expensive) subway.

Markets in the outskirts that connect towards the subway network mostly experience a drop in customers, which can be seen more salient in the South-Western areas.

Virtually all districts experience average increases in commuting costs across routes to all destinations, although with considerable heterogeneity. Figure 7 shows an origin-specific commuting cost index $\tau_o = \sum_d \lambda_{d|o} \tau_{od}$ that averages commuting cost indexes across destinations. Through this measure of commuting access shows we can see that even though central locations experience gains in service improvement due to customers switching to private routes, the overall increase in prices and potential congestion more than offset these improvements. As a result, these districts, which relied heavily on the metro are the most affected. Peripheral districts are also affected by worsened service in feeder lines and the price increase itself but the average increase in commuting costs is up to four times lower relative to central districts.

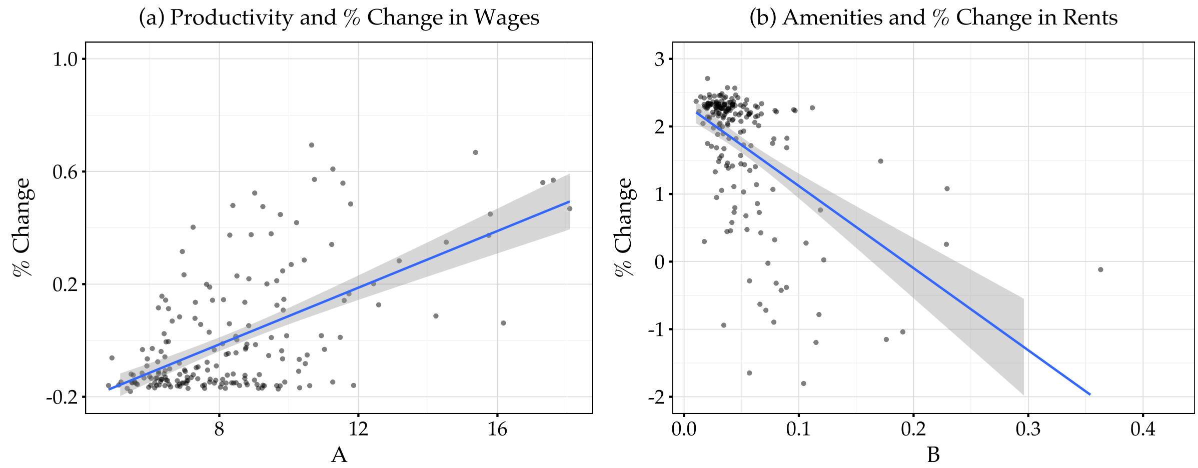
FIGURE 11. COMMUTING COST INDEX CHANGE BY ORIGIN: SUBSIDY REMOVAL



Note: Figure shows the change in the commuting cost index by origin, weighted-averaged across destinations using conditional flow shares as weights.

Economic activity becomes more decentralized as it is more costly to move within the center and towards the center (via metro). This result on the spatial distribution of economic activity is similar to the one from price deregulation, although for slightly different reasons. With price deregulation, service is improved in peripheral locations, making them more attractive. With the subsidy removal, commuting is more costly in central locations, making peripheral locations relatively more attractive. As a result, less productive (and peripheral) locations gain workers, and labor supply pressure drops wages, and residential demand in these locations drives rents up. This is shown in figure 12, where there is a positive relationship between fundamental productivity and wage changes, and a negative relationship between a measure of fundamental location-specific amenities $B_o = \sum_d B_{od}$ and rent changes.

FIGURE 12. FUNDAMENTALS AND PRICE CHANGES FOLLOWING SUBSIDY ELIMINATION



Note: Figure shows the correlation between location fundamentals and the change of local prices. Every dot is a location. Panel (a), shows that there is a positive relationship between local productivity and changes in wages. In panel (b), there is a negative relationship between local (average) amenities and changes in rents.

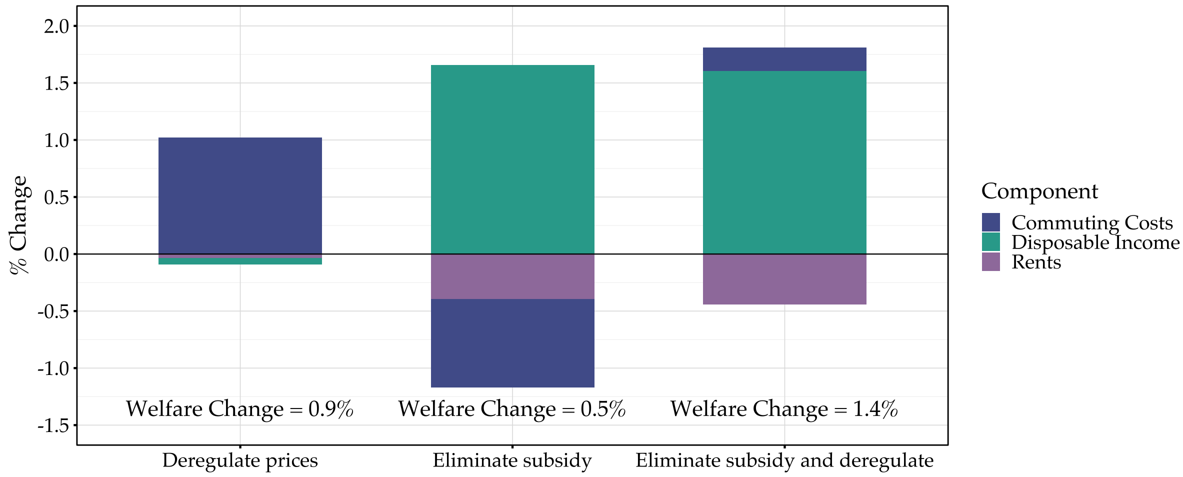
4.3 Counterfactual 3: remove all regulations and subsidies for transit prices

Finally, consider a counterfactual where we let prices to be determined in equilibrium, and the subway price reflects its true cost. This counterfactual represents a world where there are no price interventions in the economy. It is not a first-best because of congestion externalities, but it is a second-best economy in the sense that there are no other distortions introduced by the government. Figure 13 shows a comparison of the welfare change across (i) price deregulation alone, (ii) subsidy removal alone, and (iii) these two policies evaluated jointly.

By jointly letting the market set private prices and removing the subsidy, we can achieve a net welfare gain of $\approx$ 1.4%, as shown in the third column of figure 13. The welfare gains are substantially larger (roughly 1.5 to 3 times larger, respectively) when private sector is fully able to adjust through all margins, i.e. entry and prices, rather than the more limited re-optimization that takes place when price regulation remains in place. The main source of the extra gains comes from the adjustment of commuting costs: even after the large increase in metro prices, the rest of prices and service in the private sector adjust in such a way that overcomes the negative price effect.

This result implies that the welfare gains and resources saved from eliminating the subsidy alone would be amplified with the deployment of complementary policies that allow entry and prices to be allocated more efficiently across space—or policies that let at least partially reflect costs and market conditions. The design of optimal policy in this setting is left as an interesting avenue for future research.

FIGURE 13. WELFARE CHANGE DECOMPOSITION FOR DIFFERENT POLICIES



Note: Figure shows the welfare change decomposition across different policies into the main three elements of welfare: wages, rents, and commuting costs. The baseline environment consists of both price regulation, and subway fare subsidies. In the first policy, deregulation of prices, includes the subway subsidy, so the counterfactual is effectively ceteris paribus. The second policy, analogously, includes the price regulation. The third column thus presents a world where there are no regulations nor subsidies.

4.3.1 Discussion

In terms of the magnitudes reported, these are roughly comparable to those found by the literature studying infrastructure improvements in developing settings. Zárate (2024) documented a positive welfare gain of $\approx 0.6 - 0.8\%$ following the opening of a new subway line in Mexico City. Even though the model that he uses includes alternative margins such as labor reallocation from informal to formal jobs, which drives part of the gains, the model comes from the same class of models and is calibrated with essentially the same core data. He reports the range $\approx 0.6 - 0.8\%$ by turning on and off such additional margins. Tsivanidis (2019) found a positive gain of $\approx 0.6 - 2.3\%$ following the opening of a new BRT system in Bogotá; 0.6 if migration from outer Colombia is allowed, and 2.3 if is not. His model also comes from the same class of models although with slight variations in the assumptions on migration decisions. Therefore, even if the magnitudes across models are not completely directly comparable, they serve as a practical benchmark.

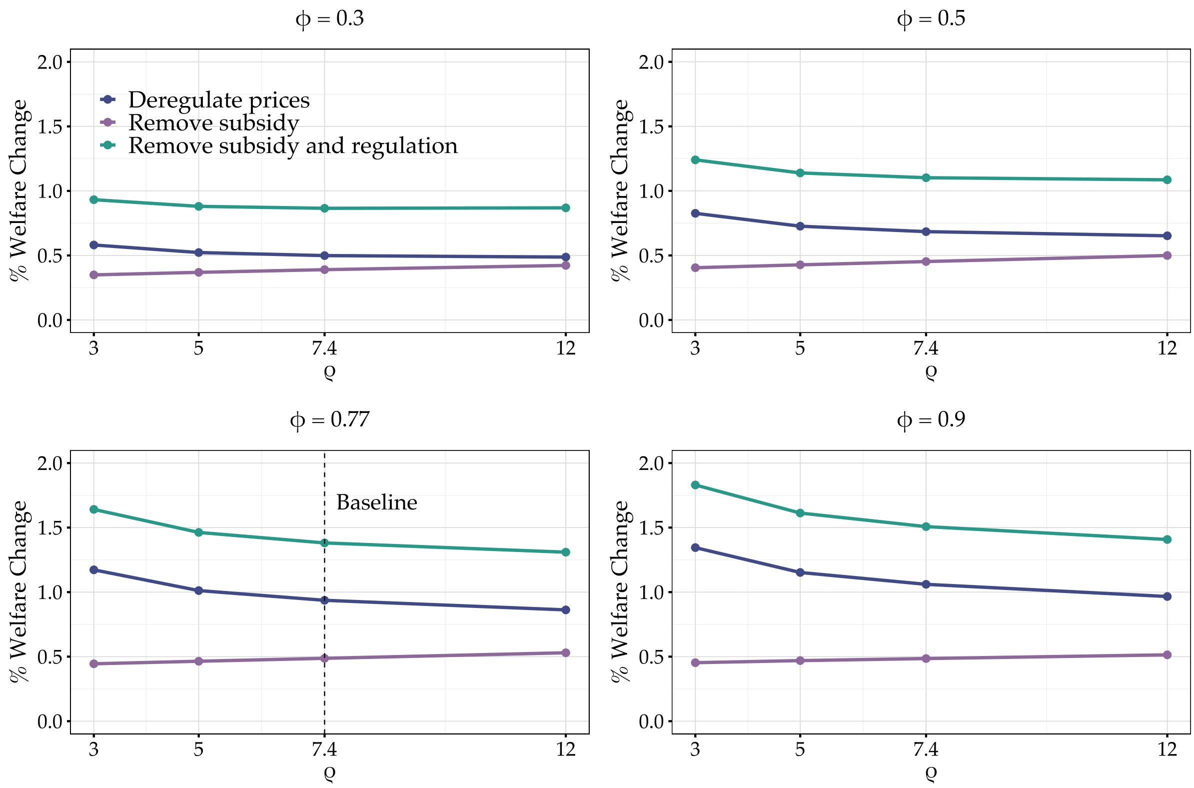
An important remark is that the analysis of the policies presented here would be unfeasible with the models in the literature: even if the supply side of transportation is completely omitted, these models abstract away from prices (even exogenously) and their corresponding income/budget effects. Furthermore, the funding of government interventions, e.g. building a subway line, is not considered within the model and this could under or over-state overall welfare effects. In particular, here it was shown that in fact a large amount of resources could be liberated from subsidy removal, and potentially allocated to welfare-enhancing uses, including an infrastructure improvement itself. Although the model presented here is rudimentary in the sense that the potential alternative uses of those resources are limited—and resources are in fact reimbursed in integrity as a zero income tax—, the quantitative exercise reveals that the decisions of how to fund transit interventions could flip the sign of welfare evaluations.

Overall, I show that non-infrastructure policies that shift prices and directly or indirectly affect private transit providers can generate substantial changes in welfare, comparable to those that the literature studying infrastructure improvements has found.

5 Results under alternative parametrizations

In this section I explore how sensitive are the welfare quantifications from the baseline parametrization. In particular, I explore lower and higher elasticities of both substitution and congestion. Figure 14 shows how welfare changes as we increase the elasticity of congestion (from left to right panel) and vary the elasticity of substitution in the horizontal axis.

Figure 14. Welfare change under alternative parametrizations



Note: Each panel holds fixed a given value of the elasticity of congestion ϕ. In each panel, I vary the elasticity of substitution across routes ρ in the horizontal axis. Each dot corresponds to the welfare change of a model with a given combination of these two parameters. The dashed line shows the baseline welfare change under the identified parameters.

First, welfare has some notable variation for the deregulation of prices policy, across different parameter configurations. If we compare scenarios with low (0.3) and high congestion (0.9), the welfare changes range from 0.5% to around 1.4%, which is more substantive relative to the other policy. But what economic mechanisms explain this variation? Note that welfare gains are much more amplified in cases where congestion is large. This is mainly because the reallocation of entry, particularly less entry (or exit) in previously congested corridors, leads to a much more substantive improvement in trip times. The elasticity of substitution across routes also plays an important role in settings with high congestion (lower panels), but virtually plays no role in settings with low congestion. With high congestion elasticity, note that large responses of agents to cost changes, e.g. with $\rho = 12$, imply smaller welfare gains. The intuition is that as commuters switch to alternative routes following small changes in commuting cost (time or price), they all try to get on the best route, congestion kicks in and this could backfire on welfare. When agents are not very responsive, on the other hand, demand is effectively spread more evenly across routes so congestion does not backfire on welfare significantly.

Regarding the elimination of the subway subsidy, the qualitative and quantitative welfare changes stay practically the same. Only in the case where congestion is very low (first panel), what matters the most for the magnitude of the welfare change is the elasticity of substitution. Comparing low (3) and high (12) values, we can see that welfare change ranges from -0.4 to around -0.5, which is not substantial in absolute terms. Overall, the welfare effects of removing both the subsidy and regulation follows roughly the same pattern as removing the regulation: when congestion forces are large (i.e. the lower panels), the variation of the welfare change coming from variation in the substitution elasticity can be as large as 0.5 percentage points. Given that we could expect the value of ρ to fall within 7 and 10—given robustness exercises explained in Section 3.5—, the overall welfare qualitative and quantitative change should remain stable.

6 CONCLUSION

This paper develops a quantitative spatial framework in which commuting costs are endogenously determined by the joint equilibrium of private minibus entry, prices, frequencies, and congestion, interacting with a public mass-transit network that is given. Two mechanisms emerge as central. First, because we allow for the possibility of multimodal trips, there is a within-route complementarity: a cost change in one leg (e.g., a metro fare increase) alters demand on connected private legs, amplifying local interventions across the network depending on the elasticity of substitution across routes. Second, because private transit operates on shared roads, there is a frequency–congestion trade-off: more entry raises frequency and reduces waits but slows trips on those links, with effects that depend on the elasticity of congestion and on how intensively markets overlap in space.

Exploiting the sudden collapse of a subway line that generated exogenous speed variation at the road-link-level, I identify both a high elasticity of route substitution and a high congestion elasticity. These key parameters, together with new data on the characteristics of the near-universe of public and private lines in Mexico City, discipline the full model.

Policy counterfactuals highlight that non-infrastructure, price-shifting levers can yield material welfare changes through general-equilibrium reallocation of flows, entry, waits, and trip times. In particular, these effects are comparable to those found by the literature when studying new public transit openings. First, deregulating private fares increases welfare by about 0.9%, largely by easing congestion and realigning prices to heterogeneous route costs, and generating service improvements mainly in local and peripheral markets. Second, eliminating a metro fare subsidy increases welfare by roughly 0.5%: the negative commuting cost effects of price increases and congestion in central areas due to riders reallocating toward private alternatives is out-weighted by increases in disposable income due to the elimination of taxes to fund the subsidy. Lastly, removing all regulations and subsidies increases welfare by 1.4% while saving fiscal resources, underscoring that network interactions and private supply responses effectively shape impacts.

The broader lesson from this analysis is that urban transit policy in mixed systems must take into account interactions at the route and network level, through modes, fares, and times. Because route legs are complements within multimodal trips and transit operators share roads, interventions can propagate across seemingly distant parts of the city. Endogeneizing commuting costs makes these spillovers visible and quantifiable. Future work could analyze what is the optimal allocation of entry across space, what policies could implement such an allocation, and the optimal design of public and private networks.

REFERENCES

ADLER, M. W., F. LIBERINI, A. RUSSO, AND J. N. V. OMMEREN (2020): "The congestion relief benefit of public transit: evidence from Rome," *Journal of Economic Geography*, 21, 397–431.

AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): "The Economics of Density: Evidence from the Berlin Wall," *Quarterly Journal of Economics*, 130, 1237–1281.

ALLEN, T. AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129, 1085–1140.

——— (2022): "The Welfare Effects of Transportation Infrastructure Improvements," *Review of Economic Studies*, 89, 2911–2957.

ALLEN, T., D. ATKIN, S. CANTILLO CLEVES, AND C. E. HERNÁNDEZ (2023): "The Traveling Trucker Problem," NBER Working Paper 31961, National Bureau of Economic Research, also in *AEA Papers and Proceedings* 114 (2024): 334–339, doi:10.1257/pandp.20241044.

ALMAGRO, M., A. BARBIERI, J. C. CASTILLO, A. HICKOK, AND T. SALZ (2024): "Optimal Urban Transportation Policy: Evidence from Chicago," Working paper.

BJÖRKEGREN, D., M. DUHAUT, A. NAGPAL, AND N. TSIVANIDIS (2025): "Public and Private Transit: Evidence from Lagos," NBER Working Paper 33899, National Bureau of Economic Research, working paper.

BORDEU, O. (2023): "Commuting Infrastructure in Fragmented Cities," Working paper, working paper.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2020): "Geography, Transportation, and Endogenous Trade Costs," *Econometrica*, 88, 657–691.

Bryan, G. T., K. Frye, and M. Morten (2025): "Spatial Economics for Low- and Middle-Income Countries," NBER Working Paper 33453, National Bureau of Economic Research, Cambridge, MA.

Conwell, L. (2024): "Privatized Provision of Public Transit," Job Market Paper; working paper.

Fuchs, S. and W. F. Wong (2024): "Multimodal Transport Networks," Working paper, CESifo Working Paper No. 11362; also Atlanta Fed Working Paper 2022-13a (rev. 2024), working paper.

Khanna, G., C. Medina, A. Nyshadham, J. Ramos, J. Tamayo, and Z. H. Tiew (2024): "Spatial Mobility, Economic Opportunity, and Crime," Working paper; add link when available.

Kreindler, G., A. Gaduh, T. Graff, R. Hanna, and B. A. Olken (2023): "Title TBD (Public Transit in Jakarta)," Working paper; fill in when finalized.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018): "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 108, 3855–3890.

Mosquera, R. (2024): "Stuck in traffic: Measuring congestion externalities with negative supply shocks," *Regional Science and Urban Economics*, 108, 104013.

Tsivanidis, N. (2019): "Evaluating the Aggregate and Distributional Impacts of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," Working paper; various revisions through 2023.

Zárate, R. D. (2024): "Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City," Working paper.

Appendix

Appendix Contents

A Model derivations

A.1 Commuting flows

This section derives the route-level choice probabilities under the nested Fréchet structure in (4)–(6) and the CES commuting cost index (8). Define the deterministic component of utility for option (o,d,r) as:

$$U_{odr} = a_{odr}\,\varepsilon_{odr}, \qquad a_{odr} \equiv \frac{B_{od}\,\tilde{w}_d}{Q_o^{\alpha_h}\,\tau_{odr}}, \qquad \tau_{odr} = \frac{P_{odr}^{\alpha_c}}{\bar{T} - t_{odr}}.$$

Workers draw multiplicative shocks ε_{odr} with joint CDF (nested Fréchet),

$$F(\vec{\varepsilon}) = \exp\left(-\sum_{o,d} \left(\sum_{r \in \mathcal{R}_{od}} \varepsilon_{odr}^{-\rho} \right)^{\theta/\rho} \right), \qquad \theta < \rho,$$

as in (6).

Fix an origin–destination pair (o,d) and consider routes $r \in \mathcal{R}_{od}$. Conditional on the realization of $\varepsilon_{odr} = \varepsilon$, the event that route r beats every other route $r' \neq r$ is

$$\{U_{odr'} \leq U_{odr} \;\; \forall r' \neq r\} \quad \Longleftrightarrow \quad \{\varepsilon_{odr'} \leq (a_{odr}/a_{odr'})\,\varepsilon \;\; \forall r' \neq r\}.$$

With i.i.d. Fréchet(ρ) shocks within (o,d), having CDF $F(x) = \exp(-x^{-\rho})$ and pdf $f(x) = \rho\, x^{-(1+\rho)} \exp(-x^{-\rho})$, the conditional probability that r is best within (o,d) equals

$$\Pr(r \mid od) = \int_0^\infty f(\varepsilon) \prod_{r' \neq r} F\left(\tfrac{a_{odr}}{a_{odr'}}\,\varepsilon \right) d\varepsilon.$$

Compute the integral using the change of variable $z = \varepsilon^{-\rho}$ (so $dz = -\rho\,\varepsilon^{-(1+\rho)}\,d\varepsilon$):

$$\begin{aligned}
\Pr(r \mid od) &= \int_0^\infty \rho\,\varepsilon^{-(1+\rho)} \exp\left(-\varepsilon^{-\rho} \right) \exp\left(-\varepsilon^{-\rho} \sum_{r' \neq r} (a_{odr'}/a_{odr})^{\rho} \right) d\varepsilon \\
&= \int_0^\infty \exp\left(-z\left[1 + \sum_{r' \neq r} (a_{odr'}/a_{odr})^{\rho}\right] \right) dz \; = \; \frac{1}{\sum_{r'} (a_{odr'}/a_{odr})^{\rho}} \; = \; \frac{a_{odr}^{\rho}}{\sum_{r'} a_{odr'}^{\rho}}.
\end{aligned}$$

Substituting $a_{odr} \propto \tau_{odr}^{-1}$ inside (o,d) gives the familiar CES share

$$\lambda_{r|od} = \frac{\tau_{odr}^{-\rho}}{\sum_{r'} \tau_{odr'}^{-\rho}} \quad . \tag{27}$$

Let the *within-pair maximum* be $Y_{od} \equiv \max_{r \in \mathcal{R}_{od}} U_{odr} = \max_r a_{odr}\varepsilon_{odr}$. For any threshold $x > 0$,

$$\{Y_{od} \leq x\} \iff \{\varepsilon_{odr} \leq x/a_{odr} \;\; \forall r\}.$$

Taking the marginal of the nested Fréchet CDF (6) for a single pair (o,d) (i.e., setting the thresholds for all other pairs to $+\infty$) yields

$$\Pr(Y_{od} \le x) = \exp\left(-\left(\sum_r (x/a_{odr})^{-\rho} \right)^{\theta/\rho} \right) = \exp\left(-x^{-\theta} \left(\sum_r a_{odr}^{\rho} \right)^{\theta/\rho} \right).$$

Hence Y_{od} is Fréchet(θ) with *scale*

$$A_{od} \ \equiv \ \left(\sum_r a_{odr}^{\rho} \right)^{\theta/\rho}.$$

Moreover, the family $\{Y_{od}\}_{od}$ is independent, since for any $\{x_{od}\}$,

$$\Pr(Y_{od} \le x_{od} \ \forall od) = \exp\left(-\sum_{od} x_{od}^{-\theta} A_{od} \right) = \prod_{od} \exp\left(-x_{od}^{-\theta} A_{od} \right).$$

Therefore the probability that pair (o, d) delivers the overall maximum across all pairs is

$$\Pr(od) = \int_0^{\infty} g_{od}(y) \prod_{(o',d') \neq (o,d)} G_{o'd'}(y)\, dy = \frac{A_{od}}{\sum_{o',d'} A_{o'd'}},$$

where $G_{od}(y) = \exp(-A_{od} y^{-\theta})$ and $g_{od}(y) = \theta A_{od} y^{-(1+\theta)} \exp(-A_{od} y^{-\theta})$. Using $a_{odr} = (B_{od}\tilde{w}_d)/(Q_o^{\alpha_h} \tau_{odr})$,

$$A_{od} = \left(\left(\tfrac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}}\right)^{\rho} \sum_r \tau_{odr}^{-\rho} \right)^{\theta/\rho} = \left(\tfrac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}} \right)^{\theta} \tau_{od}^{-\theta}, \quad \tau_{od} \equiv \left(\sum_r \tau_{odr}^{-\rho} \right)^{-1/\rho}.$$

Hence

$$\lambda_{od} = \frac{\left(\dfrac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}\ \tau_{od}} \right)^{\theta}}{\sum_{o',d'} \left(\dfrac{B_{o'd'}\tilde{w}_{d'}}{Q_{o'}^{\alpha_h}\ \tau_{o'd'}} \right)^{\theta}} \quad . \tag{28}$$

By the law of total probability and conditional independence, combining (27) and (28) gives

$$\lambda_{odr} = \lambda_{od} \times \lambda_{r|od} = \underbrace{\frac{\left(\dfrac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}\ \tau_{od}} \right)^{\theta}}{\sum_{o',d'} \left(\dfrac{B_{o'd'}\tilde{w}_{d'}}{Q_{o'}^{\alpha_h}\ \tau_{o'd'}} \right)^{\theta}}}_{\lambda_{od}} \times \underbrace{\frac{\tau_{odr}^{-\rho}}{\sum_{r'} \tau_{odr'}^{-\rho}}}_{\lambda_{r|od}} \quad . \tag{29}$$

A.2 Proof of complementarity proposition

Proof. Take logs and differentiate $\lambda_{r|od} = \tau_{odr}^{-\rho} / \sum_k \tau_{odk}^{-\rho}$ to obtain:

$$\frac{\partial \ln \lambda_{r|od}}{\partial x} = -\rho \frac{\partial \ln \tau_{odr}}{\partial x} + \rho \sum_{k \in \mathcal{R}_{od}} \lambda_{k|od} \frac{\partial \ln \tau_{odk}}{\partial x}.$$

Given expression:

$$\lambda_{r|od} = \frac{\tau_{odr}^{-\rho}}{\sum_k \tau_{odk}^{-\rho}}$$

Taking the natural log:

$$\ln \lambda_{r|od} = \ln(\tau_{odr}^{-\rho}) - \ln\left(\sum_k \tau_{odk}^{-\rho}\right)$$

$$\ln \lambda_{r|od} = -\rho \ln \tau_{odr} - \ln\left(\sum_k \tau_{odk}^{-\rho}\right)$$

Differentiating with respect to x:

$$\frac{\partial \ln \lambda_{r|od}}{\partial x} = -\rho \frac{\partial \ln \tau_{odr}}{\partial x} - \frac{\partial \ln\left(\sum_k \tau_{odk}^{-\rho}\right)}{\partial x}$$

For the second term, using the chain rule:

$$\begin{aligned}
\frac{\partial \ln\left(\sum_k \tau_{odk}^{-\rho}\right)}{\partial x} &= \frac{1}{\sum_k \tau_{odk}^{-\rho}} \cdot \sum_k \frac{\partial \tau_{odk}^{-\rho}}{\partial x} \\
&= \frac{1}{\sum_k \tau_{odk}^{-\rho}} \cdot \sum_k \left(-\rho \tau_{odk}^{-\rho-1} \frac{\partial \tau_{odk}}{\partial x}\right) \\
&= \frac{1}{\sum_k \tau_{odk}^{-\rho}} \cdot \sum_k \left(-\rho \tau_{odk}^{-\rho} \frac{1}{\tau_{odk}} \frac{\partial \tau_{odk}}{\partial x}\right) \\
&= -\rho \sum_k \frac{\tau_{odk}^{-\rho}}{\sum_j \tau_{odj}^{-\rho}} \cdot \frac{\partial \ln \tau_{odk}}{\partial x} \\
&= -\rho \sum_k \lambda_{k|od} \frac{\partial \ln \tau_{odk}}{\partial x}
\end{aligned}$$

With t_{odr} fixed, $\ln \tau_{odr} = \alpha_c \ln P_{odr} - \ln(T - t_{odr})$ so

$$\partial \ln \tau_{odr} / \partial \ln P_\varphi = \alpha_c \, \partial \ln P_{odr} / \partial \ln P_\varphi = \alpha_c \, s_{\varphi|r}$$

Substitute back to obtain (*):

$$\frac{\partial \ln \lambda_{r|od}}{\partial \ln P_\varphi} = -\rho \, \alpha_c \, s_{\varphi|r} \ + \ \rho \sum_k \lambda_{k|od} \, \alpha_c \, s_{\varphi|k} = \rho \, \alpha_c \big(\bar{s}_\varphi - s_{\varphi|r}\big)$$

For the aggregate response of $\Lambda_\varphi = \sum_{r \in \mathcal{R}_\varphi} \lambda_{r|od}$,

$$\frac{\partial \Lambda_\varphi}{\partial \ln P_\varphi} = \sum_{r \in \mathcal{R}_\varphi} \lambda_{r|od} \frac{\partial \ln \lambda_{r|od}}{\partial \ln P_\varphi} = \rho \, \alpha_c \Big(\bar{s}_\varphi \sum_{r \in \mathcal{R}_\varphi} \lambda_{r|od} - \sum_{r \in \mathcal{R}_\varphi} \lambda_{r|od} s_{\varphi|r}\Big).$$

Since $s_{\varphi|r} = 0$ for $r \notin \mathcal{R}_\varphi$, the last sum equals $\sum_k \lambda_{k|od} s_{\varphi|k} = \bar{s}_\varphi$. Therefore

$$\frac{\partial \Lambda_\varphi}{\partial \ln P_\varphi} = \rho\, \alpha_c\, \bar{s}_\varphi\, (\Lambda_\varphi - 1) \le 0,$$

because $0 \le \Lambda_\varphi \le 1$ and $\bar{s}_\varphi \ge 0$.

□

Corollary 2 (Two-route illustration). Suppose $\mathcal{R}_{od} = \{1, 2\}$ with route 1 using only φ ($P_{od,1} = P_\varphi \Rightarrow s_{\varphi|1} = 1$) and route 2 using φ and φ' ($s_{\varphi|2} = P_\varphi/(P_\varphi + P_{\varphi'}) \in (0, 1)$). Then

$$\frac{\partial \ln \lambda_{1|od}}{\partial \ln P_\varphi} = -\rho\alpha_c \Big(1 - [\lambda_{1|od} \cdot 1 + \lambda_{2|od} \cdot s_{\varphi|2}]\Big) < 0,$$

$$\frac{\partial \ln \lambda_{2|od}}{\partial \ln P_\varphi} = -\rho\alpha_c \Big(s_{\varphi|2} - [\lambda_{1|od} \cdot 1 + \lambda_{2|od} \cdot s_{\varphi|2}]\Big) \;\ge\; 0,$$

while the total φ-using mass $\Lambda_\varphi = \lambda_{1|od} + \lambda_{2|od}$ falls when P_φ rises. Hence, a subsidy to φ raises the aggregate flow on all routes that use φ and, within those, expands routes where φ is a larger component—capturing complementarity.

A.3 Bus-specific CES demand

Due to Cobb-Douglas preferences, worker ω spends a proportion α_c of his income in a CES bundle of commuting trips, so total commuting expenditure is $P_{odr} C_{odr}(\omega) = \alpha_c y_{odr}$. This bundle of trips is composed of the c_φ trips along the different markets φ that compose a route *odr* so that

$$C_{odr}(\omega) = \left(\sum_{\varphi \in odr} c_\varphi^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

Given that two markets φ within a route are **perfect complements**, then, this is a CES bundle with elasticity of substitution $\sigma = 0$. This implies that we have a Leontief bundle instead, so

$$C_{odr}(\omega) = \min(c_1, ..., c_\varphi, ...) \;\text{ as }\; \sigma \to 0$$

with an associated price index $P_{odr} = \sum_{\varphi \in odr} P_\varphi$. Optimal consumption in a Leontief bundle implies that $c_1 = ... = c_\varphi$, and since aggregate commuting expenditure is $P_{odr} C_{odr}(\omega) = \alpha_c y_{odr}$, then individual Leontief demand is given by

$$c_\varphi(\omega) = \frac{\alpha_c y_{odr}}{P_{odr}}$$

where this comes from the fact that the budget constraint is $\sum_{\varphi \in odr} P_\varphi c_\varphi = \alpha_c y_{odr}$ and $P_{odr} = \sum_{\varphi \in odr} P_\varphi$. Also, note that expenditure in trips of a given market is $P_\varphi c_\varphi(\omega) = \frac{P_\varphi}{\sum_\varphi P_\varphi} \alpha_c y_{odr}$
Now, for each market φ along the route, worker consumes a CES bundle of trips in individual

buses within that market with elasticity of substitution χ, so that

$$c_\varphi(\omega) = \left(\sum_{i\in\varphi} q_i^{\frac{\chi-1}{\chi}}\right)^{\frac{\chi}{\chi-1}}$$

With this setup we can derive a worker-level demand for individual bus driver i:

$$\begin{aligned} q_{i,\varphi}(\omega) &= p_{i,\varphi}^{-\chi} P_\varphi^{\chi-1} P_\varphi c_\varphi(\omega) \\ &= \left(\frac{p_{i,\varphi}}{P_\varphi}\right)^{-\chi} \frac{\alpha_c y_{odr}}{P_{odr}} \end{aligned}$$

where $P_\varphi \equiv \left(\sum_{i\in\varphi} p_{i,\varphi}^{1-\chi}\right)^{\frac{1}{1-\chi}}$ is the price index of market φ. To get total demand for driver i, we must aggregate across all workers that choose a route that involves using market φ. That is, we aggregate across all origins, destinations, and routes, that pass through market φ in some segment

$$\begin{aligned} q_{i,\varphi} &= \sum_{od} \sum_{r|\varphi\in r} \int_{\omega|odr} q_{i,\varphi}(\nu)\, d\nu \\ &= \sum_{od} \sum_{r|\varphi\in r} \int_{\omega|odr} \left(\frac{p_{i,\varphi}}{P_\varphi}\right)^{-\chi} \frac{\alpha_c y_{odr}}{P_{odr}}\, d\nu \\ &= \left(\frac{p_{i,\varphi}}{P_\varphi}\right)^{-\chi} \sum_{od} \sum_{r|\varphi\in r} \frac{\alpha_c y_{odr}}{P_{odr}} \int_{\omega|odr} 1\, d\nu \\ &= \left(\frac{p_{i,\varphi}}{P_\varphi}\right)^{-\chi} \underbrace{\sum_{od} \sum_{r|\varphi\in r} \frac{\alpha_c y_{odr}}{P_{odr}} \lambda_{odr} \bar{L}}_{\equiv D_\varphi} \end{aligned}$$

The last term in the bracket is the total demand for trips in the market, D_φ.

A.4 Proof of frequency-congestion trade-off proposition

Proof. Write $T_\psi(M) \equiv \sum_{\ell\in\psi} S_\ell(M)^\phi$ so $t_\psi^{\text{trip}} = \bar{t}_\psi T_\psi$ and $Freq_\psi = M_\psi/(\bar{t}_\psi T_\psi)$. Then

$$\frac{\partial \ln Freq_\psi}{\partial \ln M_\varphi} = \underbrace{\mathbf{1}\{\psi = \varphi\}}_{\text{numerator}} - \underbrace{\frac{\partial \ln T_\psi}{\partial \ln M_\varphi}}_{\text{congestion}}.$$

For the congestion term,

$$\frac{\partial T_\psi}{\partial \ln M_\varphi} = \sum_{\ell\in\psi} \phi\, S_\ell^{\phi-1} \frac{\partial S_\ell}{\partial \ln M_\varphi} = \sum_{\ell\in\psi} \phi\, S_\ell^{\phi-1}\, \mathbf{1}\{\ell \in \varphi\}\, M_\varphi = \phi \sum_{\ell\in\psi} S_\ell^\phi \frac{M_\varphi}{S_\ell} \mathbf{1}\{\ell \in \varphi\}.$$

Divide by $T_\psi = \sum_{\ell\in\psi} S_\ell^\phi$ to get

$$\frac{\partial \ln T_\psi}{\partial \ln M_\varphi} = \phi \sum_{\ell\in\psi} \frac{S_\ell^\phi}{\sum_{j\in\psi} S_j^\phi} \cdot \frac{M_\varphi}{S_\ell} = \phi \sum_{\ell\in\psi} w_{\ell|\psi}\, s_{\varphi|\ell}.$$

Therefore,

$$\frac{\partial \ln Freq_\varphi}{\partial \ln M_\varphi} = 1 - \phi \sum_{\ell\in\varphi} w_{\ell|\varphi}\, s_{\varphi|\ell} \quad \text{and} \quad \frac{\partial \ln Freq_\psi}{\partial \ln M_\varphi} = -\,\phi \sum_{\ell\in\psi} w_{\ell|\psi}\, s_{\varphi|\ell} \ \ (\psi \neq \varphi),$$

which are the stated expressions. Since $w_{\ell|\psi}$ and $s_{\varphi|\ell}$ lie in $[0,1]$ and the weights sum to one, each $\beta \in [0,1]$. □

A.5 Model with price regulation

Let the government decide that all buses must charge $\bar{p}_\varphi = p_{i,\varphi}, \forall i \in \varphi, \forall \varphi$. Let $n_\varphi^{\max} \equiv \frac{\bar{T}^d}{t_\varphi^{\text{trip}}}$ denote the maximum number of trips possible within the time endowment. Then the profit maximization problem for a given bus i in a given market φ is

$$\max_{n_{i,\varphi}} \pi_{i,\varphi} = \bar{p}_\varphi\, q_\varphi^c n_{i,\varphi} - \delta t_\varphi^{\text{trip}} n_{i,\varphi} - f_\varphi^e$$

$$\begin{aligned} \text{s.t.} \qquad & q_{i,\varphi} = M_\varphi^{\frac{\chi}{1-\chi}} D_\varphi \\ & \frac{q_{i,\varphi}}{q^c} \leq n_{i,\varphi} \\ & n_{i,\varphi} t_\varphi^{\text{trip}} \leq \bar{T}^d \end{aligned}$$

The price index is given by

$$P_\varphi = M_\varphi^{1/(1-\chi)} \bar{p}_\varphi \tag{30}$$

Profits under price regulation are

$$\pi_{i,\varphi} = \left(\bar{p}_\varphi\, q_\varphi^c - \delta\, t_\varphi^{\text{trip}}\right) n_{i,\varphi} - f_\varphi^e \tag{31}$$

where $n_{i,\varphi} = \min\left[\frac{\bar{T}^d}{t_\varphi^{\text{trip}}}, \frac{M_\varphi^{\frac{\chi}{1-\chi}} D_\varphi}{q_\varphi^c}\right]$, reflecting the fact that once a firm is at capacity, the maximum revenue they can get is utilizing all their available capacity. Entry M_φ is pinned-down by free entry, making $\pi_{i,\varphi} = 0$. Denote the margin per passenger as $m_\varphi^{\text{pax}} = \bar{p}_\varphi - \frac{\delta\, t_\varphi^{\text{trip}}}{q_\varphi^c}$ and the maximum capacity of seats in a given shift $S_\varphi = q_\varphi^c n_\varphi^{\max}$. Zero profits requires that the quantity of seats served is q_φ^{zp} such that:

$$m_\varphi^{\text{pax}} q_\varphi^{zp} = f_\varphi^e \iff q_\varphi^{zp} = \frac{f_\varphi^e}{m_\varphi^{\text{pax}}}$$

To achieve zero profits and allow the market to operate at or below capacity, we can target the per-bus quantity

$$q_{\varphi}^{tar} = \min\{q_{\varphi}^{zp}, S_{\varphi}\},$$

and given individual CES demand, the M that delivers such target is

$$M_{\varphi} = \left(\frac{D_{\varphi}}{q_{\varphi}^{tar}}\right)^{\frac{\chi-1}{\chi}}$$

B Algorithm to solve general equilibrium

Given a vector of parameters $\vec{x} = (\vec{A}, \vec{B}, \vec{H}, \vec{H^c}, \bar{L}, \bar{T}, \bar{T}^d, \alpha_h, \alpha_c, \beta, \theta, \rho, \phi, q_{\varphi}^c, \delta, \chi, \vec{t_{\varphi}}, \vec{f_{\varphi}^e})$

1. Guess initial distribution of people λ^0 and entrants M^0
2. Begin outer loop to solve contraction mapping in λ

   - **Compute economy prices and demand for transportation**. Compute all the elements necessary to calculate demand at the market level.

     – Compute trip times and wait times

     $$t_{\varphi}^{\text{trip}}(\vec{M}) = \bar{t}_{\varphi} \sum_{\ell} \left(\sum_{\varphi : \ell \in \varphi} M_{\varphi} \right)^{\phi}, \quad t_{\varphi}^{\text{wait}}(\vec{M}) = \frac{1}{2} \frac{t_{\varphi}^{\text{trip}}(\vec{M})}{M_{\varphi}}$$

     – (Market prices case) Compute price indices implied by zero-profits equations

     $$P_{\varphi} = M_{\varphi}^{-\frac{1}{\chi-1}} \left(\delta_{\varphi} + \frac{f_{\varphi}^e}{n_{i,\varphi}}\right) \frac{1}{q^c}$$

     – (Fixed prices case) Compute price indices implied definition of price index

     $$P_{\varphi} = M_{\varphi}^{-\frac{1}{\chi-1}} \bar{p}_{\varphi}$$

     – Compute commuting costs

     $$P_{odr} = \sum_{\varphi \in odr} P_{\varphi}, \quad t_{odr} = \sum_{\varphi \in odr} t_{\varphi}^{\text{wait}} + \gamma_{odr}^{\varphi} t_{\varphi}^{\text{trip}}$$

     – Compute commercial rents

     $$Q_d^c = A_d \left(\frac{1-\beta}{\beta} \frac{L_d}{H_d^c}\right)^{\beta}$$

     – Compute wages

     $$w_d = A_d \Big(\frac{\beta}{1-\beta} \frac{H_d^c}{L_d}\Big)^{1-\beta}$$

– Compute residential rents

$$Q_o = \frac{\alpha_H \sum_d \sum_r \lambda_{odr|o} w_d(\bar{T} - t_{odr}) R_o}{H_o}$$

– Compute demand

$$D_\varphi(\vec{M}) = \sum_{od} \sum_{r|\varphi \in r} \frac{\alpha_c w_d(\bar{T} - t_{odr})}{P_{odr}} \lambda_{odr} \bar{L}$$

- **Tranportation market**. Given demand, solve for updated vector of entrants M.
  – (Market prices case) Compute updated M implied by the market clearing equation

$$D_\varphi(M) = M_\varphi^{\frac{\chi}{\chi-1}} q_\varphi^c n_{i,\varphi}$$

  – (Fixed prices case) Compute updated M implied by zero profit equations

$$\pi_{i,\varphi} = \begin{cases} \bar{p}\, M^{\frac{\chi}{1-\chi}}\, D_\varphi - \delta \bar{T}^d - f_\varphi^e & \text{if } \frac{q_{i,\varphi}}{q_c} \leq n_\varphi^{\max} \\ \bar{p}\, n_\varphi^{\max}\, q_c - \delta \bar{T}^d - f_\varphi^e & \text{if } \frac{q_{i,\varphi}}{q_c} > n_\varphi^{\max} \end{cases}$$

- Update commuting flows λ with new factor prices, rents and commuting costs

$$\lambda'_{odr} = \frac{\frac{B_o w_d^\theta}{Q_o^{\alpha_h \theta} \tau_{od}^\theta}}{\sum_{od} \frac{B_o w_d^\theta}{Q_o^{\alpha_h \theta} \tau_{od}^\theta}} \times \frac{\tau_{odr}^{-\rho}}{\sum_r \tau_{odr}^{-\rho}} \qquad \text{, where} \qquad \tau_{od} \equiv \left(\sum_{r \in \mathcal{R}_{od}} \underbrace{\left(\frac{P_{odr}^{\alpha_c}}{\bar{T} - t_{odr}} \right)}_{\equiv \tau_{odr}}{}^{-\rho} \right)^{-\frac{1}{\rho}}$$

- Iterate until $|\lambda' - \lambda| < \text{tol}$

C Data and calibration

C.1 Transit network data collection

I first defined origins and destinations at the district level. Using INEGI's Encuesta Origen Destino 2017 district geographies, for a total of 192 districts out of 194, excluding the airport and one municipality in Hidalgo state that had no land use data. I merged census tract (AGEB) polygons with their 2020 census population totals and intersected AGEB centroids with EOD districts to select, for each district, the single most populated urban AGEB as the district's location. For each OD, I queried multi-alternative public-transit directions from the Google Maps Directions API at a fixed peak time. I parsed the responses to obtain alternative-level distance and duration, step-level GIS polylines, and extracted transit metadata—agency, line/short name, vehicle type, headway, number of stops—as well as any reported fares.

With the step-level "segments" in hand, I rebuilt the network by line. For each (agency, line_name), I gathered all segments observed across all ODs, de-duplicated their geometries, and applied a spatial union to recover a route line per transit line. I classified each line as *Private* when the agency matched *Sistema de Transporte Público Concesionado* (or *Corredores Concesionados*), and *Public* otherwise, enabling a direct map-based comparison of private minibus corridors and formal public modes. I then constructed provisional line attributes by collapsing segments to the line level: round-trip distance and in-vehicle time as twice the maximum segment distance or duration observed for that line (a conservative full-run proxy), and line headway as the mean reported headway across appearances. I corrected under-measured lines using a separate set of own-district trips (short shuttles that more fully reveal route extents): whenever these yielded larger implied round-trip distance or time, I replaced the provisional values. I assigned a unique `line_id`, computed $M_{\text{obs}} = t_{\text{trip}}/\text{headway}$, linked attributes back to each OD–route–segment, and calculated segment weights γ as the share of a line's round-trip distance accounted for by that segment. The resulting objects are: (i) an `sf` network of recovered line geometries with public/private tags, (ii) a line-level table with length, trip time, headway, and M_{obs}, and (iii) an OD–route–segment file with `line_id`, segment time, headway, and γ, ready for the model.

To capture the *local* network actually used within neighborhoods, I treated urban AGEB centroids as candidate origins and destinations, restricted to the same metro area, and matched each AGEB to an EOD district via point-in-polygon. For each district I targeted $\sim$50 unique within-district OD pairs (about 5% of all possible pairs). For every OD I requested multi-alternative public-transit routes (fixed peak time), parsed step-level segments, decoded polylines, and extracted the same transit metadata. I stacked all segments across trips, attached the originating AGEBs and district IDs, converted times to minutes, and exported a tidy routes dataset with segment-level travel times. Because these short trips do not traverse full lines, I intentionally did not infer line-level attributes from this sample; the goal was to recover the *within-district* network footprint and costs.

I then cleaned and harmonized the full dataset in four passes. First, I aggregated walking segments in a route to a single 'walking time' leg, removed airport OD pairs, and dropped entire OD–routes whose lines lacked frequency information. Second, I trimmed negligible lines by counting usages and discarding those used only a handful of times (threshold ≤ 3). Third, I incorporated within-district routes as $o = d$, discarded lines observed only once in that local sample, kept only lines present in the inter-district attributes (walking allowed), merged canonical lengths/headways, computed $\gamma = \text{distance}_m/\text{length}$, averaged repeated appearances by $(o, d, \text{agency}, \text{line})$ to form one segment per line, attached `line_id/headway`, and appended these to the main set. Finally, I deleted routes that were only walking. The result is a tidy, deduplicated `routes` table aligned with an updated `line_attributes` containing only viable, observed lines. This `routes` object is the collection of all the *od* choice sets.

C.2 Inversion of od-level amenities B_{od}

To match the observed total OD flows, given observed wages, rents, and commuting costs, I use the gravity equation for *od* flows:

$$\lambda_{od} = \frac{\frac{B_{od}\, w_d^\theta}{Q_o^{\alpha_h \theta}\, \tau_{od}^\theta}}{\sum_{o',d'} \frac{B_{o'd'}\, w_{d'}^\theta}{Q_{o'}^{\alpha_h \theta}\, \tau_{o'd'}^\theta}} \quad \text{where} \quad \tau_{od} = \left(\sum_r \tau_{odr}^{-\rho}\right)^{-\frac{1}{\rho}},$$

Since B_{od} is pinned-down up to a scale factor, we need to pick a global normalization. I set district 1 own's commuting amenities to 1, that is, set $B_{o^*,d^*} = 1$. This normalization choice pins down the overall scale for $\{B_{od}\}$.

$$B_{od} = \frac{\lambda_{od}}{\lambda_{11}} \frac{\frac{Q_o^{\alpha_h \theta}\, \tau_{od}^\theta}{w_d^\theta}}{\frac{Q_1^{\alpha_h \theta}\, \tau_{11}^\theta}{w_1^\theta}}$$

C.3 Inversion of productivities

Given observed wages, commercial land, and total labor, from the wage equation we just recover A_d:

$$w_d = A_d \left(\frac{\beta}{1-\beta} \frac{L_d}{H_d^c}\right)^{1-\beta}$$

D IDENTIFICATION OF PARAMETERS FROM SUBWAY SHOCK

D.1 Details of the subway Line 12 collapse

Mexico City's Metro is one of the largest and busiest rapid transit systems in North America, with 12 lines covering approximately 200–225 km of track and serving around 4.5 million passengers daily. Within this vast network, Line 12, inaugurated in 2012, is the longest line, stretching about 24.1 km and encompassing 20 stations. It carves a vital corridor across southern Mexico City, connecting the densely populated, lower-income southeastern area—primarily the Tláhuac municipality—to central districts through a combination of underground and elevated sections. For many residents in those peripheral neighborhoods, especially where other rapid transit options are scarce, Line 12 offered a crucial direct link to the city's employment and service hubs.

On the night of May 3, 2021, at around 10:20 p.m., a section of the elevated track between Tezonco and Olivos stations on Mexico City's Line 12 (Tláhuac–Mixcoac) collapsed. The failure of a supporting beam caused two cars of a moving train to fall onto Avenida Tláhuac, resulting in 26 fatalities and roughly 80 injuries. This was the deadliest accident in the Metro system in nearly fifty years. Emergency response was immediate, with federal and local agencies—including the army, navy, National Guard, Civil Protection, and the Red Cross—mobilized to rescue passengers and assist victims. The Mexico City government and the Executive Commission for Attention to Victims (CEAVI) set up information kiosks in the accident zone, hospitals, and the prosecutor's office to provide support for families. In the short run, authorities suspended Line 12 operations entirely and introduced substitute services. Around 490 public buses were deployed along Avenida Tláhuac, as well as connections to Tasqueña and Ciudad Universitaria. Despite these measures, the suspension disrupted mobility for a substantial share of daily Metro passengers; the system as a whole serves approximately 4.6 million riders per day, and Line 12 alone accounted for nearly 175,000 daily users.

An independent forensic investigation led by the Norwegian firm DNV identified a series of structural flaws as the root cause of the accident, including missing or poorly installed bolts, deficient welds, irregularities in materials, and inadequate supervision during construction. The reopening process was slow and staggered. After nearly 20 months of closure, the underground portion of Line 12 (from Mixcoac to Atlalilco) reopened in January 2023, restoring service to roughly 175,000 daily riders. However, the elevated section where the collapse occurred remained closed until 2024.

D.2 TomTom data details and processing

Details. TomTom provides traffic stats in a given road link at a given point in time via the Traffic Stats API. I acquired a sample of two months before the shock, March and April, 2021, and the same months in 2022, to avoid potential month-to-month seasonality concerns. TomTom provides moments of the distribution of speed for each road link, such as the deciles and the mean, for the given period of interest. So, for example, the distribution of speed in a given link is computed given all the cars that passed through that link at any moment between March and April.

Data processing. For practical purposes, I utilize the mean speed of each link, over the two-month period. I purchased only weekdays–the 24 hours in the day–and excluded holidays. A finer analysis could collect data on peak vs off-peak hours but due to budget constraints, I collected the full day. Some links had 0 sample of cars, so I dropped those. TomTom provides their own geographical definitions of road links, so I had to spatially match the TomTom links to the OSM links – which are the links on which the model is based and are much larger than the TomTom links (about 5-10 per OSM link). To aggregate TomTom link speed into OSM speed, I take the length-weighted average of TomTom links. I end up with a final sample of 556 OSM links, each with pre/post speed, and the change in speed.

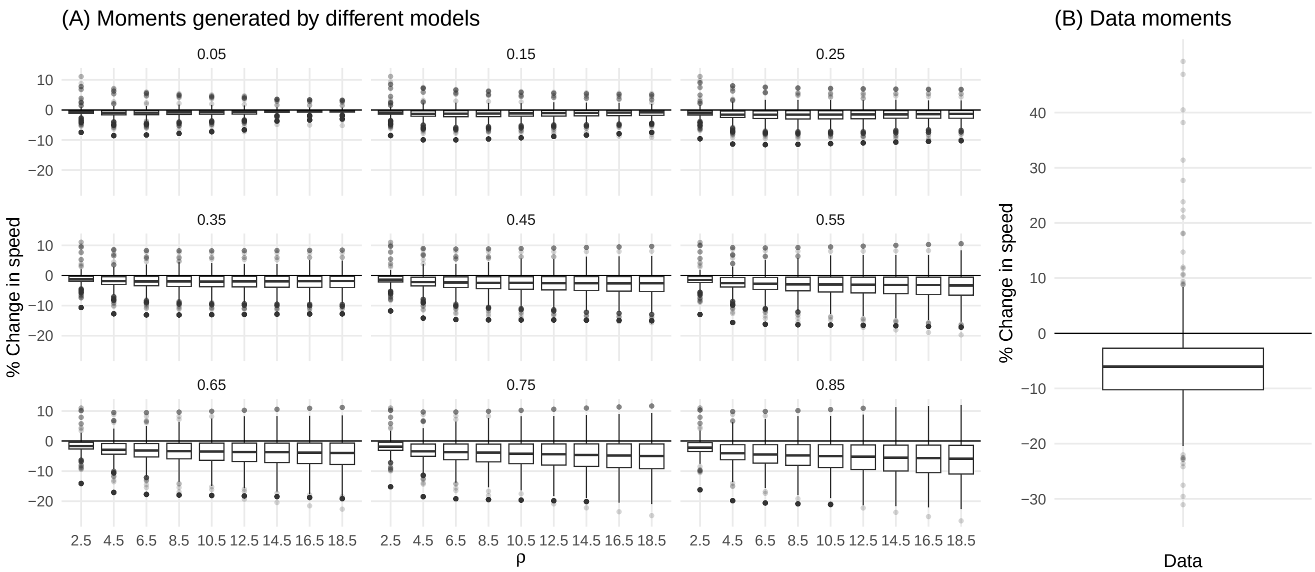
D.3 Robustness of SMM exercise

TABLE 8—IDENTIFIED PARAMETERS USING LINKS WITHIN DISTINCT SUBWAY BUFFERS

Distance Threshold	ρ	ϕ	Loss
20 km buffer	8.27	0.78	0.0059
3 km buffer	7.40	0.77	0.0058
1.5 km buffer	9.79	0.69	0.0061

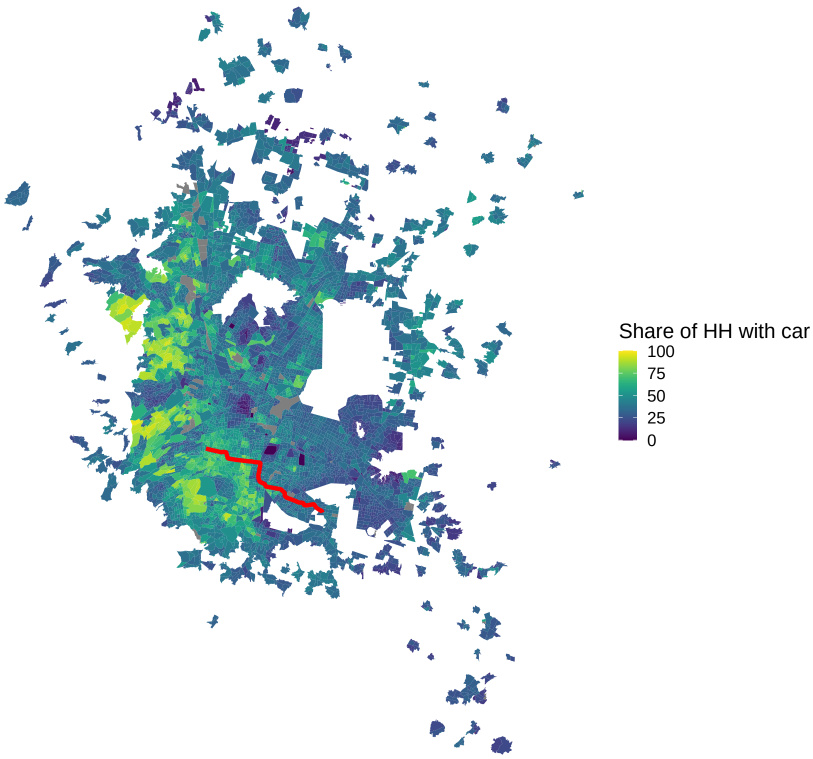
Note: Table shows identified parameters considering subsets of road links within different distance buffers from the collapsed subway line. Loss refers to the value of the objective function at the argmin.

FIGURE 15. SPEED CHANGES GENERATED BY DIFFERENT MODELS



Notes: facets are ϕ values. The box shows the median and interquartile speed change. The whiskers show the "typical range" (non-outlier min/max, 1.5xIQR).The dots are outliers.

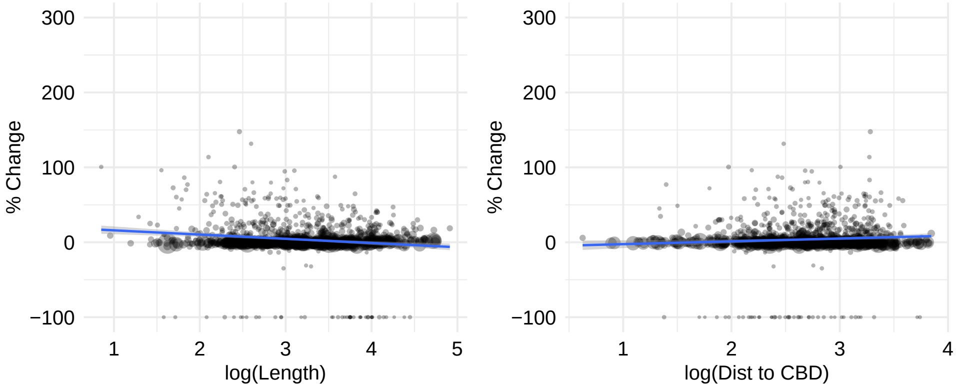
FIGURE 16. CAR OWNERSHIP BY HOUSEHOLD LIVING NEAR THE COLLAPSED SUBWAY LINE



Note: Figure shows mean car ownership by households across census tracts in the metro area. The red line is the collapsed line. Collapse occurred near the end of the line, on the South-East side, where low-income Tláhuac neighborhood is located. Although the line connects towards more affluent neighborhoods in South Mexico City, most of the users come from Tláhuac. Low car ownership in the South-West region suggests low substitution towards private cars, following the subway collapse.

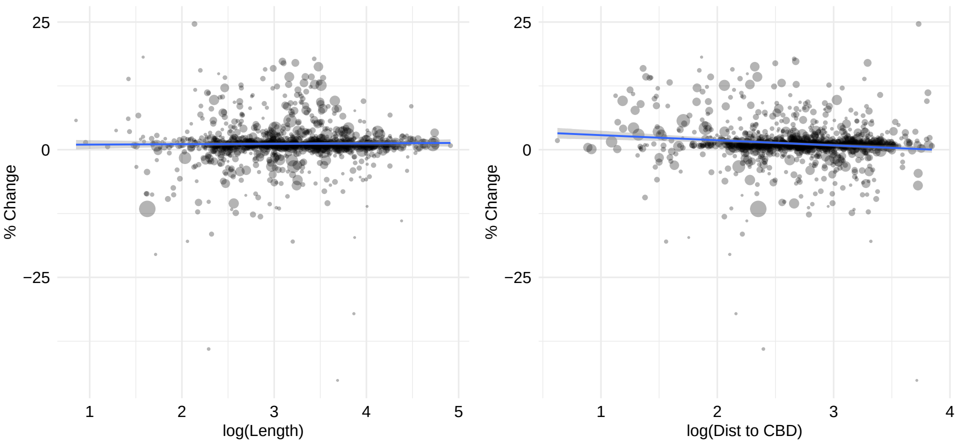
E Additional tables and figures

FIGURE 17. CORRELATION BETWEEN MARKET CHARACTERISTICS AND CHANGE IN FLOWS: PRICE DEREGULATION



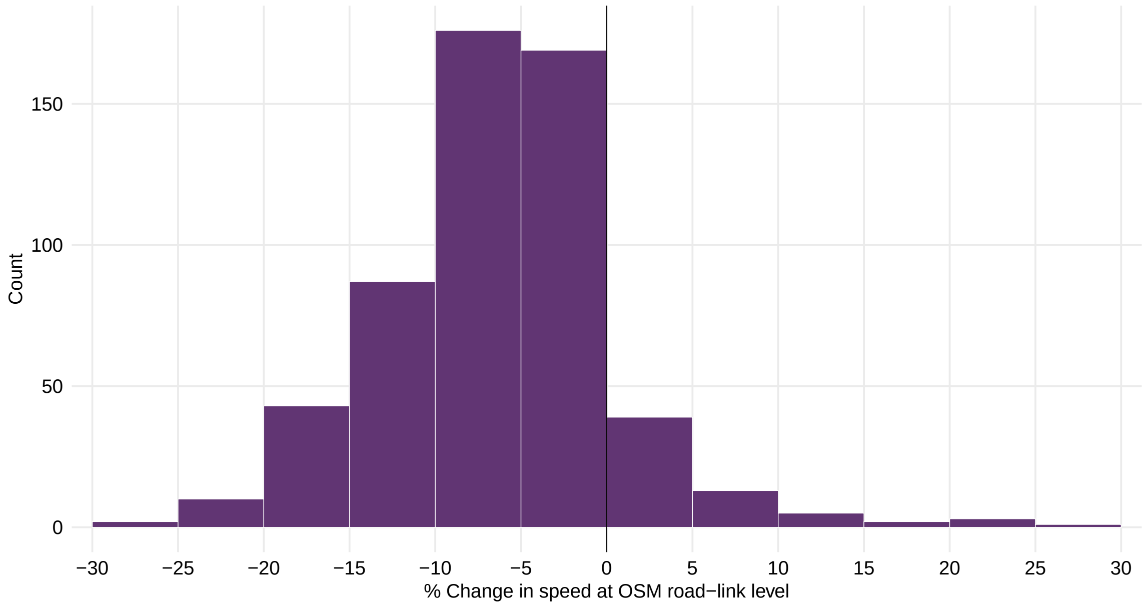
Note: Each dot represents a market. Dot size represents the pre-policy flow through that market. In the left panel, I am correlating length (kilometers) of the transit line and the change in flow. In the right panel, I am correlating the average distance of the transit line (across different segments in the line) to the central business district.

FIGURE 18. CORRELATION BETWEEN MARKET CHARACTERISTICS AND CHANGE IN FLOWS: METRO SUBSIDY REMOVAL



Note: Each dot represents a market. Dot size represents the pre-policy flow through that market. In the left panel, I am correlating length (kilometers) of the transit line and the change in flow. In the right panel, I am correlating the average distance of the transit line (across different segments in the line) to the central business district.

FIGURE 19. TOMTOM ROAD-LEVEL SPEED DATA: CHANGES PRE/POST SUBWAY SHOCK



Note: Figure shows the distribution of TomTom speed changes at the road-link level before and after the collapse of the subway line. Data was acquired from TomTom Traffic Stats API.

TABLE 9—PRICE REGULATION EXAMPLES IN PRIVATELY OPERATED MARKETS (SELECTED DEVELOPING CITIES)

Jurisdiction	**Mode / market**	**Instrument**	**Implementing authority**	**Key detail (year) / Source**
Mexico City (CDMX), Mexico	Private minibus ("transporte concesionado")	Regulated fare ladder (base + distance bands)	Secretaría de Movilidad (SEMOVI)	Tariff updated +1 MXN effective 15-Jun-2022; official fare matrix published. [S1], [S2], [S3]
Estado de México (Edomex), Mexico	Private combis/microbuses	Distance-based tariff table ("pirámide tarifaria")	Secretaría de Movilidad (Edomex)	Official table; vehicles must display the authorized "pirámide tarifaria". (2017; still referenced) [S4], [S5]
Philippines (national)	Jeepneys (PUJ)	Regulated *minimum* fare	LTFRB (national regulator)	Provisional +1 increase effective 08-Oct-2023 set min. at ₱13 (traditional) / ₱15 (modern). [S6], [S7], [S8]
Tanzania (Dar es Salaam)	Daladala (city buses)	Regulated commuter fares	LATRA (national regulator)	New city/intercity fares announced, effective 08-Dec-2023; operators carry official fare chart. [S9], [S10]
Bangladesh (cities)	City buses	Regulated *per-km* fare	BRTA (national regulator)	City-bus fare reset to Tk 2.42/km (from 2.45) per 2024 circular/reporting. [S11], [S12], [S13]
Cape Town, South Africa	Minibus taxis (MBT)	*Operator-set fares (no fixed government fare schedule)*; licensing/route oversight	Provincial Regulatory Entity (Western Cape) / City of Cape Town	MBT owners/drivers determine fares; City/Province regulate operating licences and routes; CITP reports sample fares (not a tariff). [S40], [S41], [S42]
Lagos, Nigeria	Danfo (informal minibuses)	*Temporary mandated discount*	Lagos State Gov.	During 2023–2024 palliative period the State mandated a 25% discount on commercial yellow buses; later withdrawn. [S43], [S44]
Nairobi, Kenya	Matatus (PSV)	*Proposed* fare regulation (draft legislation / county rules); currently operator-set	National: MoT/NTSA; County: Nairobi City County	2023 Bill to empower the Transport CS to set min/max fares; 2025 county draft rules include stricter fare pricing + cashless proposals (not yet final). [S45], [S46], [S47]

Note: Links for each source are available upon request.

Table 10—Fare subsidy examples in publicly operated systems (selected developing cities)

Jurisdiction	Mode / system	Instrument	Implementing authority	Key detail (year) / Source
Chile (national)	Urban bus/BRT/metro (multiple cities)	*Permanent operating subsidy* (national law)	Ministerio de Transportes y Telecomunicaciones (DTPR)	Ley 20.378 creates national subsidy to support fares and service (since 2009); legal + program pages. [S14], [S15]
Bogotá, Colombia	TransMilenio + SITP	Targeted discounts / free passes	Alcaldía / TransMilenio	Sisbén A1–B7 discounted fares; 2025 expansion to monthly *free-pass loads* for vulnerable groups. [S16], [S17], [S18]
Mexico City, Mexico	Metro; Metrobús	Zero-fare categories (gratuities)	STC Metro; Metrobús	Free access for older adults and persons with disabilities; program pages. [S19], [S20], [S21]
São Paulo, Brazil	Municipal bus (SPTrans)	Large recurring *operating subsidy*	Prefeitura de São Paulo / SPTrans	2024 subsidies R$5–6bn; 2025 projection R$6.4–6.5bn alongside fare policy updates. [S22], [S23], [S24]
Jakarta, Indonesia	TransJakarta, MRT, LRT	*Free travel* for 15 targeted groups	Provincial Government (DKI Jakarta)	Official rollout May-2025; Smart City guidance and public notices. [S25], [S26], [S27]
Delhi, India	DTC + Cluster buses	Women ride free (FFPT); smart-card rollout	GNCT of Delhi / DTC	Scheme running since 2019; 2025 shift to lifetime/smart-card for eligible residents. [S28], [S29], [S30]
Lagos, Nigeria	BRT and regulated public transport	*Temporary* fare subsidy (–50%, then –25%)	Lagos State / LAMATA	50% cut from 02-Aug-2023 ended 06-Nov-2023; partial discounts continued briefly thereafter. [S31], [S32], [S33]
Argentina (national, SUBE)	Bus/metro/train (AMBA + cities)	Social tariff (–55%, combinable with RED SUBE)	Gobierno Nacional (ANSES / Min. Transporte)	Ongoing 55% discount for eligible groups; guidance and FAQs on accumulation with RED SUBE. [S34], [S35], [S36], [S37]
Cape Town, South Africa	MyCiTi (BRT) & GABS (contracted buses)	Operating subsidies; regulated fare schedule	City of Cape Town (MyCiTi) / Western Cape Gov. (GABS contract)	Annual MyCiTi distance-band fare schedule; Golden Arrow receives operating subsidy via long-standing Provincial/National contract. [S38], [S39]

Note: Links for each source are available upon request.