

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 29, 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 and decentralized 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 transit trips involve 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, 2022](#); [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 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 calibration, deregulating private fares increases welfare by about 0.9%, removing the metro fare subsidy increases welfare by roughly 0.5%, and applying both jointly yields a net gain of around 1.4%.

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

²These 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 residence, workplace, and route decisions ([Allen and Arkolakis, 2022](#); [Bordeu, 2023](#)). Workers trade off time and income to choose potentially multimodal routes from a given route choice set. For each origin-destination 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 *markets*, e.g. take minibus A and subway B, or simply minibus 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 monetary costs of all the markets that are used along the route. Under standard assumptions about the idiosyncratic taste of workers, the commuting cost index between an origin-destination pair aggregates the ratios of price and effective work-time net of commuting with some elasticity of substitution across routes.³ Conditional on location and route decisions, workers consume—standard—final good and housing, but also a commuting good bundle across markets contained in the route choice. This generates a tractable demand system for private firms in each market.

The supply side—novel in this class of models—consists of many decentralized private markets 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 face some individual residual demand and maximize profits by choosing prices and the number of trips subject to vehicle capacity and a time endowment. Markets are heterogeneous: 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 monetary 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

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

depending on the elasticity of congestion to additional entrants and the relative presence of a market in a given road link.

Quantification with new data. I quantify the model using as setting one of the largest metro areas in the world: the Metropolitan Area of Mexico City. 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* routes, making it possible to measure service and coverage at the city scale.⁴ Furthermore, major transit disruptions in the city provide exogenous shocks to learn about route-preference and congestion parameters, and active policies—fare regulation and public subsidies—make it an ideal laboratory for mixed-system policy analysis.

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 extracted from Google’s API, which hosts a census collected by WhereIsMyTransport. To do so, I simulated Google Maps transit trips across origin–destination pairs, recovering the geography and service characteristics of roughly 80% of the transit lines in the system—around 2,000 in total—and mapped each line to OpenStreetMap links to account for road-level congestion. For each transit line, I recovered the name, kilometer length, frequency, and the time it takes to complete a full lap. Crucially, I back out the implied observed number of entrants operating each line using the definition of frequency.⁵ The model is then calibrated to match these observed outcomes.

Inspection of the data reveals important features of commuting and public-private networks that further motivate the structural framework. I summarize this in three stylized facts. i) The private network is 19 times larger (38,000 km) than the public network (2,000 km), has twice the reach of the public one, and serves more intensively peripheral locations. ii) Private lines are on average longer, faster, and more frequent relative to public lines; but the scope for congestion is salient in private lines: more than 80 private lines can share a single road link. iii) Transit represents 62% of all trips. From all transit trips, 83% are done using private transit in some leg, dwarfing the 17% of public transit, and 60% of trips that involve using private transit are multimodal.

Identification of key elasticities with experiment. Next, I identify two key elasticities exploiting plausibly exogenous variation from a natural experiment. The first is the elasticity of substitution across routes, which governs how commuters reallocate across routes when relative costs change. The second is the elasticity of congestion to additional entrants, which determines how entry affects travel time on the road.

I overcome two important challenges to identify these elasticities: data availability on commuting flows at the route level, and exogenous variation in costs. Since route-level flows are

⁴Sistema 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 corridors. They take several forms, such as minibuses of different sizes, but are mostly vans.

⁵Frequency is the inverse of the headway. The headway is the time between two vehicles.

very difficult to observe, I focus instead on road speeds, which map directly to changes in flows in the model. To obtain exogenous variation in costs, I exploit the collapse of a subway line.⁶ The subway shock exogenously increased costs for some routes, forcing commuters to substitute toward alternatives, mostly constituted by private providers. 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—compared 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. I exploit the unique features of my model to study two policies: a fare regulation in the private sector and a subway fare subsidy. Mexico City operates with ongoing price regulations and fare subsidies, so I calibrate the model under a version in which fares are set uniformly across firms within each market and subway fares are subsidized. This intervened-price setting serves as the baseline for counterfactuals.

These two policies are of both academic and policy interest because they act through under-studied mechanisms and entail salient fiscal trade-offs. Further, they reveal new 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.⁷ 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. Commuting costs decrease on average 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 and competitive pressures due to entry. Quantitatively, deregulation increases welfare by about 0.9%, decentralizing economic activity toward less-productive peripheral districts. Longer markets connecting towards central areas lose flows while more peripheral, shorter, and local markets gain flows. The intuition is that because service is improved in private markets, and these primarily serve peripheral districts, commuting within and across these districts increases.

⁶In 2021, the elevated portion of a subway line collapsed. 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.

⁷In 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.

Regarding transit subsidies, public systems—most notably the subway—operate with a flat fare that is invariant to distance or line switches. Government officials have claimed the subsidy to be as large as 72% of the true cost of a trip, with the implied gap to stated cost recovered through the city budget.⁸ 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 generates a reallocation of demand towards substitute private routes. Entry in substitute routes located in central areas increases congestion, increasing trip times but improving frequency. On the contrary, 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—where most of the subway system is located.

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 \(2024\)](#), who found a positive gain of $\approx 0.6 - 0.8\%$ following the opening of a 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.⁹ 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 work: urban transportation, endogenous trade costs, and the broader quantitative spatial literature. In urban transportation, prior work has mostly evaluated welfare from single-mode, publicly provided expansions, e.g., BRT in Bogotá ([Tsivanidis, 2019](#)), cable car in Medellín ([Khanna et al., 2024](#)), or a subway line in Mexico City ([Zárate, 2024](#)). Relatedly, [Conwell \(2024\)](#) studies boarding and queuing externalities in minibus markets in Cape Town. This paper contributes to this literature by quantitatively studying another class of interventions—price-shifting policies—that affect both the public and private sector simultaneously.

Most closely, [Björkegren et al. \(2025\)](#) examine how the private sector responds when government rolls out 13 new BRT lines along corridors already served by private providers and document declines in private frequencies and prices, consistent with substitution. I complement these results by showing that the private sector is also a large complement to mass transit on connecting markets: when mass-transit costs increase (the flip-side of their BRT expansion), entry and frequencies rise in substitute corridors but fall on connectors. These heterogeneous

⁸Actual 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>

⁹These 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.

responses highlight network interdependencies that single-mode evaluations may not capture. Another closely related paper, (Almagro et al., 2024), studies how a government sets public prices and frequencies and uses road pricing to mitigate congestion and environmental externalities in Chicago. This paper complements by endogenizing prices and frequencies through private entry, in the presence of public mass transit, and in a city-wide general-equilibrium setting.

Methodologically, 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. Abstracting away from income effects may be a reasonable assumption in richer-cities settings (Ahlfeldt et al., 2015; Monte et al., 2018) since monetary costs are negligible and time represents the main cost, but insufficient to study developing-city settings due to binding budget constraints (Bryan et al., 2025). Further, to my best knowledge, this is the first paper that provides an estimate of the elasticity of substitution across transit routes and the elasticity of congestion exploiting quasi-experimental variation from 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, e.g. Google Maps—but still preserves the tractability and efficiency to solve a large-scale spatial model with potentially thousands of markets.

2 MODEL

I develop a general equilibrium quantitative spatial model building on canonical frameworks (Ahlfeldt et al., 2015; Monte et al., 2018; Allen and Arkolakis, 2022), but adding 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 multimodal 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

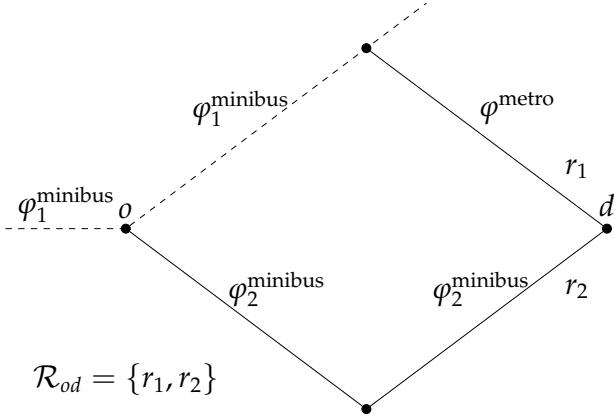
The economy is populated by an exogenous measure of workers \bar{L} , each of which has a time endowment \bar{T} to supply labor and commute. The geography in this economy consists of J locations indexed by o (origin) and d (destination), each of which has an endowment of residential \bar{H}_o and commercial \bar{H}_d^c land.

Transportation network. Travel occurs on transportation networks comprising a road network and multiple transit networks. The (directed) road network is a graph $\mathcal{G}^{\text{rd}} = (\mathcal{N}^{\text{rd}}, \mathcal{L}^{\text{rd}})$ with nodes $i \in \mathcal{N}^{\text{rd}}$ and directed links $\ell = (i, j) \in \mathcal{L}^{\text{rd}}$. Transit networks are indexed by a sector $k \in \{\text{Public}, \text{Private}\}$ and a mode $m \in \mathcal{M}_k$ (e.g., subway, train, minibus). For each (k, m) , the (directed) transit network is $\mathcal{G}^{k,m} = (\mathcal{N}^{k,m}, \mathcal{L}^{k,m})$ with links $\ell^{k,m} = (i, j) \in \mathcal{L}^{k,m}$. Nodes $\mathcal{N}^{k,m}$ represent stops or boarding points. For road-using modes—primarily private minibuses—

each transit link is embedded in the road network via a mapping $\iota^{k,m} : \mathcal{L}^{k,m} \rightarrow 2^{\mathcal{L}^{\text{rd}}}$ that assigns to $\ell^{k,m}$ the (nonempty) set of road links it traverses; for fixed-guideway modes (e.g., subway/rail), $\iota^{k,m}$ is empty.¹⁰

Transit markets. Within each transit network (k, m) there is a set of markets denoted $\Phi^{k,m}$. A market $\varphi \in \Phi^{k,m}$ is a directed path in $\mathcal{G}^{k,m}$: an ordered sequence of links. A market corresponds to what is called a “line” in practice, e.g., minibus line A, subway line B. Markets have four attributes: (i) price P_φ , (ii) trip time t_φ^{trip} , (iii) frequency F_{φ} , and (iv) wait time t_φ^{wait} . These attributes are endogenous for private markets. For public markets, I assume these attributes are given and fixed.

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.

Routes. Each location pair od is connected by a set of routes \mathcal{R}_{od} . A route $r \in \mathcal{R}_{od}$ is an ordered sequence of $K \in \mathbb{N}^+$ markets, $r = (\varphi_1, \dots, \varphi_K)$.¹¹ Note that under this definition, route choice *implies* mode choice and the order of modes. Figure 1 shows an example of a route in this setting. As can be seen, getting from one location to another might involve using either a full market or parts of several different markets. The route depicted on top uses a portion of minibus market $\varphi_1^{\text{minibus}}$, and a subway leg φ^{metro} . The route below uses a different single minibus market $\varphi_2^{\text{minibus}}$.

2.2 Demand for commuting

Workers ω have heterogeneous preferences for locations and routes and choose the (o, d, r) triplet that maximizes their utility. Preferences of a worker are defined over final good con-

¹⁰I further assume that such mapping is empty even for road-using public modes. The reason is that many of the bus-type modes have their own lanes, for example the BRT or trolleybus.

¹¹Transitioning between legs of the route may involve transfers and/or walking, which are not explicitly modeled for ease of exposition.

sumption $X(\omega)$, which will serve as numeraire, housing $H_o(\omega)$, a commuting good CES bundle $C_{odr}(\omega)$ defined at the route level, idiosyncratic amenities taste shock $\varepsilon_{odr}(\omega)$, and deterministic amenities B_{od} at the location-pair level that capture the common value that workers receive from living in o and working in d . Conditional on a (o, d, r) choice, each worker maximizes his utility by solving the following problem:

$$\max_{X, H, C} \quad B_{od} \left(\frac{X(\omega)}{\alpha_x} \right)^{\alpha_x} \left(\frac{H_o(\omega)}{\alpha_h} \right)^{\alpha_h} \left(\frac{C_{odr}(\omega)}{\alpha_c} \right)^{\alpha_c} \varepsilon_{odr}(\omega),$$

subject to

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

where w_d is the wage per unit of time, Q_o is price of land (henceforth rent), P_{odr} is the price of the commuting bundle, and Ω captures proportional income adjustments from a land portfolio and taxes.¹² Workers spend their time endowment supplying labor time n_{odr} and using the residual for commuting time t_{odr} .

The commuting good bundle $C_{odr}(\omega)$ combines consumption of goods in all of the different markets that comprise route r .¹³ In particular, conditional on the route choice, these markets become perfect complements as commuters consume trips in equal proportions. Thus, the bundle takes a Leontief form

$$C_{odr}(\omega) = \min(c_{\varphi_1}(\omega), \dots, c_{\varphi_K}(\omega)), \quad (1)$$

with an associated price index $P_{odr} = \sum_{\varphi \in odr} P_\varphi$.¹⁴ Furthermore, I assume that each market-specific worker demand $c_\varphi(\omega)$ is itself a CES bundle of commuting goods from individual firms i within a market:

$$c_\varphi(\omega) = \left(\sum_{i \in \varphi} q_i(\omega)^{\frac{\chi-1}{\chi}} \right)^{\frac{\chi}{\chi-1}}, \quad (2)$$

where χ is the elasticity of substitution across firms' goods.¹⁵

After maximizing, the indirect utility for a worker choosing (o, d, r) is

$$V_{odr}(\omega) = B_{od} \frac{\tilde{w}_d (\bar{T} - t_{odr})}{Q_o^{\alpha_h} P_{odr}^{\alpha_c}} \varepsilon_{odr}(\omega), \quad \tilde{w}_d \equiv w_d (1 + \Omega). \quad (3)$$

¹²I assume that worker's own the land in the economy, and the aggregate land portfolio is split proportionally to income. The term enters multiplicatively so that it does not distort spatial allocations to a first order. This will become clear in Sections 2.6 and 2.7, where I unpack the value of Ω .

¹³To generate intuition, one can think of these goods as seats in a unit, e.g. minibus or subway.

¹⁴One can also think of this as a CES bundle $C_{odr}(\omega) = \left(\sum_{\varphi \in odr} c_\varphi^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$ with elasticity of substitution $\sigma \rightarrow 0$.

More detailed derivations and explanations can be found in Appendix A.3.

¹⁵This assumption ensures tractability and facilitates the computation of the equilibrium, as it will become clear in the next section.

If we denote

$$\tau_{odr} \equiv \frac{P_{odr}^{\alpha_c}}{\bar{T} - t_{odr}}, \quad (4)$$

then we can rewrite (3) as a more familiar expression to standard commuting models:

$$V_{odr}(\omega) = B_{od} \frac{\tilde{w}_d}{Q_o^{\alpha_h} \tau_{odr}} \varepsilon_{odr}(\omega), \quad (5)$$

hence τ_{odr} in this expression is a micro-founded effective commuting cost measure that combines both time and price and has an intuitive interpretation: it is the monetary cost per unit of effective work time.

Commuting flows. Let the idiosyncratic taste shock $\varepsilon_{odr}(\omega)$ be distributed extreme-value type II distribution (nested Fréchet):

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

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. Each worker receives a one-time i.i.d. shock $\varepsilon_{odr}(\omega)$ and makes 1) residence and workplace decisions, and 2) route choice that maximizes their indirect utility.

Using this assumption of the distribution of indirect utility, the probability of workers choosing (o,d) is

$$\lambda_{od} = \frac{\left(\frac{B_{od} w_d}{Q_o^{\alpha_h} \tau_{od}} \right)^\theta}{\sum_o \sum_d \left(\frac{B_{od} w_d}{Q_o^{\alpha_h} \tau_{od}} \right)^\theta}, \quad (7)$$

where I collapsed the aggregation of the effective commuting costs across routes into an od -specific CES commuting index with elasticity of substitution ρ :

$$\tau_{od} \equiv \left(\sum_{r \in \mathcal{R}_{od}} \tau_{odr}^{-\rho} \right)^{-1/\rho}. \quad .^{16} \quad (8)$$

Following the law of total probability, we can further decompose the full probability that a worker will choose odr into the probability that the od pair is chosen and the probability that route r is chosen conditional on living/working in od :

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

¹⁶For detailed derivations, refer to Appendix A.1.

Finally, given that workers are mobile, in equilibrium each location-route triplet (o, d, r) must yield the same expected utility, and in particular equal to the expected utility of the economy. Denote such expected utility as the constant \bar{W} . Given my distribution assumptions about utility, welfare in this economy is

$$\bar{W} \equiv \mathbb{E}[\max V_{odr}(\omega)] = k \left(\sum_o \sum_d \left(\frac{B_{od} \tilde{w}_d}{Q_o^{\alpha_h} \tau_{od}} \right)^\theta \right)^{\frac{1}{\theta}}, \quad (10)$$

where k is a constant $k = \Gamma(\frac{\theta-1}{\theta})$, and $\Gamma(\cdot)$ is the gamma function.

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

$$P_{odr} \equiv \sum_{\varphi \in r} P_\varphi, \quad (11)$$

$$t_{odr} \equiv \sum_{\varphi \in r} \left(t_\varphi^{\text{wait}} + \gamma_{odr}^\varphi t_\varphi^{\text{trip}} \right), \quad (12)$$

where $\gamma_{odr}^\varphi \in [0, 1]$ is the share of a market φ 's trip time used by route r . For example, a route could use only half of a bus line, so $\gamma_{odr}^\varphi = 0.5$.

Demand at the market level. 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:

$$D_\varphi = \sum_{od} \sum_{r| \varphi \in r} \frac{\alpha_c \tilde{w}_d (\bar{T} - t_{odr})}{P_{odr}} \lambda_{odr} \bar{L}. \quad ^{17} \quad (13)$$

2.3 Supply of transportation

I make three main assumptions about the transportation market structure:¹⁸

1. Markets' geographies are exogenously defined.
2. There is free entry into each market.
3. Private markets are segmented: there is a potential mass of entrant firms in each market, and once they enter they cannot switch.

Firm's demand in a market φ . Recall that workers consume a CES bundle of commuting goods from individual firms in each market with elasticity of substitution χ , as shown in equation (2). Aggregating demand for a firm across all workers that consume in that market, we

¹⁷Detailed derivations in Appendix A.3

¹⁸These assumptions are discussed in the discussion section 2.9 right after the model.

get an individual residual CES demand for firm i .¹⁹

$$q_{i,\varphi} = \left(\frac{p_{i,\varphi}}{P_\varphi} \right)^{-\chi} D_\varphi. \quad (14)$$

Firm's problem. Individual firms decide whether to enter the market by paying a fixed cost of entry f_φ^e , which is denominated in terms of the final good. Upon entry, firms maximize profits by choosing the price $p_{i,\varphi}$ and quantity of trips $n_{i,\varphi}$ that maximize their profits, subject to a capacity constraint q_φ^c and a time endowment \bar{T}^d .²⁰ Firms face an individual CES demand, given by equation (14). The problem of the firm is:

$$\begin{aligned} \max_{p_{i,\varphi}, n_{i,\varphi}} \pi_{i,\varphi} &= \underbrace{\left(p_{i,\varphi} \frac{q_{i,\varphi}}{n_{i,\varphi}} - \delta_\varphi t_\varphi^{\text{trip}} \right) n_{i,\varphi} - f_\varphi^e}_{\text{Per-trip revenue}} \\ \text{s.t. } q_{i,\varphi} &= \left(\frac{p_{i,\varphi}}{P_\varphi} \right)^{-\chi} D_\varphi, \\ \frac{q_{i,\varphi}}{n_{i,\varphi}} &\leq q_\varphi^c, \\ n_{i,\varphi} t_\varphi^{\text{trip}} &= \bar{T}^d. \end{aligned}$$

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

$$p_{i,\varphi} = \left(\frac{D_\varphi}{q_\varphi^c n_{i,\varphi}} \right)^{\frac{1}{\chi}} P_\varphi. \quad (15)$$

The number of trips are such that the firm exhausts its time endowment, given the time to complete each trip:

$$n_{i,\varphi} = \frac{\bar{T}^d}{t_\varphi^{\text{trip}}}. \quad (16)$$

The price index of the commuting good in this market is

$$P_\varphi \equiv \left(\sum_{i \in \varphi} p_{i,\varphi}^{1-\chi} \right)^{\frac{1}{1-\chi}} = P_\varphi \left(\frac{D_\varphi}{q_\varphi^c n_{i,\varphi}} \right)^{\frac{1}{\chi}} M_\varphi^{\frac{1}{1-\chi}}, \quad (17)$$

where M_φ denotes the mass of entrants. From equation (17), it follows that the following

¹⁹Detailed derivations in Appendix A.3

²⁰To generate intuition, one can think of a firm as a minibus, or a driver that owns a minibus, that once it enters a market drives around their corresponding market geographic delimitation. Further, one can think of the capacity as the number of seats in a minibus, and the time endowment as a shift.

²¹This 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.

market-clearing relation must hold:

$$D_\varphi(\mathbf{M}) = M_\varphi^{\chi/(\chi-1)} q_\varphi^c n_{i,\varphi}, \quad \forall \varphi \in \Phi^{\text{Priv}}. \quad (18)$$

This equation states that the total capacity in the market—the product of the number of entrants, and the capacity and the number of trips of each entrant—must be equal to demand. 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 (18) describes a system of $|\Phi^{\text{Priv}}|$ equations, where Φ^{Priv} is the set of private markets.

Free entry. Because all firms are homogeneous, entrants enter each market until everyone's profits are driven to zero. Imposing zero profits $\pi_{i,\varphi} = 0$ implies that:

$$P_\varphi = M_\varphi^{-\frac{1}{\chi-1}} \frac{t_\varphi^{\text{trip}}}{\bar{T}^d} \left(\frac{\delta_\varphi \bar{T}^d + f_\varphi^e}{q_\varphi^c} \right). \quad (19)$$

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, we can determine the rest of the service characteristics. For quantification purposes, I assume that the time it takes a service unit to complete a lap, i.e. the trip time, is a function of the total number of entrants 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

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

Trip time in a given market is then a function of a market-specific time shifter \bar{t}_φ and congestion across all the road links that define that market:

$$t_\varphi^{\text{trip}}(\mathbf{M}) = \bar{t}_\varphi \sum_\ell S_\ell(\mathbf{M})^\varphi. \quad (20)$$

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

$$\text{Freq}_\varphi(\mathbf{M}) \equiv \frac{M_\varphi}{t_\varphi^{\text{trip}}(\mathbf{M})}. \quad (21)$$

Assuming that buses arrive uniformly at each stop, wait time from the perspective of a worker is

$$t_\varphi^{\text{wait}}(\mathbf{M}) = \frac{1}{2} \frac{1}{\text{Freq}_\varphi(\mathbf{M})}. \quad (22)$$

This functional form implies that a worker that arrives at a bus stop will wait on average half of

the headway between buses.²² Importantly, note that these market-specific outcomes depend not only on the market's own entry, but on the entry of all markets that are related to the market through congestion.

2.4 Residential land markets

In equilibrium, the housing price Q_o adjusts such that supply of housing equals demand of housing in each location:

$$\bar{H}_o Q_o = \alpha_h Y_o R_o, \quad \forall o \in J. \quad (23)$$

where $Y_o \equiv \sum_d \sum_r \lambda_{odr|o} \tilde{w}_d (\bar{T} - t_{odr})$ 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 (23) equates aggregate expenditure for housing in each location to the income generated in the housing rental market.

2.5 Final good and input 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:

$$y_d = A_d \left(\frac{L_d}{\beta} \right)^\beta \left(\frac{H_d^c}{1-\beta} \right)^{1-\beta}. \quad (24)$$

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. In equilibrium, the supply of commercial land equals the demand, $\bar{H}_d^c = H_d^c$, therefore wages in each location are given by

$$w_d = A_d \left(\frac{\beta}{1-\beta} \frac{\bar{H}_d^c}{L_d} \right)^{1-\beta}, \quad (25)$$

and commercial rents are given by

$$Q_d^c = A_d \left(\frac{1-\beta}{\beta} \frac{L_d}{\bar{H}_d^c} \right)^\beta. \quad (26)$$

²²The 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, for example, many minibus-owner associations impose these sort of controls. To obtain this, we can assume some process of arrival of buses with a parameter $Freq(\cdot)$ that controls its mean and variance. For example, in the different case of a Poisson process with parameter $\lambda = Freq(\cdot)$, because a Poisson random variable is memoryless, we would have that $t_\varphi^{\text{wait}} = \frac{1}{Freq_\varphi(M_\varphi)}$, or that people wait on average the full headway between buses.

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:

$$\dot{y}_{odr} = y_{odr} (1 + \nu), \quad \nu = \frac{I}{\sum_{odr} y_{odr} \lambda_{odr} \bar{L}}.$$

Note that I assumed that land income is proportionally distributed to labor income such that spatial allocations are not affected, to a first order.²³

2.7 Government funding to provide fare subsidies in public transit

I allow for the possibility of fare subsidies among public transit markets, which I assume are funded by income taxes to commuters. Let \mathcal{S} denote the total subsidy:

$$\mathcal{S} \equiv \sum_{\varphi_{\text{public}}} (P_{\varphi_{\text{public}}}^* - P_{\varphi_{\text{public}}}) D_{\varphi_{\text{public}}},$$

where $P_{\varphi_{\text{public}}}^*$ denotes the price in absence of subsidies, and $P_{\varphi_{\text{public}}}$ 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 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

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

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

²³The 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.

captures the proportional income adjustments from land income revenue and taxes.

2.8 General equilibrium

An equilibrium in this economy is a collection $\{w_j, Q_j, Q_j^c\}_{j \in J}$ of factor prices across locations, a collection $\{P_\varphi, t_\varphi^{\text{trip}}, t_\varphi^{\text{wait}}\}_{\varphi \in \Phi}$ of transportation-market prices and times, an allocation of entrants $\{M_\varphi\}_{\varphi \in \Phi}$, an allocation of residents and effective labor $\{R_j, L_j\}_{j \in J}$, a welfare constant \bar{W} , and land-revenue and tax constants (ν, η) that, given fundamentals $\{A_j, B_{jj'}\}_{(j,j') \in J}$, and land endowments $\{\bar{H}_j, \bar{H}_j^c\}_{j \in J}$ satisfy the equilibrium conditions defined below.

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

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

commuting flow gravity equations hold

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

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

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

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

$$D_\varphi(\mathbf{M}) = M_\varphi^{\frac{\chi}{\chi-1}} q_\varphi^c \frac{\bar{T}^d}{t_\varphi^{\text{trip}}}.$$

3. Residential land markets clear:

$$Q_o = \frac{\alpha_h Y_o R_o}{H_o}.$$

4. Commercial land markets clear:

$$Q_d^c = A_d \left(\frac{1-\beta}{\beta} \frac{L_d}{\bar{H}_d^c} \right)^\beta.$$

5. Labor markets clear:

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

6. Government runs a balanced budget:

$$\eta \sum_{odr} y_{odr} (1 + \nu) \lambda_{odr} \bar{L} = \sum_{\varphi_{\text{public}}} (P_{\varphi_{\text{public}}}^* - P_{\varphi_{\text{public}}}) D_{\varphi_{\text{public}}}$$

7. Final good market clears.

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

2.9 Discussion of the model's insights and assumptions

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 previous work 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 . By totally differentiating expression (4),

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

we can see that any intervention that affects either times or prices can affect commuting costs directly and indirectly through changes in entry incentives—in the presence of private providers. For example, when we improve or build infrastructure ([Tsivanidis, 2019](#); [Zárate, 2024](#); [Khanna et al., 2024](#)), we effectively shift times directly within the structural framework, via τ . If we consider a fare subsidy, for example, it would have a direct effect on the price as well. These effects would be captured in the third term of the above. What my framework adds, then, is the possibility of further equilibrium adjustments through network-wide changes in costs and entry incentives. In particular, entry can affect i) prices through network complementarities and competitive pressure, and ii) congestion and frequencies on the road. These effects correspond to the first two terms in the above.

To understand better how these effects come into play, I highlight two key mechanisms that govern how interactions across markets take place: a i) route-cost substitution/complementarity effect and a ii) frequency/congestion trade-off.

Regarding the first, note that because a route is composed by one or more markets, conditional on the route choice they become perfect complements. Given that agents do not care about individual costs of markets but about the aggregate route cost, a cost change in one market will directly impact the flow through other markets within the route, further impacting entry.

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

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

and the elasticity of the aggregate probability across all routes that use such market with respect 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. These changes in flows ultimately affect entry through changes in demand.

Regarding the second, due to the presence of congestion on the road 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

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

Then:

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

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

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

Proof. See Appendix A.4. \square

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 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.

Taking stock, most of the new action in this proposed framework—relative to standard models—comes from these two key mechanisms. These mechanisms are in turn governed predominantly by the elasticity of substitution across routes ρ and the elasticity of congestion ϕ . Thus, the value of these parameters will be crucial to assess any policy intervention.

Discussion of assumptions. Within the model I interpret a private transportation firm as a driver–vehicle (minibus) unit that provides service over a day/shift within a fixed market (line) φ . Unlike Conwell (2024) and Björkegren et al. (2025), who study queuing with fixed origin-destination ends of trips, I assume a loop operation along the geography of the market:

vehicles continuously circulate and pick up and drop off passengers along the line. With this vision in mind, I interpret the individual firm residual demand as the total number of ‘seats’ or ‘trips’ demanded throughout a day or a shift.

On the demand side, commuters allocate expenditure between non-transport consumption and a commuting composite good. I assume Cobb–Douglas at the top tier and a CES aggregator within the commuting composite across routes and then markets, which yields tractable market-level demand $D_\varphi(\mathbf{M})$ for service on market φ , and an individual firm CES residual demand $q_{i,\varphi}(\cdot)$ that depends exclusively on market-specific aggregates (D_φ, P_φ) that are taken as constants from a single-firm’s perspective.

This structure is not meant to suggest that residents derive utility from ‘seats’ or ‘trips’ per se; rather, it parsimoniously rationalizes nontrivial spending on mobility and delivers a computationally tractable demand system at the firm level, which allows me to solve for the general equilibrium across a large number of markets. To see this, note that if we define a fixed-point operator by solving for M_φ in equation (18):

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

then we have an update map for the vector of entrants that facilitates the computation of the model by applying usual fixed-point algorithms.

In terms of the assumptions of the supply side, that the geography of markets is exogenous and fixed, that there is free entry into each market, and that private markets are segmented, I have some reasons to assume that. The first assumption implies that the actual corridors or roads through which minibuses, trains, or subways 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.²⁴ The second and third assumptions allow me to preserve tractability and facilitate 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.²⁵ The model’s fixed cost of entry, f_φ^e , can be thus interpreted as the annualized cost of the vehicle and permit, while the firm’s profit maximization problem captures the driver’s daily operational decisions, with the zero-profit condition reflecting the long-run outcome in a competitive market for drivers.

²⁴This 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.

²⁵Minibuses 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.

3 EMPIRICAL SETTING

I use the Metropolitan Area of Mexico City as empirical setting to quantify the model and study policy counterfactuals. In this section first I introduce the setting and describe features of the city that make it an appealing setting. Then I introduce three stylized facts on the spatial supply of transit, the characteristics and arrangement of private lines, and commuting patterns on mode usage. These facts further motivate the framework and highlight role of public and private transit in a megacity, with private providers at the core. Lastly, I provide some institutional context on the market structure of the private sector and ongoing fare regulations and subsidies.

3.1 Setting introduction: Metropolitan Area of Mexico City

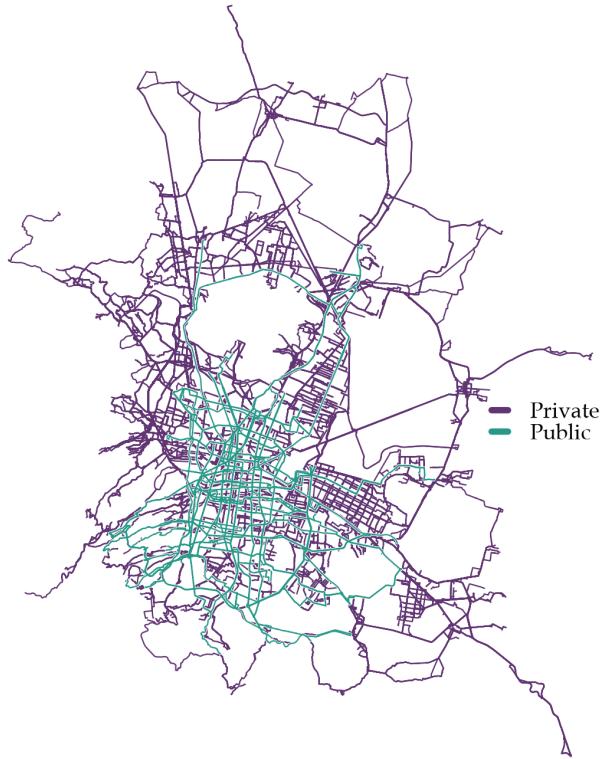
The Metropolitan Area of Mexico City, home to more than 22 million people, combines a large scale mixed transit system, with substantial heterogeneity across neighborhoods in terms of access to transit and demographics. Mexico City shares many features of the large urban agglomerations around the world, e.g. throughout South America, Africa, South-East Asia. 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. 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). Mexico City operates extensive subway and BRT networks with many lines. By contrast, the State of Mexico—despite housing more than half of the area’s population—has only one BRT line and one subway line, with just 11 of the system’s 192 subway stations.

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 commuting towards Mexico City. Daily mobility in the metro area relies on both the public system—Metro, Metrobus, Tren Ligero, Cablebus, RTP buses, and trolleybus—and a vast network of privately operated *micros/combis/colectivos*. These small, flexible minibuses extend coverage where mass transit is sparse, particularly in the outskirts, but are also available in central areas of the city. I provide a rich characterization of these transit networks in the following subsection.

3.2 Stylized facts about the transit network and commuting

I present three stylized facts that describe (i) the spatial distribution of private and public transit, (ii) the characteristics of transit lines and arrangement, and (iii) commuting patterns by mode usage. These facts remark the importance of considering public and private transit, their characteristics and interactions, and the potential scope for externalities such as congestion when considering a structural framework in a city with such mixed systems.

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.

Fact 1: On the spatial supply of transit. *The private network is 19 times larger (38,000 km) than the public network (2,000 km), has twice the reach of the public one, and serves more intensively peripheral locations.*

Figure 2 shows the public and private transit lines in the metro area. As can be seen, the outskirts of the city rely heavily—and in some locations almost exclusively—on private providers, while more central areas enjoy a mix of both the public and private system.

Furthermore, table 1 shows the overall length and coverage of both networks. Comparing both columns it is very salient how much broad the private network is. The total length is 19 times larger for the private relative to the public. Also, in terms of reach, the private network reaches at least two-thirds of the more than 5,000 census tracts in the area, while public transit reaches only a third. If we measure reach by comparing the share of census tracts with at least one public or private line, the private network reaches a little more than twice the number of tracts, 68% compared to 29%. Interestingly, a third of census tracts have at least 10 lines while there is not a single tract with the same amount of public lines. This statistic speaks to how dense and extensive private networks can be.

TABLE 1—TRANSIT NETWORK COVERAGE BY TYPE OF TRANSIT IN MEXICO CITY

Metric	Private	Public
Total network length (km)	38,307	2,019
Network length density (total km length / total city area in km ²)	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: Total city area was calculated summing the area of the union of census tracts. Census tracts are INEGI's definition of AGEB (Área Geoestadística Básica). 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.

Fact 2: On characteristics and overlap of private lines. *Private lines are on average longer, faster, and more frequent relative to public lines; but the scope for congestion is salient in private lines: more than 80 private lines can share a single road link.*

Table 2 shows the descriptive statistics for the near-universe of private and public lines. It can be seen that on average private lines are larger, but the time it takes to complete a full-length trip is shorter, relative to public lines. Also, the time between units, i.e., the headway, is shorter for private lines, implying more frequency of service. An interesting fact is that the distribution of these characteristics is quite broad. For example, there are very short private lines (2 km) and very long ones (136 km), spanning trip times that range from 7 minutes to 295 minutes ≈ 5 hours, or headways ranging from a minute to half an hour. These facts highlight the large variation that there can be in line characteristics, so it is important to account for this heterogeneity in costs and service in a structural framework.

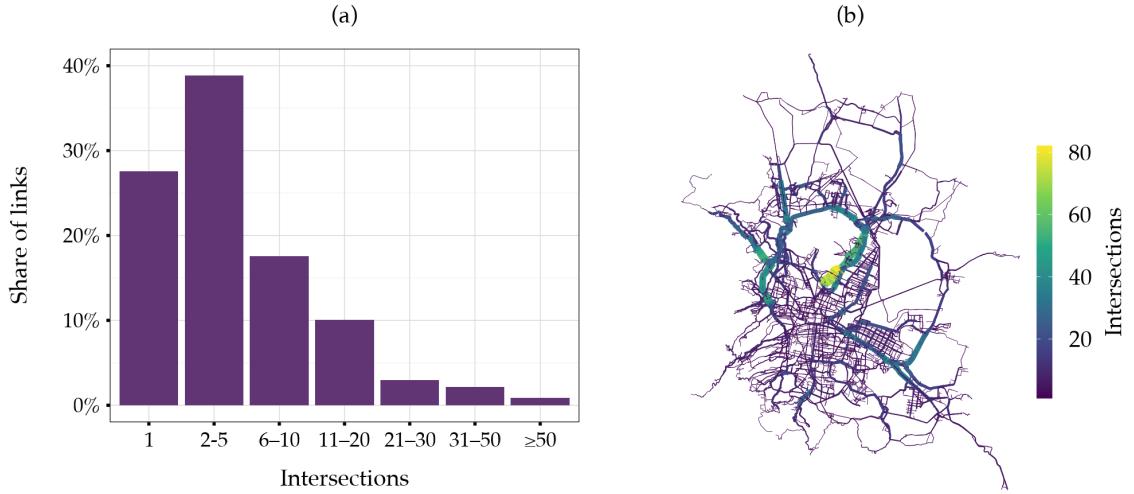
TABLE 2—DESCRIPTIVE STATISTICS BY TYPE OF TRANSIT LINES

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.

Furthermore, the scope for congestion externalities is large due to the arrangement of private lines. Figure 3 shows the distribution of intersections of private lines across all road links in the metropolitan area. Panel (a) shows a histogram where it can be seen that most links carry between 2 and 5 private lines, with a mean of 6, but many links carry tens of lines. Panel (b) shows the spatial distribution represented by color and thickness of lines. As can be seen, there is substantial heterogeneity across links: some other roads carry up to 80 private lines. This occurs mostly on large roads that connect from peripheral areas towards the center. This fact suggests that the scope for congestion externalities is large: entry in lines that pass through these busy links can affect many other lines through congestion.

FIGURE 3. INTERSECTIONS OF PRIVATE TRANSIT LINES AND ROAD LINKS



Note: Panel (a) shows a histogram of the number of intersections of private transit lines across all OpenStreetMap links in the metropolitan area. There are a total of 23,577 OSM links in the metropolitan area. Panel (b) the spatial distribution of intersections, colored by the number of private 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.

Fact 3: On commuting patterns on mode usage. *Transit represents 62% of all trips. From all transit trips, 83% are done using private transit in some leg, dwarfing the 17% of public transit, and 60% of trips that involve using private transit are multimodal.*

Table 3 classifies commuting trips by public/private mode usage, using a sample of detailed trip-level data that is representative of 6.4 million commuting trips. Many things are worth remarking. First, almost two-thirds of all trips (excluding walk-only) use transit, almost doubling the use of private car, which represents 35% of all trips. Within this sample of commuters, 45% of households report owning a car. Compared to the share of trips made using a car, this suggests that car substitution with transit is likely limited. Further, the table shows that 83% of transit trips rely on a minibus in some leg, and that 60% of these trips are comprised of potentially many modes and legs. These facts remark the importance of the private network and its role in the overall network. Minibuses are not only used as a last-mile service but rather as

a central vehicle that carries most of commuting—half of all trips in the economy.

TABLE 3—COMMUTING TRIPS ON A TYPICAL WEEKDAY IN 2017

Trips	Number	Share (%)
Total	60,559	100.00
Walk-only	8,435	13.90
Without 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 from INEGI's Encuesta Origen Destino 2017, which is a representative survey that describes trips made by commuters, with detailed information about the legs of each trip. This sample is representative of 6.4 million trips. Share are within the immediate parent 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.

3.3 Institutional context of market structure and fares

In this section I provide institutional context on the specific market structure of the private sector, which is broadly similar to other countries, as well as details about fare regulation and public transit subsidies.

Market structure of the private sector. In Mexico City's metropolitan area, private transit operates under a concession system known as *Sistema de Transporte Público Concesionado*. In this arrangement, the government gives permits—concessions—to private providers and allows them to operate in predetermined corridors. Bus owners in the majority of corridors often organize in associations, although a small number of them are formally constituted as firms. The government issues individual bus permits to these bus-owners, who decide whether to operate the vehicle themselves or 'rent' the vehicle to a driver. The day-to-day operation is thus carried out by individual drivers who typically do not own the vehicles they operate. The driver rents a minibus from an owner or association, and is required to pay a fixed daily quota (the so-called *la cuota*). The driver collects fares in cash throughout the day, and once the quota has been covered, the 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: either the State of Mexico or Mexico City. Minibuses rely predominantly on cash payment, so fare integration with the rest of the network is lacking.²⁶ The fare consists in

²⁶There are a handful of private operators that recently started accepting a mobility card, only within Mexico City only.

a base fare plus an additional per-km increase after some distance threshold. These two fare components are slightly different across state borders. 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, 6 pesos for trips between 5 and 12 km, and 6.50 pesos beyond 12 km. In practice, all trips are virtually charged the same amount, with little distance variation. The only notable fare difference is thus across state borders.

Public transit subsidies. Mexico City’s government subsidizes all its transit systems but most notably the subway.²⁷ The subway’s flat fare has been kept at 5 pesos per trip since 2013. This flat fare is invariant to the number of legs or distance traveled, so switching lines is free. 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 subway reported 1.17 billion trips, so a subsidy of 13 pesos per trip amounts to 15.21 billion pesos, or around 830 million US dollars. This represents roughly two-fifths of the total transportation budget and around 5% of the total local government’s expenditure.²⁸

4 QUANTIFICATION

To quantify the model 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 or vehicles, frequency, length (kilometers), and trip time, i.e. the time it takes to complete a full-length trip. These data are often very hard to get in developing countries, as private and often informal transit providers 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. Second, in addition to the transit network, we need the more standard data sources such as granular wages, rents, population, and commuting flows.

In this section I will first describe the network data collection, and then discuss the identification of the key elasticities and calibration of the rest of the parameters.

4.1 Transit network and route choice sets

Mexico City’s MA is one of the few developing megacities in the world that has detailed data on the private transit network.²⁹ Using the Google Maps Directions API, I simulated transit trips across all 192² origin–destination pairs to recover the geography and characteristics of roughly 80% of the transit routes in the system—around 2,000 in total. Characteristics include the name of the line, kilometer length, the time it takes to complete a full-length trip, and fre-

²⁷Sytems include Metro, Metrobús (BRT), Tren Ligero, Trolleybus, Buses RTP.

²⁸From *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.

²⁹A 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.

quency. Importantly, I backed out the number of units or vehicles operating in each line using the definition of frequency—the mass of vehicles divided by the trip time. Table 2 provides some descriptive statistics. An interesting fact is that private lines are on average longer in kilometer length, 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 1. 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. Finally, To account for own-district commuting, I simulated additional within-district trips to better capture local networks.³⁰

Route choice sets. For each od pair I take \mathcal{R}_{od} to be the finite set of alternatives returned by Google Maps and treat it as fixed for equilibrium computation. In the resulting sets, the average number of route alternatives across Google’s od choice sets is 4.6, ranging from a single alternative up to six. With these assumptions, I depart from optimal routing frameworks ([Allen and Arkolakis, 2022](#); [Fuchs and Wong, 2024](#); [Bordeu, 2023](#)) that either explicitly enumerate infinite paths [Allen and Arkolakis \(2022\)](#), or find the optimal routes and modes at each node recursively [Fuchs and Wong \(2024\)](#).³¹ 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 network information that describes stop nodes, switch nodes, and transfer rules—which is unavailable.³² Google’s algorithm already encodes stop locations, complex transfer penalties, timed transfers, and minimum walk times—unobservable elements. Therefore Google’s routes plausibly proxy the otherwise infinite route choice set ([Allen and Arkolakis, 2022](#)).

Second, recursive routing methods ([Fuchs and Wong, 2024](#)) would undermine tractability given my model’s assumptions. In this paper, the authors obtain tractability by specifying recursive routing with multiplicative (iceberg) edge-level costs and a nested mode choice, where the effective edge cost aggregates mode-specific costs. Under these assumptions, any market-level price index enters multiplicatively into every route that uses that edge, and by recursion, propagates to all continuations. As a result, in my model, a firm’s residual demand would depend on the full vector of prices and path compositions across all markets, undermining the market separability, and thus the computational tractability. By contrast, I work with an exogenous, finite \mathcal{R}_{od} and additively separable route-level costs. Firm-level demand thus depends only on a firm’s own price, and the market-specific price index and demand: a computationally tractable system as described in section 2.9.

³⁰In 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. I direct the interested reader to Appendix C.1 for more details about the collection of data and processing.

³¹Such recursive routing models find optimal paths by solving a Bellman-type problem at each node of the transport graph. This approach is less tractable in this setting due to the model’s additive cost structure and the need to maintain market-level separability for the firm’s problem.

³²Such data is found in a General Transit Feed Specification (GTFS), which is a standardized data format that stores all transit information. The source GTFS from Google is not publicly available.

Road network. To account for congestion at the road-link level, 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.

4.2 Identification of elasticities with natural experiment

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. There are two key challenges to identify these parameters: data availability and exogenous variation in costs. It is very hard 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 I focus on a mapping of flows: we can observe how traffic—and in particular speed—changes at the road-link level following a change in the cost of some routes. The second challenge is to obtain plausibly exogenous variation in route costs. To overcome this challenge, I exploit exogenous variation generated by the collapse of a subway line.

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

Mapping experiment to the model. The goal is then, to simulate a subway shock in the model ($P_{\text{metro}} \rightarrow \infty$) 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, and ultimately road-link-level speed. Then, I compare the speed moments generated by the model to the data, and find the vector of parameters (ρ, ϕ) that best fits the data. 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.³⁴ Figure 4 shows the distribution of the changes in speed for a sample of road links near the subway line. Average speed across links decreased $\approx 6\%$.³⁵

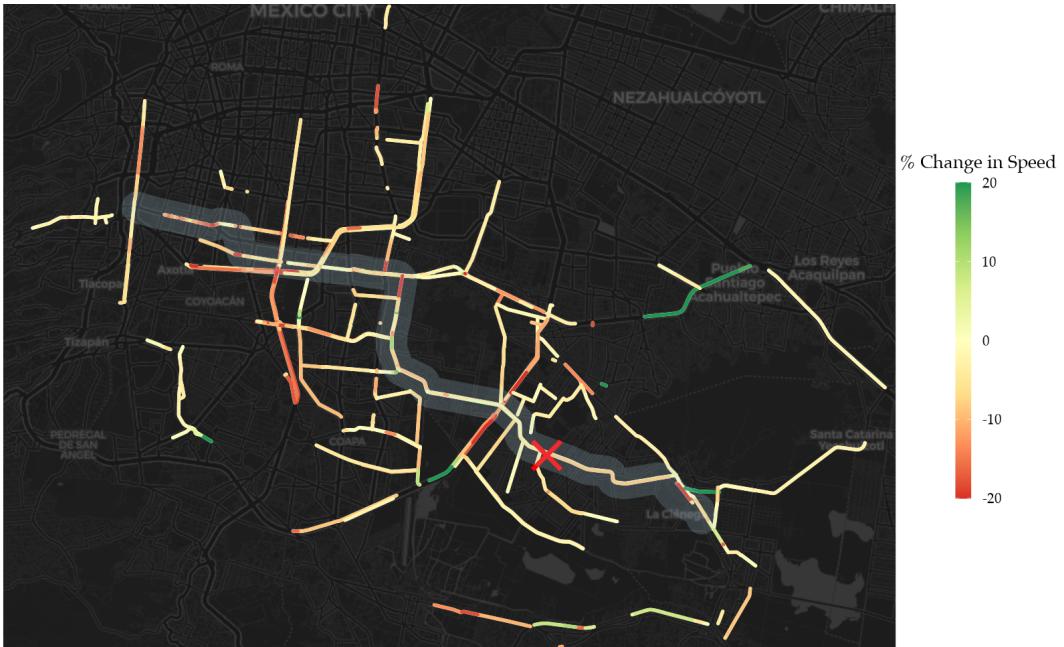
What moments are most informative about substitution of routes ρ relative to congestion ϕ ? The core intuition of the identification is the following. After the shock, on a road link that is part of the *only* remaining route for an *od* pair, any change in speed must primarily reflect the congestion effect (ϕ) of all diverted commuters crowding onto that single option. In contrast, on links that are part of *od* pairs with *multiple* remaining routes, the relative changes in speed

³³More details about this shock can be found in Appendix D.1

³⁴Full details on the data and processing, can be found in Appendix D.2

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

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 red represents the site of the collapse. The subway line is depicted as the shaded buffer. Data was obtained from TomTom Traffic Stats API.

across those alternatives also reveal how commuters substitute between them (ρ).

To try to isolate variation that identifies congestion and substitution, I thus 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. After the collapse of the subway these routes become impassable, so the set of alternatives is reduced. Depending on the substitution strength ρ , the flow of commuters will distribute 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 remove variation coming from substitution of routes, given that by definition, there is no substitution in a single route.³⁶ 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.³⁷

Table 4 summarizes the identified elasticities and the model fit, which does a reasonable job at

³⁶Note that we cannot fully purge the variation coming from ρ because the flow on these singleton routes still depends on the pre-shock level of the conditional route flow $\lambda_{r|od}$, which in the model is a function of the substitution elasticity.

³⁷Although I make robustness check for 1.5, 3, and 20km buffers, and parameters do not change substantially. Table 8 in the shows these results.

TABLE 4—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.

matching the target quantile moments. Almost all the differences between model and target moments fall within a 2% error, with the exception of the largest 0.9 quantile for non-singleton links, that has a 73% error.

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 just as sensitive.

Limitations of this identification approach. In this exercise, I attempted to identify the elasticity 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. 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—located in the South-East and 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 than the Western and South-Western tracts of the city, that roughly range between 50% and 100% of the vehicle ownership share.³⁸ This is suggestive evidence that car-related action is likely limited.

³⁸Figure 15 in Appendix D.3 shows the spatial heterogeneity.

4.3 Data sources and calibration of parameters

I assemble a rich battery of microdata to calibrate the rest of the parameters. Data for wages comes from the INEGI Economic Census 2019; residential and commercial land use data comes from the Urban Planning Ministries of Mexico City and the State of Mexico (SEIDUVI and SEIDU); residential rents from INFONAVIT (transaction-level home sales, 2018–2020); geographies, population, and commuting flows from the 2017 Encuesta Origen–Destino (EOD); expenditure shares from the 2018 Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH); time endowments from the 2019 Encuesta Nacional de Uso del Tiempo (ENUT).

Geography, population, and commuting flows. I define a location to be the districts from the most recent origin-destination travel survey, Encuesta Origen Destino. I set the size of the economy $\bar{L} = 3,608,702$ 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. I average tract-level workplace wages from the Economic Census to the EOD district level. To obtain residential and commercial land I aggregate from the tract level to the district level. The price of residential land comes from aggregating transaction-level home prices sold from 2018 to 2020 to the district level.

Fundamentals inversion. To invert the model to obtain fundamental productivities A_d and 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 via transit $\alpha_c = 0.09$ using data from INEGI's Encuesta Nacional de Gasto e Ingreso de los Hogares (ENIGH) 2018.³⁹ 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. I take the value of $\theta = 2$ from [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, I develop a version of the model where prices of individual firms $p_{i,\varphi}$ cannot adjust, and instead are fixed exogenously to a uniform price \bar{p}_φ . In Appendix A.5 I describe this

³⁹The data describes several expenditure categories for housing and transport. For housing, I include the categories related to rent and mortgage payments. For transport, I include the transit expenditure and exclude car-related expenses.

model, although the only difference is that the price cannot adjust, so all adjustment happens through entry, trip time and wait adjustment. This version of the model corresponds to the ‘observed world’, and cost parameters are calibrated to replicate the observed outcomes in this world: entry, frequency, trip time, and price.

To calibrate the fixed costs of entry f_φ^e , vehicle capacities q_φ^c , 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 firm could perform in a day (based on trip times and firms’ time endowment) and compare it to the number of trips required to satisfy demand at a baseline capacity $q_0^c = 15$ 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 $\{q_\varphi^c\}_\varphi$ 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_φ^e 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.

Elasticity of substitution across firms. I set $\chi = 15$ to approximate the fact that minibuses are likely highly substitutable while retaining computational tractability.⁴⁰

TABLE 5—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_φ^c	Capacity constraint		Rescaled to match observed entrants and demand
γ_{odr}^φ	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_φ^e	Fixed costs of entry		To match observed p, M, t from Google Maps
\tilde{t}_φ	Trip time shifters		To match observed travel times from Google Maps
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).

Trip time shifters and shares of market usage. From the information provided in the Google

⁴⁰Results do not change quantitatively or qualitatively if we set, $\chi \in [10, 20]$.

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. For example, if I tell Google how to get from A to B, it will tell me for each alternative, which line I need to use, for how much time, and the overall length in meters of that portion of the trip. With this information I can compute, for each route, the share of market usage γ_{odr}^φ . So, for example, if a route alternative uses a market for only 25% of the full length of that market, then $\gamma_{odr}^\varphi = 0.25$. Then, having recovered the trip time (t_φ^{trip}) 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}_φ that exactly match the observed time in that market, given the observed number of entrants.

Table 5 summarizes the parameter calibration.

5 ENHANCING WELFARE THROUGH PRICE-SHIFTING POLICIES

I exploit the unique features of my model to evaluate the welfare and spatial effects of two policies currently present in this setting: a uniform fare regulation in the private sector and a subway subsidy. These policies are not unique to the Mexican context; rather, they are applied in many other contexts with subtle differences.⁴¹

5.1 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. That is, $p_{i,\varphi} = \bar{p}_\varphi, \forall i \in \varphi$, where \bar{p}_φ is set exogenously and fixed. The only difference between this version of the model with respect to the general framework is that price adjustments are shut down, so all the adjustment comes from entry, trip and wait times.⁴² Now, allow the full adjustment of prices in general equilibrium, that is, according to equations (15) and (19) in the model.

Welfare effects. Welfare increases $\approx 0.9\%$ by removing the fare regulation. This effect is driven mostly by a generalized decrease in commuting costs, behind which there is substantial heterogeneity across markets. To see this, first let me decompose the change in the welfare expression given in equation 10 into the contribution of each of its elements as

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

where each term represents the contribution of each element to the total change in welfare: the change in disposable income, rents, and commuting costs. Note that a decrease in τ means that commuting costs decrease, so contributes positively to welfare. Table 6 shows the decomposition of welfare following price deregulation in the private sector. Mostly all of the change in welfare is coming from a decrease in commuting costs, which in turn are driven mostly by

⁴¹ Examples include Argentina, Bangladesh, Brazil, Chile, Colombia, India, Indonesia, Mexico, Philippines, South Africa, Tanzania. For a more detailed description of these examples, see Tables 9 and 10 in Appendix E.

⁴² Such extension of the model is explained in Appendix A.5.

a reduction in prices, but also by reductions in trip times and wait times, representing 80%, 18%, and 2% respectively of the overall commuting cost contribution. There is relatively little action coming from disposable income and rent adjustment, contributing negatively 5% and 3% respectively, 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.

TABLE 6—WELFARE DECOMPOSITION: DEREGULATE PRICES

Component	Percent (%) change
$\Delta\bar{W}$	0.94
$\Delta\tau$	1.03
ΔP	0.81
Δt^{trip}	0.19
Δt^{wait}	0.02
$\Delta\tilde{w}$	-0.05
ΔQ	-0.03

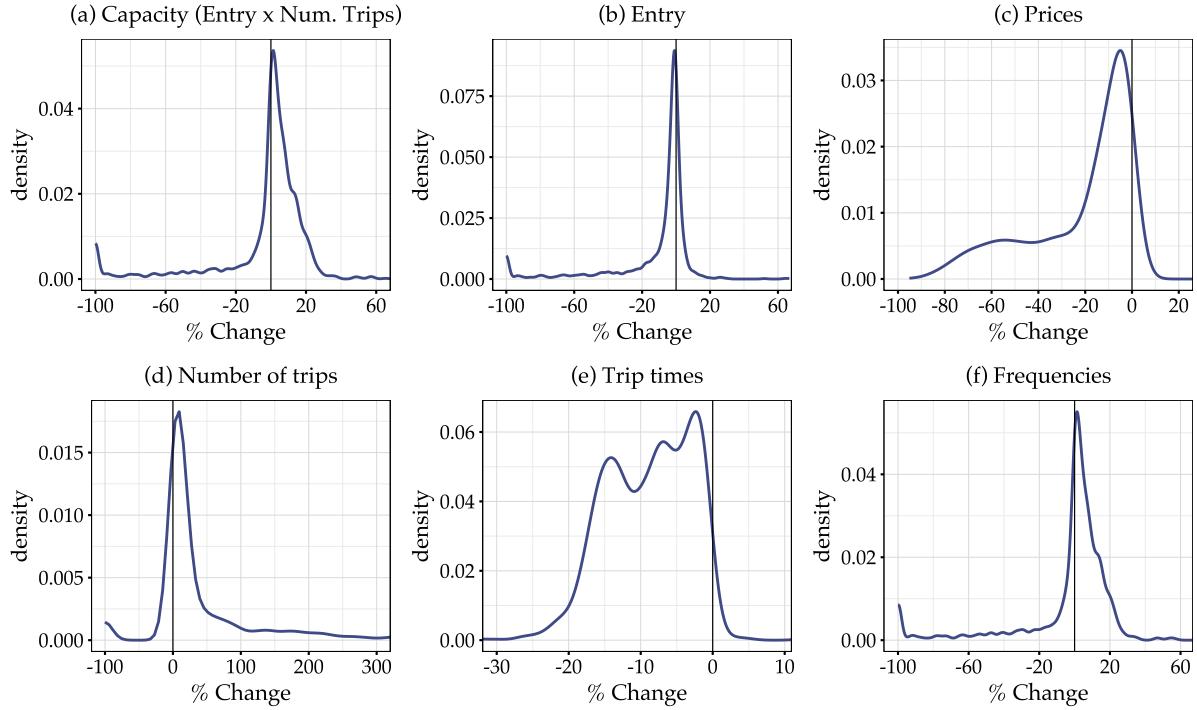
Note: Table shows the welfare change decomposition following the policy change into the contributions of each of its elements: commuting costs and the subcomponents price, trip and wait time; disposable income, and rents.

For example, one could start by stating that deregulation allows a reallocation of service, distinguishing between the number of vehicles on the road (entrants) and the actual service they provide (market capacity). This framing would preempt potential reader confusion and immediately highlight one of the model’s most interesting trade-offs: fewer buses can, in equilibrium, provide more service if they can move faster.

Effects across markets. Figure 5 shows substantial heterogeneity across markets’ characteristics. Deregulation allows for a reallocation of market capacity, which includes the number of vehicles on the road (entrants) and the actual capacity they provide: the product of individual capacity and the number of trips they complete. Panel (b) shows that there are both increases and decreases in entry across markets, but mostly falls. However, the number of trips that each entrant completes predominantly increases, as shown in panel (d). Combining both entry and number of trips, in panel (a) it can be observed that there are overall increases in market capacity. Further, given that by market clearing capacity reflects demand, we can interpret this result as capacity increases to meet surging demand.

Prices mostly fall with substantial heterogeneity, as shown in panel (c), with most of changes concentrated within a 10 percentage-point fall. A few markets experience slight increases in prices. Not shown in panel (c), however, are infinite increases in the price indexes from a reduced number of markets with zero entry. These markets become deserted as demand readjusts, so there are no incentives to enter. This is reflected in panel (b), where entry drops 100%. Trip times, shown in panel (e), follow the same pattern as prices as they increase slightly in a reduced amount of markets but mostly decrease across markets, improving in the order of up to 30%. Wait times improve in the majority of markets due to frequencies mostly improving, as shown in panel (f). Summing up, price and time decreases is what explains the overall decrease in commuting costs.

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 capacity in the market, which is composed of the total amount of entrants (b) multiplied by the number of trips (d) that each unit completes during their time endowment or shift. For the number of trips, for visualization purposes I zoomed-in up to the x-axis limit of 300, although there little extra mass beyond this threshold due to extreme outliers. Panel (c) shows the changes in price indices, i.e. the CES aggregates of individual prices, and excludes infinite values from markets with zero entry. Panel (e) shows the changes in the time it takes to complete a trip. Panel (f) shows the changes in frequencies, which are defined as the mass of entrants over trip time.

What are the economic mechanisms that explain these changes? Trip times fall when congestion eases with fewer entrants. Perhaps counterintuitively, capacity increases despite fewer entrants because faster trips allow each entrant to complete more trips per shift, meeting higher demand where it arises. This result highlights one interesting model trade-off: fewer buses can, in equilibrium, provide more service if they can move faster. Because the firm's marginal cost to complete a trip depends primarily on trip time, this leads to decreases in prices, outweighing the negative price effect due to less entry. Given the relatively large value of the elasticity of substitution across routes, as prices and trip times fall in some markets more than others commuters shift towards these better alternatives, resulting in a small number of markets being left with little demand. This is why we observe some entry and capacity going towards zero. Service frequency is improved on average. The frequency-congestion trade-off dominates in favor of commuters: while there are on average less entrants, alleviated congestion on the road improves trip times, resulting in increased frequencies and therefore lower waits.

Which markets gain and lose customers? Figure 6 shows the change in flows across markets. It can be seen that peripheral markets are the ones that primarily experience an increase in flows,

FIGURE 6. CHANGE IN FLOWS ACROSS MARKETS FOLLOWING FARE DEREGULATION



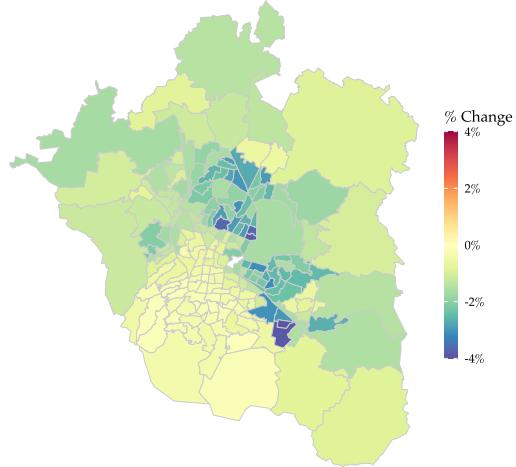
Note: Figure shows the change in flows at the market level (aggregating across all routes that use a given market), following the policy counterfactual in the model. The thickness of lines represents the absolute value of the percent change of the flow, and the color denotes whether it was a positive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

particularly in the Eastern part of the city. On the other hand, markets that are located in more central areas, and some of the markets that connect towards central areas, lose customers. This is explained because central markets' prices are relatively higher following the deregulation. Because commuters are sensitive to changes in relative route costs, and the periphery is primarily served by private markets that improved service, it becomes relatively cheaper to commute within and across these peripheral districts. Further, these peripheral markets that gain flows are on average shorter in length than markets that go towards the center or that are located in the center.⁴³ This suggests that because these markets are shorter and thereby have lower trip times, the cost to operate them is also lower, which is reflected in relatively lower prices.

Commuting flows increase among outskirt districts and overall economic activity becomes more decentralized, as commuting becomes relatively cheaper for peripheral 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 peripheral locations tend to be less productive and with lower amenities relative to central locations, as

⁴³This is shown in additional figure 16 in Appendix E

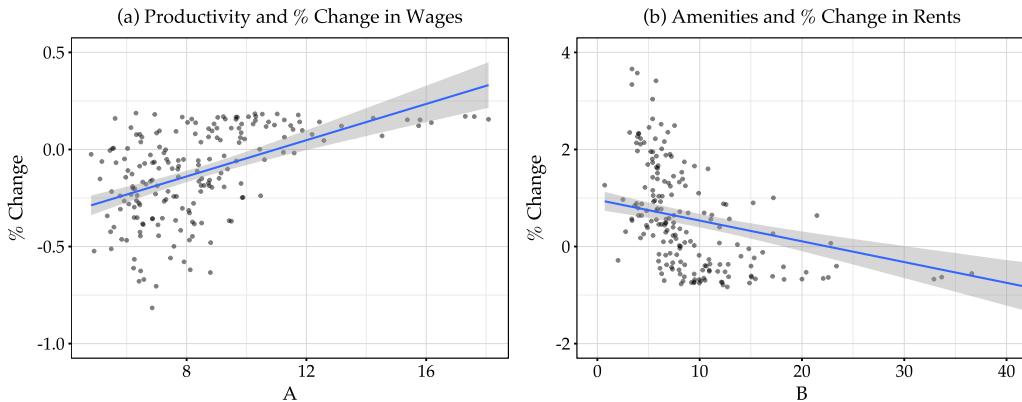
FIGURE 7. COMMUTING COST INDEX CHANGE FOLLOWING DEREGULATION



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

reflected by the fundamental productivity A and a measure of local amenities $B_o = \sum_d B_{od}$.⁴⁴ 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 pressure, and on the other hand that rents increased as well in those locations due to housing demand pressure. The overall wage and rent effects on welfare are very limited, as shown before, though.

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.

Taken together, the results from this exercise suggest that removing price regulations could

⁴⁴The spatial distribution of the productivity and amenity measures can be found in figure 18 in Appendix E.

improve efficiency in the economy by realigning prices with costs, improving service characteristics such as prices, frequencies, and trip times; and lowering congestion. In terms of the commuting access benefits, these are mainly enjoyed by residents in the outskirts of the city—locations characterized by lower productivity.

5.2 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, remove a 72% subway fare subsidy.⁴⁵ That is, increase the price of the subway from $P_{\varphi_{\text{metro}}}$ to $P_{\varphi_{\text{metro}}}^*$, where the latter represents the price without the subsidy and the former the actual price, and $\frac{P_{\varphi_{\text{metro}}}^* - P_{\varphi_{\text{metro}}}}{P_{\varphi_{\text{metro}}}^*} = 0.72$. Further, the removal of the subsidy implies that there is no longer a need to fund such subsidy, so the income tax rate is set to $\eta = 0$.

Welfare Effects. Welfare increases $\approx 0.5\%$ following the subsidy removal. Relative to deregulating private-sector fares, the welfare change in this counterfactual is driven by substantial changes in all three welfare components: commuting costs, disposable income, and rents. Table 7 shows the decomposition of welfare into these components. Commuting costs contribute negatively to welfare in 0.77 percentage points, primarily affected by the direct increase in subway prices (0.72 pp) but notably also by an increase in trip times (0.05 pp). Note that in a model without the endogenous adjustment of entry and times, the extra $5/77 = 6.4\%$ adjustment of commuting costs via times would be overlooked. Further, even if wait times would seem to not contribute whatsoever to the change in commuting costs, as I will explain briefly, this is because there is a large heterogeneity in service responses across markets that happen to cancel each other out in the aggregate.

TABLE 7—WELFARE DECOMPOSITION: SUBSIDY REMOVAL

Component	Percent (%) change
$\Delta \bar{W}$	0.49
$\Delta \tau$	-0.77
ΔP	-0.72
Δt^{trip}	-0.05
Δt^{wait}	0.00
$\Delta \tilde{w}$	1.67
ΔQ	-0.40

Note: Table shows the welfare change decomposition following the policy change into the contributions of each of its elements: commuting costs and the subcomponents price, trip and wait time; disposable income, and rents.

Perhaps counterintuitively, welfare increases slightly in the economy following the subsidy removal because the increase in disposable income due to a zero tax rate outweighs the negative effect of commuting costs and rents. As residents get the subsidy burden rebated, they are able to consume more of all goods in the economy, which is captured in the welfare expression

⁴⁵This is the magnitude of the subsidy reported by authorities, which 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.

via real income—disposable income divided by rents, adjusted by commuting costs. This increase in disposable income, however, is what contributes to the increase of 0.40 pp in rents. These general equilibrium effects have not been documented before. None of the recent papers studying new transit infrastructure (Tsivanidis, 2019; Zárate, 2024; Bordeu, 2023; Khanna et al., 2024) consider the general equilibrium effects of actually funding such infrastructure improvements. The results presented here suggest that it matters how we fund transit interventions for welfare assessments.

Effects across markets. The aggregate welfare effects, however, hide substantial heterogeneous impacts across markets and space. 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. 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 alternatives to subway-using routes 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:

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

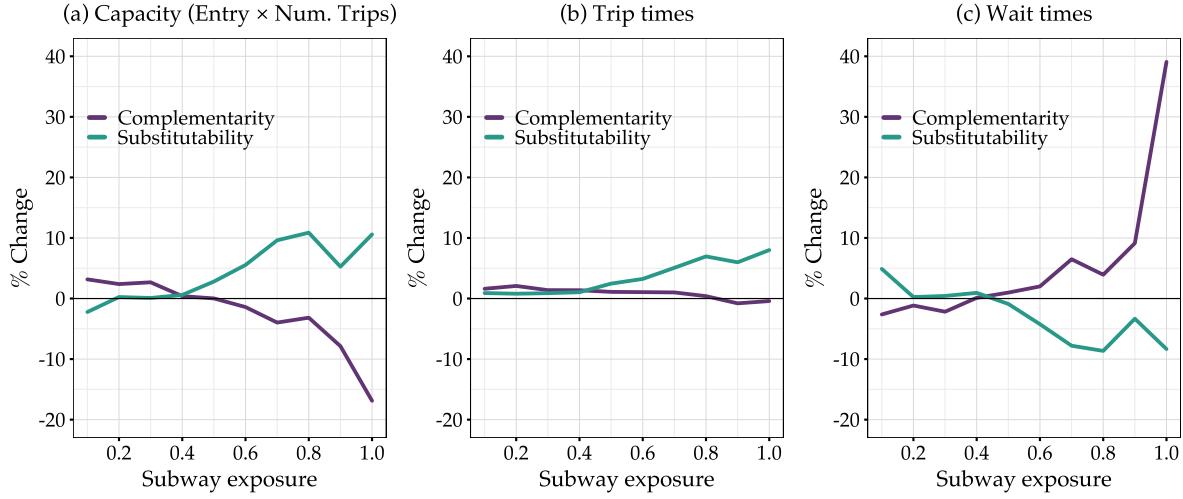
$$\text{Expo}_{\varphi}^{\text{comp}} = \frac{\sum_{(o,d)} \sum_{r \in \mathcal{R}_{od}} \mathbf{1}\{\varphi \in r\} \mathbf{1}\{M(r) = 1\}}{\sum_{(o,d)} \sum_{r \in \mathcal{R}_{od}} \mathbf{1}\{\varphi \in r\}}$$

- *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., if the od set 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 shows the effects across private markets on capacity, 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

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.

can be quite substantial for more exposed markets.

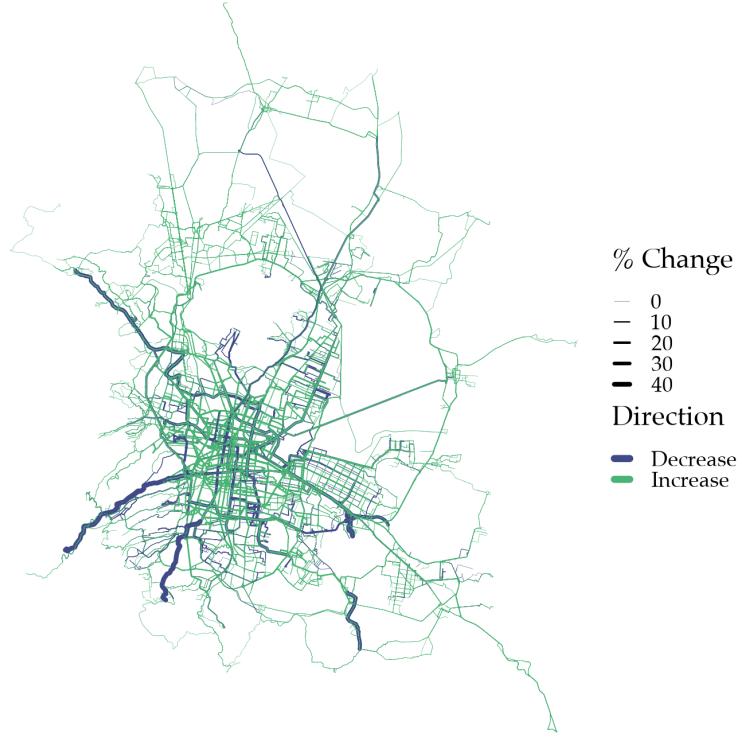
Substitute markets experience a gain in customers due to relative prices changing and substitute towards alternative routes, leading up to 10% increases in entry. Entry, however, leads to congestion, which is manifested in an increase in trip times up to 7%. Frequencies improve as well in these markets, improving wait times by around 10% as well. These effects are consistent with [Björkegren et al. \(2025\)](#), who find the analogous effect in private markets in Lagos following the introduction of a BRT system: public entry displaces private markets, so frequencies drop and wait times surge.⁴⁶

Complementary markets, on the other hand, experience a drop in demand leading to capacity reductions of up to 17%. Most notably, wait times in highly exposed markets increase up to 40% as they become deserted and frequencies drop. Because of the cost-complementarity with the subway, these exposed markets experience a drop in demand that is proportional to the large elasticity of substitution across routes. 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. 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. These results complement [Björkegren et al. \(2025\)](#), which by design emphasize corridor-level over city-wide responses. Moreover, these findings suggest that private transit is not only a last-mile provider but instead can complement and substitute public transit in meaningful ways.

In terms of the spatial effects across markets, figure 10 shows the change in flows across mar-

⁴⁶Introducing a new line is analogous to *decreasing* the price of this public mode.

FIGURE 10. CHANGE IN FLOWS ACROSS MARKETS FOLLOWING SUBSIDY REMOVAL



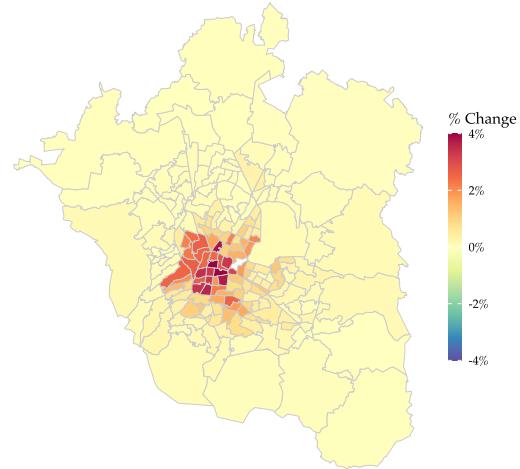
Note: Figure shows the change in flows at the market level (aggregating across all routes that use a given market), following the policy counterfactual in the model. The thickness of lines represents the absolute value in the percent change of the flow, and the color denotes whether it was a positive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

kets. As can be seen, markets located in central areas are the ones that gain customers, reflecting substitution away from the now more expensive subway-using routes. The lines in central areas that appear with decreasing flows are precisely those corresponding to the 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.

Economic activity becomes more decentralized as a result. This is mainly because central areas are served by the subway and also are prone to congestion due to the presence of private markets. Then, as the subway price increases and substitution towards private markets gener-

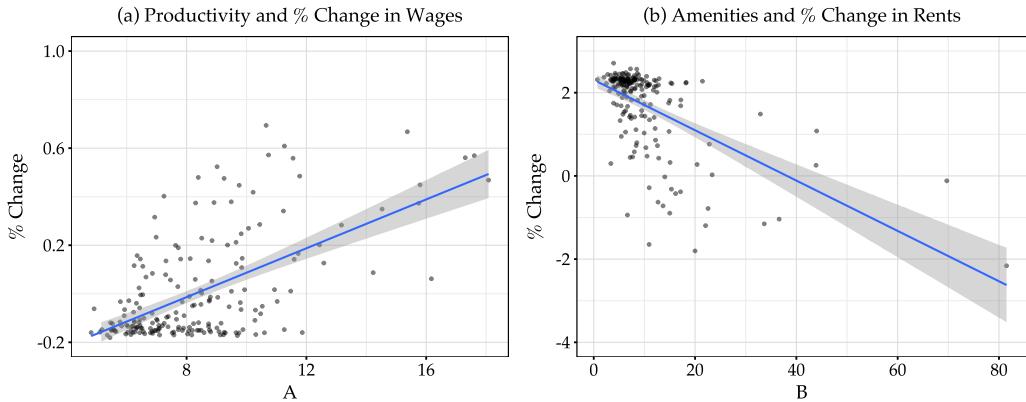
FIGURE 11. COMMUTING COST INDEX CHANGE BY ORIGIN: SUBSIDY REMOVAL



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

ates congestion, it becomes relatively more costly to move within the center and towards the center. 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 both as residences and workplaces. As a result, less productive, peripheral locations, gain workers, labor supply pressure reduces 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 $B_0 = \sum_d B_{od}$ and rent changes.

FIGURE 12. FUNDAMENTALS AND PRICE CHANGES FOLLOWING SUBSIDY ELIMINATION



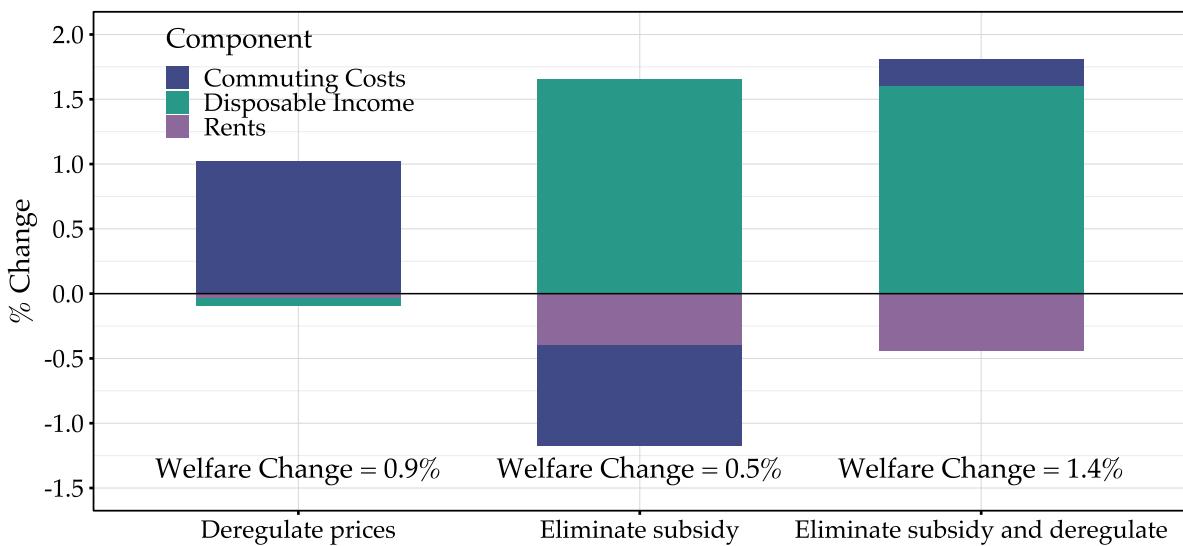
Note: Figure shows the correlation between location fundamentals and the change of local prices. Every dot is a location. Panel (a), shows that there is a positive relationship between local productivity and changes in wages. In panel (b), there is a negative relationship between local (average) amenities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

5.3 Counterfactual 3: remove private fare regulation and subway fare subsidy

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. Note that the first two bars that refer to the first two counterfactuals, correspond to the numbers presented in tables 6 and 7.

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. This magnitude is substantially larger relative to implementing the two policies alone; roughly 1.5 to 3 times, respectively. Intuitively, this is because after removing the subsidy there is also a more flexible private adjustment through all margins, i.e., entry, prices, and times, rather than the more limited adjustment 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 attributes in the private sector adjust in such a way that overcomes the negative price effect.

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.

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 with 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.

5.3.1 Discussion

In terms of the welfare 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 quantitative spatial 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 it is not. His model also comes from the same class of models although with slight variations in the assumptions on migration decisions. Therefore, their results on welfare 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.

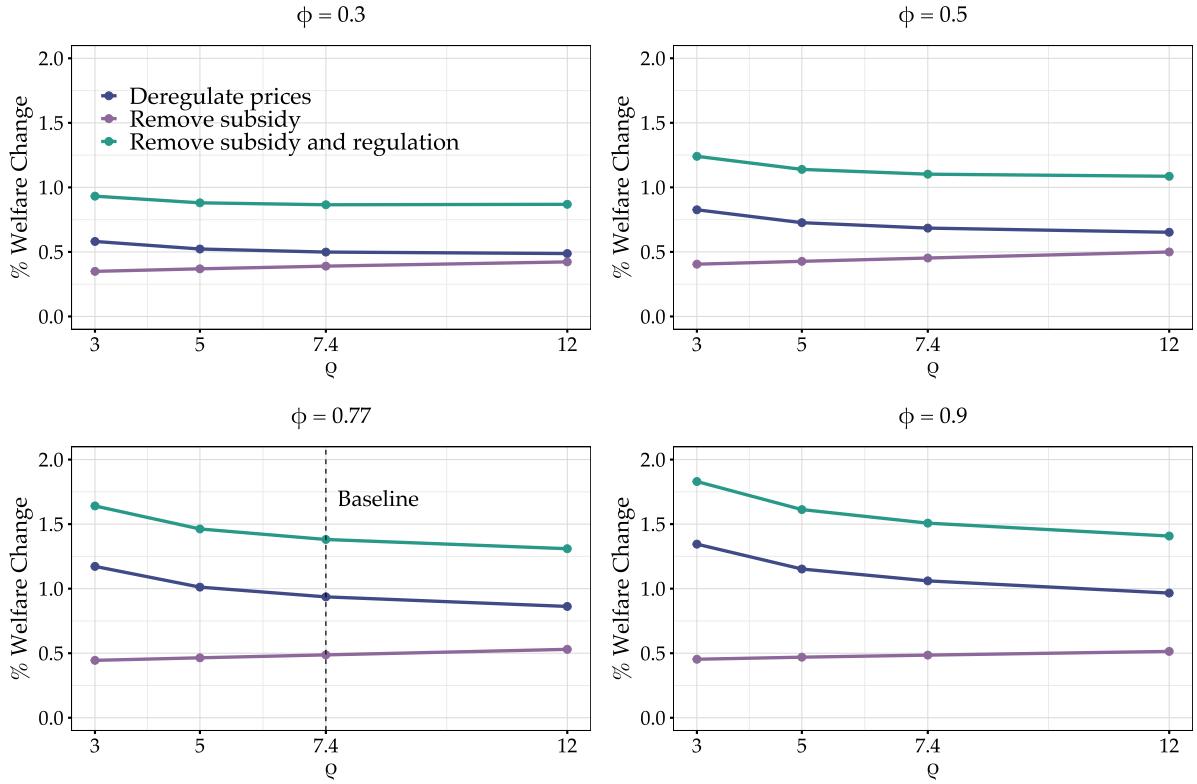
Overall, I show that non-infrastructure policies that shift prices and directly or indirectly affect private transit providers can generate substantial changes in welfare, comparable to those that the literature studying infrastructure improvements has found.

6 RESULTS UNDER ALTERNATIVE PARAMETRIZATIONS

In this section I explore the sensitivity of the welfare changes reported in the previous section. 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.

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

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.

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 following the robustness exercises performed in Section 4.2, the overall welfare qualitative and quantitative change should remain stable.

7 CONCLUSION

This paper develops a quantitative spatial framework in which commuting costs are endogenously determined due to a private transit sector that interacts with a broader public network, and responds to demand. Two mechanisms emerge as central. First, because of the nature of multimodal trips, there is a within-route complementarity: a cost change in one leg affects demand on connected legs depending on the elasticity of substitution across routes, potentially amplifying local interventions across the network. 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 the effect depending on the elasticity of congestion. 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 adjustments in entry, prices, and times. First, deregulating private fares increases welfare by about 0.9%, largely by decreases in prices that realign with heterogeneous costs, service improvements in frequency and waits, and improved trip times through reduced congestion. Second, removing a subway fare subsidy increases welfare by roughly 0.5% as the direct negative price effects on commuting cost and increases in congestion due to riders reallocating toward private alternatives is out-weighted by increases in disposable income due to a rebate of the subsidy burden. Lastly, removing both the fare regulation and subway subsidy increases welfare by 1.4% while saving fiscal resources, underscoring that network interactions and private supply responses effectively shape impacts. Although these policies act on different parts of the network—so their aggregate welfare effects are nearly additive—each relies on strong within-policy interactions.

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 prices and time costs. 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 (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, CE-Sifo 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	50
A.1 Commuting flows and commuting cost index	50
A.2 Proof of complementarity proposition	52
A.3 Firm-specific CES demand	53
A.4 Proof of frequency-congestion trade-off proposition	54
A.5 Model with price regulation	55
B Algorithm to solve general equilibrium	56
C Data and calibration	57
C.1 Transit network data collection	57
C.2 Inversion of od-level amenities B_{od}	59
C.3 Inversion of productivities	59
D Identification of parameters from subway shock	59
D.1 Details of the subway Line 12 collapse	59
D.2 TomTom data details and processing	60
D.3 Robustness of SMM exercise	61
E Additional tables and figures	62

A MODEL DERIVATIONS

A.1 Commuting flows and commuting cost index

This section derives the route-level choice probabilities under the nested Fréchet structure in (3)–(6) and the CES commuting cost index (8). Define the deterministic component of utility for option (o, d, r) as:

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

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

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

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

$$\{U_{odr'} \leq U_{odr} \ \forall r' \neq r\} \iff \{\varepsilon_{odr'} \leq (a_{odr}/a_{odr'}) \varepsilon \ \forall r' \neq r\}.$$

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

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

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

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

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

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

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 thresh-

olds for all other pairs to $+\infty$) yields

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

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 $\{x_{od}\}$,

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

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 $a_{odr} = (B_{od}\tilde{w}_d)/(Q_o^{\alpha_h}\tau_{odr})$,

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

Hence

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

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

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

This result can also be directly obtained (and perhaps more easily) by applying the Generalized Extreme Value Theorem, using the nested Fréchet as the correlation function.

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:

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

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 (*):

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

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

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

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

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

because $0 \leq \Lambda_\varphi \leq 1$ and $\bar{s}_\varphi \geq 0$.

□

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

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

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

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 Firm-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

$$C_{odr}(\omega) = \min(c_1, \dots, c_\varphi, \dots) \text{ as } \sigma \rightarrow 0$$

with an associated price index $P_{odr} = \sum_{\varphi \in odr} P_\varphi$. Optimal consumption in a Leontief bundle implies that $c_1 = \dots = c_\varphi$, 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 $T_\psi(M) \equiv \sum_{\ell \in \psi} S_\ell(M)^\psi$ 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,

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

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

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 $n_\varphi^{\max} \equiv \frac{T^d}{t_\varphi^{\text{trip}}}$ 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

$$\begin{aligned} \max_{n_{i,\varphi}} \pi_{i,\varphi} &= \bar{p}_\varphi q_\varphi^c n_{i,\varphi} - \delta t_\varphi^{\text{trip}} n_{i,\varphi} - f_\varphi^e \\ \text{s.t.} \quad q_{i,\varphi} &= M_\varphi^{\frac{\chi}{1-\chi}} D_\varphi \\ \frac{q_{i,\varphi}}{q^c} &\leq n_{i,\varphi} \\ n_{i,\varphi} t_\varphi^{\text{trip}} &\leq \bar{T}^d \end{aligned}$$

The price index is given by

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

Profits under price regulation are

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

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 $\vec{x} = (\vec{A}, \vec{B}, \vec{H}, \vec{H}^c, \vec{L}, \vec{T}, \vec{T}^d, \alpha_h, \alpha_c, \beta, \theta, \rho, \phi, q_\varphi^c, \delta, \chi, \vec{t}_\varphi, \vec{f}_\varphi^e)$

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

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

- (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

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

- 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

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

- **Transportation 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

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

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

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

- 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 AGEBS 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 ~ 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 recover A_d from the inverse demand equation for labor:

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

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 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. This introduction of public buses is not explicitly taken into account in the model, but rather it is implicitly absorbed into the private entry margin. This is because there are multiple private lines that coincide with the corridor of these emergency lines, therefore from the perspective of the model it does not matter whether a bus is labeled private or public as long as it captures the increase in flows. 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.

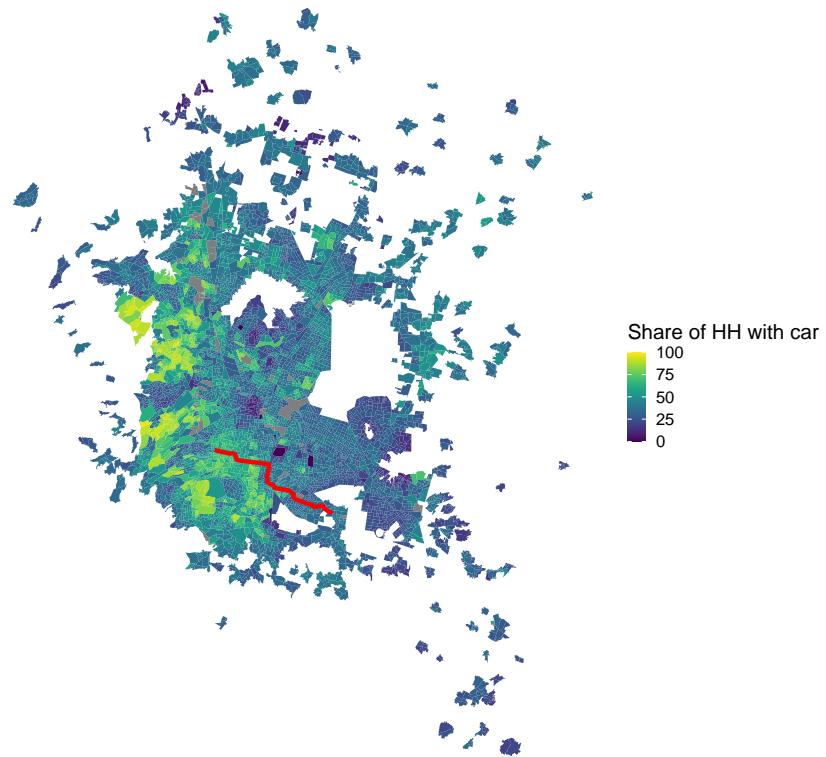
D.3 Robustness of SMM exercise

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.

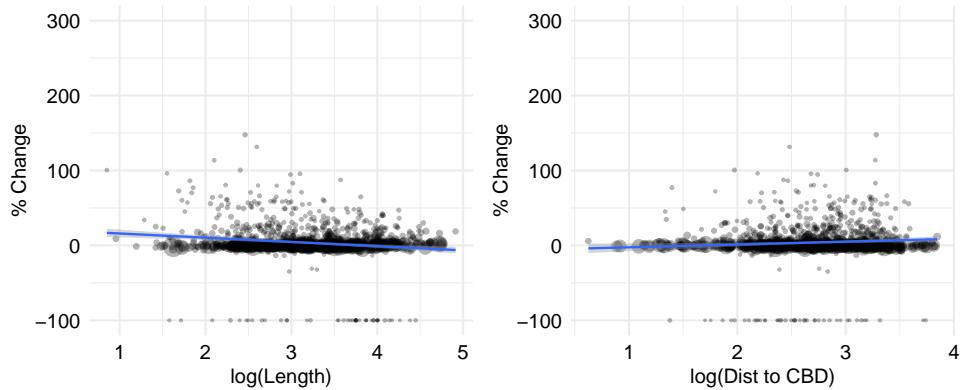
FIGURE 15. CAR OWNERSHIP BY HOUSEHOLD LIVING NEAR THE COLLAPSED SUBWAY LINE



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-East region suggests low substitution towards private cars, following the subway collapse.

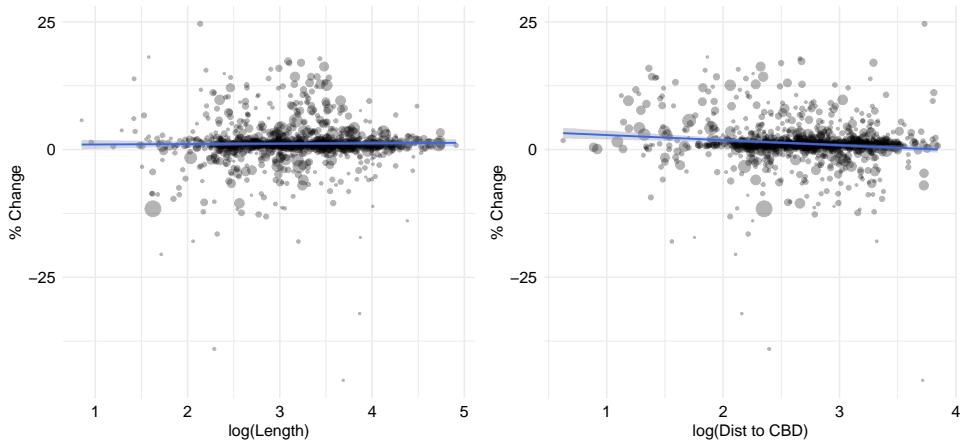
E ADDITIONAL TABLES AND FIGURES

FIGURE 16. CORRELATION BETWEEN MARKET CHARACTERISTICS AND CHANGE IN FLOWS:
PRICE DEREGULATION



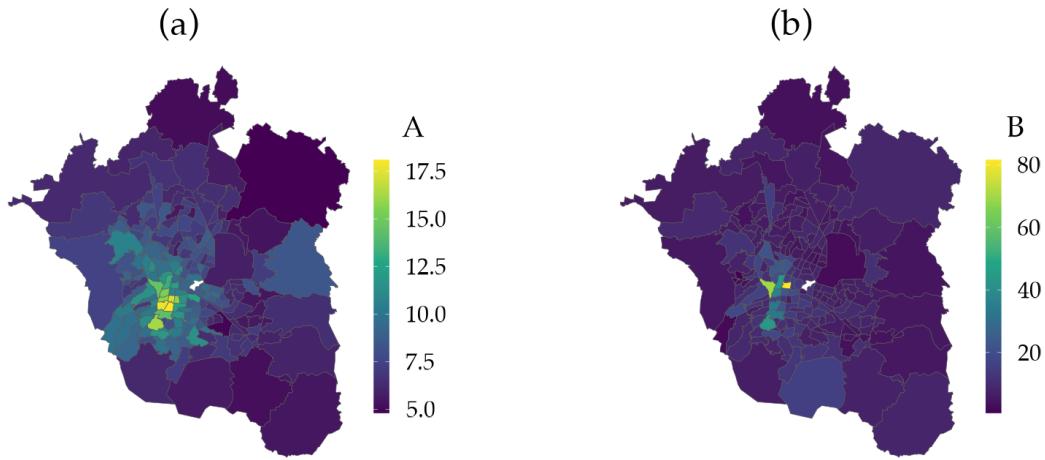
Note: Each dot represents a market. Dot size represents the pre-policy flow through that market. In the left panel, I am correlating length (kilometers) of the transit line and the change in flow. In the right panel, I am correlating the average distance of the transit line (across different segments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

FIGURE 17. CORRELATION BETWEEN MARKET CHARACTERISTICS AND CHANGE IN FLOWS:
METRO SUBSIDY REMOVAL



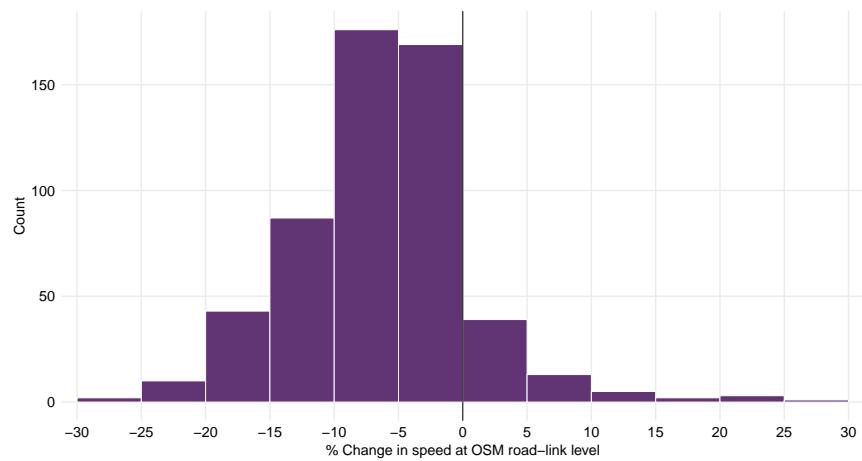
Note: Each dot represents a market. Dot size represents the pre-policy flow through that market. In the left panel, I am correlating length (kilometers) of the transit line and the change in flow. In the right panel, I am correlating the average distance of the transit line (across different segments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

FIGURE 18. SPATIAL DISTRIBUTION OF FUNDAMENTALS



Note: Figure shows the distribution of location fundamental measures of productivity A_d and amenities $B_o = \sum_d B_{od}$.

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 <i>minimum</i> fare	LTFRB (national regulator)	Provisional +1 increase effective 08-Oct-2023 set min. at P13 (traditional) / P15 (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 <i>per-km</i> 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)	<i>Operator-set fares (no fixed government fare schedule); licensing/route oversight</i>	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)	<i>Temporary mandated discount</i>	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)	<i>Proposed fare regulation (draft legislation / county rules); currently operator-set</i>	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)	<i>Permanent operating subsidy</i> (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 <i>free-pass loads</i> 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)	<i>Large recurring operating subsidy</i>	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	<i>Free travel</i> 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	<i>Temporary fare subsidy</i> (–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.