Title

Essays in Trade, Urban, and Spatial Economics

ISBN

9798180196873

Author

Mosqueda Juarez, Jordan

Publication Date

2026-07-24

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays in Trade, Urban, and Spatial Economics

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Economics

by

Jordán Mosqueda Juárez

Committee in charge:

Professor Fabian Eckert, Co-Chair
Professor Marc Muendler, Co-Chair
Professor Nir Jaimovich
Professor Craig McIntosh
Professor Fabian Trottner

2026

The dissertation of Jordán Mosqueda Juárez is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2026

DEDICATION

This dissertation is dedicated to my mother, Gabriela, and my father, Jaime, who have given me everything I needed to pursue academic success at the highest level. Their love, support, and commitment to my education are invaluable to me, and there is nothing I could do to thank them enough.

I also dedicate this dissertation to my wife, Emiliana, who has loved and encouraged me through the most difficult times. This dissertation marks the end of a beautiful part of my life and the beginning of a new chapter in our lives.

Finally, I dedicate this dissertation to my family and friends, who have always been there to support me and bring me happiness and joy through the difficult moments that a PhD can bring.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am deeply grateful to Professors Fabian Eckert and Fabian Trottner for their guidance, encouragement, and constant support throughout this dissertation. Whenever I hit a wall, they helped me think through it with patience and clarity. I am especially thankful for the many weekly meetings, held without fail, that shaped both the ideas in this dissertation and the way I learned to approach research.

I am also grateful to Professor Marc Muendler for serving as co-chair of my committee and for his guidance and support throughout the program. I thank Professors Nir Jaimovich and Craig McIntosh for their time, advice, and teachings, and for serving on my committee.

I thank my coauthors Leticia Juárez, Román Zárate, Alfonso Cebreros, and Daniel Ramos-Menchelli. Working with them has made this dissertation better and has taught me a great deal about how to do research collaboratively. Chapter 2, in full, is coauthored with Leticia Juárez and Román Zárate. The dissertation author was the primary author of this chapter.

Chapter 3, in full, is coauthored with Alfonso Cebreros and Daniel Ramos-Menchelli. The dissertation author was the primary author of this chapter.

VITA

2017	Bachelor of Arts in Economics, Instituto Tecnológico Autónomo de México (ITAM)
2018–2020	Economist, Directorate General for Economic Research, Banco de México
2020–2026	Doctor of Philosophy in Economics, University of California San Diego
2021–2025	Teaching Assistant, University of California San Diego
2022–2023	Research Assistant, University of California San Diego
2024	Visiting Summer Ph.D. Researcher, Banco de México
2026	Visiting Fellow Ph.D. Researcher, Center for U.S.-Mexican Studies, University of California San Diego

FIELDS OF STUDY

Major Field: Economics

Studies in International Trade, Urban Economics, and Spatial Economics

Professors Fabian Eckert and Marc Muendler

ABSTRACT OF THE DISSERTATION

Essays in Trade, Urban, and Spatial Economics

by

Jordán Mosqueda Juárez

Doctor of Philosophy in Economics

University of California San Diego, 2026

Professor Fabian Eckert, Co-Chair
Professor Marc Muendler, Co-Chair

This dissertation studies how transportation costs and frictions shape economic activity within and across cities. Across three essays, it examines how commuting costs, transit networks, and trade frictions affect where people live and work, how firms and households connect to markets, and how public policy changes welfare.

The first chapter develops a quantitative urban model in which private and public transit jointly determine commuting costs within a city. We investigate how private providers' decisions

shape commuting costs considering complementarities with the public network, and the welfare gains and spatial consequences of policies that directly shift prices such as fare regulation and subsidies. 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.

The second chapter, coauthored with Leticia Juárez and Román Zárate, studies optimal transit networks in developing countries. Combining a spatial model with network and commuting data, it shows how budget-constrained governments should allocate transit infrastructure across public mass transit and privately operated services. The analysis highlights the additional welfare gains from accounting for private network design rather than expanding only public mass transit.

The third chapter, coauthored with Alfonso Cebreros and Daniel Ramos-Menchelli, examines domestic trade frictions created by crime. Using confidential data from Mexican transport firms, agricultural markets, and administrative crime records, it measures how crime exposure along shipping routes raises trade costs and affects prices, market access, and welfare. The chapter shows that crime can make locations economically distant even when they are geographically close.

Together, the chapters show that transportation costs are not exogenous outcomes purely determined by technological or infrastructure constraints. They are shaped by infrastructure, congestion, network design, and institutional conditions, and they play a central role in determining the spatial organization of economic activity.

CHAPTER 1

Equilibrium Commuting Costs: The Role of Private and Public Transit

1.1 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

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

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

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%.

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) by trading off time and income across potentially multimodal routes from a given route choice set. For each origin-destination pair, some route alternatives may involve single rides, while some other may involve multiple legs. 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

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

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

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

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

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

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 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

[6]In 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.

setting serves as the baseline for counterfactuals.

These two policies are of both academic and policy interest because they act through understudied 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.

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

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

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

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.

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

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 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.

1.2 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.

1.2.1 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 $\iota^{k,m} : \mathcal{L}^{k,m} \to 2^{\mathcal{L}^{\text{rd}}}$ 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

[10]I 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.

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 $Freq_\varphi$, 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.

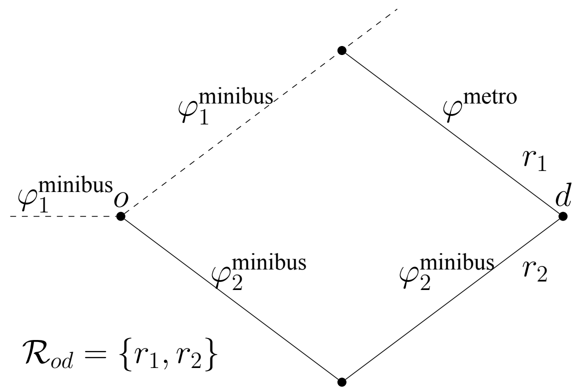


Figure 1.1: Illustration of alternative routes between o and d.

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 1.1 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 $\varphi_2^{\text{minibus}}$.

1.2.2 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 consumption $X(\omega)$, 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

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

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

$$C_{odr}(\omega) = \min(c_{\varphi_1}(\omega), \ldots, c_{\varphi_K}(\omega)), \tag{1.1}$$

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

[12] I 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 1.2.6 and 1.2.7, where I unpack the value of Ω.

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

[14] One 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 $\sigma \to 0$. More detailed derivations and explanations can be found in Appendix 1.A.3.

market:

$$c_{\varphi}(\omega) = \left(\sum_{i \in \varphi} q_i(\omega)^{\frac{\chi - 1}{\chi}} \right)^{\frac{\chi}{\chi - 1}}, \tag{1.2}$$

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

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

$$V_{odr}(\omega) = B_{od} \frac{\tilde{w}_d(\bar{T} - t_{odr})}{Q_o^{\alpha_h} P_{odr}^{\alpha_c}} \varepsilon_{odr}(\omega), \quad \tilde{w}_d \equiv w_d(1 + \Omega). \tag{1.3}$$

If we denote

$$\tau_{odr} \equiv \frac{P_{odr}^{\alpha_c}}{\bar{T} - t_{odr}}, \tag{1.4}$$

then we can rewrite (1.3) as a more familiar expression to standard commuting models:

$$V_{odr}(\omega) = B_{od} \frac{\tilde{w}_d}{Q_o^{\alpha_h} \tau_{odr}} \varepsilon_{odr}(\omega), \tag{1.5}$$

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):

$$F(\vec{\varepsilon}) = \exp\left[- \sum_{o,d} \left(\sum_{r \in \mathcal{R}_{od}} \varepsilon_{odr}^{-\rho} \right)^{\theta/\rho} \right]. \tag{1.6}$$

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.

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

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

$$\lambda_{od} = \frac{\left(\frac{B_{od}\,\tilde{w}_d}{Q_o^{\alpha_h}\,\tau_{od}}\right)^{\theta}}{\sum_o \sum_d \left(\frac{B_{od}\,\tilde{w}_d}{Q_o^{\alpha_h}\,\tau_{od}}\right)^{\theta}}, \tag{1.7}$$

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

$$\tau_{od} \equiv \left(\sum_{r \in \mathcal{R}_{od}} \tau_{odr}^{-\rho}\right)^{-1/\rho}.^{16} \tag{1.8}$$

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:

$$\lambda_{odr} = \underbrace{\frac{\left(\frac{B_{od}\,\tilde{w}_d}{Q_o^{\alpha_h}\,\tau_{od}}\right)^{\theta}}{\sum_o \sum_d \left(\frac{B_{od}\,\tilde{w}_d}{Q_o^{\alpha_h}\,\tau_{od}}\right)^{\theta}}}_{\lambda_{od}} \times \underbrace{\frac{\tau_{odr}^{-\rho}}{\sum_r \tau_{odr}^{-\rho}}}_{\lambda_{r|od}}. \tag{1.9}$$

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

$$\bar{W} \equiv \mathbb{E}[\max V_{odr}(\omega)] = k \left(\sum_o \sum_d \left(\frac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}\tau_{od}}\right)^{\theta}\right)^{\frac{1}{\theta}}, \tag{1.10}$$

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

Commuting costs at the route level. Given that routes are comprised of potentially many markets

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

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:

$$P_{odr} \equiv \sum_{\varphi \in r} P_{\varphi}, \tag{1.11}$$

$$t_{odr} \equiv \sum_{\varphi \in r} \left(t_{\varphi}^{\text{wait}} + \gamma_{odr}^{\varphi} \, t_{\varphi}^{\text{trip}} \right), \tag{1.12}$$

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:

$$D_{\varphi} = \sum_{od} \sum_{r|\varphi \in r} \frac{\alpha_c \tilde{w}_d (\bar{T} - t_{odr})}{P_{odr}} \lambda_{odr} \bar{L}.^{17} \tag{1.13}$$

1.2.3 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 (1.2). Aggregating demand for a firm across all workers that consume in that market, we get an individual

[17]Detailed derivations in Appendix 1.A.3

[18]These assumptions are discussed in the discussion section 1.2.9 right after the model.

residual CES demand for firm i:[19]

$$q_{i,\varphi} = \left(\frac{p_{i,\varphi}}{P_\varphi}\right)^{-\chi} D_\varphi. \tag{1.14}$$

Firm's problem. Individual firms decide whether to enter the market by paying a fixed cost of entry f^e_φ, 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^c_φ and a time endowment $\bar{T}^d$.[20] Firms face an invidiual CES demand, given by equation (1.14). The problem of the firm is:

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

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

$$p_{i,\varphi} = \left(\frac{D_\varphi}{q^c_\varphi n_{i,\varphi}}\right)^{\frac{1}{\chi}} P_\varphi.\text{[21]} \tag{1.15}$$

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

$$n_{i,\varphi} = \frac{\bar{T}^d}{t^{\text{trip}}_\varphi}. \tag{1.16}$$

[19]Detailed derivations in Appendix 1.A.3

[20]To 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.

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

The price index of the commuting good in this market is

$$P_\varphi \equiv \left(\sum_{i\in\varphi} p_{i,\varphi}^{1-\chi}\right)^{\frac{1}{1-\chi}} = P_\varphi \left(\frac{D_\varphi}{q_\varphi^c n_{i,\varphi}}\right)^{\frac{1}{\chi}} M_\varphi^{\frac{1}{1-\chi}}, \tag{1.17}$$

where M_φ denotes the mass of entrants. From equation (1.17), it follows that the following market-clearing relation must hold:

$$D_\varphi(\mathbf{M}) = M_\varphi^{\chi/(\chi-1)}\, q_\varphi^c\, n_{i,\varphi}, \quad \forall\varphi \in \Phi^{\text{Priv}}. \tag{1.18}$$

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 **M** because entry determines commuting costs, as described in detail below, so equation (1.18) 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:

$$P_\varphi = M_\varphi^{-\frac{1}{\chi-1}} \frac{t_\varphi^{\text{trip}}}{\bar{T}^d} \Big(\frac{\delta_\varphi \bar{T}^d + f_\varphi^e}{q_\varphi^c}\Big). \tag{1.19}$$

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 $\bar{t}_\varphi$ and congestion across all the road links that define that market:

$$t_\varphi^{\text{trip}}(\mathbf{M}) = \bar{t}_\varphi \sum_\ell S_\ell(\mathbf{M})^\phi. \tag{1.20}$$

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

$$Freq_\varphi(\mathbf{M}) \equiv \frac{M_\varphi}{t_\varphi^{\text{trip}}(\mathbf{M})}. \tag{1.21}$$

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

$$t_\varphi^{\text{wait}}(\mathbf{M}) = \frac{1}{2}\frac{1}{Freq_\varphi(\mathbf{M})}. \tag{1.22}$$

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.

[22]The particular process that I am assuming is a uniform arrival of buses, which is one the simplest processes to model bus arrivals. This seems plausible if we think that buses are coordinated to some degree in terms of schedule. In the Mexican context, 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 $Freq(.)$ that controls its mean and variance. For example, in the different case of a Poisson process with parameter $\lambda = Freq(.)$, 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.

1.2.4 Residential land markets

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

$$\bar{H}_o Q_o = \alpha_h Y_o R_o, \quad \forall o \in J. \tag{1.23}$$

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 (1.23) equates aggregate expenditure for housing in each location to the income generated in the housing rental market.

1.2.5 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^c_d. The production function is given by:

$$y_d = A_d \Big(\frac{L_d}{\beta}\Big)^{\beta} \Big(\frac{H^c_d}{1-\beta}\Big)^{1-\beta}. \tag{1.24}$$

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, $\bar{H}^c_d = H^c_d$, therefore wages in each location are given by

$$w_d = A_d \Big(\frac{\beta}{1-\beta} \frac{\bar{H}^c_d}{L_d}\Big)^{1-\beta}, \tag{1.25}$$

and commercial rents are given by

$$Q^c_d = A_d \left(\frac{1-\beta}{\beta} \frac{L_d}{\bar{H}^c_d}\right)^{\beta}. \tag{1.26}$$

1.2.6 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

$$I^C \equiv \sum_d Q^c_d H^c_d,$$

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:

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

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]

1.2.7 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:

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

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

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

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

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

and therefore $\Omega = \nu - \eta(1+\nu)$, presented in the commuter's budget constraint in Section 1.2.2, captures the proportional income adjustments from land income revenue and taxes.

1.2.8 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 $\bar{W}$:

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

commuting flow gravity equations hold

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

the number of residents in each location satisfies

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

and the effective labor units in each location satisfy

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

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:

$$w_d = A_d \Big(\frac{\beta}{1-\beta}\frac{\bar{H}_d^c}{L_d}\Big)^{1-\beta}.$$

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 1.B.

1.2.9 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 (1.4),

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

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 1.1 (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

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

and the elasticity of the aggregate probability across all routes that use such market with respect to the price is

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

Proof. See Appendix 1.A.2. □

Proposition 1.1 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 1.2 (Frequency–congestion trade-off). *Fix a market φ with entrants M_φ and hold $M_{-\varphi}$ fixed. Define the link-level share of φ on link ℓ and the ψ-specific congestion weights*

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

Then:

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

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

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

Proof. See Appendix 1.A.4. □

Proposition 1.2 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 1.2 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 1.1 (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 turned 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 $q_{i,\varphi}(.)$ 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 (1.18):

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

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—

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

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^e_φ, 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.

1.3 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.

1.3.1 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

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

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.

1.3.2 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.

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 1.2 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.

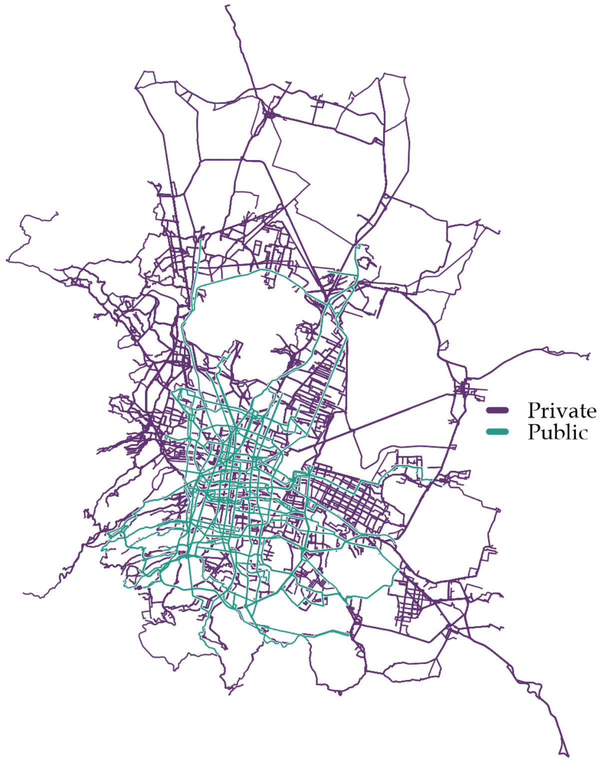


Figure 1.2: Public and private transit lines in Mexico City M.A.

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

Furthermore, table 1.1 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 1.1: 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 km2)	16.07	0.85
Average number of lines passing through a census tract	9.05	0.59
Share of census tracts with at least 1 line	0.68	0.29
Share of census tracts with at least 5 lines	0.46	0.02
Share of census tracts with at least 10 lines	0.29	0.00

Note: 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 1.2 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 $\approx$ 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.

Furthermore, the scope for congestion externalities is large due to the arrangement of private lines. Figure 1.3 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

Table 1.2: 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.

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.

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 1.3 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

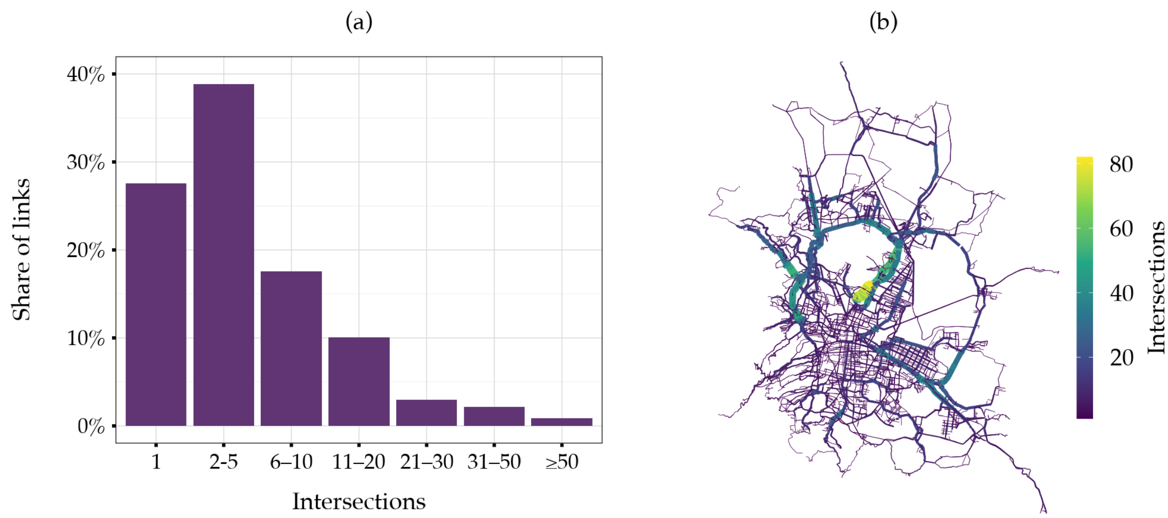


Figure 1.3: Intersections of private transit lines and road links

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.

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.

1.3.3 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

Table 1.3: Commuting trips 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.

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 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

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

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 billon trips, so a subsidy of 13 pesos per trip amounts to 15.21 billion pesos, or around 830 million US dollars. This represents roughly two-fifths of the total transportation budget and around 5% of the total local government's expenditure.[28]

1.4 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.

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

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

1.4.1 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 192^2 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 frequency. 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 1.2 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 1.1. 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

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

[30] In this model, the commuting costs to do own-district commuting are not zero (or one if one thinks of iceberg commuting costs). Commuting costs are positive, so even if you travel within the district, you spend some time and money. Importantly, own-district flows should be allocated as demand to some market. I direct the interested reader to Appendix 1.C.1 for more details about the collection of data and processing.

[31] Such recursive routing models find optimal paths by solving a Bellman-type problem at each node of the transport

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 1.2.9.

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.

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.

[32]Such 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.

1.4.2 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 ($P_{\text{metro}} \to \infty$) and compute the equilibrium response in demand and, in turn, supply and trip times (speed). The demand response implies a relocation of commuters towards alternative routes, affecting service provision, 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 1.4 shows the distribution of

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

[34] Full details on the data and processing, can be found in Appendix 1.D.2

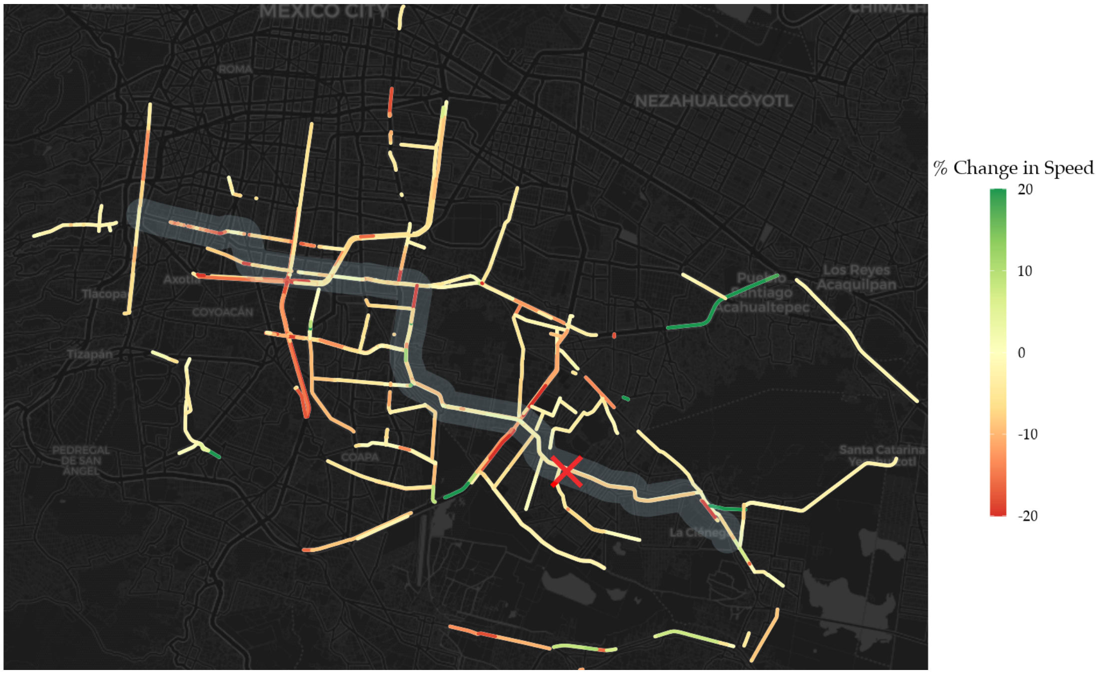


Figure 1.4: TomTom road-level speed data: changes pre/post subway shock

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.

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 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

[35]The distribution of speed changes is shown in Appendix 1.E figure 1.19.

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 1.4: 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.

Table 1.4 summarizes the identified elasticities and the model fit, which does a reasonable job at 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

[36]Note 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.

[37]Although I make robustness check for 1.5, 3, and 20km buffers, and parameters do not change substantially. Table 1.8 in the shows these results.

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.

1.4.3 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

[38]Figure 1.15 in Appendix 1.D.3 shows the spatial heterogeneity.

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 1.C.2 and Appendix 1.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 $\bar{T}^d = 24$.

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

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 $\bar{p}_{\varphi}$. In Appendix 1.A.5 I describe this 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^e_{φ}, vehicle capacities q^c_{φ}, and the marginal cost per trip δ, I target the observed equilibrium outcomes–entry, frequency, trip times, and prices. I obtain firm-level demand from equilibrium route-level demand. Then, compute the number of trips a 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 $q^c_0 = 15$ 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 $\{q^c_{\varphi}\}_{\varphi}$ 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^e_{φ} are calibrated as the residual profits that rationalize the zero-profit condition under free entry. This procedure jointly pins down entry costs, capacities, and marginal operating costs in a way that is consistent with observed prices, demand, and technological constraints on service provision.

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

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

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

Table 1.5: 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^c_φ	Capacity constraint		Rescaled to match observed entrants and demand
γ^φ_{odr}	Shares of market usage across routes		Google Maps
χ	Elasticity of substitution across firms (buses)	15	
δ	Marginal cost per time unit	36.4	Calibrated to get nonnegative fixed costs of entry
f^e_φ	Fixed costs of entry		To match observed p,M,t from Google Maps
$\bar{t}_\varphi$	Trip time shifters		To match observed travel times from Google Maps
$\bar{T}^d$	Time endowment of drivers	24 hours	

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

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^φ_{odr}. So, for example, if a route alternative uses a market for only 25% of the full length of that market, then $\gamma^\varphi_{odr} = 0.25$. Then, having recovered the trip time (t^{trip}_φ) that each unit has to complete in a lap in a given market, and observing the number of entrants in each market, I can recover the trip time shifters $\bar{t}_\varphi$ that exactly match the observed time in that market, given the observed number of entrants.

Table 1.5 summarizes the parameter calibration.

1.5 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]

1.5.1 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 $\bar{p}_\varphi$ 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 (1.15) and (1.19) 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 1.10 into the contribution of each of its elements as

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

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 1.6 shows the decomposition

[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 1.9 and 1.10 in Appendix 1.E.

[42] Such extension of the model is explained in Appendix 1.A.5.

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 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 1.6: 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 1.5 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.

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 1.6 shows the change in flows across markets. It can be seen that peripheral markets are the ones that primarily experience an increase in flows, particularly in the Eastern part of the city. On the other hand, markets that are located in more

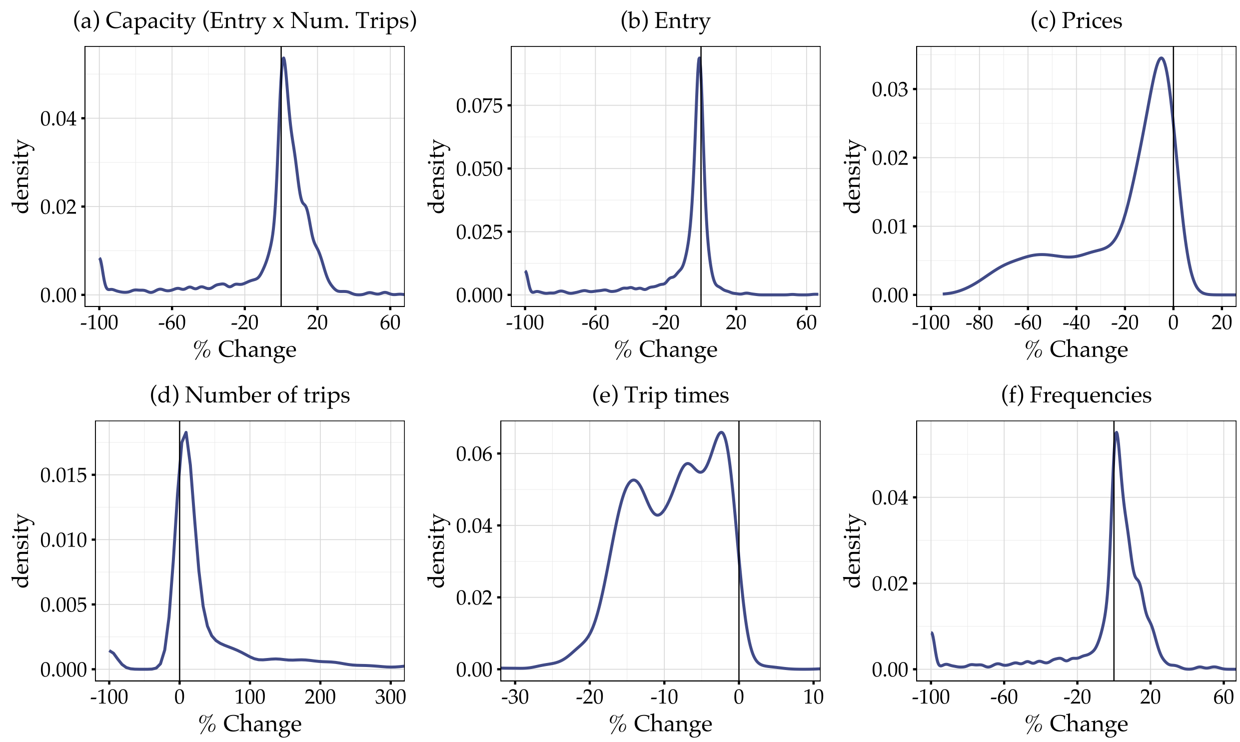


Figure 1.5: Market characteristics following price deregulation

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.

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.

[43]This is shown in additional figure 1.16 in Appendix 1.E

Figure 1.6: 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.

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 1.7 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 reflected by the fundamental productivity A and a measure of local amenities $B_o = \sum_d B_{od}$.[44] Figure 1.8 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

[44]The spatial distribution of the productivity and amenity measures can be found in figure 1.18 in Appendix 1.E.

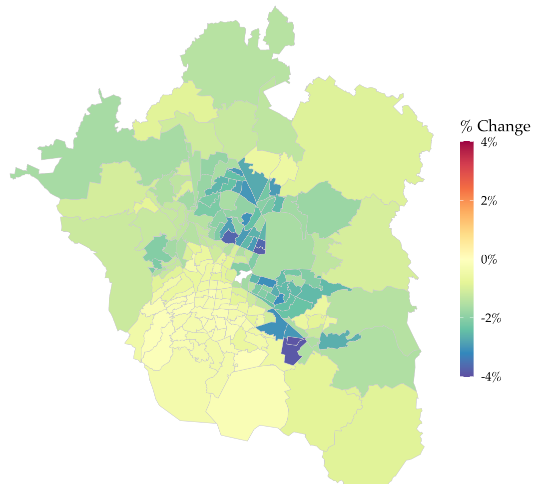


Figure 1.7: Commuting cost index change following deregulation

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

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 1.8: 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 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.

1.5.2 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 1.7 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.

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 via real income—disposable income divided by rents, adjusted by commuting costs. This increase in

[45]This 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.

Table 1.7: 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.

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:

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

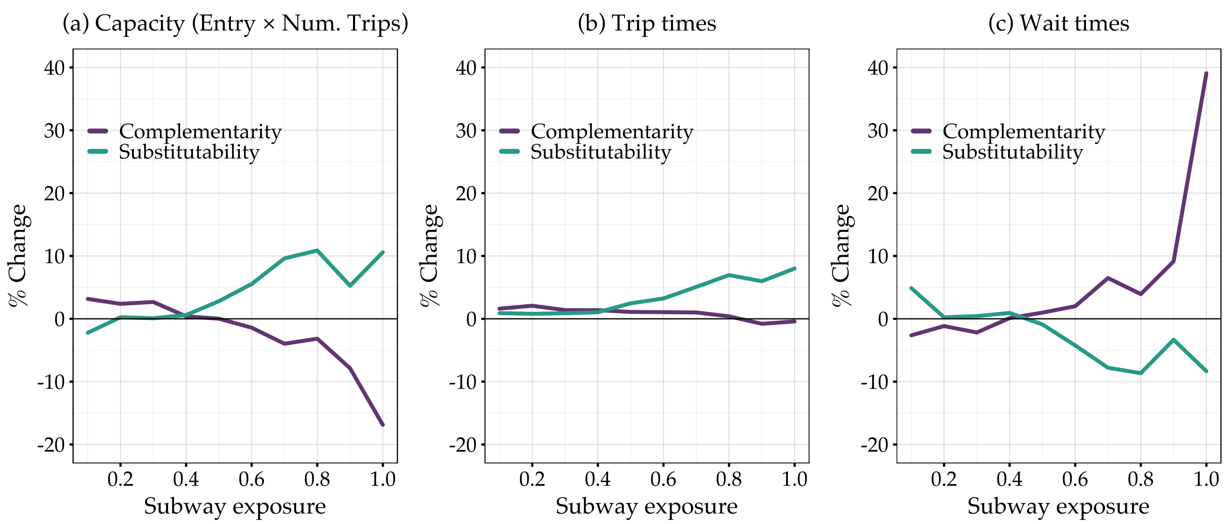


Figure 1.9: Effects of subsidy removal by metro exposure

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.

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 1.9 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 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

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

Figure 1.10: 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.

transit in meaningful ways.

In terms of the spatial effects across markets, figure 1.10 shows the change in flows across markets. 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 1.7 shows an origin-specific commuting cost index $\tau_o = \sum_d \lambda_{d|o} \tau_{od}$ that averages commuting cost indexes across destinations.

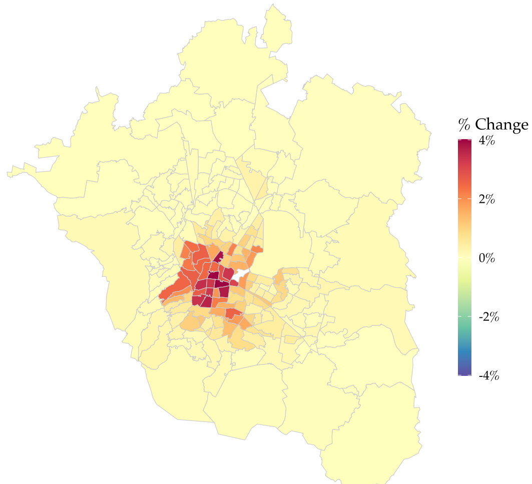


Figure 1.11: Commuting cost index change by origin: subsidy removal

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

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 generates 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

1.12, where there is a positive relationship between fundamental productivity and wage changes, and a negative relationship between a measure of fundamental location-specific amenities $B_o = \sum_d B_{od}$ and rent changes.

Figure 1.12: 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.

1.5.3 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 1.13 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 1.6 and 1.7.

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 1.13. 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.

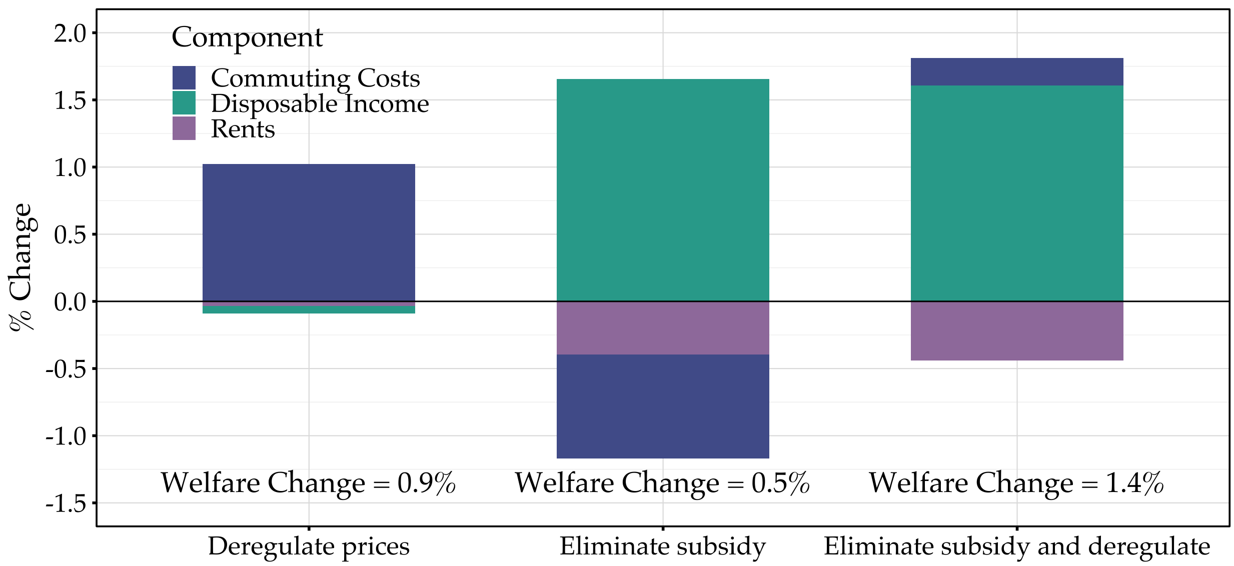


Figure 1.13: Welfare change decomposition for different policies

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.

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 quantative 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 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.

1.6 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 1.14 shows how welfare changes as we increase the elasticity of congestion (from left to right panel) and vary the elasticity of substitution in the horizontal axis.

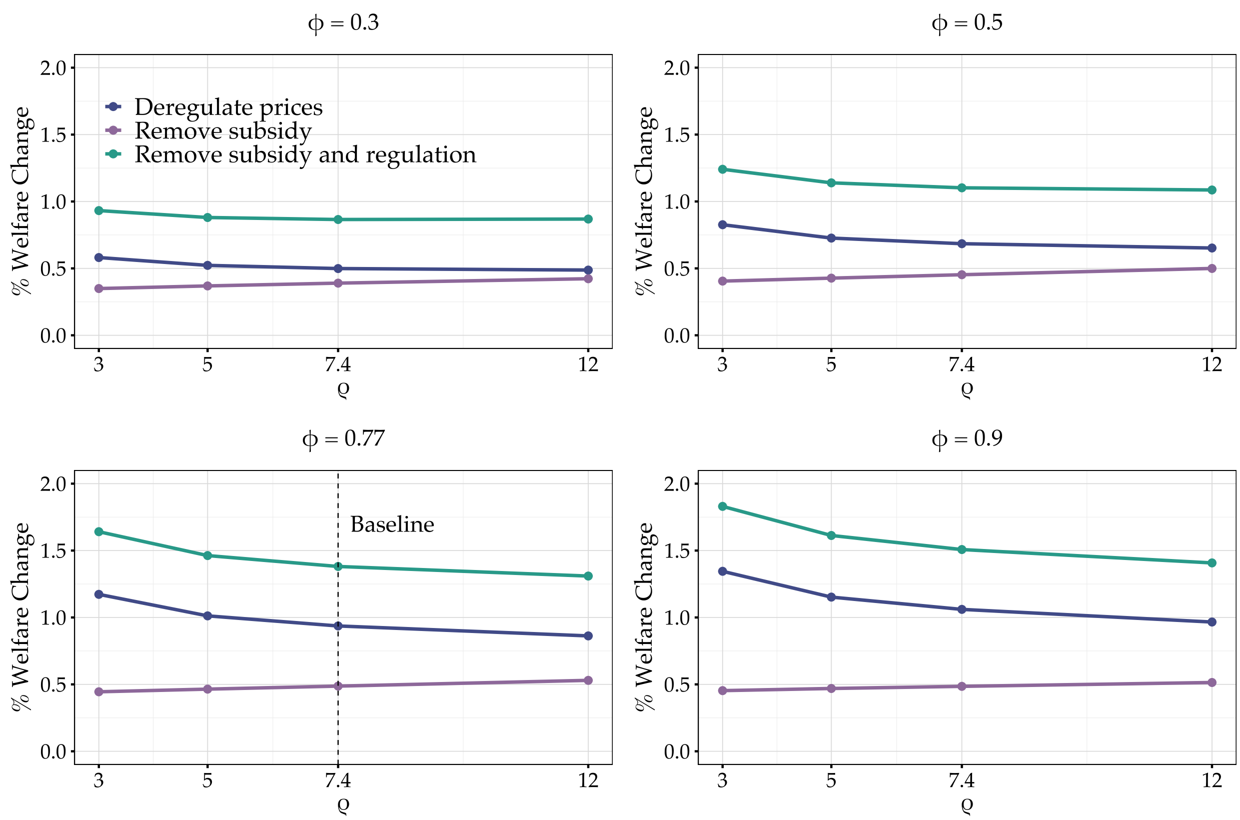


Figure 1.14: Welfare change under alternative parametrizations

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

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

Regarding the elimination of the subway subsidy, the qualitative and quantitative welfare changes stay practically the same. Only in the case where congestion is very low (first panel), what matters the most for the magnitude of the welfare change is the elasticity of substitution. Comparing low (3) and high (12) values, we can see that welfare change ranges from 0.4 to around 0.5, which is not substantial in absolute terms. Overall, the welfare effects of removing both the subsidy and regulation follows roughly the same pattern as removing the regulation: when congestion forces are large (i.e. the lower panels), the variation of the welfare change coming from variation in the substitution elasticity can be as large as 0.5 percentage points. Given that we could expect the value of ρ to fall within 7 and 10 following the robustness exercises performed in Section 1.4.2, the overall welfare qualitative and quantitative change should remain stable.

1.7 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.

Appendix

1.A Model derivations

1.A.1 Commuting flows and commuting cost index

This section derives the route-level choice probabilities under the nested Fréchet structure in (1.3)–(1.6) and the CES commuting cost index (1.8). Define the deterministic component of utility for option (o, d, r) as:

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

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

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

as in (1.6).

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

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

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

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

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

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

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

$$\lambda_{r|od} = \frac{\tau_{odr}^{-\rho}}{\sum_{r'} \tau_{odr'}^{-\rho}} \quad . \tag{1.27}$$

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 $x > 0$,

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

Taking the marginal of the nested Fréchet CDF (1.6) for a single pair (o, d) (i.e., setting the thresholds for all other pairs to $+\infty$) yields

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

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

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

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

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

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

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

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

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

Hence

$$\lambda_{od} = \frac{\left(\dfrac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}\,\tau_{od}} \right)^{\theta}}{\sum_{o',d'} \left(\dfrac{B_{o'd'}\tilde{w}_{d'}}{Q_{o'}^{\alpha_h}\,\tau_{o'd'}} \right)^{\theta}} \quad . \tag{1.28}$$

By the law of total probability and conditional independence, combining (1.27) and (1.28) gives

$$\lambda_{odr} = \lambda_{od} \times \lambda_{r|od} = \underbrace{\frac{\left(\dfrac{B_{od}\tilde{w}_d}{Q_o^{\alpha_h}\,\tau_{od}} \right)^{\theta}}{\sum_{o',d'} \left(\dfrac{B_{o'd'}\tilde{w}_{d'}}{Q_{o'}^{\alpha_h}\,\tau_{o'd'}} \right)^{\theta}}}_{\lambda_{od}} \times \underbrace{\frac{\tau_{odr}^{-\rho}}{\sum_{r'} \tau_{odr'}^{-\rho}}}_{\lambda_{r|od}} \quad . \tag{1.29}$$

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.

1.A.2 Proof of complementarity proposition

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

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

Given expression:

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

Taking the natural log:

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

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

Differentiating with respect to x:

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

For the second term, using the chain rule:

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

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

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

Substitute back to obtain (*):

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

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

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

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

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

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

$\square$

Corollary 1.2 (Two-route illustration)**.** *Suppose $\mathcal{R}_{od} = \{1, 2\}$ with route 1 using only φ ($P_{od,1} = P_\varphi \Rightarrow s_{\varphi|1} = 1$) and route 2 using φ and φ' ($s_{\varphi|2} = P_\varphi/(P_\varphi + P_{\varphi'}) \in (0, 1)$). Then*

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

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

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

1.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

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

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

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

where this comes from the fact that the budget constraint is $\sum_{\varphi \in odr} P_\varphi c_\varphi = \alpha_c y_{odr}$ and $P_{odr} = \sum_{\varphi \in odr} P_\varphi$. Also, note that expenditure in trips of a given market is $P_\varphi c_\varphi(\omega) = \frac{P_\varphi}{\sum_\varphi P_\varphi} \alpha_c y_{odr}$

Now, for each market φ along the route, worker consumes a CES bundle of trips in individual buses within that market with elasticity of substitution χ, so that

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

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

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

where $P_\varphi \equiv \left(\sum_{i \in \varphi} p_{i,\varphi}^{1-\chi} \right)^{\frac{1}{1-\chi}}$ is the price index of market φ. To get total demand for driver

i, we must aggregate across all workers that choose a route that involves using market φ. That is, we aggregate across all origins, destinations, and routes, that pass through market φ in some segment

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

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

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

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

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

For the congestion term,

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

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

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

Therefore,

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

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

1.A.5 Model with price regulation

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

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

$$\text{s.t.} \qquad q_{i,\varphi} = M_{\varphi}^{\frac{\chi}{1-\chi}} D_{\varphi}$$

$$\frac{q_{i,\varphi}}{q^c} \leq n_{i,\varphi}$$

$$n_{i,\varphi} t_{\varphi}^{\text{trip}} \leq \bar{T}^d$$

The price index is given by

$$P_{\varphi} = M_{\varphi}^{1/(1-\chi)} \bar{p}_{\varphi} \tag{1.30}$$

Profits under price regulation are

$$\pi_{i,\varphi} = \left(\bar{p}_{\varphi}\, q_{\varphi}^{c} - \delta\, t_{\varphi}^{\text{trip}}\right) n_{i,\varphi} - f_{\varphi}^{e} \tag{1.31}$$

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 $m^{\text{pax}}_{\varphi} = \bar{p}_{\varphi} - \frac{\delta\, t^{\text{trip}}_{\varphi}}{q^{c}_{\varphi}}$ and the maximum capacity of seats in a given shift $S_{\varphi} = q^{c}_{\varphi} n^{\text{max}}_{\varphi}$. Zero profits requires that the quantity of seats served is q^{zp}_{φ} such that:

$$m^{\text{pax}}_{\varphi} q^{zp}_{\varphi} = f^{e}_{\varphi} \iff q^{zp}_{\varphi} = \frac{f^{e}_{\varphi}}{m^{\text{pax}}_{\varphi}}$$

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

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

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

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

1.B Algorithm to solve general equilibrium

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

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

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

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

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

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

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

- Compute commuting costs

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

- Compute commercial rents

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

- Compute wages

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

- Compute residential rents

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

- Compute demand

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

- **Tranportation market**. Given demand, solve for updated vector of entrants M.

– (Market prices case) Compute updated M implied by the market clearing equation

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

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

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

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

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

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

1.C Data and calibration

1.C.1 Transit network data collection

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

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

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

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

1.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 $\{B_{od}\}$.

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

1.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:

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

1.D Identification of parameters from subway shock

1.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.

1.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.

1.D.3 Robustness of SMM exercise

Table 1.8: 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.

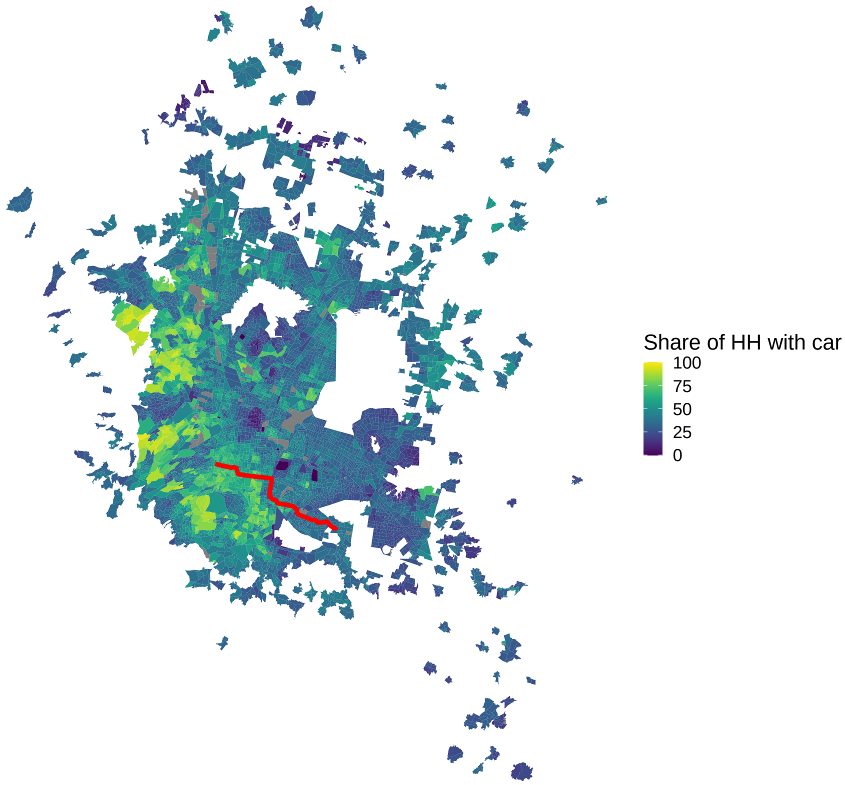


Figure 1.15: Car ownership by household living near the collapsed subway line

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.

1.E Additional tables and figures

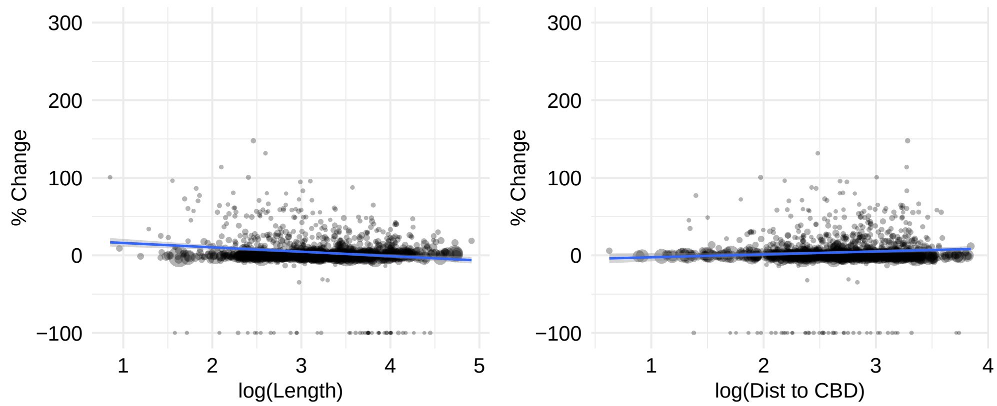


Figure 1.16: Correlation between market characteristics and change in flows: price deregulation

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.

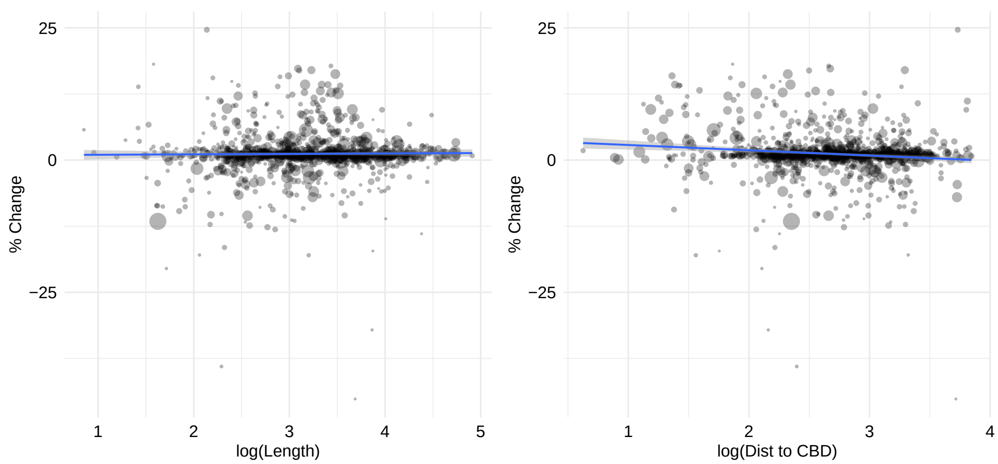


Figure 1.17: Correlation between market characteristics and change in flows: metro subsidy removal

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.

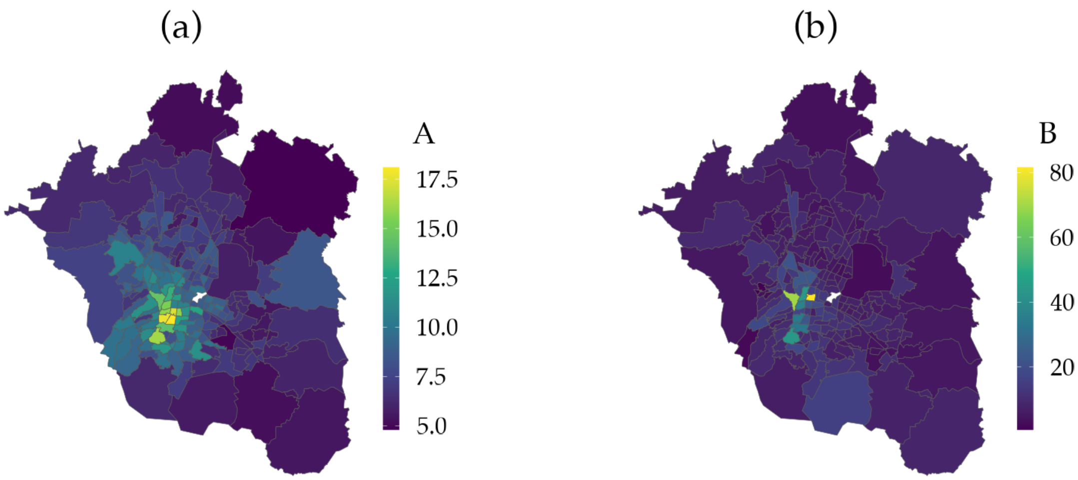


Figure 1.18: Spatial distribution of fundamentals

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

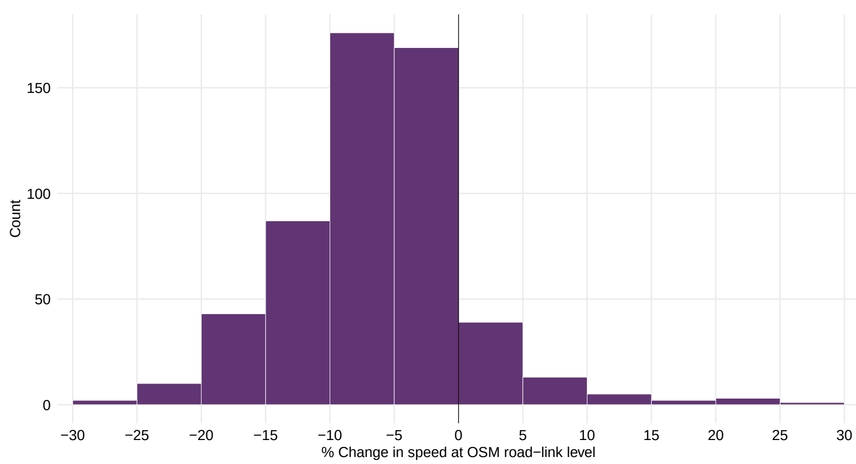


Figure 1.19: TomTom road-level speed data: changes pre/post subway shock

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 1.9: Price regulation examples in privately operated markets (selected developing cities)

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

Note: Links for each source are available upon request.

Table 1.10: Fare subsidy examples in publicly operated systems (selected developing cities)

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

Note: Links for each source are available upon request.

Bibliography

Adler, Martin W, Federica Liberini, Antonio Russo, and Jos N. van Ommeren, "The congestion relief benefit of public transit: evidence from Rome," *Journal of Economic Geography*, 12 2020, *21* (3), 397–431.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Quarterly Journal of Economics*, 2015, *130* (3), 1237–1281.

Allen, Treb and Costas Arkolakis, "The Welfare Effects of Transportation Infrastructure Improvements," *Review of Economic Studies*, 2022, *89* (6), 2911–2957.

_, David Atkin, Santiago Cantillo Cleves, and Carlos Eduardo Hernández, "The Traveling Trucker Problem," NBER Working Paper 31961, National Bureau of Economic Research December 2023. Also in *AEA Papers and Proceedings* 114 (2024): 334–339, doi:10.1257/pandp.20241044.

Almagro, Milena, Andrea Barbieri, Juan Camilo Castillo, Arianna Hickok, and Timo Salz, "Optimal Urban Transportation Policy: Evidence from Chicago," 2024. Working paper.

Björkegren, Daniel, Maël Duhaut, Aishwarya Nagpal, and Nick Tsivanidis, "Public and Private Transit: Evidence from Lagos," NBER Working Paper 33899, National Bureau of Economic Research May 2025. Working paper.

Bordeu, Olivia, "Commuting Infrastructure in Fragmented Cities," Working Paper 2023. Working paper.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou, "Geography, Transportation, and Endogenous Trade Costs," *Econometrica*, 2020, *88* (2), 657–691.

Bryan, Gharad T., Kyra Frye, and Melanie Morten, "Spatial Economics for Low- and Middle-Income Countries," NBER Working Paper 33453, National Bureau of Economic Research, Cambridge, MA February 2025.

Conwell, Lucas, "Privatized Provision of Public Transit," 2024. Job Market Paper; working paper.

Fuchs, Simon and Woan Foong Wong, "Multimodal Transport Networks," Working Paper, CESifo Working Paper No. 11362; also Atlanta Fed Working Paper 2022-13a (rev. 2024) 2024. Working paper.

Khanna, Gaurav, Catalina Medina, Anant Nyshadham, Juan Ramos, Juan Tamayo, and Zi Hao Tiew, "Spatial Mobility, Economic Opportunity, and Crime," 2024. Working paper; add link when available.

Kreindler, Gabriel, Arya Gaduh, Tyler Graff, Rema Hanna, and Benjamin A. Olken, "Title TBD (Public Transit in Jakarta)," 2023. Working paper; fill in when finalized.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

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

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

Zárate, Román David, "Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City," 2024. Working paper.

CHAPTER 2

Optimal Transit Networks in Developing Countries

Coauthored with Leticia Juárez and Román Zárate.

2.1 Introduction

Urban transit systems in developing cities face a fundamental allocation problem: how scarce public resources should be allocated between publicly provided mass transit and privatized, often informal, transit provision. This question is quantitatively important: privately operated minibuses today provide between 50 and 100 percent of urban transit in many developing-world cities (Conwell, 2024), yet the optimal design of mixed public-private transit networks remains poorly understood. Different allocation choices can have profound implications for urban structure, welfare, and commuting patterns by shaping congestion, environmental externalities, and the overall cost of commuting between any two locations (Mosqueda, 2026). These issues are particularly salient in developing countries, where rapid urbanization and tight fiscal constraints limit governments' ability to expand mass transit networks.

The central question of this paper is how a planner should allocate public and private investment in a network. The planner faces a key trade-off. Mass transit technologies are fixed-cost intensive, requiring large upfront investments but featuring low marginal operating costs. By contrast, private transit has higher marginal operating costs and entails externalities such as congestion, but requires little upfront investment. When governments cannot afford to extend mass transit to the entire city, these second-best technologies become an attractive alternative to connect underserved areas and meet mobility needs.

Despite its policy relevance, addressing this question poses three main challenges. First, the transit network design problem is computationally intractable: the solution space grows exponentially with city size, and the problem exhibits both substitution and complementarity forces—as routes may complement each other when they connect, but substitute when they compete for riders—which can preclude finding a global optimum (Kreindler et al., 2023). Second, observed transit networks emerge from decentralized decisions by commuters and private operators, so evaluating optimal policy requires a framework that can jointly account for network design, commuting behavior, and general equilibrium responses. Third, data on actual private routes are scarce throughout the developing world. In the majority of cities, service is informal; therefore, trip geographies and service characteristics are rarely collected centrally.

In this paper we address these challenges by, first, providing a general spatial equilibrium framework that delivers the planner's optimal public and private networks, and, second, quantifying it by leveraging a unique dataset that describes the universe of public and private lines in the Metropolitan Area of Mexico City. We then use this quantitative model to study where and how much a planner should invest in these two technologies. Counterfactual analyses show that optimally expanding the network using both public and private transit can deliver additional welfare gains of around 18% on top of the gains generated by expanding the public network in isolation, suggesting that developing cities could benefit substantially from joint planning of public and private networks.

Framework. We extend the optimal transport network framework of Fajgelbaum and Schaal (2020), which studies goods trade across regions, to within-city commuting with heterogeneous transit technologies. In our model, the planner chooses i) optimal consumption and labor allocations across space, ii) commuting flows on each edge, and iii) the mix of public and private infrastructure on each link, subject to land, production, and infrastructure budget constraints. The main innovation is to allow each edge to be served by different transport technologies that differ in both operating and construction costs. We assume that commuting costs are a function of technology-specific shifters that capture differences in marginal costs of service provision, which reflects the fact that transporting a worker by subway is substantially cheaper than by minibus. We also assume that the planner faces a network-building constraint in which she allocates a fixed budget across technologies with different construction costs—reflecting that drilling a tunnel is more expensive than adapting street curbs.

We understand private infrastructure as a bundle that includes the minibuses themselves, service capacity, operating permits, curbs, stops, terminals, and dedicated or priority lane access, while public infrastructure refers to fixed mass-transit assets such as subway, BRT, and rail capacity.

The key modeling choice is that transport costs depend on a continuous mix of technologies, which preserves convexity and makes the problem tractable. Under a log-linear specification of commuting costs, we show that the optimal public-private infrastructure mix on each edge depends exclusively on the trade-off between the relative marginal cost of operation and construction across technologies: even if public transit is more efficient to operate, edges where tunneling is costly, for example, will optimally rely on private provision.

Quantification. The Metropolitan Area of Mexico City provides a natural laboratory to study these trade-offs. Its transit system combines extensive formal public infrastructure with a large private sector, operates under binding fiscal constraints, and exhibits strong complementarities between public and private provision—three features that are central to the model. Private transit is quantitatively dominant: more than half of all trips are made using a minibus.[1] This extensive private network plays a central role in connecting peripheral areas largely underserved by mass transit infrastructure, reflecting the high fixed costs and fiscal constraints that limit public system expansion.

Public and private transit are also deeply intertwined along commuting paths: the majority of transit users rely on private transit for at least one leg of their journey, and most of these trips combine private and public services (Mosqueda, 2026). At the same time, private provision generates congestion externalities, with more than 80 private lines overlapping on many road links (Mosqueda, 2026). Together, these features make Mexico City a particularly informative setting to study how a planner should optimally allocate public and private transit technologies across a network, and how their interaction shapes commuting costs, congestion, and welfare in equilibrium.

[1]We calculate this from a travel survey, INEGI Encuesta Origen-Destino 2017.

To bring the model to the data, we combine a unique set of high-resolution data sources for Mexico City that jointly characterize the transit network, commuting environment, and spatial economic structure. The core input is a detailed General Transit Feed Specification (GTFS) collected by the private firm "Where Is My Transport". This dataset provides a unified representation of the city's full transit system, including both formal public mass transit—metro, BRT, and suburban rail—and private informal services such as *combis* and *micros*, with complete route-level information on network topology, frequencies, and travel times.

To motivate the analysis, we use these data to document two sets of facts about Mexico City's transit system. Private transit accounts for 85 percent of all route-kilometers in the network—nearly six times the coverage of public mass transit—and this share rises from 68 percent near the CBD to 100 percent in the outermost periphery. Public and private modes are not substitutes but are tightly linked within individual trips: 33 percent of all transit commutes combine both modes in a single journey, and combi-to-metro transitions alone make up half of all mode-to-mode switches. On average, commuters on these mixed trips spend 37 minutes on private transit for access and egress—comparable to the in-vehicle time on public segments—and 81 percent of metro stations have at least one combi stop within 100 meters. These patterns show that private transit serves a structural first-and-last-mile role, so that any evaluation of public investment must account for how it reshapes the private trips that feed into it.

We then calibrate our model. We use the GTFS dataset to construct a simplified transit network that consists of 427 nodes and 1,187 edges. Each node in the network is richly characterized. We leverage spatially granular microdata on wages, population, employment, and commuting flows from INEGI, as well as detailed land-use information from the Secretaría de Desarrollo Urbano. Together, the network and administrative microdata allow us to calibrate the two key primitives of the model: mode-specific operating costs and infrastructure construction costs. We calibrate the building and operating cost parameters to the equilibrium in which the population distribution and the network are fixed to the observed ones. We identify higher fixed investment costs to build public transit relative to private transit, but lower marginal operating costs. The calibrated framework thus enables us to quantitatively evaluate counterfactual infrastructure allocations.

Welfare effects of transit expansions. We study a set of counterfactual exercises to assess the gains from budget-feasible expansions of the transit system. The counterfactuals differ in the extent to which the government accounts for private transport infrastructure when designing the network. In particular, we consider two cases: i) one where the planner expands the network using only public transit, and ii) another where the planner considers both public and private investments. In both scenarios, we consider a 50% infrastructure budget increase and compare the two scenarios against the benchmark observed equilibrium where the network is fixed to the observed one.

We show that welfare gains from network expansion depend critically on how private transit is incorporated into the planning problem. When public infrastructure is designed without accounting for joint optimal expansions of the private network, the resulting public-only expansions deliver smaller welfare gains. We find that optimally expanding the public network alone delivers a 1.33% welfare gain, but the welfare gain under joint public-private planning is approximately 20% larger, reaching 1.59%. These results highlight that a significant share of the potential benefits from transit investment arises not from expanding public infrastructure alone, but from integrating private transport into network design decisions.

In terms of the spatial distribution of investments, we find that both public and private investments aim to better connect previously poorly connected productive hubs located on the outskirts of the city. We also show that new investments attempt to connect peripheral nodes to one another and to the existing transit infrastructure, which is predominantly concentrated in central areas of the city.

Taken together, the paper provides a unified framework for analyzing and computing optimal transit networks in budget-constrained cities. Rather than treating informal transit as a temporary fix or a policy failure, the analysis shows how its provision emerges endogenously from the planner's optimal choices and how its interaction with public transit shapes urban structure, commuting flows, and welfare.

Related literature. This paper relates to a growing literature on optimal transportation networks and infrastructure investment in general equilibrium. Several contributions study network design by optimizing over transport links or transport costs (Alder, 2025; Tarasov and Felbermayr, 2014; Allen and Arkolakis, 2022), typically treating transport technologies or bilateral costs as primitives and focusing on marginal or decentralized adjustments around an existing network. Our problem is also conceptually related to optimal transport network design in non-economic settings, which studies network formation under flow and cost constraints (Bernot et al., 2009). More closely related are recent papers that analyze optimal transport infrastructure and policy in spatial equilibrium models with endogenous flows and congestion (Felbermayr and Tarasov, 2022; Allen and Arkolakis, 2022; Almagro et al., 2024). Closest to our approach, Fajgelbaum and Schaal (2020) develops a quantitative framework for optimal transport networks in spatial equilibrium that allows for network-wide welfare evaluation; we build on their framework and extend it to a setting with heterogeneous transport technologies, adapting it from goods trade to within-city commuting and to the coexistence of public mass transit and decentralized private transit.

This paper relates to the literature on urban transit infrastructure and welfare in developing cities. More broadly, our analysis connects to the literature on aggregate welfare losses from misallocation (Rogerson and Restuccia, 2004; Hsieh and Klenow, 2009) and to work emphasizing spatial misallocation arising from geographic frictions or spatially targeted policies (Desmet and Rossi-Hansberg, 2013; Adamopoulos et al., 2022; Asturias et al., 2014; Fajgelbaum et al., 2019; Hsieh and Moretti, 2019). Within the urban context, existing work provides causal evidence that large-scale public transit investments can generate substantial welfare gains in congested urban environments, primarily through reductions in commuting costs and changes in residential sorting (Tsivanidis, 2019). Complementary studies evaluate the welfare effects of specific single-mode, publicly provided transit expansions, including cable car investments in Medellín (Khanna et al., 2024) and subway expansions in Mexico City (Zárate, 2024). In contrast to this literature, which typically evaluates isolated public transit projects or abstracts from the coexistence of multiple transport modes, our contribution is to study welfare and misallocation in a unified framework where public mass transit and decentralized private or informal transit jointly determine commuting patterns and welfare under binding fiscal constraints.

A large body of research studies how changes in transportation infrastructure and transport costs affect economic activity and productivity. Some papers estimate the effects of road expansion on productivity and industrial outcomes in the United States (Fernald, 1999), regional economic activity along highway corridors (Chandra and Thompson, 2000; Baum-Snow, 2007; Duranton et al., 2014), access to railways in India (Donaldson, 2018), and expressway construction in China (Faber, 2014). This literature highlights the central role of transportation networks in shaping regional growth and spatial economic outcomes. Our approach complements these studies by focusing on the optimal allocation of transit technologies within cities rather than on realized expansions of a single mode, allowing us to quantify how infrastructure choices interact with population distribution, congestion, and economic activity in general equilibrium.

This paper also contributes to the literature on the coexistence of formal and informal urban transport systems in developing cities. Recent work shows that public transit investments interact strongly with incumbent private transit, shaping prices, waiting times, entry, and welfare through equilibrium responses of private operators (Conwell, 2024; Björkegren et al., 2025). In particular, public entry can crowd out or reallocate private supply, with welfare effects that depend on commuters' tradeoffs between prices and waiting times and on private operators' responses along connected routes (Björkegren et al., 2025), while frictions internal to informal transit—such as queuing, scale economies, and market power—play a central role in determining equilibrium outcomes and policy effectiveness (Conwell, 2024). Closely related work examines optimal public transport networks and commuters' responses to waiting times and service frequency in large developing-country cities, abstracting from informal provision while highlighting the importance of system-wide equilibrium forces (Kreindler et al., 2023). We complement this literature by developing a quantitative spatial equilibrium model in which a budget-constrained planner chooses the optimal allocation of public mass transit and decentralized private/informal transit; the planner's solution is disciplined by an estimated elasticity of substitution between modes that governs their complementarity in equilibrium.

2.2 Model

We build on Fajgelbaum and Schaal (2020) and extend their framework for optimal road networks and trade between locations to a city setting with commuting and transit networks. The key difference is that the planner chooses mode-specific flows and mode-specific infrastructure, particularly public and private transit infrastructure, so an edge in the network is no longer summarized by one aggregate commuting flow and one infrastructure measure. The output of the planner's problem thus gives an allocation of residents, workers, and mode-specific flows and infrastructure across locations.

Environment. The economy consists of a discrete set of locations $\mathcal{J} = \{1, \dots, J\}$ and has a fixed number of workers of measure $\bar{L}$. Locations are arranged on a directed graph $(\mathcal{J}, \mathcal{E})$, where $\mathcal{E}$ denotes the set of edges.

Workers may live and work in different locations because they can commute. The mass of workers living and working in a given location j is denoted by L_j^R and L_j^L, respectively. Worker commuting between locations generates a commuting flow L_{jk} across any ordered node pair (j, k). Workers can only travel through connected locations. A connection is defined by the neighbor set of nodes; that is, any location j has neighbors $k \in \mathcal{N}(j)$. To reach any $k' \notin \mathcal{N}(j)$, workers must travel through a sequence of connected locations.

Let $\mathcal{M} \equiv \{\text{pub}, \text{pri}\}$ denote the set of transit technologies. For each edge $(j, k) \in \mathcal{E}$ and mode $m \in \mathcal{M}$, let $L_{jk}^m \geq 0$ denote the commuting flow from j to k using mode m. The aggregate flow on the physical edge is

$$L_{jk} \equiv L_{jk}^{\text{pub}} + L_{jk}^{\text{pri}}. \tag{2.1}$$

Analogously, the aggregate infrastructure on each edge is composed of public and private infrastructure:

$$I_{jk} \equiv I_{jk}^{\text{pub}} + I_{jk}^{\text{pri}}. \tag{2.2}$$

The planner therefore chooses not only how many workers move across each edge, but also which technology or type of investment to use on that edge.

Transport technology and commuting costs. Commuting consumes real resources, measured in units of labor. One can think of this as a literal loss of the daily time endowment, exhaustion, fatigue, or any other aspect of commuting that results in a loss of productive units of labor. Let T_{jk}^{m} denote the labor-equivalent commuting resource cost generated by mode-m flow on edge (j,k).

We assume that private transit is congestible. This captures the idea that, in practice, lower private-transit capacity implies longer waiting times because of boarding and matching externalities (Conwell, 2024). Let I_{jk}^{pri} denote private-transit infrastructure on edge (j,k). Following Fajgelbaum and Schaal (2020), we parameterize the private commuting resource cost on link (j,k) as

$$T_{jk}^{\text{pri}} = \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma}. \tag{2.3}$$

Here $\delta_{jk}^{\tau,\text{pri}}$ captures geographic characteristics that raise per-unit commuting costs, and θ_{pri} is a technology-specific shifter that captures differences in the marginal operating cost of private transit. The parameter β is the elasticity of congestion, while γ governs the elasticity with which private infrastructure reduces commuting costs.

Similarly, for public transit, traveling on edge (j,k) uses a constant amount of resources given by

$$T_{jk}^{\text{pub}} = \delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}} L_{jk}^{\text{pub}}, \tag{2.4}$$

where parameter $\delta_{jk}^{\tau,\text{pub}}$ captures link-specific public-transit frictions, while θ_{pub} captures the marginal operating cost of the public technology.

We assume that public transit has no congestion but has capacity constraints. Public ridership must satisfy

$$L_{jk}^{\text{pub}} \leq \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}, \qquad \forall (j,k) \in \mathcal{E}, \tag{2.5}$$

where $\kappa_{jk}^{\text{pub}} > 0$ converts public infrastructure into public passenger capacity.[2] Infrastructure does

[2]This is the minimal convex way to make the public network endogenous while preserving the assumption that

not lower the per-rider public operating cost in (2.4); it expands the set of feasible public flows.

The total commuting resource cost on a link is the sum of mode-specific resource costs:

$$T_{jk} = T_{jk}^{\text{pub}} + T_{jk}^{\text{pri}}. \tag{2.6}$$

This additive formulation is central to preserving convexity, which will allow a much more efficient computational solution to the model. We will explain this in detail in the next section.

Building costs. We assume that building investment requires the use of an exogenous resource K that is in fixed aggregate supply in the economy. This resource is the government's budget. The cost of building infrastructure varies across links and technologies, reflecting geographic heterogeneity. Building one unit of public infrastructure on link jk requires $\delta_{jk}^{I,\text{pub}}$ units of K, while building one unit of private infrastructure requires $\delta_{jk}^{I,\text{pri}}$ units. The link-specific costs $\delta_{jk}^{I,\text{pub}}$ and $\delta_{jk}^{I,\text{pri}}$ capture how geographic features—such as terrain, geology, existing urban density, and land acquisition costs—differentially affect the cost of building each type of infrastructure. Aggregating across all links in the economy, we obtain the following resource constraint:

$$\sum_{(j,k)\in\mathcal{E}} \left[\delta_{jk}^{I,\text{pub}}\, I_{jk}^{\text{pub}} + \delta_{jk}^{I,\text{pri}}\, I_{jk}^{\text{pri}}\right] \;\leq\; K. \tag{2.7}$$

Flow constraint. At each location, effective labor demand must be feasible given local residents and incoming commuters. Because commuting consumes labor resources, outgoing flows enter the constraint together with their mode-specific resource costs:

$$\underbrace{L_j^L}_{\text{Local workers}} + \underbrace{\sum_{k\in\mathcal{N}(j)}\sum_{m\in\mathcal{M}} L_{jk}^m}_{\text{Outgoing commuters}} + \underbrace{\sum_{k\in\mathcal{N}(j)}\sum_{m\in\mathcal{M}} T_{jk}^m}_{\text{Commuting resource costs}} \leq \underbrace{L_j^R}_{\text{Residents}} + \underbrace{\sum_{i\in\mathcal{N}(j)}\sum_{m\in\mathcal{M}} L_{ij}^m}_{\text{Incoming commuters}}. \tag{2.8}$$

The right-hand side is the amount of labor available at node j: residents plus commuters arriving from neighboring nodes. The left-hand side is the amount of labor used at node j: local employment,

public transit does not create congestion.

commuters sent to neighboring nodes, and labor-equivalent commuting losses.

Preferences. Workers derive utility from consumption of a homogeneous freely-traded final consumption good C and a non-traded good H, i.e., residential land or housing. Utility of an individual worker who consumes c units of the final good and h units of land is

$$U(c_j, h_j) = c_j^{\alpha} h_j^{1-\alpha}.$$

Production and land. There are only two goods in this economy: a freely-traded final consumption good and a non-traded good. Each location produces the final good with a Cobb-Douglas technology combining productivity, labor, and commercial land. That is,

$$C_j = A_j L_j^{L\psi} H_j^{c1-\psi}$$

Regarding the non-traded good, land, we assume that each location has a fixed supply of residential and commercial land that we denote by H_j and $H^c{}_j$, respectively.

2.2.1 Planner's Problem

The planner maximizes welfare by choosing consumption and labor allocations across locations, commuting flows across edges, and infrastructure investments in each link of the transit network. Formally, the planner solves

$$\max_{\substack{\{c_j, h_j, L_j^L, L_j^R\}_{j \in \mathcal{J}}, \\ \{L_{jk}^m\}_{(j,k)\in\mathcal{E},\, m\in\mathcal{M}}, \\ \{I_{jk}^{\text{pub}}, I_{jk}^{\text{pri}}\}_{(j,k)\in\mathcal{E}}}} u$$

subject to the constraints described below.

Spatial equilibrium. The planner guarantees every resident at least utility u, so that no location offers strictly lower welfare:

$$u \leq U(c_j, h_j) \qquad \forall\, j \in \mathcal{J}.$$

Balanced commuting flows. At each location, the supply of workers—residents plus incoming commuters—must cover demand—local workers plus outgoing commuters net of commuting costs:

$$\underbrace{L_j^L}_{\text{Local workers}} + \underbrace{\sum_{k \in \mathcal{N}(j)} \sum_{m \in \mathcal{M}} L_{jk}^m}_{\text{Outgoing commuters}} + \underbrace{\sum_{k \in \mathcal{N}(j)} \sum_{m \in \mathcal{M}} T_{jk}^m}_{\text{Commuting resource costs}} \leq \underbrace{L_j^R}_{\text{Residents}} + \underbrace{\sum_{i \in \mathcal{N}(j)} \sum_{m \in \mathcal{M}} L_{ij}^m}_{\text{Incoming commuters}}.$$

Commuting costs. The loss of resources on edge (j, k) depends on the modal composition of infrastructure, commuter volume, and total capacity:

$$T_{jk}^{\text{pri}} = \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma}, \qquad T_{jk}^{\text{pub}} = \delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}} L_{jk}^{\text{pub}}.$$

Public capacity. For every edge (j, k), public ridership is bounded by public passenger capacity:

$$L_{jk}^{\text{pub}} \leq \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}.$$

Population, goods, and land market clearing. Aggregate population is fixed at $\bar{L}$, aggregate final-good consumption cannot exceed aggregate production, and housing consumption is bounded by local residential land:

$$\sum_{j \in \mathcal{J}} L_j^R = \bar{L}, \qquad \sum_{j \in \mathcal{J}} c_j L_j^R \leq \sum_{j \in \mathcal{J}} C_j, \qquad h_j L_j^R \leq H_j \quad \forall j \in \mathcal{J}.$$

Network-building constraint. Total infrastructure spending is bounded by the aggregate budget K, and investment on each edge must respect pre-existing capacity bounds:

$$\sum_{(j,k)\in\mathcal{E}} \left[\delta_{jk}^{I,\text{pub}}\, I_{jk}^{\text{pub}} + \delta_{jk}^{I,\text{pri}}\, I_{jk}^{\text{pri}}\right] \;\leq\; K,$$

$$\underline{I}_{jk}^{\text{pub}} \leq I_{jk}^{\text{pub}} \leq \bar{I}_{jk}^{\text{pub}}, \qquad \underline{I}_{jk}^{\text{pri}} \leq I_{jk}^{\text{pri}} \leq \bar{I}_{jk}^{\text{pri}} \qquad \forall\,(j,k) \in \mathcal{E}.$$

Non-negativity. All allocation variables are non-negative:

$$c_j, h_j, L_j^R, L_j^L \geq 0, \qquad L_{jk}^{\text{pub}}, L_{jk}^{\text{pri}} \geq 0, \qquad I_{jk}^{\text{pub}}, I_{jk}^{\text{pri}} \geq 0.$$

The planner's problem can be thought of as a maximization problem in steps, where the planner first chooses i) the allocations of consumption, residents, and labor in each location, then ii) commuting flows across all location pairs, and finally iii) the optimal mix of public-private transit infrastructure investments along the network.

2.2.2 Optimal commuting flows

We now turn to analyze the main outcomes and economic mechanisms of the model. The central implication is that the optimal network is governed by comparative advantage across transit technologies. Public transit is attractive on links where its low operating cost and high passenger capacity compensate for the infrastructure resources required to build it. Private transit is attractive where its lower construction cost dominates, especially on low-volume links where congestion costs remain small. Thus, spatial variation in the optimal public-private network is driven by the interaction between operating costs, building costs, and equilibrium commuting volumes.

As in Fajgelbaum and Schaal (2020), the optimal flow allocation can be represented through a potential field. The multipliers on the local balanced-flow constraints determine the direction and intensity of commuting flows. Let P_j^F denote the multiplier on the balanced-flow constraint at node

j, and let P^L denote the shadow value of aggregate labor. Define

$$v_j \equiv P_j^F + P^L.$$

The term $P_k^F - P_j^F$ is the marginal gain from moving one additional worker from node j toward node k. The term v_j is the value of one unit of labor-equivalent resources used in commuting from node j.

For public transit, the resource cost is linear:

$$T_{jk}^{\text{pub}} = \delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}} L_{jk}^{\text{pub}}.$$

Let η_{jk}^{pub} be the multiplier on the public-capacity constraint

$$L_{jk}^{\text{pub}} \leq \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}.$$

The first-order condition for public flow is the no-arbitrage condition

$$P_k^F - P_j^F \leq (P_j^F + P^L)\delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}} + \eta_{jk}^{\text{pub}}, \tag{2.9}$$

with equality if $L_{jk}^{\text{pub}} > 0$. Thus the planner sends workers through public transit only when the value of moving a worker from j to k covers both the operating cost of public transit and the shadow cost of scarce public capacity.

Unlike private transit, public transit has constant marginal operating cost. Therefore (2.9) does not deliver an invertible formula for L_{jk}^{pub}. Instead, it determines whether public transit is active on the link. The level of public flow is pinned down jointly by network flow balance and the public-capacity constraint.

For private transit, the resource cost is

$$T_{jk}^{\text{pri}} = \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}} \left(L_{jk}^{\text{pri}} \right)^{1+\beta} \left(I_{jk}^{\text{pri}} \right)^{-\gamma}.$$

The social marginal resource cost of private commuting is

$$MC_{jk}^{\text{pri}} = (1+\beta)\delta_{jk}^{\tau,\text{pri}}\theta_{\text{pri}}\left(L_{jk}^{\text{pri}}\right)^{\beta}\left(I_{jk}^{\text{pri}}\right)^{-\gamma}. \tag{2.10}$$

The private-flow no-arbitrage condition is

$$P_k^F - P_j^F \le (P_j^F + P^L)(1+\beta)\delta_{jk}^{\tau,\text{pri}}\theta_{\text{pri}}\left(L_{jk}^{\text{pri}}\right)^{\beta}\left(I_{jk}^{\text{pri}}\right)^{-\gamma},$$

with equality if $L_{jk}^{\text{pri}} > 0$. Since the private marginal cost is strictly increasing in private flow when $\beta > 0$, this condition can be inverted:

$$L_{jk}^{\text{pri}} = \left[\frac{1}{1+\beta}\frac{\left(I_{jk}^{\text{pri}}\right)^{\gamma}}{\delta_{jk}^{\tau,\text{pri}}\theta_{\text{pri}}}\frac{\max\{P_k^F - P_j^F, 0\}}{P_j^F + P^L}\right]^{1/\beta}. \tag{2.11}$$

Private flows therefore increase with the destination potential $P_k^F - P_j^F$ and with private infrastructure, and decrease with operating costs and congestion.

2.2.3 Optimal transit network

Given mode-specific commuting flows, the planner chooses public and private infrastructure. Let P^K be the multiplier on the network-building constraint.

Public infrastructure. Public infrastructure creates public passenger capacity. It does not directly reduce the per-rider operating cost in (2.4). Let η_{jk}^{pub} denote the multiplier on the public-capacity constraint

$$L_{jk}^{\text{pub}} \le \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}.$$

For an interior public-infrastructure choice, the first-order condition is

$$\underbrace{\eta_{jk}^{\text{pub}}\kappa_{jk}^{\text{pub}}}_{\text{Marginal benefit of public capacity}} = \underbrace{P^K\delta_{jk}^{I,\text{pub}}}_{\text{Marginal cost of public infrastructure}}. \tag{2.12}$$

This condition states that the marginal benefit of the additional public flow allowed by this capacity must be equal to the opportunity cost of using the scarce network-building resource on that edge.

Combining (2.9) and (2.12), an active public edge with interior infrastructure satisfies

$$P_k^F - P_j^F = (P_j^F + P^L)\delta_{jk}^{\tau,\text{pub}}\theta_{\text{pub}} + \frac{P^K \delta_{jk}^{I,\text{pub}}}{\kappa_{jk}^{\text{pub}}}. \tag{2.13}$$

The first term on the right-hand side is the public operating cost of carrying one more rider, valued in units of the planner's objective. The second term is the infrastructure cost of creating one more unit of public passenger capacity. Thus public transit is used on links where the value of moving workers across the link is large enough to cover both operating costs and the infrastructure required to serve them. Further, if public infrastructure has no lower bound and the upper bound is slack, then public capacity binds on every active public edge:

$$L_{jk}^{\text{pub}} = \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}. \tag{2.14}$$

Otherwise, the planner could reduce I_{jk}^{pub} without changing public flows, freeing resources for other links. Hence, in this model public investment is proportional to the public flow it serves:

$$I_{jk}^{\text{pub}} = \frac{L_{jk}^{\text{pub}}}{\kappa_{jk}^{\text{pub}}}.$$

Private infrastructure. Private infrastructure works differently. It does not impose a hard capacity constraint; instead, it reduces the congestion cost of private commuting. For an interior choice of private infrastructure, the first-order condition is

$$\underbrace{(P_j^F + P^L)\gamma\delta_{jk}^{\tau,\text{pri}}\theta_{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma-1}}_{\text{Marginal reduction in private commuting costs}} = \underbrace{P^K \delta_{jk}^{I,\text{pri}}}_{\text{Marginal cost of private infrastructure}}. \tag{2.15}$$

The planner builds private infrastructure up to the point where the reduction in congestion costs equals the opportunity cost of the network-building resource. The benefit of private infrastructure is larger on links with more private flow, worse private commuting frictions, and a higher value of labor resources at the origin. It is smaller when private infrastructure is expensive to build or when the aggregate infrastructure budget is tight.

Solving (2.15) gives

$$I_{jk}^{\text{pri},*} = \left[\frac{\gamma (P_j^F + P^L) \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}} \left(L_{jk}^{\text{pri}} \right)^{1+\beta}}{P^K \delta_{jk}^{I,\text{pri}}} \right]^{1/(1+\gamma)} . \tag{2.16}$$

This is the mode-specific analogue of the optimal infrastructure condition in Fajgelbaum and Schaal (2020). Private investment rises with private flow and with the value of reducing commuting losses, and falls with construction costs and the shadow price of the network budget. Importantly, this expression makes explicit that optimal private infrastructure depends crucially on relative marginal operating and building costs.

2.2.4 Technology choice and the cost trade-off

The previous conditions characterize public and private infrastructure separately. We can combine them to recover the main economic trade-off behind the optimal network. Public transit has a low and constant operating cost, but carrying public riders requires public capacity. Private transit does not face a hard capacity constraint, but private riders generate congestion, so the marginal cost of private service rises with the utilization of private infrastructure. The planner therefore compares two ways of producing mobility on each link: public capacity, which has a constant generalized unit cost, and private transit, whose optimized marginal cost depends on operating costs, construction costs, and congestion. The following proposition states the main trade-off.

Proposition 2.1 (Technology choice and cost trade-offs)**.** *Consider a link (j,k) with slack infrastructure bounds. Suppose $\gamma > 0$ and $\beta \geq \gamma$. Define*

$$V_j \equiv P_j^F + P^L,$$

$$A_{jk}^{\text{pub}} \equiv \delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}}, \qquad A_{jk}^{\text{pri}} \equiv \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}},$$

and let

$$D_{jk}^{\text{pub}} \equiv P^K \delta_{jk}^{I,\text{pub}}, \qquad D_{jk}^{\text{pri}} \equiv P^K \delta_{jk}^{I,\text{pri}}$$

denote the shadow cost of building one unit of public and private infrastructure on link (j,k)*. If technology-specific commuting costs are given by* (2.3) *and* (2.4)*, then, on an active public edge, public capacity binds and the generalized unit cost of public service is*

$$\mathfrak{c}_{jk}^{\text{pub}} = \underbrace{V_j A_{jk}^{\text{pub}}}_{\textit{Operating cost}} + \underbrace{\frac{D_{jk}^{\text{pub}}}{\kappa_{jk}^{\text{pub}}}}_{\textit{Capacity-building cost}}, \tag{2.17}$$

and the generalized unit cost of private service is

$$\mathfrak{c}_{jk}^{\text{pri}}\left(L_{jk}^{\text{pri}}\right) = (1+\beta)\gamma^{-\gamma/(1+\gamma)} \left(V_j A_{jk}^{\text{pri}}\right)^{1/(1+\gamma)} \left(D_{jk}^{\text{pri}}\right)^{\gamma/(1+\gamma)} \left(L_{jk}^{\text{pri}}\right)^{(\beta-\gamma)/(1+\gamma)}. \tag{2.18}$$

If both modes are used on the same link, their generalized marginal costs are equal:

$$\mathfrak{c}_{jk}^{\text{pub}} = \mathfrak{c}_{jk}^{\text{pri}}\left(L_{jk}^{\text{pri}}\right). \tag{2.19}$$

Proof. See Appendix 2.C.2. □

Proposition 2.1 characterizes the condition that determines the optimal mix of public and private infrastructure in the network. Equation (2.17) shows that public transit is attractive when it has a low operating cost, A_{jk}^{pub}, or when public capacity is cheap to create, $\delta_{jk}^{I,\text{pub}}/\kappa_{jk}^{\text{pub}}$. Equation (2.18) shows that private transit is attractive when private operating costs and private construction costs are low.

The main economic trade-off in the model is therefore that public transit is a low-operating-cost technology that requires costly capacity, while private transit's marginal cost rises with congestion but may compensate with low investment costs. The planner thus assigns flows and infrastructure so that, on links where both modes are active, the generalized marginal cost of public capacity equals the optimized generalized marginal cost of private congestible service.

2.2.5 Convexity and multiplicity

The planner's problem is a large-scale network design problem. In principle, this problem may have many local optima, since infrastructure choices on one edge affect the returns to infrastructure on other edges through equilibrium commuting flows. Convexity is therefore important for both economic and computational reasons. Economically, it guarantees that the counterfactual network we compute is a global optimum rather than a local improvement around a particular initial guess. Computationally, it implies that the first-order conditions are sufficient and that different numerical solutions, if they arise, can be interpreted as either numerical error or as multiple global optima rather than as different local optima. Thus, convexity guarantees uniqueness of the counterfactual welfare level, while strict convexity would additionally guarantee uniqueness of the allocation.

We now state the condition under which our mode-specific formulation is convex.

Proposition 2.2 (Convexity of the planner's problem)**.** *Suppose that private commuting costs are given by*

$$T_{jk}^{\text{pri}} = \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}} \left(L_{jk}^{\text{pri}} \right)^{1+\beta} \left(I_{jk}^{\text{pri}} \right)^{-\gamma},$$

public commuting costs are given by

$$T_{jk}^{\text{pub}} = \delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}} L_{jk}^{\text{pub}},$$

and public infrastructure enters through the linear capacity constraint

$$L_{jk}^{\text{pub}} \leq \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}.$$

If $\beta \geq \gamma \geq 0$, then the mode-specific transport block is convex. Consequently, the full planner's problem is convex, up to the standard homothetic aggregate-variable transformation in the mobile-

labor case,[3] *because all other constraints in the problem are linear or convex and the objective is linear in u. If $\beta = \gamma > 0$, the private transport block is convex but not strictly or strongly convex.*

Proof. See Appendix 2.C.3. □

Note that the relevant object is the total commuting resource cost that enters the balanced-flow constraint. For the congestible private technology, suppressing edge subscripts, the link-level resource cost is

$$T^{\text{pri}}(L, I) = AL^{1+\beta} I^{-\gamma}, \qquad A > 0. \tag{2.20}$$

Intuitively, the parameter γ governs the returns to infrastructure, while β governs congestion. If infrastructure is too effective relative to congestion, then the transport technology has increasing returns: more infrastructure lowers costs, lower costs attract more flows, and more flows increase the incentive to invest in the same link. The condition $\beta \geq \gamma$ rules out this increasing-returns force by requiring congestion to be strong enough relative to the returns to infrastructure. The public technology, on the other hand, is convex by construction. Public operating costs are linear in public ridership, and public infrastructure enters through the linear capacity constraint.

The full planner's problem inherits the convexity of the transport block. The objective is linear in the welfare level u. The population constraint, network-building constraint, public-capacity constraint, infrastructure bounds, and non-negativity constraints are linear. The production technology and the utility constraint are concave-side feasibility constraints under an aggregate-variable representation.

The boundary case $\beta = \gamma$ deserves special attention. In that case,

$$T^{\text{pri}}(L, I) = AL^{1+\beta} I^{-\beta}.$$

[3] By aggregate-variable representation, we mean rewriting individual consumption and housing choices as aggregate objects, for example $X_j = c_j L_j^R$ and $R_j = h_j L_j^R$, and using the perspective form implied by homothetic preferences. This removes bilinear terms such as $c_j L_j^R$ and $h_j L_j^R$ from the resource constraints. The per-capita notation in the text is expositional; the computational formulation uses the aggregate-variable version.

This function is homogeneous of degree one:

$$T^{\text{pri}}(tL, tI) = tT^{\text{pri}}(L, I) \qquad \text{for all } t > 0.$$

Hence the function is linear along rays $(L, I) \mapsto (tL, tI)$. This implies that the private transport block is not strictly convex. As a result, different allocations of private flow and private infrastructure can deliver the same congestion ratio and the same welfare value.

Finally, if $\gamma > \beta$, the private transport block is non-convex. This is the increasing-returns case emphasized by Fajgelbaum and Schaal (2020). In that case, the first-order conditions are no longer sufficient for a global optimum, and the computed network may depend on the initial condition of the solver.

As we show in the next sections, we calibrate these parameters using detailed network data and obtain $\beta > \gamma$; thus, the baseline equilibrium and counterfactuals guarantee a unique global solution.

2.3 Illustrative example in a stylized city

To build intuition and understand the key insights from the model, consider a stylized city that consists of a simple grid of 5-by-5 locations arranged on a two-dimensional plane. In this city, all nodes are equally productive except for the center node, which is twice as productive. All locations have the same endowment of the non-traded good, land. The size of the city is $\bar{L} = 1{,}000$ labor units. Locations are connected via a network in which each node has up to 8 neighbors (horizontal, vertical, and diagonal connections). The planner allocates a budget $K = 1{,}000$ across edges, choosing both the level of infrastructure investment and the share devoted to public versus private transit on each link. The baseline parameterization in this economy is

- Transport congestion parameters: $\beta = 0.8$ and $\gamma = 0.4$.

- Transport technology parameters: $\theta_{\text{pub}} = 0.1$ and $\theta_{\text{pri}} = 0.5$.

- Infrastructure costs: $\delta_I^{\text{pub}} = 5 \cdot \delta_{jk}$, and $\delta_I^{\text{pri}} = 1 \cdot \delta_{jk}$.

Here δ_{jk} is a Euclidean distance matrix normalized to 1 for vertical and horizontal edges. To build intuition, we set the marginal cost parameter of the private technology to be five times larger than that of the public technology and, conversely, the fixed cost of building public technology to be five times larger than that of the private technology.

Figure 2.1 illustrates the baseline equilibrium of the model in this environment, which helps clarify the main economic mechanisms at work. Although all locations are symmetric in terms of land endowments and network connectivity, the higher productivity of the central node induces a strong spatial reallocation of both residents and infrastructure. Population concentrates disproportionately at the center and, to a lesser extent, in nearby nodes, as shown in panel (a), reflecting the trade-off between commuting costs and access to higher productivity. The planner responds endogenously by concentrating public investment on the central corridors, while private links form a thinner access layer through the inner grid and toward the outer ring (panel (b)). As a result, commuting flows exhibit a pronounced radial pattern toward the central node, with the largest flows on the final approaches to the center (panel (c)). Together, these panels highlight how small exogenous productivity differences can generate large endogenous spatial asymmetries in population density, infrastructure provision, and commuting patterns.

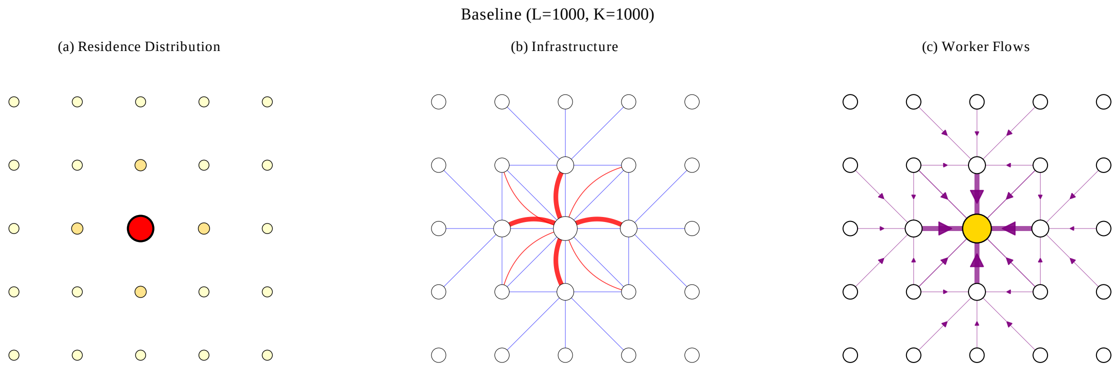


Figure 2.1: Equilibrium output: baseline example

Note: The figure illustrates the baseline equilibrium in a stylized 5-by-5 spatial economy. Nodes represent locations and edges denote commuting links between neighboring locations (horizontal, vertical, and diagonal). In panel (a), node size and color intensity capture the equilibrium residential distribution, with darker and larger nodes indicating higher population density. Panel (b) displays the optimal allocation of infrastructure investment across links. Red lines represent public infrastructure; blue lines represent private infrastructure. Thicker and darker edges correspond to higher total investment levels. Panel (c) shows equilibrium commuting flows, where arrow direction indicates the direction of worker movements and arrow thickness is proportional to the volume of commuters. The central node is twice as productive as all other locations. Total labor supply is fixed at $L = 1000$ and total infrastructure spending is constrained by $K = 1000$.

2.3.1 Comparative Statics

In this subsection, we examine how the network is designed when i) we increase the planner's budget, ii) we change the marginal cost parameters, or iii) we change the marginal building-cost parameters. In the following exercises, we describe the intuition behind each of these scenarios.

Increasing budget K. How does the network change when we increase the planner's budget? In particular, we multiply the baseline budget by factors of 1.2, 1.5, and 2. Figure 2.2 shows the optimal transit network for different values of K. The baseline case is shown in panel (a). First, public infrastructure (red curves) is concentrated on the links connecting the center to its immediate neighbors, reflecting high commuting demand toward the productive center. As K increases, these public links expand mainly on the intensive margin: the red curves around the CBD become thicker, while public investment remains concentrated on the highest-flow corridors. Second, private infrastructure (blue lines) provides a low-cost connectivity layer around the public core. For $K = 1000$, 1200, and 1500, this layer has a sparse star-and-inner-grid footprint that channels access toward the center. At $K = 2000$, private infrastructure expands across many more horizontal, vertical, and diagonal links in the outer grid, filling out the network where the fixed cost of public transit is not justified. Thus, additional resources do not lead the planner to build public transit everywhere; they deepen public capacity near the center and use private transit to extend the extensive margin of connectivity.

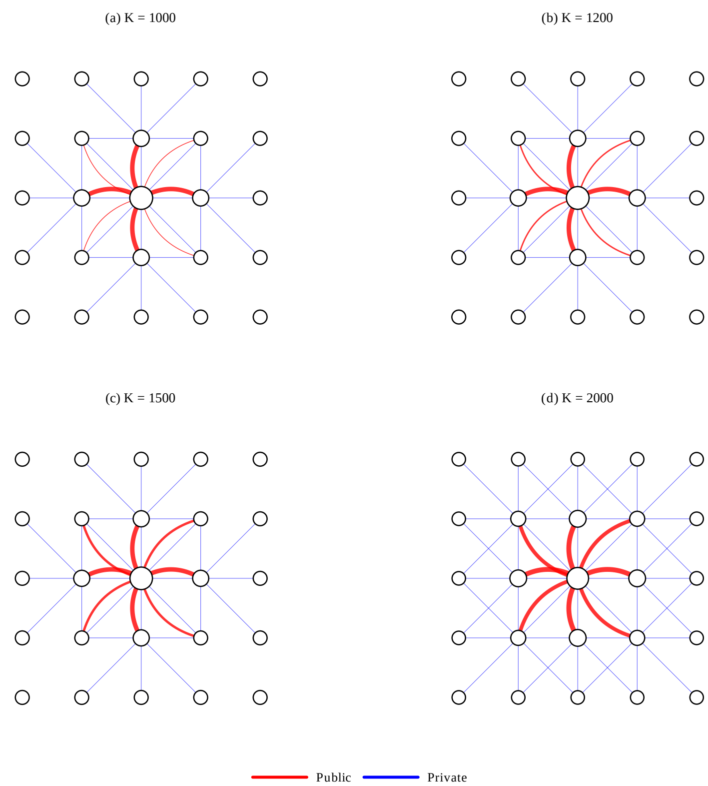


Figure 2.2: Optimal network for different budgets K

Note: The figure shows the optimal infrastructure allocation for a 5-by-5 grid city with a single CBD at the center. Red curved lines represent public transit infrastructure; blue straight lines represent private transit infrastructure. Line width is proportional to infrastructure investment, normalized globally across all panels to enable direct comparison. Node size reflects the number of residents at each location. Baseline calibration: $\bar{L} = 1{,}000$, $\beta = 0.8$ (congestion), $\gamma = 0.4$ (infrastructure elasticity), $\alpha = 0.7$ (consumption share), $\theta_{\text{pub}} = 0.1$, $\theta_{\text{pri}} = 0.5$, $\delta^I_{\text{pub}} = 5\delta$, $\delta^I_{\text{pri}} = \delta$.

Increasing marginal costs θ. How does the network change when the private technology becomes relatively cheaper or more expensive to operate? Relative to the baseline scenario where $\theta_{\text{pri}} = 0.5$, here we compare $\theta_{\text{pri}} = 0.1$ and $\theta_{\text{pri}} = 1.0$. Figure 2.3 shows the networks under these parameters in panels (a) and (b), respectively. When private transit is cheap to operate ($\theta_{\text{pri}} = 0.1$, panel (a)), private infrastructure expands across much of the grid and becomes thick on some central and diagonal links. Public transit remains important on the highest-demand corridors, but the planner can rely more heavily on private links to extend network coverage. When private transit becomes more expensive ($\theta_{\text{pri}} = 1.0$, panel (b)), the private footprint contracts back toward the inner grid and the central public corridors become relatively more important. In both cases, public transit is concentrated where flows are thickest, while private transit provides flexible connectivity where the construction cost advantage is valuable. This illustrates a key insight: the optimal technology mix depends not on absolute costs but on the relative efficiency of public versus private provision.

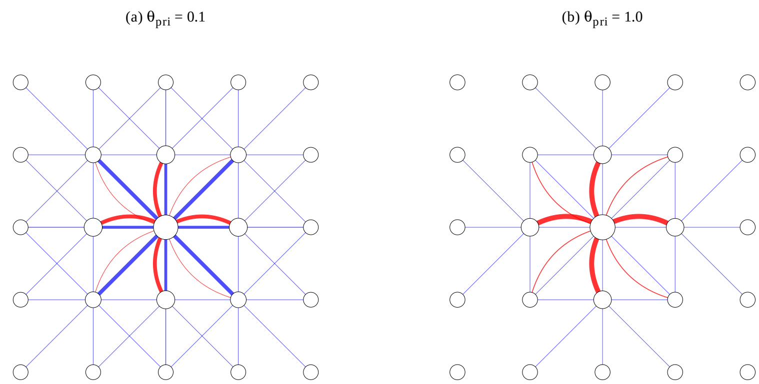


Figure 2.3: Optimal network for different marginal costs

Note: The figure shows the optimal infrastructure allocation for a 5-by-5 grid city with a single CBD at the center. Red curved lines represent public transit infrastructure; blue straight lines represent private transit infrastructure. Line width is proportional to infrastructure investment, normalized globally across all panels to enable direct comparison. Node size reflects the number of residents at each location. Baseline calibration: $\bar{L} = 1{,}000$, $\beta = 0.8$ (congestion), $\gamma = 0.4$ (infrastructure elasticity), $\alpha = 0.7$ (consumption share), $\delta^I_{\text{pub}} = 5\delta$, $\delta^I_{\text{pri}} = \delta$.

Increasing building costs δ. How does the network change when the public technology becomes relatively cheaper or more expensive to build? Holding $\delta^I_{\text{pri}} = \delta$ fixed, Figure 2.4 compares a case where public infrastructure is as cheap to build as private infrastructure ($\delta^I_{\text{pub}} = \delta$) to the baseline case where public infrastructure is five times more expensive ($\delta^I_{\text{pub}} = 5\delta$). When public infrastructure is cheap to build (panel (a)), the planner builds a much denser mixed network: public transit is thick on central corridors and also appears on additional links around the inner grid and the outer edges, while private infrastructure creates broad connectivity across the city. When public infrastructure is more expensive (panel (b)), investment is much more concentrated: public transit remains on the central spokes, and private transit appears as a sparse access layer around the center and toward the outer ring. The planner therefore preserves public transit for the routes where its operating-cost advantage is most valuable and relies less on it when construction costs rise.

The key insight from this exercise illustrates the main trade-offs of the model: construction costs (δ) and operating costs (θ) interact to determine the optimal technology mix. Public transit is used where its low operating cost can justify the fixed cost of capacity, while private transit provides a lower-cost way to preserve connectivity on links where demand is not large enough to support public investment. The quantitative relevance of these mechanisms depends on the empirical magnitudes of the cost parameters and the structure of the observed network. We document these features for Mexico City in the next section before turning to calibration and counterfactuals.

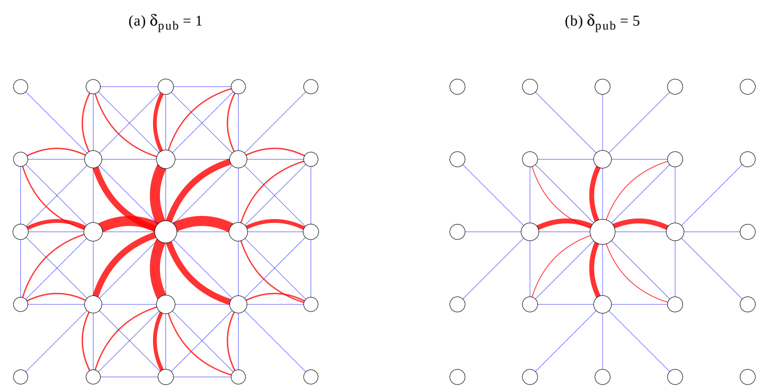


Figure 2.4: Optimal network for different building costs

Note: The figure shows the optimal infrastructure allocation for a 5-by-5 grid city with a single CBD at the center. Red curved lines represent public transit infrastructure; blue straight lines represent private transit infrastructure. Line width is proportional to infrastructure investment, normalized globally across all panels to enable direct comparison. Node size reflects the number of residents at each location. Baseline calibration: $\bar{L} = 1{,}000$, $\beta = 0.8$ (congestion), $\gamma = 0.4$ (infrastructure elasticity), $\alpha = 0.7$ (consumption share), $\theta_{\text{pub}} = 0.1$, $\theta_{\text{pri}} = 0.5$.

2.4 Private Transit in Developing Cities: Facts from Mexico City

We now turn to Mexico City to document that the model's central ingredients—the co-existence of private and public transit provision and the complementarity between modes—are quantitatively salient features of the data, and to motivate the calibration strategy of Section 2.5.

In developing cities, informal transit is not a transitional fix but a structural feature of urban mobility: privately operated minibuses and colectivos today provide between 50 and 100 percent of urban transit trips in many cities (Conwell, 2024), yet their interaction with public mass transit is poorly understood. Recent evidence shows these modes are far from independent—public investments reshape prices, waiting times, and route coverage through equilibrium responses of private operators (Björkegren et al., 2025; Conwell, 2024)—so studying them in isolation risks missing the most important channels through which infrastructure shapes welfare.

Mexico City is an ideal setting to address these questions. It combines one of the largest public rail networks in the Western Hemisphere with an extensive web of informally operated minibuses, operates under binding fiscal constraints, and benefits from unusually rich microdata on both modes (Mosqueda, 2026). Before documenting the set of facts that discipline the model's calibration and motivate the counterfactual exercises in Section 2.6, we describe the data sources that underlie our analysis.

2.4.1 Data

We exploit a unique dataset that describes the universe of private and public transit routes and combine it with granular administrative data.

WhereIsMyTransport data. Public and privatized transport supply is described using a GTFS (General Transit Feed Specification) dataset produced by WhereIsMyTransport for 2018. A GTFS is a standardized dataset specification composed of a set of interrelated text files that provide a structured description of transit systems. These files define route identifiers, stop locations, trip schedules, service calendars, and operational attributes, allowing for a spatially explicit representation of the transport network that can be linked to geographic information systems.

In addition to the usual formal public transport services such as subway and BRT, the WhereIsMyTransport GTFS feed incorporates private providers such as *microbuses* and *combis*, which are often underrepresented or absent in official transit datasets. For these services, the dataset reports route geographies, stop locations, operator or association identifiers, service frequencies, and weekly schedules. As a result, the GTFS feed provides a comprehensive description of both public and private transport supply in the metropolitan area, enabling consistent measurement of network structure and spatial coverage across modes.

Table 2.1: GTFS Line Characteristics by Transit Technology

Statistic	Public	Private
Lines	96	2,191
Stops per line	26.1	15.1
Line length (km)	15.5	13.1
Line speed (km/h)	24.8	15.8
Time between stops (min)	1.9	4.3
Average peak headway (min)	5.9	9.7
Average lowest scheduled headway (min)	5.6	8.2
Route-km share (%)	4.9	95.1

Note: Notes: The table summarizes the transit routes and schedules in the 2018 WhereIsMyTransport General Transit Feed Specification (GTFS) data for the Mexico City metropolitan area. A line is a distinct route pattern in the schedule data, so branches and opposite directions of the same named route may be counted separately. Private lines are informal colectivo and minibus services; public lines are formal rail, bus rapid transit, trolleybus, cable, and suburban rail services, following the same classification used in the calibration. Line length is measured along the mapped route geometry. Stops, length, speed, and time between stops are first computed at the scheduled-trip level and then averaged to the line level using scheduled service frequency as weights; table entries report averages across lines. Average peak headway is the average time between departures during weekday 06:00–10:00 service. Average lowest scheduled headway is the shortest scheduled time between departures observed for each line over the service week, averaged across lines. Route-km share is each technology's share of total classified line kilometers.

Table 2.1 summarizes the resulting transit supply measures by ownership type. The private network accounts for the overwhelming majority of route coverage, with 2,191 line patterns and 95 percent of classified route-kilometers, compared with 96 public line patterns. Public lines are faster on average and have shorter headways, while private lines are more spatially extensive and operate with fewer stops per line and longer times between stops. These differences highlight the central role of private transit in providing network coverage, despite its lower scheduled speeds and less frequent service relative to formal public transport.

INEGI data. We rely on two key datasets from Instituto Nacional de Estadística y Geografía (INEGI). First, the Encuesta Origen-Destino (EOD) 2017 for the Zona Metropolitana del Valle de México is a large-scale household travel survey that records commuting flow patterns within the metropolitan area. The EOD provides trip-level microdata, including trip origin and destination, departure and arrival times, trip purpose, sequence of transport modes used, including both public and private modes, monetary travel costs by mode, and walking time at the end of the trip.

Second, we use data from the Economic Census 2019, which describes wages at the firm (and workplace) level for the universe of firms in Mexico.

Urban Planning Ministry data. Residential and commercial land use data come from the Urban Planning Ministries of Mexico City and the State of Mexico (Secretaría de Desarrollo Urbano e Infraestructura (SEIDUVI) and Secretaría de Desarrollo Urbano y Vivienda (SEDUVI)). These datasets describe land use at the tract, block, and residential/commercial unit levels.

2.4.2 Fact 1: Private Transit Dominates Network Coverage and Peripheral Connectivity

The first fact is about the scale of private transit relative to public transit. Private modes account for the large majority of route-kilometers and origin–destination pairs served in developing cities (Conwell, 2024; Kreindler et al., 2023), yet this is often overlooked in policy discussions that focus on formal infrastructure. Understanding the size of private transit is necessary for any welfare analysis of network design, since it determines the baseline connectivity that commuters face and how much public investment can improve access.

Table 2.2: Comparison between private and public transportation in usage and route-kilometers covered

Metric	Private	Public
Share of total route km covered	0.85	0.15
Share of trips that use this transport	0.53	0.33
Share of trips that **only** use this transport	0.32	0.12
Avg share of trip time	0.35	0.19

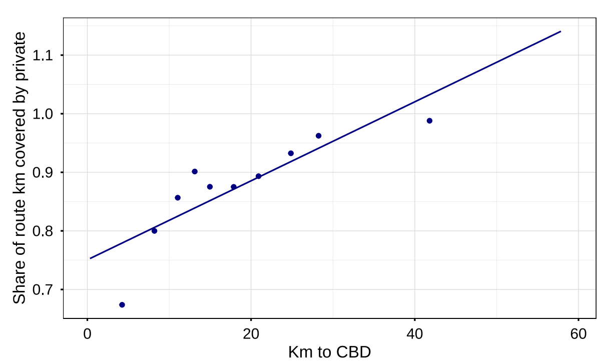


Figure 2.5: Distance to the CBD and share of km covered by private transportation

Table 2.2 shows the relative size of private and public transit in the Mexico City metropolitan area. Private transit accounts for 85 percent of total route-kilometers in the network, compared to just 15 percent for public mass transit—a ratio of nearly 6 to 1. In absolute terms, the private network covers 28,373 route-kilometers versus 5,002 for formal public transit. Despite this large geographic footprint, private transit carries 53 percent of all transit trips while public transit carries 33 percent, reflecting that public infrastructure serves high-density, high-frequency corridors that concentrate demand. Among trips that use only one mode, 32 percent use only private transit versus 12 percent that use only public transit. On average, private transit accounts for 35 percent of total trip time, compared to 19 percent for public transit, showing that private modes make up a large share of the actual travel experience even on multimodal journeys.

The geographic distribution of coverage is closely tied to distance from the city center. Figure 2.5 shows a strong positive relationship between distance to the CBD and the share of route-kilometers covered by private transit: further from the center, private transit becomes increasingly dominant, filling gaps that public infrastructure does not reach.

Figures 2.6 and 2.7 make this spatial pattern visible across the metropolitan area. The left panel shades each district of the ZMVM by the share of route-kilometers operated by private transit (combis) and overlays formal public transit lines—Metro, BRT (Metrobús), and commuter rail (Tren Suburbano)—in purple. Two patterns stand out. First, formal public infrastructure is essentially confined to a roughly 15 km radius around the CBD: virtually all Metro, BRT, and rail corridors fall inside the innermost concentric ring. Second, the periphery of the metropolitan area is overwhelmingly served by private operators: most peripheral districts are shaded at 90 percent or more private coverage, and many reach 100 percent. The right panel quantifies this center–periphery gradient by decomposing total route-kilometers into distance bands and reporting the private share within each. Even in the innermost 0–5 km ring—the part of the city best served by formal infrastructure—private operators already account for 68 percent of route-kilometers. The private share rises monotonically with distance: from 78 percent in the 5–10 km band, to over 90 percent beyond 20 km, and to 100 percent in the outermost 35 km+ ring. Together, the two figures show that the dominance of private transit is not a city-wide average masking a balanced mix—it is a structural feature of the periphery, consistent with the model's prediction that the high fixed costs of public transit lead a budget-constrained planner to concentrate mass-transit investment on high-density central corridors and to leave peripheral areas to private provision (Mosqueda, 2026).

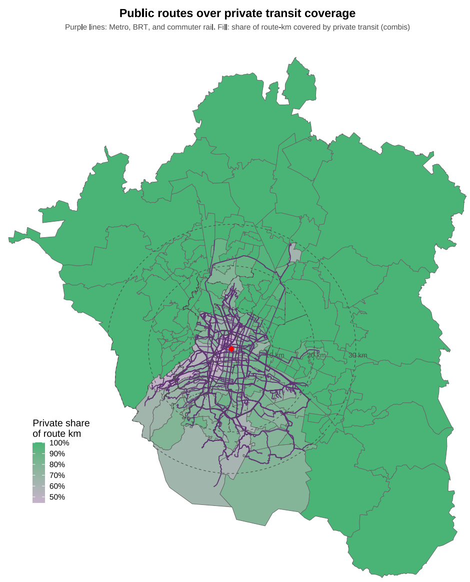


Figure 2.6: Spatial Distribution of Private Transit Coverage

Notes: Each district is shaded by the share of route-kilometers served by private transit (combis). Purple lines indicate formal public transit routes (Metro, BRT, and commuter rail). Dashed circles mark 10, 20, and 30 km from the CBD (red diamond).

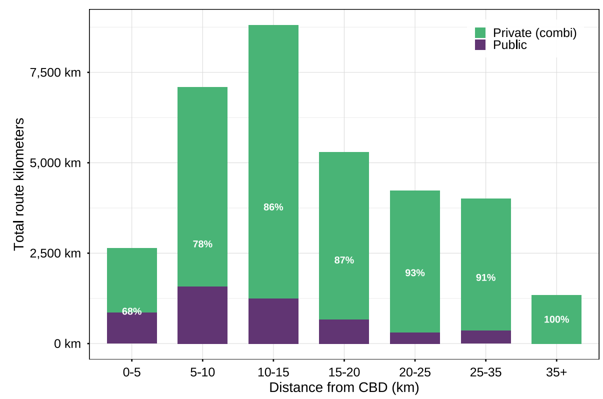


Figure 2.7: Route-Kilometers by Distance from CBD

Notes: Each bar shows total route-kilometers by distance band from the CBD, decomposed into private (combi) and public transit. Percentages indicate the private share within each band.

2.4.3 Fact 2: Private and Public Transit Are Strongly Complementary

A key feature of Mexico City's transit system is that public and private modes do not simply substitute for one another: they are tightly linked within individual commuting trips. This complementarity matters for understanding the welfare effects of transit policy, since any expansion of public infrastructure will affect not only the travelers who switch to it directly but also the private transit trips that feed into and out of it (Björkegren et al., 2025). How much commuters combine modes along a single journey also determines how operating cost differences between technologies show up in aggregate welfare—a key input to the model's calibration.

Tsivanidis (2019) documents substantial mode-combining behavior in Bogotá following the TransMilenio BRT expansion, where informal feeders absorbed much of the access demand created by new trunk lines. Similar patterns appear in the context of cable car expansions in Medellín (Khanna et al., 2024) and subway expansions in Mexico City (Zárate, 2024). Our data allow us to describe this complementarity more precisely and at finer spatial detail than prior work.

Table 2.3: Complementarity of public and private transit

Metric	Value
Share of transit trips: public only	0.18
Share of transit trips: private + public (mixed)	0.33
P(uses private | uses public)	0.65
Avg access+egress time on private (min)	37.34
Share metro stations with >= 1 combi stop (100 m)	0.81
Share metro stations with >= 2 combi stops (100 m)	0.64
Avg combi stops per metro station (100 m)	2.44

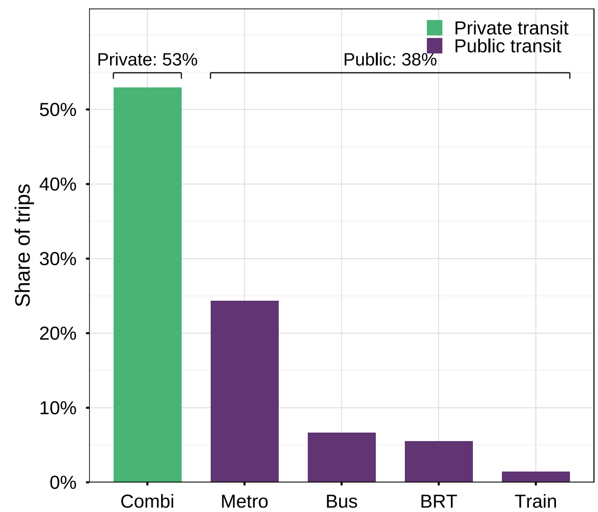


Figure 2.8: Share of trips by transportation mode

Table 2.3 documents several dimensions of this complementarity. Among all transit trips in the metropolitan area, 33 percent combine private and public services within a single journey. Conditioning on trips that use public transit at least once, 65 percent also involve private transit, meaning that reaching public infrastructure almost always requires at least one private transit leg. On average, commuters on mixed trips spend about 37 minutes on private transit for access and egress combined—a number that is similar to or exceeds the in-vehicle time on many public transit segments, meaning that private transit operating costs are incurred across nearly the full length of multimodal commutes. Figure 2.8 adds to this picture: while combis account for the largest share of individual transit trips (53 percent), metro and BRT together account for 38 percent, and the overlap between these categories in individual journeys is large, showing that the two technologies serve complementary rather than competing roles.

Figures 2.9 and 2.10 dig into the structure of this complementarity at both the trip and the infrastructure level. The left panel reports the transition matrix of consecutive mode pairs within trips that use more than one mode (excluding walking segments). The matrix is highly asymmetric and concentrated on a small number of cells. By far the single largest cell is combi-to-metro, which alone accounts for 49.6 percent of all observed mode-to-mode transitions. Combi-to-BRT (5.9 percent) and combi-to-bus (4.9 percent) follow, and the dominant reverse transition is metro-to-combi (14.4 percent). The pattern is therefore not one of commuters switching between competing public modes, but of private transit feeding commuters into a public trunk service—typically the Metro—and then absorbing them again on the return leg. The right panel shows that this behavioral pattern has a direct counterpart at the infrastructure level: it maps every Metro station in Mexico City and colors it by the number of combi stops located within 100 meters. Eighty-one percent of Metro stations have at least one combi stop in their immediate vicinity, 64 percent have two or more, and the average is 2.44 combi stops per station. This spatial co-location is not a coincidence: private feeder services have grown organically around mass-transit access points, providing the physical interface that makes combi-to-metro transitions possible. Taken together, the two figures show that public and private transit in Mexico City are linked both behaviorally—commuters routinely chain them within a single journey—and infrastructurally—private stops cluster around public stations. Any evaluation of public-network investment that ignores this complementarity will understate both its benefits, since new public capacity will inherit a dense private feeder network, and its reach, since most new riders will arrive on a combi.

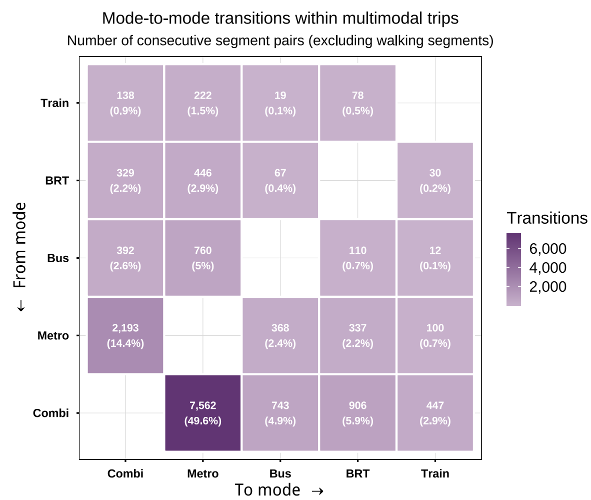


Figure 2.9: Mode-to-Mode Transitions within Multimodal Trips

Notes: Each cell reports the number and share of consecutive segment pairs transitioning from the row mode to the column mode within multimodal trips, excluding walking segments. Darker shading indicates higher frequency. Weekday work commutes, EOD 2017.

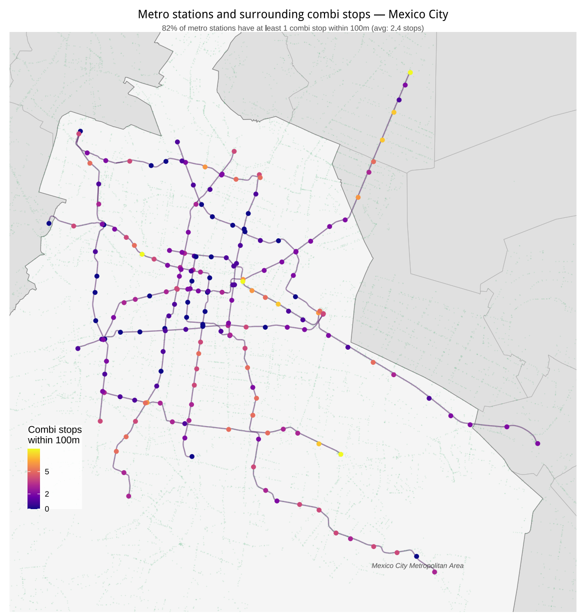


Figure 2.10: Metro Stations and Surrounding Combi Stops

Notes: Each dot represents a Metro station, colored by the number of combi stops within 100 meters. Eighty-one percent of stations have at least one combi stop nearby, with an average of 2.44 stops per station.

These patterns have two important implications for the model. First, they support modeling public and private transit as jointly shaping network-level commuting costs, since individual journeys routinely combine private access and egress with public trunk segments. Second, they imply that the welfare gains from public transit investment come in part through their effect on the private transit trips that connect commuters to public infrastructure—a channel that would be missed entirely in a single-mode framework.

Together, these two facts show that private transit in Mexico City is not a marginal or residual phenomenon: it accounts for the large majority of network coverage (Fact 2.4.2) and plays a structural first-and-last-mile role in multimodal commuting (Fact 2.4.3). They also highlight the central tension that motivates the model: private transit fills gaps that public infrastructure cannot afford to serve, yet it does so at higher social cost and with greater congestion externalities than an unconstrained planner would choose (Mosqueda, 2026). Quantifying the welfare cost of this arrangement, and the gains from moving toward an optimal mix of public and private investment, is the task of the remainder of the paper.

2.5 Calibration

2.5.1 Building a simplified network

Transit systems are more complex to represent than standard road networks. GTFS data contain many overlapping routes across multiple modes, transfer points, direction-specific patterns, service-frequency differences, and mode/operator heterogeneity, all of which generate a very high-dimensional network object. A simplified representation is therefore useful: it preserves the broad spatial structure and baseline service patterns while keeping the state space manageable for computation.

We build a simplified transport network from GTFS route shapes by overlaying the metropolitan study area with a regular hexagonal grid of 5-km flat-to-flat hexagons. Figure 2.11 shows the simplified network by transit type. Each hexagon is treated as a node, located at its centroid.

Hexagons are convenient because they provide a uniform spatial partition, avoid dependence on irregular administrative boundaries, and yield a relatively isotropic local neighborhood structure (each interior cell has the same number of immediate neighbors). This makes adjacency, distance calculations, and spatial aggregation of fundamentals more stable and transparent. The grid is clipped to the metropolitan boundary, and we compute an undirected geometric adjacency graph by linking hexagons that share a border. This geometric graph defines the set of feasible links in the model. The final simplified graph consists of 427 nodes and 1,187 edges.

GTFS information is then used to identify which neighboring hex-to-hex links are served by transit and to construct baseline infrastructure measures. In broad terms, we map observed routes onto the hex grid, translate them into movements across adjacent cells, and then aggregate these movements into edge-level transit intensity measures. This gives us a simplified representation of existing service on the network, including route counts split into formal and informal services. This produces two related objects: i) an adjacency matrix over all neighboring hexagons—the set of links the model may invest in—and ii) GTFS-based edge attributes, including lower bounds for the observed public and private infrastructure.

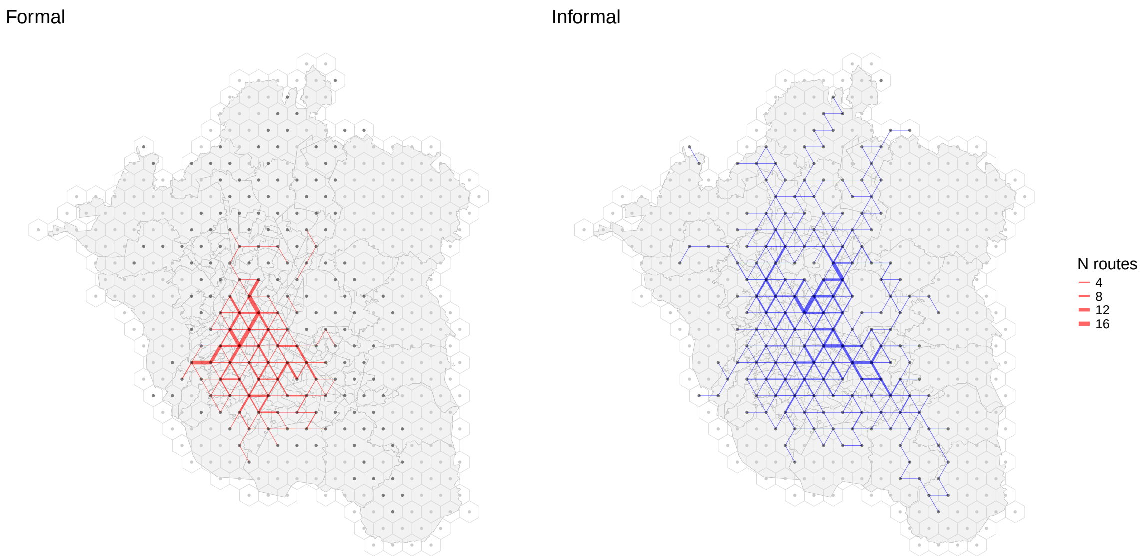


Figure 2.11: Simplified networks by public/private type

Note: The figure shows the simplified transit networks by public or private type. The city was discretized into symmetric hexagons, with centroids representing the nodes in the network. Neighboring nodes are connected if at least one transit line passes between them. The thickness of the lines represents the number of transit routes that connect two given nodes. Network data come from WhereIsMyTransport's GTFS file.

2.5.2 Parameter calibration

Marginal costs of operation θ. To calibrate the marginal cost parameters θ_{pub} and θ_{priv}, we exploit the granular detail of the GTFS file. In particular, we use speed as a proxy for the efficiency of each technology and its social marginal cost. Speed reflects the quality of the infrastructure and the units that run through it, as well as the scope for potential congestion. Using the GTFS, we compute the speed of all trips recorded for each individual transit line, and we take the average speed of public and private lines. We normalize $\theta_{\text{pub}} = 1$ because these parameters are shifters in our commuting cost specification and only their relative levels matter. Taking the ratio of average speeds across public and private lines yields $\theta_{\text{priv}} = 1.64$. This suggests that minibuses are 64% more costly to operate than public options such as the subway, BRT, and train.

Building costs δ. To calibrate the construction parameters δ_{pub} and δ_{priv}, we follow Fajgelbaum and Schaal (2020) and use the government budget constraint (2.7) to recover the δ parameters that are consistent with the existing (observed) infrastructure. First, we define the infrastructure measures

I_{jk}^{pri} and I_{jk}^{pub} to be the total number of private and public transit lines, respectively, connecting any two nodes in the network.[4] Second, we assume building costs are constant scalars that vary only across public and private modes, but are otherwise constant across links. Third, because only the relative levels of δ_{pub} and δ_{priv} matter, we normalize $\delta_{\text{pub}} = 1$ and recover the δ_{priv} that is consistent with a given budget K. Given the levels of our infrastructure measures, we set $K = 2{,}000$ and recover $\delta_{\text{priv}} = 0.1$. This suggests that building private infrastructure is 10 times cheaper than building public infrastructure.

Elasticity of congestion. We take the value of the elasticity of congestion from Mosqueda (2026). We set $\beta = 0.77$. In his paper, this elasticity corresponds to the elasticity of road speed with respect to an additional minibus on the road and is estimated using exogenous variation generated from the collapse of a major subway line. The interpretation of this elasticity maps directly to our interpretation of private transit infrastructure generating congestion.

Infrastructure elasticity. To calibrate the elasticity of private commuting costs with respect to infrastructure, γ, we use cross-sectional variation in EOD district-pair commuting flows and GTFS-based measures of private transit supply. For each directed district pair, we construct the flow of work trips that use private transit in at least one leg, and we measure private infrastructure using the density of private GTFS routes along plausible paths connecting the origin and destination districts. We then estimate a PPML gravity-style regression of private commuting flows on the log private infrastructure index, controlling for origin and destination fixed effects, distance, and public-transit infrastructure. The coefficient on private infrastructure, $\widehat{b}_I$, maps into the model parameter through $\widehat{\gamma} = \beta\widehat{b}_I$. Using the externally calibrated value $\beta = 0.77$, the benchmark estimate $\widehat{b}_I = 0.097$ implies $\widehat{\gamma} = 0.075$. Across alternative specifications that vary the dependent variable, the private-infrastructure measure, and the set of controls, the estimates are generally positive, with implied values of γ ranging from roughly 0.05 to 0.11. The benchmark PPML estimate is positive but imprecise; estimates based on positive-flow log-linear specifications and alternative measures

[4] Fajgelbaum and Schaal (2020) use road lanes as their infrastructure measure.

of private route availability are more statistically precise. We therefore interpret the exercise as disciplining the order of magnitude of γ, rather than as causal identification. Since the benchmark value is positive, small relative to β, and satisfies the convexity restriction $\gamma < \beta$, we use $\gamma = 0.075$ in the benchmark and report robustness over a grid of values below β.

Commuting cost shifters. We calibrate the link-specific commuting cost shifters $\delta_{jk}^{\tau,\text{pub}}$ and $\delta_{jk}^{\tau,\text{pri}}$ by inverting the commuting cost equations in the model. These parameters capture geographic and network frictions that make commuting more or less costly across links, holding fixed the technology-specific cost shifters and the congestion/infrastructure elasticities. In particular, using the EOD we observe commuting flows and travel times across origin-destination pairs, and using GTFS we measure the corresponding public and private transit infrastructure. Given calibrated values of θ_{pub}, θ_{pri}, β, and γ, we choose $\delta_{jk}^{\tau,\text{pub}}$ and $\delta_{jk}^{\tau,\text{pri}}$ so that equations (2.4) and (2.3) exactly reproduce observed commuting resource costs. Thus, public shifters rationalize observed public travel times given public flows, while private shifters rationalize observed private travel times given private flows and private infrastructure. We first perform this inversion at the EOD district-pair level, where travel times and flows are observed, and then map the resulting shifters to the model's network edges. The calibrated δ^{τ} objects therefore summarize the residual spatial heterogeneity in commuting costs that is not explained by mode technology, congestion, or measured infrastructure. The median values of these jk matrices are $\delta_{jk}^{\tau,\text{pri}} = 0.00026$, and $\delta_{jk}^{\tau,\text{pub}} = 0.13320$.

Public infrastructure capacity. To calibrate κ, which captures how many people in a given link can actually travel using one unit of public infrastructure, we target the *Encuesta Origen-Destino* aggregate share of commuters using public (17%) and private (83%) transit. We set $\kappa = 924$.

Land and productivity. To obtain a measure of H and H^c, we rely on land use data that show residential and commercial land available at the tract level. We then aggregate to the hexagon units that define our network. For productivity A, we compute average wages at the census tract level using data from the Economic Census, which reports wages at the firm level. We then aggregate into our hexagon geographic units and map average wages directly to productivity.

2.6 Welfare effects of transit expansions

In this section, we study the gains from budget-feasible expansions of the transit system. In particular, given the current network and spatial equilibrium, we assess where it is optimal to invest in transit infrastructure and how these investments differ across transit technologies. We depart from the baseline equilibrium in which the public and private networks are the observed ones. In our calibration, we assume that the current network was built with budget K, and the counterfactuals presented here increase K by 50%.

To highlight and assess the role of the private technology, which has received limited attention in the urban transit interventions literature, we perform two counterfactuals: i) one where the planner considers only public transit infrastructure investments, and ii) another where the planner considers both public and private infrastructure. Comparing these two scenarios allows us to assess the potential additional welfare gains of investing in private transit relative to public transit alone.

Table 2.4: Welfare gains of transit expansions

Scenario	Δ Welfare (%)	Public budget share (%)	Public infr. share (%)	Public flow share (%)
(1) Build public only	1.334	100.000	15.452	56.246
(2) Build public and private	1.594	97.448	14.976	39.027

Note: The counterfactuals are relative to the baseline, which is the equilibrium with the observed network with no new investment ($K = 0$). Both scenarios consider a 50% expansion of the government's budget K.

Welfare gains from network expansions. Table 2.4 shows the welfare gains from increasing the baseline budget by 50% and optimally expanding the current network. Focusing on the welfare gain column, the first row shows a 1.33% gain from expanding the network when the planner implements only public transit investments, while the second row shows a 1.59% gain when the planner can invest in both technologies. The comparison suggests that private transit can deliver an extra marginal welfare gain of 0.26 percentage points, or roughly a 20% additional welfare gain. This magnitude is quantitatively important: governments in developing cities could benefit

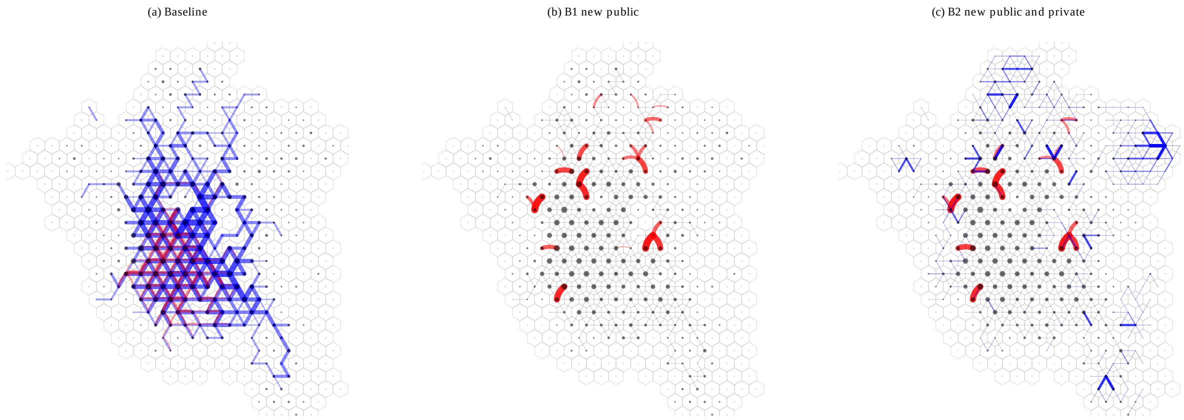


Figure 2.12: Optimal transit expansions

Note: Note: The figure shows two counterfactual scenarios of expanding transit infrastructure relative to the current network. We consider a 50% increase in budget K. The B1 scenario shows the optimal expansion when the planner uses public transit exclusively. The B2 scenario shows the optimal expansion when the planner can use both public and private transit. Red lines indicate new public infrastructure and blue lines indicate new private infrastructure. Grey lines represent existing infrastructure. Node size represents the distribution of population by residence, with larger nodes indicating higher population density.

substantially by considering the joint design of the network, rather than the common recipe of expanding and evaluating public mass transit in isolation.

To put the magnitude of these gains in perspective, in the developing-country setting, Graff (2024) found that reorganizing Africa's entire road network would result in a welfare gain of 1.3%, and that optimally expanding it by a tenth would result in a 0.8% gain. Our results align overall with these magnitudes, suggesting they are economically meaningful.

Spatial effects. Where should the planner expand the network? Figure 2.12 shows the spatial distribution of counterfactual investments. The left panel shows the baseline public and private networks. The middle panel shows the optimal expansion when the planner can only add public infrastructure. The right panel shows the optimal expansion when the planner can expand both public and private transit. In all panels, grey lines represent existing infrastructure and node size represents the distribution of population by residence.

The public-only expansion in the middle panel is not simply a radial expansion around the city center. Instead, the planner places new public infrastructure in a small number of high-return corridors where large commuting flows or productive destinations make the high fixed cost of public investment worthwhile. A first set of investments appears in the northwest, around the Tlalnepantla–Naucalpan–Atizapán corridor. These locations form a productive satellite-city area with limited integration into the existing mass-transit network. A second set of investments appears around the airport (East), which is another important employment and connectivity node. A third set appears toward Ecatepec in the northeast. Unlike the northwest corridor and the airport, this area is not among the most productive locations, but it attracts and generates very large commuting flows. The planner therefore invests there because improving access for a large mass of commuters generates large aggregate gains even when destination productivity is not especially high.

Comparing the public-only expansion with the mixed public–private expansion highlights the role of technology choice. When the planner can use both technologies, public investment becomes more selective: it is used on links where the combination of thick flows, productive destinations, and network centrality justifies the high fixed cost of public infrastructure. Private investment, by contrast, expands much more widely across the periphery. This reflects the lower cost and greater flexibility of the private technology. The private network is especially useful for extending coverage to peripheral areas, feeding commuters into the existing network, and connecting locations where demand is valuable but not large enough to justify public infrastructure. Thus, the optimal mixed network does not replace public transit with private transit everywhere. Rather, it assigns public infrastructure to the thickest and most central corridors, while using private transit to provide a lower-cost extensive margin of connectivity.

2.6.1 Discussion

Our model and quantitative results abstract from private automobile mode choice and therefore should be interpreted as a model of optimal transit network design, not of the full multimodal urban transport equilibrium. This abstraction is deliberate. In Mexico City, the empirically central and under-studied margin for transit planning is the interaction between formal public trunk infrastructure and extensive private/informal feeder services, which jointly determine a large share of transit connectivity and commuting costs. Incorporating private cars would add important substitution and road-congestion channels, but would require a substantially richer model of mode choice, car ownership, and shared road capacity. We therefore view our results as characterizing the optimal allocation of transit technologies conditional on the broader automobile environment. To the extent that improved transit also draws travelers from cars, our estimates may understate the broader welfare gains from coordinated transit planning. The welfare results presented here can therefore be interpreted as a lower bound.

2.6.2 Robustness to alternative parameters

Appendix 2.B reports sensitivity exercises that vary the infrastructure budget, the marginal operating cost of private transit, and the construction cost of private infrastructure. In each exercise, we compare the same two counterfactuals as above: B1 allows only public-transit expansion, while B2 allows the planner to expand both public and private transit. Overall, the welfare patterns are stable across these perturbations. The mixed public–private counterfactual always delivers larger gains than the public-only counterfactual, and the magnitudes remain close to those in the main exercise.

Table 2.5 shows that welfare gains increase monotonically with the size of the budget expansion. Increasing ΔK from 10 to 75 percent raises the public-only welfare gain from 0.35 to 1.86 percent and the mixed public–private gain from 0.62 to 2.11 percent. The incremental gain from allowing private investment is positive at all budget levels and remains approximately stable, at around one-quarter of a percentage point. Thus, larger budgets raise welfare in both planning problems, but the role of private transit remains quantitatively similar: it provides an additional margin of network expansion beyond what can be achieved by public infrastructure alone.

The cost sensitivity exercises reinforce the model's technology-choice mechanism. Table 2.6 increases θ_{pri}, making private transit more expensive to operate. As private operating costs rise, the welfare gain from the mixed counterfactual converges toward the public-only counterfactual: the incremental B2–B1 gain declines from 0.38 percentage points when $\theta_{\text{pri}} = 1.05$ to

0.04 percentage points when $\theta_{\text{pri}} = 10$. Hence, when private transit becomes sufficiently costly to operate, the additional value of allowing private investment becomes small. Table 2.7 varies the private construction-cost shifter. Although total welfare gains rise in this exercise because the calibrated budget also rises, the incremental gain from allowing private investment falls from 0.31 to 0.21 percentage points as private infrastructure becomes more expensive to build. This pattern is consistent with the same trade-off: private transit contributes most when it combines low fixed costs with acceptable marginal operating costs.

Taken together, the sensitivity analysis suggests that the main quantitative conclusion is not driven by a narrow parameter choice. The level of welfare gains changes with the budget and with private-transit cost parameters, but the qualitative ranking of counterfactuals is robust. Public investment remains the main source of gains on high-demand corridors, while allowing private investment adds a smaller but positive welfare gain by extending the network on links where public infrastructure is too costly relative to demand.

2.7 Conclusion

This paper studies how a budget-constrained planner should optimally allocate public and private transit technologies across an urban transportation network. The central trade-off is that mass transit technologies are fixed-cost intensive but have low marginal operating costs, while private transit requires little upfront investment but involves higher marginal costs and generates congestion externalities. Understanding how to optimally combine these technologies is particularly important in developing cities, where fiscal constraints limit the expansion of mass transit infrastructure and private transit plays a central role in providing commuting mobility.

We develop a quantitative spatial equilibrium framework in which a planner chooses commuting flows, spatial allocations, and the technology composition of transit infrastructure across the network. The model delivers a simple characterization of the optimal public-private technology mix: the optimal technology deployed on each link depends on the trade-off between relative marginal operating costs and infrastructure construction costs across technologies. Public transit is optimally deployed on high-demand corridors where its low operating costs can be amortized over large flows, while private transit provides baseline connectivity across the rest of the network where building mass transit infrastructure is too expensive relative to demand.

To bring the model to the data, we construct a new dataset that combines GTFS data describing the full universe of public and private transit routes with administrative microdata on commuting flows, wages, and land use for Mexico City. The GTFS data allow us to build a unified representation of the city's transit network, including informal and private routes that are typically missing from official datasets, and to map transit supply into a simplified network structure suitable for quantitative analysis. Using these data, we document two key facts that motivate the framework: private transit dominates network coverage and peripheral connectivity, and public and private transit are strongly complementary within individual commuting trips.

Quantitatively, we find that the welfare gains from transit expansions depend critically on whether private transit is incorporated into the planning problem. When public infrastructure is expanded without accounting for the private network, welfare gains are substantially smaller. Allowing the planner to condition public investment on the existing private network improves outcomes, while jointly planning public and private infrastructure yields the largest welfare gains. These results imply that a significant share of the benefits from transit investment arises not only from expanding public infrastructure itself, but from optimally integrating private transit into the network design.

Taken together, the results suggest that private transit in developing cities should not be viewed solely as an inefficient or transitional sector, but rather as a technology that can play an important role in the optimal transit system when governments face binding fiscal constraints. The optimal policy is therefore not to replace private transit everywhere with mass transit, but to jointly design both systems so that public infrastructure serves high-capacity corridors while private transit provides flexible connectivity across the rest of the city. More broadly, the paper highlights that infrastructure policy in developing cities should be understood as a problem of optimal technology allocation under fiscal constraints, rather than as a choice between formal and informal systems.

Finally, an important avenue for future research is to explore the distributional consequences of designing private and public networks. We show that optimal network expansions should improve connectivity between peripheral locations and central infrastructure, potentially benefiting low-income workers who reside in such peripheral locations. An analysis of the distributional impacts is therefore important for assessing the equity-efficiency trade-offs in the design of transit networks.

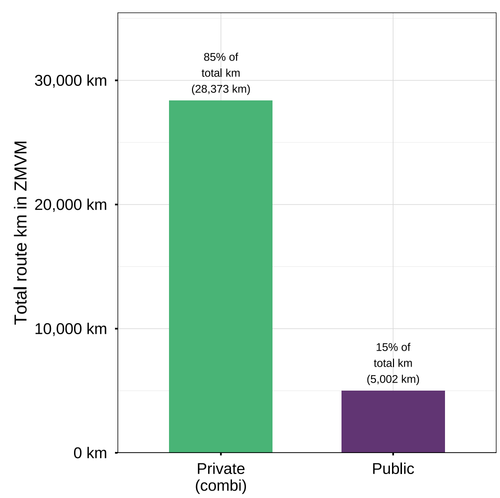
Appendix

2.A Transit network in Mexico

2.A.1 Private Transit and Network Coverage

This section presents additional aggregate evidence on the dominance of private transit operators (*combis*) in the ZMVM's transit network, complementing the spatial and distance-band evidence reported in Section 2.4.2.

Figure 2.13 summarizes aggregate network coverage across the entire ZMVM. Private transit accounts for 28,373 route-kilometers, or 85% of the total network. Formal public transit contributes 5,002 route-kilometers (15%). These figures underscore a key feature of the ZMVM's transit system: despite the visibility of large-scale public infrastructure such as the Metro, the backbone of the network in terms of spatial coverage is overwhelmingly composed of privately operated combi routes.



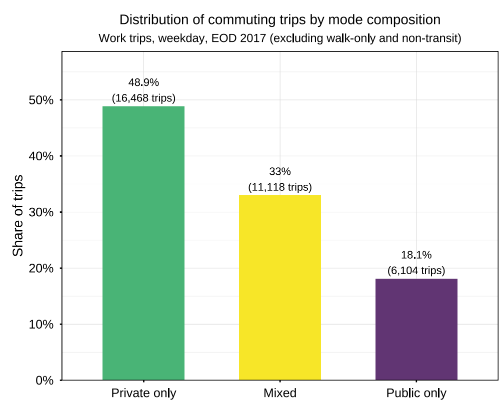
Notes: The figure shows total route-kilometers in the ZMVM by operator type. Private transit (combis) accounts for 28,373 km (85%) and formal public transit for 5,002 km (15%).

Figure 2.13: Aggregate Network Coverage: Private vs. Public Transit

2.A.2 Additional Evidence on Public–Private Transit Complementarity

This section presents additional evidence supporting the complementarity between public and private transit documented in Section 2.4.3. We decompose commuting trips by mode composition and characterize how trip attributes vary across trip types. The mode-to-mode transition heatmap and the metro–combi proximity map appear in the main text as Figures 2.9 and 2.10.

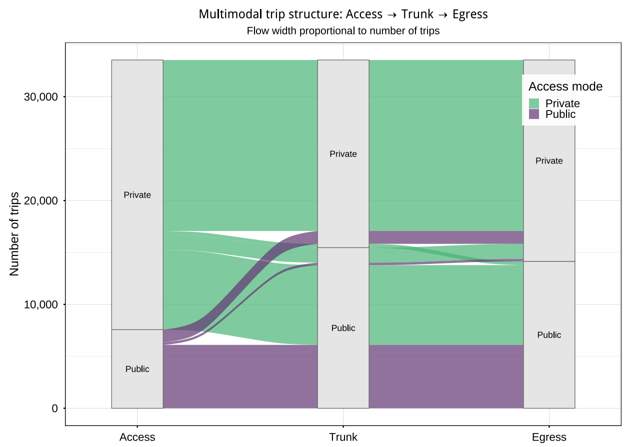
Trip composition. Figure 2.14 classifies all weekday work commutes in the 2017 EOD (excluding walk-only and non-transit trips) into three categories: private only, mixed (combining public and private modes), and public only. Private-only trips account for 48.9 percent of commutes (16,468 trips), mixed trips for 33 percent (11,118 trips), and public-only trips for 18.1 percent (6,104 trips). The large share of mixed trips confirms that public and private transit are routinely combined within individual journeys rather than serving as substitutes.



Notes: Weekday work commutes from the 2017 EOD, excluding walk-only and non-transit trips. Trips are classified as private only (all segments served by combis), public only (all segments served by metro, BRT, bus, or train), or mixed (combining both).

Figure 2.14: Distribution of Commuting Trips by Mode Composition

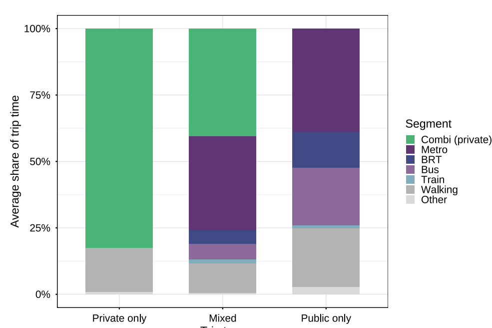
Figure 2.15 traces the structure of multimodal trips by decomposing each journey into three stages—access, trunk, and egress—and tracking whether each stage is served by private or public transit. The dominant pattern is clear: private transit serves as the primary access and egress mode, feeding commuters into and out of public trunk services. The most common multimodal configuration involves private access, public trunk, and private egress, consistent with the first-and-last-mile role of combis documented in the main text.



Notes: Alluvial diagram decomposing multimodal trips into access, trunk, and egress stages. Flow widths are proportional to the number of trips. Green denotes private (combi) and purple denotes public transit.

Figure 2.15: Multimodal Trip Structure: Access, Trunk, and Egress

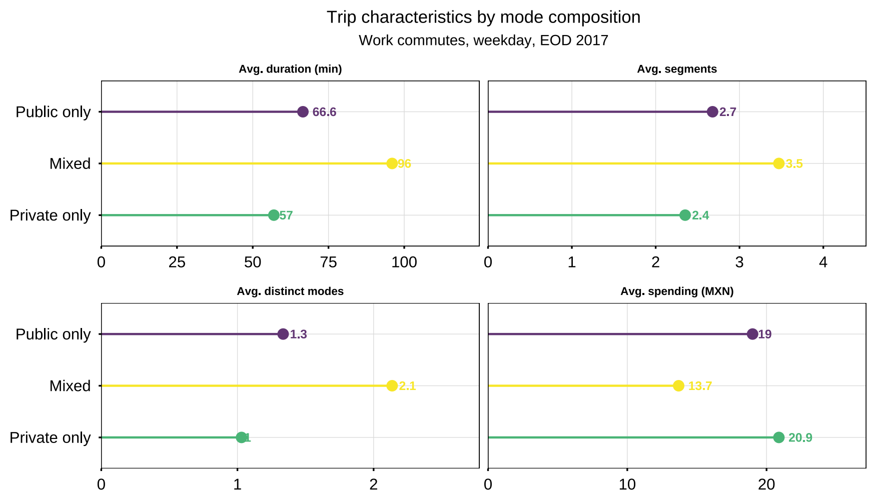
Time decomposition and trip characteristics. Figure 2.16 decomposes average trip time by mode segment across the three trip types. Private-only trips are dominated by combi time, while mixed trips allocate substantial shares to both combi and public modes (metro, BRT, bus, and train), confirming that private transit absorbs a large fraction of total travel time even on trips that use public infrastructure. Public-only trips are split primarily between metro and BRT segments.



Notes: Stacked bars show the average share of total trip time spent on each mode segment, separately for private-only, mixed, and public-only trips. Weekday work commutes, EOD 2017.

Figure 2.16: Average Share of Trip Time by Mode Segment

Figure 2.17 summarizes four trip attributes by mode composition. Mixed trips are the longest (96 minutes on average), involve the most segments (3.5) and distinct modes (2.1), and have intermediate monetary cost (13.7 MXN). Private-only trips are shorter (57 minutes) and less complex (2.4 segments, 1 mode) but the most expensive (20.9 MXN), reflecting higher per-kilometer costs of combi service. Public-only trips fall between the two in duration (66.6 minutes) and cost (19 MXN). These patterns are consistent with a system in which commuters combine modes to exploit the speed advantages of public trunk services while relying on private transit for spatial coverage—at the cost of longer and more complex journeys.



Notes: Average duration (minutes), number of segments, number of distinct modes, and monetary spending (MXN) for each trip type. Weekday work commutes, EOD 2017.

Figure 2.17: Trip Characteristics by Mode Composition

2.A.3 Productivity map

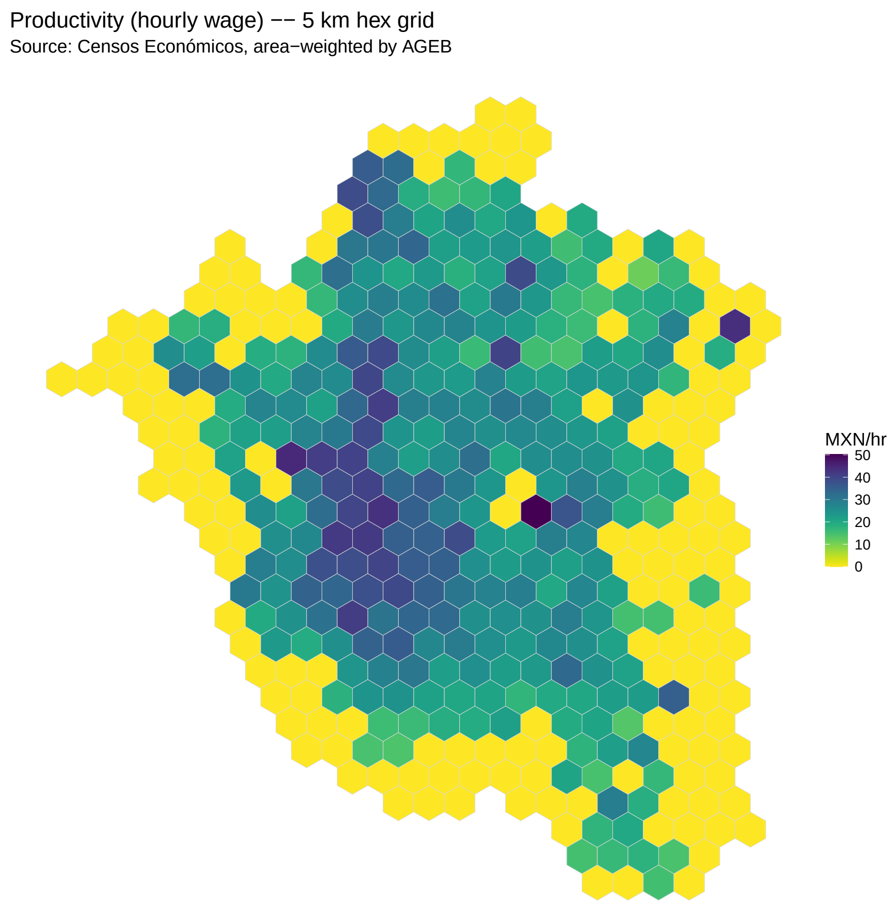


Figure 2.18: Spatial Distribution of Productivity in Mexico City

2.B Sensitivity analysis

Table 2.5: Sensitivity analysis: welfare across budget increases ($K_{\text{calib}} = 2000$)

ΔK (%)	K_B	W^{Base}	W^{B1}	W^{B2}	ΔW^{B1} (%)	ΔW^{B2} (%)	$\Delta W^{\text{B2-B1}}$ (%)
10	200	35.6793	35.8055	35.8988	0.3538	0.6154	0.2606
25	500	35.6793	35.9445	36.0380	0.7434	1.0056	0.2602
50	1000	35.6793	36.1553	36.2481	1.3342	1.5945	0.2568
75	1500	35.6793	36.3417	36.4319	1.8567	2.1094	0.2481

Table 2.6: Sensitivity analysis: welfare across θ_{pri}

θ_{pri}	W^{Base}	W^{B1}	W^{B2}	ΔW^{B1} (%)	ΔW^{B2} (%)	$\Delta W^{\text{B2-B1}}$ (%)
1.05	35.7588	36.2295	36.3655	1.3164	1.6967	0.3754
2.04	35.6508	36.1295	36.2060	1.3427	1.5572	0.2117
3.04	35.6133	36.0959	36.1489	1.3551	1.5039	0.1468
4.03	35.5950	36.0794	36.1197	1.3610	1.4740	0.1116
5.03	35.5843	36.0694	36.1020	1.3633	1.4549	0.0904
6.02	35.5773	36.0636	36.0902	1.3670	1.4417	0.0737
7.02	35.5724	36.0593	36.0818	1.3688	1.4321	0.0625
8.01	35.5688	36.0561	36.0755	1.3702	1.4248	0.0539
9.01	35.5660	36.0537	36.0707	1.3714	1.4191	0.0471
10.00	35.5637	36.0518	36.0668	1.3724	1.4145	0.0415

Table 2.7: Sensitivity analysis: welfare across δ_{pri} (with $K_B = 1.5 \times K_{\text{calib}}$)

δ_{pri}	K_{calib}	K_B	W^{Base}	W^{B1}	W^{B2}	ΔW^{B1} (%)	ΔW^{B2} (%)	$\Delta W^{\text{B2-B1}}$ (%)
0.0201	1147.2	1720.8	35.6793	35.9777	36.0908	0.8364	1.1533	0.3143
0.0515	1478.8	2218.3	35.6793	36.0490	36.1497	1.0364	1.3185	0.2792
0.0828	1810.5	2715.7	35.6793	36.1173	36.2124	1.2278	1.4943	0.2632
0.1142	2142.1	3213.2	35.6793	36.1835	36.2751	1.4132	1.6699	0.2531
0.1455	2473.8	3710.7	35.6793	36.2486	36.3370	1.5956	1.8435	0.2440
0.1768	2805.4	4208.1	35.6793	36.3084	36.3937	1.7634	2.0024	0.2349
0.2082	3137.1	4705.6	35.6793	36.3649	36.4484	1.9218	2.1557	0.2295
0.2395	3468.7	5203.1	35.6793	36.4204	36.5010	2.0773	2.3031	0.2211
0.2709	3800.4	5700.5	35.6793	36.4715	36.5499	2.2205	2.4402	0.2150
0.3022	4132.0	6198.0	35.6793	36.5210	36.5982	2.3591	2.5755	0.2114

2.C Model derivations

2.C.1 First-order conditions of the planner's problem

This appendix derives the first-order conditions for the mode-specific formulation. Define

$$A_{jk}^{\text{pub}} \equiv \delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}}, \qquad A_{jk}^{\text{pri}} \equiv \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}}.$$

Then

$$T_{jk}^{\text{pub}} = A_{jk}^{\text{pub}} L_{jk}^{\text{pub}}, \qquad T_{jk}^{\text{pri}} = A_{jk}^{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma}.$$

Let $\eta_{jk}^{\text{pub}} \geq 0$ be the multiplier on public capacity, written as $\kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}} - L_{jk}^{\text{pub}} \geq 0$. The Lagrangian is

$$
\begin{aligned}
\mathcal{L} = u &+ \sum_{j\in\mathcal{J}} \omega_j L_j^R \big(U(c_j, h_j) - u\big) \\
&+ \sum_{j\in\mathcal{J}} P_j^F \left(L_j^R + \sum_{i\in\mathcal{N}(j)} \sum_{m\in\mathcal{M}} L_{ij}^m - L_j^L - \sum_{k\in\mathcal{N}(j)} \sum_{m\in\mathcal{M}} L_{jk}^m - \sum_{k\in\mathcal{N}(j)} \sum_{m\in\mathcal{M}} T_{jk}^m \right) \\
&+ P^R \left(\bar{L} - \sum_{j\in\mathcal{J}} L_j^R \right) \\
&+ P^L \left(\bar{L} - \sum_{j\in\mathcal{J}} L_j^L - \sum_{(j,k)\in\mathcal{E}} \sum_{m\in\mathcal{M}} T_{jk}^m \right) \\
&+ P^C \left(\sum_{j\in\mathcal{J}} A_j L_j^L - \sum_{j\in\mathcal{J}} c_j L_j^R \right) + \sum_{j\in\mathcal{J}} P_j^H \left(H_j - h_j L_j^R \right) \\
&+ P^K \left(K - \sum_{(j,k)\in\mathcal{E}} \left[\delta_{jk}^{I,\text{pub}} I_{jk}^{\text{pub}} + \delta_{jk}^{I,\text{pri}} I_{jk}^{\text{pri}} \right] \right) \\
&+ \sum_{(j,k)\in\mathcal{E}} \eta_{jk}^{\text{pub}} \left(\kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}} - L_{jk}^{\text{pub}} \right) \\
&+ \sum_{j\in\mathcal{J}} \mu_j^c c_j + \sum_{j\in\mathcal{J}} \mu_j^h h_j + \sum_{j\in\mathcal{J}} \mu_j^R L_j^R + \sum_{j\in\mathcal{J}} \mu_j^L L_j^L \\
&+ \sum_{(j,k)\in\mathcal{E}} \nu_{jk}^{\text{pub}} L_{jk}^{\text{pub}} + \sum_{(j,k)\in\mathcal{E}} \nu_{jk}^{\text{pri}} L_{jk}^{\text{pri}} \\
&+ \sum_{(j,k)\in\mathcal{E}} \sum_{m\in\mathcal{M}} \underline{\mu}_{jk}^{I,m} \left(I_{jk}^m - \underline{I}_{jk}^m \right) + \sum_{(j,k)\in\mathcal{E}} \sum_{m\in\mathcal{M}} \overline{\mu}_{jk}^{I,m} \left(\overline{I}_{jk}^m - I_{jk}^m \right). \qquad (2.21)
\end{aligned}
$$

The first-order conditions with respect to consumption, housing, residents, and local labor are

$$[c_j] \qquad \omega_j L_j^R U_c(c_j, h_j) + \mu_j^c = P^C L_j^R, \qquad \forall j, \tag{2.22}$$

$$[h_j] \qquad \omega_j L_j^R U_h(c_j, h_j) + \mu_j^h = P_j^H L_j^R, \qquad \forall j, \tag{2.23}$$

$$[L_j^L] \qquad -P_j^F - P^L + P^C A_j + \mu_j^L = 0, \qquad \forall j, \tag{2.24}$$

$$[L_j^R] \qquad \omega_j\big(U(c_j,h_j)-u\big)+P_j^F-P^R-P^Cc_j-P_j^Hh_j+\mu_j^R=0, \qquad \forall j. \tag{2.25}$$

Mode-specific flow conditions. For private flow, differentiating with respect to L_{jk}^{pri} gives

$$[L_{jk}^{\text{pri}}] \qquad P_k^F-P_j^F-(P_j^F+P^L)\frac{\partial T_{jk}^{\text{pri}}}{\partial L_{jk}^{\text{pri}}}+\nu_{jk}^{\text{pri}}=0. \tag{2.26}$$

Together with complementary slackness, this implies

$$P_k^F-P_j^F \le (P_j^F+P^L)(1+\beta)A_{jk}^{\text{pri}}\left(L_{jk}^{\text{pri}}\right)^{\beta}\left(I_{jk}^{\text{pri}}\right)^{-\gamma}, \qquad = \text{ if } L_{jk}^{\text{pri}}>0. \tag{2.27}$$

When the private mode is active, solving for L_{jk}^{pri} yields (2.11).

For public flow, differentiating with respect to L_{jk}^{pub} gives

$$[L_{jk}^{\text{pub}}] \qquad P_k^F-P_j^F-(P_j^F+P^L)A_{jk}^{\text{pub}}-\eta_{jk}^{\text{pub}}+\nu_{jk}^{\text{pub}}=0. \tag{2.28}$$

Thus

$$P_k^F-P_j^F \le (P_j^F+P^L)A_{jk}^{\text{pub}}+\eta_{jk}^{\text{pub}}, \qquad = \text{ if } L_{jk}^{\text{pub}}>0. \tag{2.29}$$

Public infrastructure. Differentiating with respect to I_{jk}^{pub} gives

$$[I_{jk}^{\text{pub}}] \qquad \eta_{jk}^{\text{pub}}\kappa_{jk}^{\text{pub}}-P^K\delta_{jk}^{I,\text{pub}}+\underline{\mu}_{jk}^{I,\text{pub}}-\overline{\mu}_{jk}^{I,\text{pub}}=0. \tag{2.30}$$

An interior public-infrastructure choice therefore satisfies (2.12). Since $\eta_{jk}^{\text{pub}}>0$ only when public capacity binds, public infrastructure is built because it relaxes the feasible-flow constraint for public riders, not because it lowers public riders' operating cost.

Private infrastructure. Differentiating with respect to I_{jk}^{pri} gives

$$[I_{jk}^{\text{pri}}] \qquad -(P_j^F+P^L)\frac{\partial T_{jk}^{\text{pri}}}{\partial I_{jk}^{\text{pri}}}-P^K\delta_{jk}^{I,\text{pri}}+\underline{\mu}_{jk}^{I,\text{pri}}-\overline{\mu}_{jk}^{I,\text{pri}}=0. \tag{2.31}$$

Because

$$\frac{\partial T_{jk}^{\text{pri}}}{\partial I_{jk}^{\text{pri}}} = -\gamma A_{jk}^{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma-1}, \tag{2.32}$$

an interior solution satisfies (2.15), which yields (2.16).

2.C.2 Proof of Proposition 2.1

Fix an edge (j,k) and suppress edge subscripts whenever this creates no ambiguity. Define

$$V_j \equiv P_j^F + P^L, \qquad A^{\text{pub}} \equiv \delta_{jk}^{\tau,\text{pub}} \theta_{\text{pub}}, \qquad A^{\text{pri}} \equiv \delta_{jk}^{\tau,\text{pri}} \theta_{\text{pri}},$$

and let

$$B^{\text{pub}} \equiv P^K \delta_{jk}^{I,\text{pub}}, \qquad B^{\text{pri}} \equiv P^K \delta_{jk}^{I,\text{pri}}$$

denote the shadow cost of building one unit of public and private infrastructure on link (j,k). Let

$$\Delta P_{jk} \equiv P_k^F - P_j^F$$

be the destination potential on the edge. We prove the result for an interior solution with positive public and private flows and slack infrastructure bounds. Corner solutions are characterized by the corresponding complementary slackness inequalities.

The part of the Lagrangian that depends on the mode-specific flows and infrastructure on edge (j,k) is

$$\begin{aligned}
\mathcal{L}_{jk} = {} & \Delta P_{jk} \left(L_{jk}^{\text{pub}} + L_{jk}^{\text{pri}}\right) \\
& - V_j \left[A^{\text{pub}} L_{jk}^{\text{pub}} + A^{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma}\right] \\
& - B^{\text{pub}} I_{jk}^{\text{pub}} - B^{\text{pri}} I_{jk}^{\text{pri}} \\
& - \eta_{jk}^{\text{pub}} \left(L_{jk}^{\text{pub}} - \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}\right),
\end{aligned}$$

where $\eta_{jk}^{\text{pub}} \geq 0$ is the multiplier on the public-capacity constraint

$$L_{jk}^{\text{pub}} \leq \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}.$$

First consider public transit. The first-order condition with respect to public flow is

$$\frac{\partial \mathcal{L}_{jk}}{\partial L_{jk}^{\text{pub}}} = \Delta P_{jk} - V_j A^{\text{pub}} - \eta_{jk}^{\text{pub}} = 0,$$

whenever $L_{jk}^{\text{pub}} > 0$. Therefore,

$$\Delta P_{jk} = V_j A^{\text{pub}} + \eta_{jk}^{\text{pub}}. \tag{2.33}$$

The first-order condition with respect to public infrastructure is

$$\frac{\partial \mathcal{L}_{jk}}{\partial I_{jk}^{\text{pub}}} = \eta_{jk}^{\text{pub}} \kappa_{jk}^{\text{pub}} - B^{\text{pub}} = 0.$$

Thus,

$$\eta_{jk}^{\text{pub}} = \frac{B^{\text{pub}}}{\kappa_{jk}^{\text{pub}}} = \frac{P^K \delta_{jk}^{I,\text{pub}}}{\kappa_{jk}^{\text{pub}}}. \tag{2.34}$$

Since $B^{\text{pub}} > 0$ and $\kappa_{jk}^{\text{pub}} > 0$, equation (2.34) implies $\eta_{jk}^{\text{pub}} > 0$. By complementary slackness, public capacity must bind on an active public edge:

$$L_{jk}^{\text{pub}} = \kappa_{jk}^{\text{pub}} I_{jk}^{\text{pub}}. \tag{2.35}$$

Substituting (2.34) into (2.33), the generalized unit cost of public service is

$$\mathfrak{c}_{jk}^{\text{pub}} = V_j A^{\text{pub}} + \frac{B^{\text{pub}}}{\kappa_{jk}^{\text{pub}}}. \tag{2.36}$$

Returning to primitives,

$$\mathfrak{c}_{jk}^{\text{pub}} = (P_j^F + P^L)\delta_{jk}^{\tau,\text{pub}}\theta_{\text{pub}} + \frac{P^K \delta_{jk}^{I,\text{pub}}}{\kappa_{jk}^{\text{pub}}}.$$

The first term is the operating cost of serving one more public rider, valued in the planner's numeraire. The second term is the shadow cost of the public infrastructure required to create one more unit of public passenger capacity.

Now consider private transit. The first-order condition with respect to private infrastructure is

$$\frac{\partial \mathcal{L}_{jk}}{\partial I_{jk}^{\text{pri}}} = V_j \gamma A^{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma-1} - B^{\text{pri}} = 0.$$

Solving for private infrastructure gives

$$I_{jk}^{\text{pri},*} = \left[\frac{\gamma V_j A^{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta}}{B^{\text{pri}}}\right]^{1/(1+\gamma)}. \tag{2.37}$$

In primitives, this is

$$I_{jk}^{\text{pri},*} = \left[\frac{\gamma (P_j^F + P^L)\delta_{jk}^{\tau,\text{pri}}\theta_{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{1+\beta}}{P^K \delta_{jk}^{I,\text{pri}}}\right]^{1/(1+\gamma)}.$$

The first-order condition with respect to private flow is

$$\frac{\partial \mathcal{L}_{jk}}{\partial L_{jk}^{\text{pri}}} = \Delta P_{jk} - V_j (1+\beta) A^{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma} = 0,$$

whenever $L_{jk}^{\text{pri}} > 0$. Hence,

$$\Delta P_{jk} = V_j (1+\beta) A^{\text{pri}} \left(L_{jk}^{\text{pri}}\right)^{\beta} \left(I_{jk}^{\text{pri}}\right)^{-\gamma}. \tag{2.38}$$

Substituting (2.37) into (2.38) yields the optimized generalized marginal cost of private service:

$$\mathfrak{c}_{jk}^{\text{pri}}\left(L_{jk}^{\text{pri}}\right) = (1+\beta)\gamma^{-\gamma/(1+\gamma)}\left(V_j A^{\text{pri}}\right)^{1/(1+\gamma)}\left(B^{\text{pri}}\right)^{\gamma/(1+\gamma)}\left(L_{jk}^{\text{pri}}\right)^{(\beta-\gamma)/(1+\gamma)}. \tag{2.39}$$

Returning to primitives,

$$\mathfrak{c}_{jk}^{\text{pri}}\left(L_{jk}^{\text{pri}}\right) = (1+\beta)\gamma^{-\gamma/(1+\gamma)}\left[(P_j^F + P^L)\delta_{jk}^{\tau,\text{pri}}\theta_{\text{pri}}\right]^{1/(1+\gamma)}\left[P^K\delta_{jk}^{I,\text{pri}}\right]^{\gamma/(1+\gamma)}\left(L_{jk}^{\text{pri}}\right)^{(\beta-\gamma)/(1+\gamma)}.$$

Finally, if both public and private modes are used on the same edge, then both first-order conditions must hold with equality and both modes deliver the same marginal benefit ΔP_{jk}. Combining (2.33) and (2.38), after substituting the infrastructure first-order conditions, gives

$$\mathfrak{c}_{jk}^{\text{pub}} = \mathfrak{c}_{jk}^{\text{pri}}\left(L_{jk}^{\text{pri}}\right). \tag{2.40}$$

This proves the proposition.

The expression also shows how the technology comparison depends on the curvature of the private technology. If $\beta > \gamma$, then

$$\frac{\beta-\gamma}{1+\gamma} > 0,$$

so the optimized private marginal cost is increasing in private flow. Public transit then has a threshold interpretation: private transit is attractive for low-flow links, but high-flow corridors eventually justify public capacity. If $\beta = \gamma$, then

$$\mathfrak{c}_{jk}^{\text{pri}} = (1+\beta)\beta^{-\beta/(1+\beta)}\left[(P_j^F + P^L)\delta_{jk}^{\tau,\text{pri}}\theta_{\text{pri}}\right]^{1/(1+\beta)}\left[P^K\delta_{jk}^{I,\text{pri}}\right]^{\beta/(1+\beta)},$$

which is independent of private flow. In that boundary case, both public and private technologies have constant effective unit-cost indices, and equality of these indices can generate multiple optimal public-private allocations with the same welfare value.

2.C.3 Proof of Proposition 2.2

We prove Proposition 2.2. Consider first the private transport block. Suppressing edge subscripts, write

$$T(L, I) = AL^{1+\beta} I^{-\gamma}, \qquad A > 0.$$

The Hessian of T with respect to (L, I) is

$$\nabla^2 T(L, I) = A \begin{pmatrix} \beta(1+\beta)L^{\beta-1}I^{-\gamma} & -\gamma(1+\beta)L^{\beta}I^{-\gamma-1} \\ -\gamma(1+\beta)L^{\beta}I^{-\gamma-1} & \gamma(1+\gamma)L^{1+\beta}I^{-\gamma-2} \end{pmatrix}.$$

The leading principal minor is non-negative when $\beta \geq 0$. The determinant is

$$\det \nabla^2 T(L, I) = A^2(1+\beta)\gamma(\beta-\gamma)L^{2\beta}I^{-2\gamma-2}.$$

Therefore $\nabla^2 T(L, I)$ is positive semidefinite whenever $\beta \geq \gamma \geq 0$. Hence the private transport block is convex.

If $\beta = \gamma > 0$, then

$$T(L, I) = AL^{1+\beta}I^{-\beta}$$

is homogeneous of degree one:

$$T(tL, tI) = tT(L, I).$$

Thus the function has zero curvature along rays $(L, I) \mapsto (tL, tI)$ and cannot be strictly or strongly convex. This proves that the private block is weakly convex at $\beta = \gamma$.

The public block is linear in public flow,

$$T^{\text{pub}}(L^{\text{pub}}) = A^{\text{pub}} L^{\text{pub}},$$

and the public-capacity constraint

$$L^{\text{pub}} \leq \kappa^{\text{pub}} I^{\text{pub}}$$

is linear. The network-building constraint, infrastructure bounds, non-negativity constraints, and population constraint are also linear.

It remains to note that the balanced-flow constraint is convex once the transport block is convex. For each node j, the constraint can be written as

$$L_j^L + \sum_{k \in \mathcal{N}(j)} \sum_{m \in \mathcal{M}} L_{jk}^m + \sum_{k \in \mathcal{N}(j)} \sum_{m \in \mathcal{M}} T_{jk}^m - L_j^R - \sum_{i \in \mathcal{N}(j)} \sum_{m \in \mathcal{M}} L_{ij}^m \leq 0.$$

This is a sum of linear terms and convex transport resource costs. Therefore it is convex. Since all remaining constraints are linear or convex, and the objective is linear in u, the full planner's problem is convex under the standard aggregate-variable representation. With mobile labor and homothetic preferences, the same argument delivers the usual quasiconvex formulation used by Fajgelbaum and Schaal (2020).

Bibliography

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia, "Misallocation, selection, and productivity: A quantitative analysis with panel data from China," *Econometrica*, 2022, *90* (3), 1261–1282.

Alder, Simon, "Chinese roads in India: The effect of transport infrastructure on economic development," *Journal of International Economics*, 2025, p. 104140.

Allen, Treb and Costas Arkolakis, "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 2022, *89* (6), 2911–2957.

Almagro, Milena, Andrea Barbieri, Juan Camilo Castillo, Arianna Hickok, and Timo Salz, "Optimal Urban Transportation Policy: Evidence from Chicago," 2024. Working paper.

Asturias, Jose, Manuel García Santana, and Roberto Ramos, *Misallocation, internal trade, and the role of transportation infrastructure*, CEMFI (Centro de Estudios Monetarios y Financieros) Madrid, Spain, 2014.

Baum-Snow, Nathaniel, "Did highways cause suburbanization?," *The quarterly journal of economics*, 2007, *122* (2), 775–805.

Bernot, Marc, Vicent Caselles, and Jean-Michel Morel, *Optimal transportation networks: models and theory*, Springer, 2009.

Björkegren, Daniel, Alice Duhaut, Geetika Nagpal, and Nick Tsivanidis, "Public and Private Transit: Evidence from Lagos," Technical Report, National Bureau of Economic Research 2025.

Chandra, Amitabh and Eric Thompson, "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional science and urban economics*, 2000, *30* (4), 457–490.

Conwell, Lucas, "Privatized Provision of Public Transit," 2024.

Desmet, Klaus and Esteban Rossi-Hansberg, "Urban accounting and welfare," *American Economic Review*, 2013, *103* (6), 2296–2327.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

Duranton, Gilles, Peter M Morrow, and Matthew A Turner, "Roads and Trade: Evidence from the US," *Review of Economic Studies*, 2014, *81* (2), 681–724.

Faber, Benjamin, "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Fajgelbaum, Pablo D. and Edouard Schaal, "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 2020, *88* (4), 1411–1452.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar, "State taxes and spatial misallocation," *The Review of Economic Studies*, 2019, *86* (1), 333–376.

Felbermayr, Gabriel J and Alexander Tarasov, "Trade and the spatial distribution of transport infrastructure," *Journal of Urban Economics*, 2022, *130*, 103473.

Fernald, John G, "Roads to prosperity? Assessing the link between public capital and productivity," *American economic review*, 1999, *89* (3), 619–638.

Graff, Tilman, "Spatial inefficiencies in Africa's trade network," *Journal of Development Economics*, 2024, *171*, 103319.

Hsieh, Chang-Tai and Enrico Moretti, "Housing constraints and spatial misallocation," *American economic journal: macroeconomics*, 2019, *11* (2), 1–39.

_ **and Peter J Klenow**, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448.

Khanna, Gaurav, Catalina Medina, Anant Nyshadham, Juan Ramos, Juan Tamayo, and Zi Hao Tiew, "Spatial Mobility, Economic Opportunity, and Crime," 2024. Working paper; add link when available.

Kreindler, Gabriel, Arya Gaduh, Tilman Graff, Rema Hanna, and Benjamin A Olken, "Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta," Technical Report, National Bureau of Economic Research 2023.

Mosqueda, Jordan, "Equilibrium Commuting Costs: The Role of Private and Public Transit," 2026. Job Market Paper; working paper.

Rogerson, Richard and Diego Restuccia, "Policy distortions and aggregate productivity with heterogeneous plants," in "2004 Meeting Papers" number 69 Society for Economic Dynamics 2004.

Tarasov, Alexander and Gabriel Felbermayr, "Trade and the Spatial Distribution of Transport Infrastructure," 2014.

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

Zárate, Román David, "Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City," 2024. Working paper.

Chapter Acknowledgement. Chapter 2, in full, is coauthored with Leticia Juárez and Román Zárate. The dissertation author was the primary author of this chapter.

CHAPTER 3

Domestic Trade Frictions: Crime

Coauthored with Alfonso Cebreros and Daniel Ramos-Menchelli.

3.1 Introduction

Crime makes locations economically farther apart. A producing location can be geographically close to a market but distant in economic terms if the corridor between them exposes carriers to extortion, theft, or violence. These frictions are particularly salient in developing countries, where weak institutions enable "middlemen" to extract rents (Atkin and Khandelwal, 2020). In such an environment, transport firms may reroute, travel longer distances, spend more on prevention, reduce service, or charge higher delivered prices. Thus, market integration, specialization, and real income may suffer as a result of domestic trade frictions (Donaldson, 2018; Sotelo, 2020; Gollin and Rogerson, 2014; Atkin and Khandelwal, 2020; Atkin and Donaldson, 2022).

This paper asks how crime along domestic shipping routes affects transport costs, goods prices, and welfare. We study these questions in the context of Mexico, where crime is a first-order concern. Roughly one out of four firms reports being a victim of crime, and direct economic losses associated with crime are estimated to be 0.51 percent of GDP.[1] In particular, we ask: What is the effect of crime on transport firms' costs and agricultural goods prices? What are the aggregate welfare effects of crime as a route-level trade friction? We present evidence that crime raises transport firms' costs and, although noisier, evidence that crime affects prices. Plugging the empirical estimates into a spatial trade model with intermediaries, we find that removing this friction could increase population-weighted municipality welfare by 0.63%, with substantial heterogeneity across space. The gains are driven by reductions in trade costs and prices, thereby increasing trade integration

[1] Encuesta Nacional de Victimización a Empresas (INEGI), 2023.

across regions.

We contribute by connecting three strands of research that speak to different parts of the same economic mechanism. Domestic trade studies show that transport frictions shape prices, trade costs and flows, and welfare within countries (Donaldson, 2018; Sotelo, 2020). Transportation-market studies show that trade costs are shaped by intermediaries, route choice, and network structure (Brancaccio et al., 2020; Allen et al., 2024). Work on crime, corruption, and extortion shows that predation can act as a cost of moving goods and be passed through to firms and consumers (Olken and Barron, 2009; Sequeira and Djankov, 2014; Brown et al., 2025). What remains less understood is how criminal risk on domestic roads affects transport firms, delivered prices, and aggregate welfare once shipments can reallocate across routes and markets.

We address this gap studying these questions theoretically and empirically. Theoretically, we develop a spatial trade framework with transportation intermediaries that ship goods across crime-exposed routes along the road network. Each route has a non-crime component, such as distance or travel time, and a crime component, which we empirically measure using the locations traversed by the route. Crime raises route-specific shipping costs, and intermediaries aggregate route services into an origin-destination transport cost. This structure makes two forces explicit: crime increases bilateral trade costs when it makes the routes connecting an origin and destination more expensive, and the magnitude depends on the ability of intermediaries to substitute across routes. If routes are poor substitutes, local crime shocks translate more directly into delivered prices.

The model serves three purposes. First, it formalizes route-level crime as a domestic trade friction in an economy where intermediaries choose among feasible routes. Second, it serves as a guide for the empirical analysis of transportation costs and agricultural prices. Third, it provides the quantitative structure for counterfactuals. Once crime is mapped into route-level transport costs using the empirical estimates, we can study how much trade costs fall, how trade flows reallocate, and how welfare changes when route crime is reduced or removed.

We combine several sources of Mexican administrative and spatial data for the empirical analysis. The core firm-level evidence comes from the confidential *Encuesta Anual de Transportes*

(EAT) from *INEGI*, which consists of three panel series, 2015–2018, 2017–2021, and 2021–2022, that report transport firms' outcomes such as revenue, cost, employment, and shipping outcomes such as trips, distance traveled, shipment values, vehicle stocks, and their main origin-destination markets. To the best of our knowledge this is the first paper that exploits this dataset. We combine these data with administrative municipality-level crime records from the Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública. We also link transport firms to the *Encuesta Nacional de Victimización de Empresas* (ENVE), which reports direct victimization, economic losses, and prevention spending. Finally, to study the effect of crime on prices we exploit wholesale agricultural prices from *Sistema Nacional de Integración de Mercados* (SNIIM), covering January 2000 through February 2026, that include origin-destination prices of fruits and vegetables sold in the main wholesale markets in the country, as well as agricultural output data from *Servicio de Información Agroalimentaria y Pesquera* (SIAP) to identify producing municipalities.

We begin by constructing a shift-share crime exposure measure for the empirical analysis on transportation firms and agricultural prices. We focus on extortion, homicides, and transport-related robbery crimes because these crimes are among the most reported firm-related crimes (ENVE), and likely capture the presence of organized crime. We construct exposure measures individually for each crime, and for a crime bundle. First, using the road network, we simulate potential route alternatives across all municipalities. The "shares" are thus given by the probability of choosing a route, using the inverse of driving time, which is purely a function of the geography of the road network – arguably orthogonal to crime. We then aggregate crime counts at the municipality level to the route level using the municipalities that each route crosses. We compute the "shifts" as percent changes in crime in a given year with respect to a baseline period, 2011–2013, when crime was at lower levels and more stable relative to the latter part of the decade. This exposure design follows the logic of the shift-share literature: credibility comes primarily from predetermined, geography-based route shares interacted with later crime shifts, rather than from contemporaneous firm or price choices (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2025).

For the firm-level analysis, we aggregate the origin-destination-year shift-share measure to the firm level using the firm's reported origin-destination pairs, ranked by the share of revenue. We then run firm-level regressions with firm and year fixed effects, which control for unobserved time-invariant firm characteristics and common temporal shocks across firms. The results point to a cost-side channel. A 100% increase in bundled crime exposure is associated with 2.9% higher total costs and 2.4% higher fuel costs, but not with a meaningful increase in insurance costs. The magnitudes differ by type of crime exposure, with extortion providing the most precise and coherent evidence: higher extortion exposure raises total costs and fuel costs, reduces shipment values, and increases total distance traveled and the number of trips completed. These margins suggest a story where rerouting and delays could be leading to higher operating costs. We therefore interpret the firm-level evidence as support for the transportation-cost mechanism, with the clearest patterns coming from extortion and broader crime exposure rather than from every individual crime measure.

In addition to the reduced-form analysis, we provide complementary descriptive evidence of how crime affects firms. We do so by analyzing a subset of transportation firms that appear in both the transport EAT and victimization EAT surveys. While the EAT reveals how exposure is associated with operating outcomes, ENVE records whether the firm suffered any type of crime and the direct economic burden of victimization and prevention. Using the matched sample, we show that, among victimized matched firms in a given year, extortion, cargo robbery, and vehicle theft are common; incident counts can be large; and direct economic losses and prevention spending are substantial relative to operating costs and revenue. This evidence shows that the burden of crime for affected transport firms can be economically large, potentially an order of magnitude larger than regular taxes.

The second part of the empirical analysis studies wholesale prices of fruits and vegetables. We estimate two main specifications that exploit spatial and temporal variation. The first specification includes crop-year, origin-crop, and destination-crop fixed effects, exploiting crime variation across origin-destination pairs. Intuitively, it compares the price of an avocado produced in Michoacán and sold in Mexico City relative to the price of the same avocado in Guadalajara. We find that doubling the crime bundle exposure leads to 1.2% higher prices. The second specification is directly implied by the model, and includes a much more restrictive set of fixed effects, effectively exploiting within origin-destination pair time variation — using origin-crop-year, destination-crop-year, and origin-destination fixed effects. These fixed effects remove changes in local supply and demand crop-level shocks, as well as time invariant characteristics of the corridors traversed. We find that increasing crime exposure leads to a similar-magnitude positive change in prices (0.6%), although the estimates in these highly saturated specifications are imprecise. We therefore interpret this evidence as consistent with some price pass-through from route exposure, although additional evidence from more goods, e.g. manufactured goods and non-tradables would be useful to establish a general channel.

We then move to quantifying the welfare effects of this friction with our structural framework. We discipline the route-crime-to-cost semi-elasticity with the estimates from the firm-level analysis, set the geography of the model using SNIIM and SIAP locations and the actual national road network, and calibrate fundamental parameters using labor, wages, output and value added data from the 2024 Economic Census (INEGI). We find that counterfactually removing crime from all routes reduces mean route-level trade costs by 2.9% and destination tradable price indices by 2.2%. In a benchmark where modeled food goods account for 38% of household expenditure, the implied population-weighted municipality welfare gain is 0.63%. In terms of the spatial effects, these welfare gains are far from uniform: municipalities in the 90th percentile of gains experience increases of 1.69%, while those in the 10th percentile gain 0.16%. The largest visible gains appear in parts of Veracruz, Yucatán, Hidalgo, Durango, Sonora, Sinaloa, and Puebla, consistent with the idea that route-crime reductions matter most where exposed corridors account for an important part of delivered costs.

Taken together, the paper contributes by showing empirically and structurally how crime can operate as a network trade friction, and by quantifying the aggregate effects of this friction. It extends the domestic trade-cost literature by studying an institutional and security-related barrier to market integration, rather than a purely physical or infrastructural one (Donaldson, 2018; Sotelo, 2020). It also contributes to the transport-intermediation literature by observing the firms that move goods and measuring the operating margins through which route risk appears, building on work that shows how intermediaries, route choice, and transport-market structure shape effective trade costs (Brancaccio et al., 2020; Allen et al., 2024). Finally, it contributes to the literature on crime and economic activity by studying the reverse of the more common question: instead of asking how economic shocks affect crime (Dell, 2015; Dell et al., 2019; Dix-Carneiro et al., 2018; Khanna et al., 2025), it asks how crime feeds back into the real economy by raising the cost of moving goods across space. This mechanism is motivated by evidence that predation and corruption distort transport and distribution: Olken and Barron (2009) document structured extortion on Indonesian trucking routes, Sequeira and Djankov (2014) show that port corruption induces firms to reroute shipments, and Brown et al. (2025) show that gang extortion in El Salvador is passed through to delivery fees, retail prices, and consumer welfare. This paper brings these insights into a domestic spatial-trade framework in which local insecurity can affect prices and welfare through the entire road network.

The policy implication is that road security should be treated as a form of trade facilitation. The results suggest that crime along freight corridors raises transport firms' operating costs and, through this channel, may affect prices and welfare. Policy should therefore prioritize economically central shipping routes, particularly corridors with high exposure to extortion and limited route substitutes, rather than focusing exclusively on crime at origins or destinations. In this sense, reducing predation on transport networks can operate like an improvement in domestic infrastructure by lowering the effective cost of connecting markets.

The rest of the paper proceeds as follows. Section 3.2 describes crime in Mexico. Section 3.3 develops the spatial trade model with transport intermediaries and route-level crime. Section 3.4 describes the firm, price, production, crime, and road-network data. Section 3.5 presents the exposure construction and empirical specifications. Section 3.6 reports the firm-level, wholesale-price, and matched EAT–ENVE evidence. Section 3.7 explains how the model is calibrated to the Mexican geography. Section 3.8 quantifies the welfare cost of route crime through the shipping channel. Section 3.9 concludes.

3.2 Context: Crime in Mexico

Mexico is a natural setting for studying crime as an economic friction because insecurity is large, persistent, and directly relevant for firms. Crime is not only a household welfare concern or a local public-safety outcome; it is a recurrent feature of the environment in which establishments hire workers, buy inputs, move goods, and choose where to operate. Around 27% of firms reported in *Encuesta Nacional de Victimización de Empresas* to be victims of crime in 2023, equivalent to 1.3 million victimized establishments and 2.9 million crimes against firms. Insecurity is therefore not a rare disruption faced by a small set of firms; it is a common business condition.

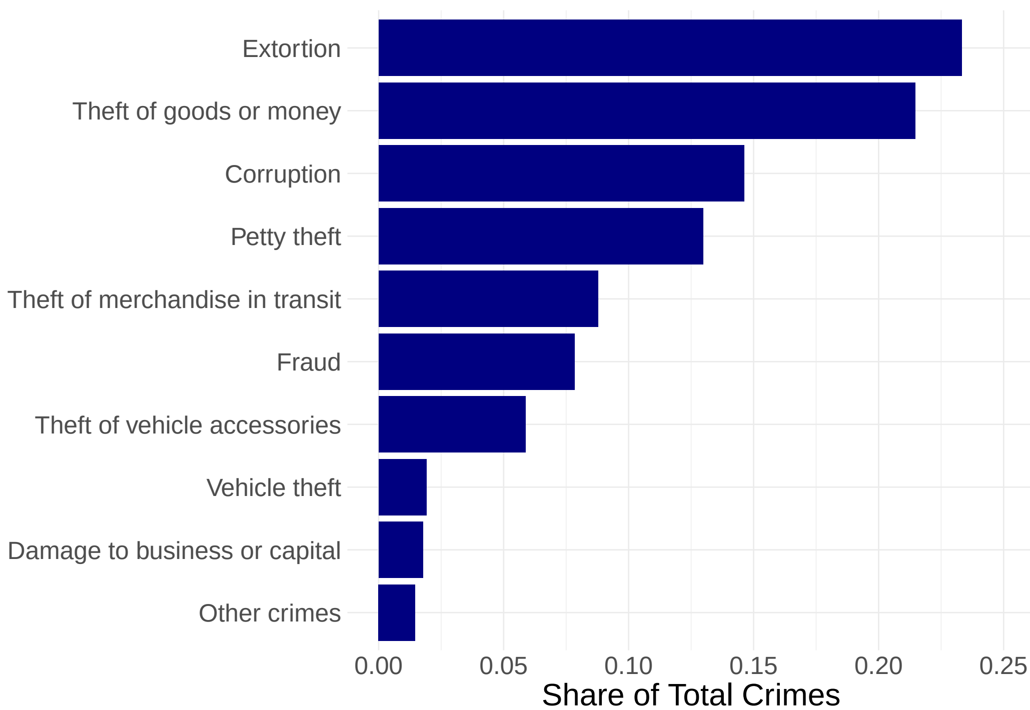


Figure 3.1: Reported crimes against firms

Notes: The figure reports shares of estimated crimes against firms by crime type. Source: INEGI, *Encuesta Nacional de Victimización de Empresas* (ENVE), 2019 public aggregate tabulations.

The presence of drug cartels and large-scale organized crime makes the Mexican business environment particularly difficult. Figure 3.1 shows the distribution of reported crimes. Of all reported crimes, extortion and theft of goods or money account for largest shares of reported crimes, while corruption, vehicle theft, and theft of merchandise in transit are also visible categories.

These crimes are particularly important because they directly entail costs in the transportation and distribution sectors.

The economic relevance of crime is also visible in its costs. In this same survey, ENVE, it is estimated that the direct cost of insecurity and crime for economic units in 2023 was equivalent to 0.51 percent of GDP. This measure includes direct losses and preventive expenditures.

Further, the incidence of crime affects firms across the size distribution. Figure 3.2 shows the top three reported crimes by firm size. For micro and small firms, the top crimes are mainly extortion, theft of goods or money, and corruption. For large firms, theft of merchandise in transit appears among the top three crimes, together with vehicle theft and extortion. This pattern is consistent with a simple exposure logic: larger firms operate broader supplier and customer networks, move larger shipment values, and rely more intensively on logistics. Their exposure to crime is therefore not confined to the establishment, but rather it travels with their goods.

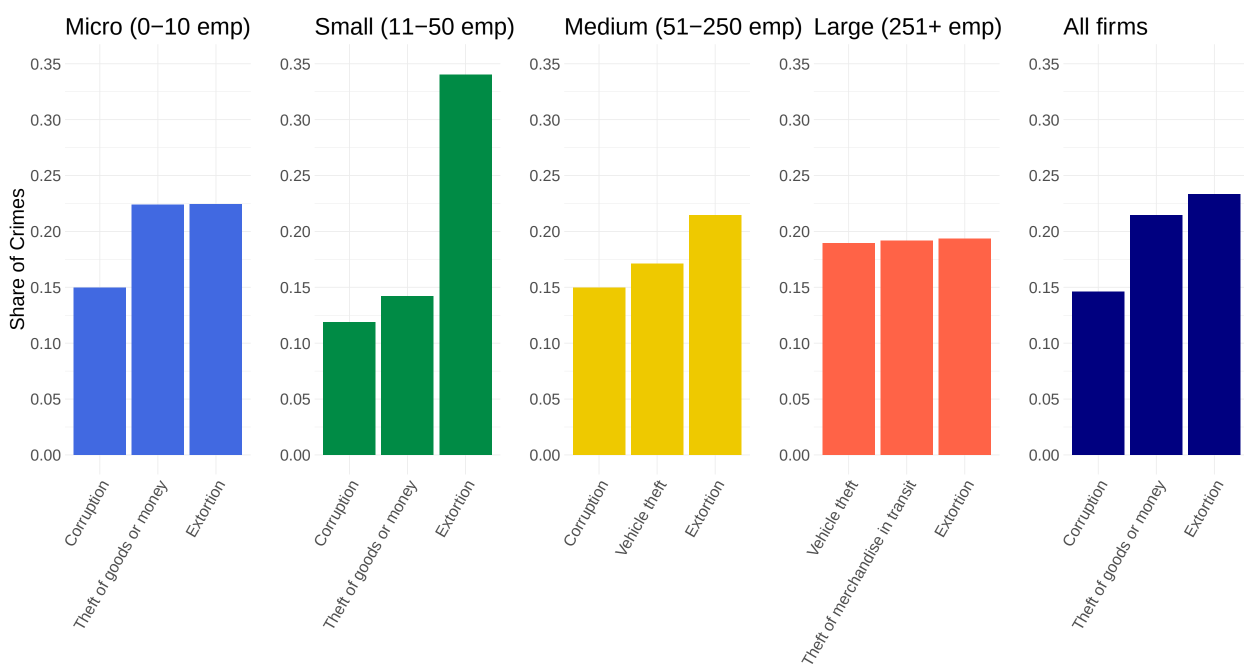


Figure 3.2: Top reported crimes by firm size, public ENVE 2019

Notes: The figure reports the top three crime categories within each firm-size group. Shares are within firm-size category. Source: INEGI, *Encuesta Nacional de Victimización de Empresas* (ENVE), 2019 public aggregate tabulations.

Mexico is also a good setting to study agricultural prices. Production of fruits and vegetables is highly specialized across space, while consumption is dispersed across large destination markets; as a result, the same crop is often shipped from concentrated producing regions to multiple wholesale markets through different road corridors. This creates spatial variation to study route-level frictions, and connects directly to a literature showing that domestic trade costs shape agricultural prices, market integration, and welfare (Donaldson, 2018; Sotelo, 2020; Atkin and Donaldson, 2022). Mexico is especially useful because SNIIM reports high-frequency wholesale prices by product, origin, and destination market, and SIAP production data make it possible to identify the producing areas behind each reported origin. These features allow us to map price differences across markets to the road corridors over which goods are likely shipped. We will explain in more detail these data in the following sections.

3.3 Model

This section lays out a spatial model of agricultural trade with transport intermediaries. Goods are produced in locations, moved through a road network with multiple feasible routes, and consumed in destination locations. Crime enters the benchmark model as a route-level shipping friction: insecurity along a corridor raises the effective cost of moving goods through that corridor. The purpose of the model is to connect route-level crime exposure to delivered prices, trade flows, and welfare in a general-equilibrium setting.

3.3.1 Environment

Goods and space. There are K agricultural goods indexed by k and a set of locations. We use o when a location is the origin of a shipment and d when a location is the destination of a shipment. In the general-equilibrium counterfactual, origins and destinations are therefore two roles played by the same set of inhabited locations. Time is indexed by t. For each origin–destination pair (o, d) there is a set of feasible road routes $\mathcal{R}_{od}$, indexed by r.

The model therefore has three spatial objects:

1. **Origins.** Production takes place in locations o, which differ in agricultural productivity, land endowments, wages, population, and access to the road network.
2. **Destinations.** The same locations appear as destinations d when residents buy delivered agricultural varieties. Destination expenditure is local income, not an exogenous market shifter.
3. **Routes.** Each origin–destination link can be served through several alternative corridors. These routes differ in travel time, road quality, geography, and the crime exposure encountered along the way.

Agents. There are three types of agents. First, competitive agricultural producers in origin o produce good k using labor and land. Second, transport intermediaries buy goods at the farm gate and deliver them to destination locations. Third, residents in each destination location aggregate delivered origin varieties through CES demand. The model also keeps track of local income, cost of living, and population, which are the objects needed to compute real-income welfare.

Within-period sequence. Within each period t, producers choose inputs and supply output at the farm gate, intermediaries organize delivery from each origin to each destination through the road network, residents spend local income on tradables and non-tradables, and local wages, prices, and population adjust to clear markets.

3.3.2 Production

Technology. Origin o produces good k under constant returns to scale with labor and fixed land:

$$Y_{ot}^{k} = A_{ot}^{k} \left(L_{ot}^{k}\right)^{\mu} \left(\bar{H}_{o}^{k}\right)^{1-\mu}, \tag{3.1}$$

where A_{ot}^{k} is productivity, L_{ot}^{k} is labor employed in that crop and origin, $\bar{H}_{o}^{k}$ is fixed agricultural land, and $\mu \in (0,1)$ is the labor share.

Factor prices and unit cost. Under perfect competition, producers sell at unit cost. Let w_{ot} denote the local wage and let r^k_{ot} denote the rental price of land used in origin o and crop k. Cost minimization implies the unit production cost

$$c^k_{ot} = \frac{1}{A^k_{ot}} \left(\frac{w_{ot}}{\mu}\right)^{\mu} \left(\frac{r^k_{ot}}{1-\mu}\right)^{1-\mu}. \tag{3.2}$$

The farm-gate cost is increasing in wages and land rents and decreasing in productivity. Because land is fixed, origins with scarcer effective land face higher marginal cost.

3.3.3 Transport Intermediation and Route-Level Trade Costs

The role of intermediaries. For each origin–destination pair (o, d) and good k, a transport intermediary buys the good at the farm-gate cost c^k_{ot} and delivers it to destination d. The intermediary does not take the road network as a black box. Instead, it faces an explicit routing problem: different corridors imply different delivery costs because they differ in physical difficulty and in exposure to crime.

Route-specific generalized shipping cost. Each route $r \in \mathcal{R}_{od}$ is characterized by route-specific crime exposure κ_{odrt}. This scalar object summarizes the insecurity faced along corridor r in period t and is taken as primitive in the model. The intermediary's route-level generalized cost has two parts: a non-crime component that reflects travel time and other predictable shipping difficulty, and a crime component that could capture, for example, extortion, robbery risk, and defensive expenditures. We write

$$g_{odrt} = \delta\, \ell_{odr} + \phi\, \kappa_{odrt}, \tag{3.3}$$

where ℓ_{odr} captures time-invariant route characteristics such as travel time, distance, or infrastructure quality, and $\delta > 0$ is the elasticity of trade costs with respect to non-crime route difficulty. The parameter $\phi > 0$ is the semi-elasticity of the transport wedge with respect to route crime.

Route-level transport wedge. The generalized shipping cost induces a route-specific cost-equivalent ad valorem wedge

$$\tau_{odrt} = \exp(\delta\, \ell_{odr} + \phi\, \kappa_{odrt})\,. \tag{3.4}$$

This wedge is the object that enters delivered marginal cost. It should be interpreted as a price-equivalent shipping multiplier: it summarizes all per-unit route-specific delivery frictions in one term, including distance, fuel, labor, maintenance, security costs, delay, and crime risk. When $\ell_{odr} = 0$ and $\kappa_{odrt} = 0$, the route is frictionless and $\tau_{odrt} = 1$.

The intermediary's shipping problem. To deliver one effective unit from origin o to destination d in period t, the intermediary combines route-specific shipping services. Let x_{odrt} denote the amount of route-r shipping service used. The intermediary solves the unit-cost problem

$$\tau_{odt} = \min_{\{x_{odrt}\}_{r\in\mathcal{R}_{od}}} \sum_{r\in\mathcal{R}_{od}} \tau_{odrt} x_{odrt} \quad \text{s.t.} \quad \left[\sum_{r\in\mathcal{R}_{od}} \eta_{odr} \left(x_{odrt}\right)^{\frac{\zeta-1}{\zeta}}\right]^{\frac{\zeta}{\zeta-1}} \geq 1, \tag{3.5}$$

where $\eta_{odr} > 0$ is a route quality weight and $\zeta > 1$ is the elasticity of substitution across route services. This problem is best interpreted as the aggregate representation of a continuum of intermediaries with heterogeneous operational fit across routes. Some shipments use one corridor, others use another, and the aggregate delivery technology summarizes that allocation in a tractable way.

Bilateral trade cost. The dual of the shipping problem is the origin–destination delivery wedge

$$\tau_{odt} = \left[\sum_{r\in\mathcal{R}_{od}} \eta_{odr} \left(\tau_{odrt}\right)^{1-\zeta}\right]^{\frac{1}{1-\zeta}}. \tag{3.6}$$

This is the reduced-form bilateral trade cost that enters the rest of the model. Importantly, it is not assumed directly. It is generated by the intermediary's routing problem.

Route shares. Conditional on serving the origin–destination pair, the share of shipping service allocated to route r is

$$\omega_{odrt} = \eta_{odr} \left(\frac{\tau_{odrt}}{\tau_{odt}} \right)^{1-\zeta}. \tag{3.7}$$

Routes with lower crime and lower physical difficulty attract more traffic. The parameter ζ governs how easily intermediaries can reallocate across corridors. When ζ is high, traffic diverts strongly toward safer or faster routes. When ζ is low, the intermediary is effectively captive to the available network.

Delivered marginal cost and price. Once the shipping problem is solved, the delivered marginal cost of one unit of good k from origin o to destination d in period t is simply

$$mc^k_{odt} = c^k_{ot} \cdot \tau_{odt}. \tag{3.8}$$

If the intermediary sector is competitive, the delivered price equals delivered marginal cost. More generally, under isoelastic demand the intermediary charges a constant product-specific markup $\mu^k_I \geq 1$, so that

$$p^k_{odt} = \mu^k_I \, c^k_{ot} \cdot \tau_{odt}. \tag{3.9}$$

This is the key object that links route-level crime to delivered prices. Taking logs gives

$$\ln p^k_{odt} = \ln \mu^k_I + \ln c^k_{ot} + \ln \tau_{odt}. \tag{3.10}$$

Thus, any empirical delivered-price regression must absorb origin–product–time production costs and isolate the part of the bilateral delivery wedge that varies with route crime.

3.3.4 Demand and Price Aggregation

Origin aggregation. Destinations combine origin varieties through an Armington CES structure. For each good k, the expenditure share of destination d on origin o is

$$S_{odt}^{k} = \beta_{od}^{k} \left(\frac{p_{odt}^{k}}{P_{dt}^{k}} \right)^{1-\sigma}, \qquad P_{dt}^{k} = \left[\sum_{o} \beta_{od}^{k} \left(p_{odt}^{k} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \tag{3.11}$$

where S_{odt}^{k} is the expenditure share of origin o in destination d for good k, β_{od}^{k} is an origin preference or quality shifter, and $\sigma > 1$ is the elasticity of substitution across origins.

Goods aggregation. At the top tier, destinations aggregate across agricultural goods:

$$P_{dt}^{T} = \left[\sum_{k} \alpha_{d}^{k} \left(P_{dt}^{k} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}}, \tag{3.12}$$

where α_{d}^{k} is the taste weight on good k and $\rho > 1$ is the elasticity of substitution across goods.

Expenditure and trade flows. Let I_{dt} denote total income in destination location d and let α denote the expenditure share on the modeled tradable goods. Tradable expenditure is

$$E_{dt}^{T} = \alpha I_{dt}. \tag{3.13}$$

In the simplest closed version with the same inhabited locations as origins and destinations, no outside sector, and no land-rent feedbacks, local income is labor income and $I_{dt} = w_{dt} L_{dt}$. More generally, I_{dt} includes all local factor income, including land rents when the fixed factor is active, and any outside-sector income used to close destination demand in the quantitative implementation.

Let s_{dt}^{k} be the expenditure share of good k in destination d,

$$s_{dt}^{k} = \alpha_{d}^{k} \left(\frac{P_{dt}^{k}}{P_{dt}^{T}} \right)^{1-\rho}. \tag{3.14}$$

Nominal expenditure by destination d on good k from origin o is then

$$X_{odt}^{k} = S_{odt}^{k} s_{dt}^{k} E_{dt}^{T}. \tag{3.15}$$

This is the object that closes the goods market and links destination income to origin revenues.

3.3.5 Additional Equilibrium Blocks

Non-tradable sector. Each location also has a non-tradable sector such as local services, storage, and housing. For an origin o, output is produced with labor and a fixed factor:

$$Y_{ot}^{NT} = A_{ot}^{NT} \left(L_{ot}^{NT}\right)^{1-\xi} \bar{H}_{o}^{\xi}, \tag{3.16}$$

where A_{ot}^{NT} is non-tradable productivity, L_{ot}^{NT} is labor in the local sector, and $\bar{H}_o$ is fixed non-tradable land or space. The associated non-tradable price is denoted p_{ot}^{NT}.

Cost of living. Residents consume modeled tradables and a residual local or outside good. Let P_{ot}^{T} denote the tradable price index relevant for households in origin o, and let $\alpha \in (0,1)$ be the expenditure share on modeled tradables. The local cost-of-living index is

$$\mathcal{P}_{ot} = \left(P_{ot}^{T}\right)^{\alpha} \left(p_{ot}^{NT}\right)^{1-\alpha}. \tag{3.17}$$

Population. Population can respond to real income and local amenities. Let B_{ot} denote the amenity level of origin o, $\mathcal{P}_{ot}$ its cost-of-living index, and $\nu > 0$ the mobility parameter. A convenient reduced-form expression for the population allocation is

$$L_{ot} \propto \bar{T}_o \left(\frac{B_{ot}\, w_{ot}}{\mathcal{P}_{ot}}\right)^{\nu}, \tag{3.18}$$

where $\bar{T}_o$ is an exogenous location weight. When ν is small, migration is sluggish and population is nearly fixed. When ν is large, people reallocate more strongly toward high-real-wage, high-amenity locations.

3.3.6 Equilibrium and Benchmark Cases

Equilibrium. Given route characteristics, land endowments, demand shifters, and baseline local fundamentals, a competitive equilibrium in each period consists of agricultural labor allocations, wages, non-tradable prices, bilateral trade costs, route shares, delivered prices, destination price indices, income, expenditure, and population levels such that:

1. producers minimize costs and sell at unit cost,
2. intermediaries solve the shipping problem in Eq. (3.5),
3. residents allocate expenditure according to the CES demand system,
4. destination tradable expenditure equals local tradable income under balanced trade,
5. labor, land, and goods markets clear, and
6. population satisfies the mobility condition in Eq. (3.18).

Welfare in this model is given by real income. For any inhabited location j, let I_{jt} denote total local income, L_{jt} local population, and $\mathcal{P}_{jt}$ the cost-of-living index in Eq. (3.17). Location welfare is

$$V_{jt} = \frac{I_{jt}/L_{jt}}{\mathcal{P}_{jt}}. \tag{3.19}$$

When local income is only labor income, this reduces to $V_{jt} = w_{jt}/\mathcal{P}_{jt}$.

The key balanced-trade restriction is that destination expenditure is generated inside the model. In the benchmark with fixed population, an outside good, and labor-only agricultural production,

$$I_{dt} = I_{dt}^X + \sum_o \chi_{do} w_{ot} L_{ot}, \qquad E_{dt}^T = \alpha I_{dt}, \qquad w_{ot} L_{ot} = \sum_{d,k} X_{odt}^k. \tag{3.20}$$

where I_{dt}^X is outside-sector income and χ_{do} maps income from inhabited locations to destination expenditure nodes. If the destination set equals the set of inhabited locations, χ_{do} is the identity matrix. With land or non-tradables active, the same logic applies using total local factor income rather than labor income alone. Appendix 3.C derives the closure and the wage fixed point in detail.

3.4 Data

We combine firm surveys, wholesale price records, agricultural production data, municipal crime data, and the road network. The firm data identify how transport intermediaries respond when their shipment portfolios become more exposed to crime. The price data identify whether those route-level shocks also appear in delivered agricultural prices. The remaining data sources locate production, destination markets, and crime along the road corridors that connect them.

3.4.1 The EAT transport-firm panel

Our firm-level evidence comes from the confidential *Encuesta Anual de Transportes* (EAT), which surveys land-transport firms in SCIAN 48–49. The survey is organized into three series, corresponding to the 2008, 2013, and 2018 questionnaire frames. We use the linked EAT panel for the years in which firms can be merged to our route-level crime measures. For each firm-year, EAT reports a rich set of operating outcomes, including total revenue, total costs, fuel expenditures, insurance expenditures, employment, total trips, total distance, total tons moved, total vehicles, and the total value shipped.

Table 3.1: EAT Panel Coverage by Survey Series

Series	First year	Last year	Firm-years	Mean firms/year	Max firms/year
Serie 2008	2015	2018	1,513	378	553
Serie 2013	2017	2021	1,696	424	424
Serie 2018	2021	2022	530	265	266

Note: Notes: Coverage is computed from the EAT firm-year sample that can be matched to OD-level crime exposure. Serie 2013 is our preferred subpanel because the number of firms is constant across years.

EAT is particularly useful for our setting because firms report their main origin-destination (OD) markets and the revenue share associated with each one. Each firm reports up to four top OD pairs. These reported revenue shares are the outer weights that allow us to aggregate route-level crime exposure into a firm-level measure. Table 3.1 summarizes the coverage of the three series, while Table 3.2 reports the distribution of the main firm outcomes used in the paper.

Our preferred estimating sample is Serie 2013. Relative to the other two frames, it delivers the cleanest panel structure for a fixed-effects design: it covers four years, from 2017 through 2021, and contains the same 424 firms in every year. By contrast, Serie 2008 exhibits substantial changes in the number of observed firms over time, and Serie 2018 covers only two years. We therefore treat Serie 2013 as the main subpanel and use the pooled sample as a complement.

Table 3.2: Selected Descriptive Statistics in the EAT Panel

Variable	Observations	Mean	P25	Median	P75
Total revenue	3,739	306,617	44,668	153,743	349,232
Shipping revenue	3,738	296,938	43,120	145,424	342,172
Total costs	3,739	180,855	26,643	92,373	212,618
Fuel expenditure	3,209	75,272	11,692	39,633	91,000
Insurance expenditure	3,721	6,553	309	2,768	7,633
Employment	3,739	211	29	100	257
Total vehicles	3,672	244	20	98	296
Total trips	3,316	31,387	3,646	12,096	32,292
Total distance	3,316	12,654,968	1,589,705	5,263,200	15,000,600
Total tons shipped	3,316	585,627	54,318	207,360	603,654
Value shipped	2,531	425,406	0	6,500	175,864

Note: Notes: This table reports the pooled distribution of the main firm outcomes used in the EAT analysis after constructing the firm-level route-crime exposure measure. Monetary amounts are reported in the survey units used by EAT. All descriptive moments are rounded to whole numbers. In the regression analysis we work with the logarithm of these outcomes.

3.4.2 Crime and road network data

Our crime measures combine municipality-year outcomes from the Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP) with the Mexican road network. The empirical crime bundle contains transporter robbery, extortion, and homicide, measured at the municipality-year level over 2011–2025. These are the crimes most closely connected to route security and the broader violence environment faced by firms and shipments.

Figure 3.3 plots annual national totals for the bundle and for each individual crime. As it can be seen, the aggregate bundle is relatively stable and at lower levels between 2011 and 2013, and then it increases towards the end of the decade. This is the temporal variation that we will exploit, as we will explain in the following section.

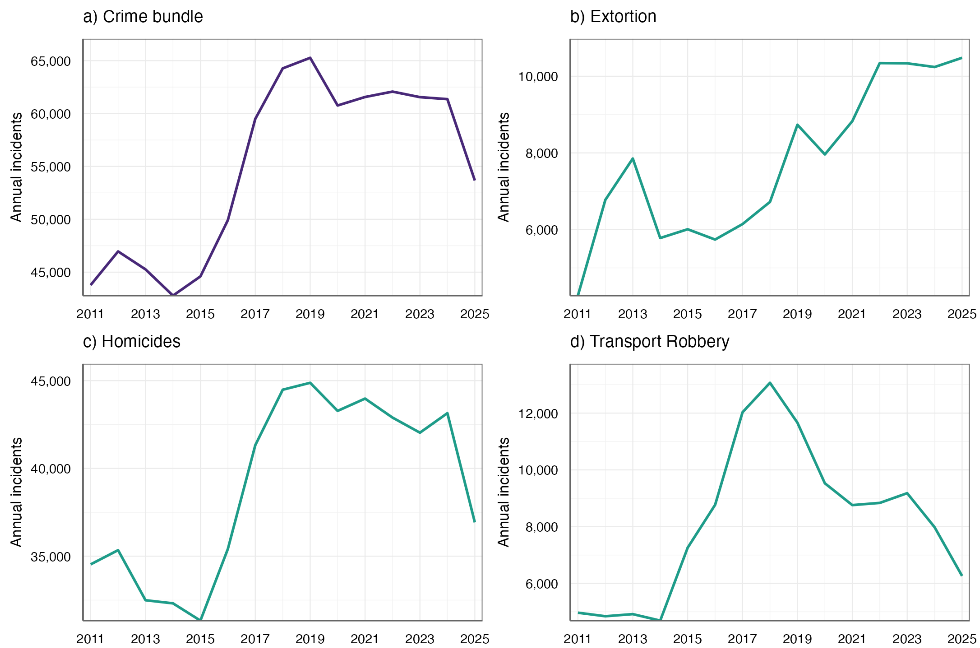


Figure 3.3: Temporal variation of crime

Notes: The figure reports annual national incident totals for the crime measures used to construct route-level exposure. Panel a reports the three-crime bundle; panel b reports extortion; panel c reports homicide; and panel d reports transporter robbery. Sources: Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP), municipal criminal incidence records.

Crime is far from uniform across space. Figure 3.4 maps municipalities by total incidents per road kilometer over 2011–2025, separately for the crime bundle and for each component. The distribution is highly uneven across space, with high-exposure municipalities appearing mostly in central Mexico, around major metro areas such as Monterrey and Mexico City, Sinaloa, and in border cities such as Tijuana and Matamoros.

We construct feasible origin-destination routes between the markets reported by EAT firms and between SIAP-weighted agricultural origins and the wholesale markets observed in SNIIM. The route construction uses the Red Nacional de Caminos road network, keeping the geography fixed so that time variation in exposure comes from crime rather than from changes in the road map. We then overlay municipality-level crime outcomes on the municipalities traversed by each route to obtain route-year measures for each crime type. This route-level panel is the raw input for the exposure objects used in the empirical strategy. The next section describes how route-year crime is aggregated into OD-level and firm-level exposure measures for the EAT and SNIIM regressions.

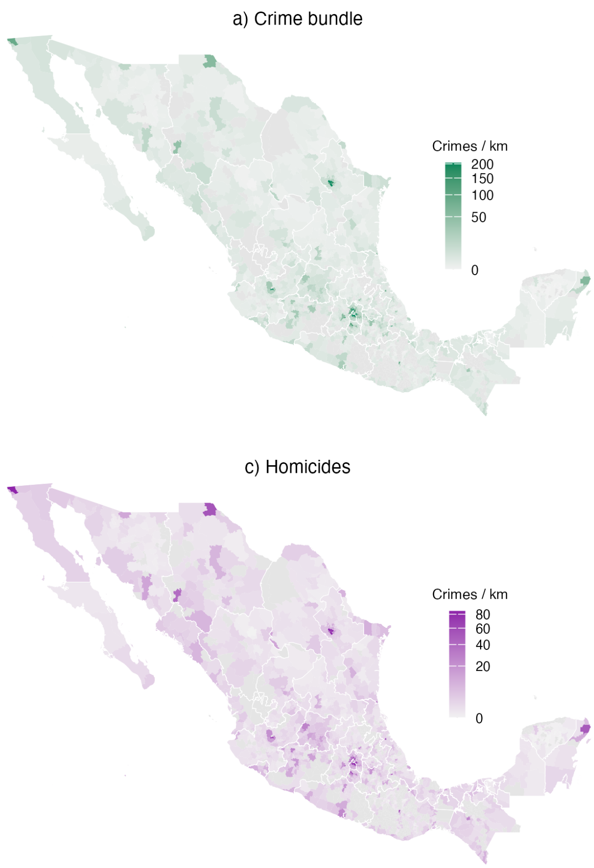


Figure 3.4: Municipality-level crime exposure per road kilometer, 2011–2025

Notes: Each panel reports total incidents over 2011–2025 divided by municipal road kilometers. The crime bundle is the sum of transporter robbery, extortion, and homicide. Color scales are panel-specific square-root scales capped at the 99th percentile to preserve within-panel variation in the presence of extreme values. Municipalities without valid road-kilometer denominators are shown in gray. Sources: Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP), municipal criminal incidence records; Red Nacional de Caminos; INEGI, 2010 municipal boundaries.

3.4.3 Enterprise victimization survey (ENVE)

The *Encuesta Nacional de Victimización de Empresas* (ENVE) is INEGI's enterprise victimization survey. It records whether economic units were victims of crime, the types of crimes they experienced, the number of incidents, whether incidents were reported to authorities, direct monetary losses, and spending on prevention. ENVE is therefore useful for measuring the firm-side costs of insecurity that do not appear in administrative crime records: extortion, theft, vehicle theft, theft of merchandise in transit, prevention spending, and the large gap between crimes experienced and crimes reported.

In the paper, ENVE plays a complementary role. Public ENVE aggregates motivate the setting by showing that insecurity is widespread across Mexican firms and that transport-relevant crimes are especially salient for larger firms. We also link ENVE to EAT when firm identifiers allow it, which gives a smaller matched sample used to interpret the EAT cost regressions: the match helps separate direct losses from preventive spending and gives a scale for the crime costs reported by transport firms.

3.4.4 Wholesale prices and agricultural origins

SNIIM wholesale prices. Our price evidence comes from the *Sistema Nacional de Información e Integración de Mercados* (SNIIM), which records wholesale prices for fruits and vegetables in Mexican wholesale markets. Each quote identifies a product, commercial presentation, origin, destination market, date, and reported price. The cleaned daily file contains about 15 million price observations from January 2000 through February 2026, covering 222 products and 60 destination markets. For the empirical price exercise, we use the frequent wholesale price and aggregate the daily records to product–presentation–origin–destination monthly cells. The yearly panel used in the main price table then averages monthly log prices within product–origin–destination–year cells and averages product varieties within the same crop so that the price outcome has the same crop–origin–destination–year grain as the route-crime exposure.

Figure 3.5 illustrates the time-series variation in four high-coverage products after removing broad seasonal and annual movements. The residual price series show that there remains substantial within-route variation once common calendar-month and year means are removed. Appendix Figure 3.9 summarizes the same residual variation across destination markets over time, which is useful for assessing whether price gaps across markets are stable or widen in periods when transport conditions deteriorate.

An interesting fact across these two figures is that dispersion of prices seems to increase over time, and particularly after 2010. This increase in dispersion coincides with the rise in crime during the 2010 decade. This is of course not causal evidence, but a useful fact to recall when identifying more formally the effect of crime on prices in the following section.

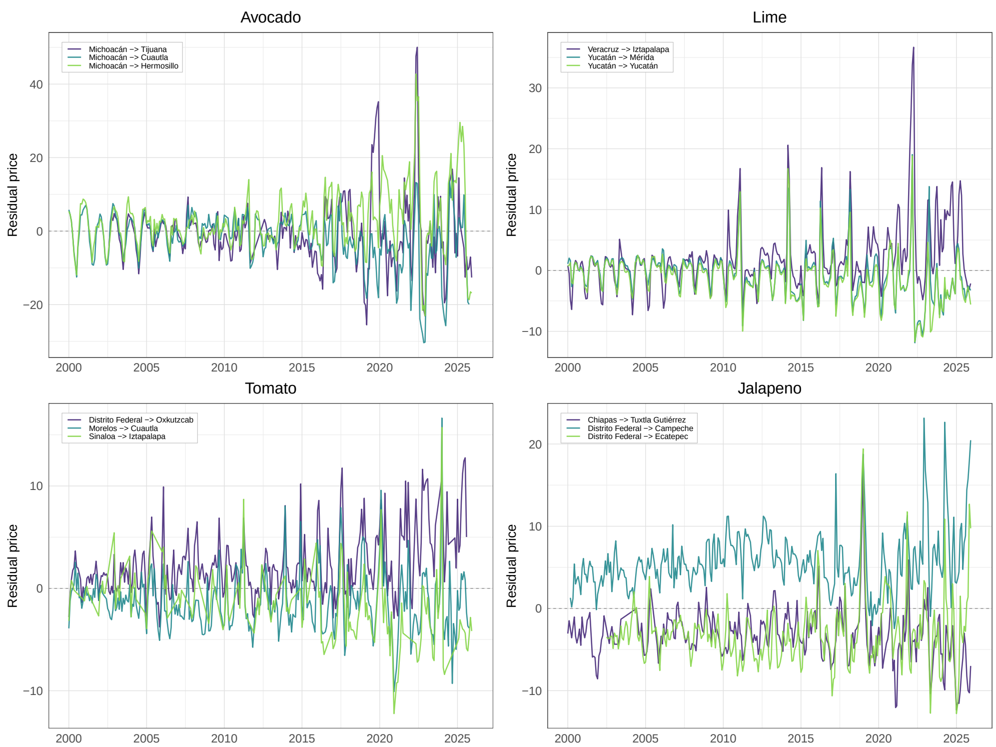


Figure 3.5: Residual wholesale price series for selected SNIIM products

Notes: Each panel plots monthly residual frequent wholesale prices for one product. Lines correspond to the three origin–destination pairs with the most observations for that product. Residual prices remove product-specific calendar-month and year means and add back the product mean. Prices are nominal Mexican pesos per kilogram.

Two features of SNIIM matter for the design. First, destinations are specific wholesale markets, not just states, so we keep destination markets distinct when building routes and estimating price regressions. Second, SNIIM origins are reported as origin states or origin labels, not as precise farm municipalities. This means that SNIIM can identify the state from which a product is shipped, but an additional production source is needed to locate the relevant origin within that state.

Agricultural production (SIAP). We use the *Cierre de la Producción Agrícola* from SIAP to locate agricultural production within SNIIM origin states. SIAP reports annual municipality-level production by crop for Mexico. We map SNIIM product names to SIAP crop names; the matched set contains 22 SIAP crops, including the main fruits and vegetables used in the price exercise. These data enter the paper in two ways. First, production-weighted crop-state centroids provide geographic origin points for routing SNIIM shipments through the road network. Second, municipality shares of state-crop production provide the weights used to aggregate municipality-level route exposure into state-origin by destination-market by crop exposure.

The production weights are predetermined relative to the crime variation used in the price regressions. We use pre-Calderón production geography to construct static crop-state centroids and use the 2005 SIAP distribution to compute the municipality shares. This is important because the weights should describe where production was located before the rise in violence, rather than allowing the origin weights themselves to adjust to later crime shocks. Figure 3.6 summarizes this concentration. For most matched crops, production within a state is concentrated enough that a production-weighted origin is informative, and the leading production municipalities are relatively stable over time.

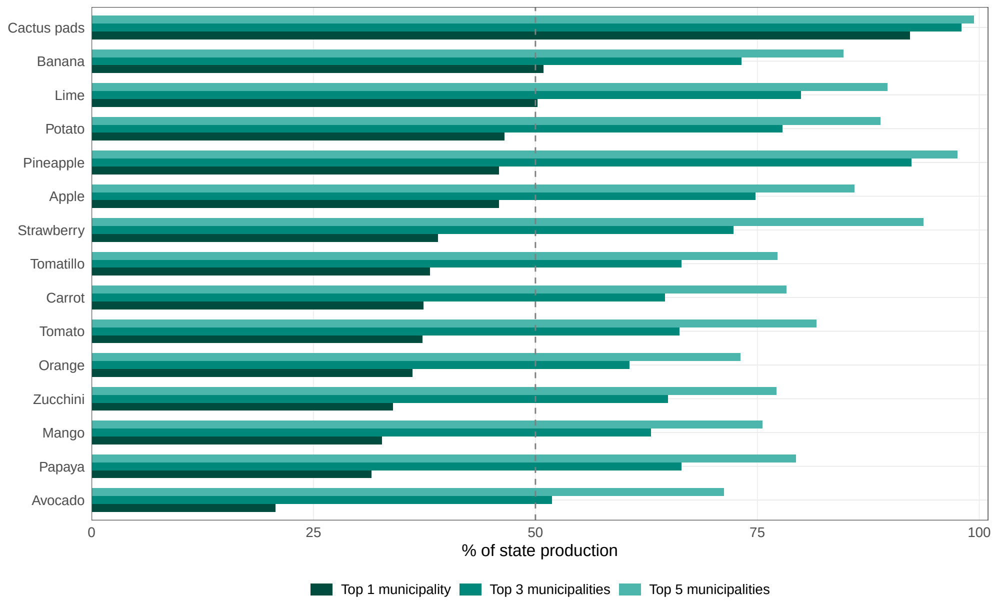


Figure 3.6: Within-state concentration of SIAP production

Notes: Bars show the production-weighted share of state-crop output produced by the largest one, three, or five municipalities within each producing state for selected matched crops. Shares are computed from 2005–2006 municipal production and averaged across producing states using state-crop production as weights. The dashed vertical line marks 50 percent of state production.

3.5 Empirical Strategy

The model provides the empirical roadmap for the paper. It says that crime matters because goods are moved through routes, not because they originate or end in a single place. That route-based logic leads to two empirical questions. First, do transport firms whose shipment portfolios are more exposed to crime experience higher costs and adjust their operations? Second, do destination markets reached through more crime-exposed corridors face higher wholesale prices? We organize the empirical strategy around those two questions.

The regressions below should be read as model-guided empirical tests rather than as exact estimating equations for every structural parameter. The model disciplines the exposure measure, the relevant margins, and the sign predictions. We then use the estimated magnitudes as inputs and moments for the quantitative calibration.

3.5.1 Measuring Exposure to Crime

A route-based notion of exposure. In the model, what matters for trade is not only crime at the endpoint locations, but crime encountered along the corridor that connects the origin and the destination. Our empirical exposure measure follows that logic. Firms ship across space, pass through multiple municipalities, and can often use more than one feasible route between the same origin and destination. Exposure to crime should therefore be measured at the route level and then aggregated up using the geography of the shipment.

Route-level crime intensity. For each route r connecting origin o to destination d, we compute route-level exposure as the average crime intensity across the interior municipalities traversed by that route:

$$\kappa_{odrt} = \frac{1}{|M_{odr}^{int}|} \sum_{m \in M_{odr}^{int}} \rho_{mt}, \tag{3.21}$$

where ρ_{mt} is the municipality-year crime bundle, standardized by road kilometers, and M_{odr}^{int} denotes the set of interior municipalities on route r. This is the empirical counterpart of the scalar route-level crime object in the model.

Figure 3.7 illustrates the spatial viaration of crime across different route alternatives from selected origins. For example, focusing on panel c) Monterrey, it can be seen that a firm that ships from Monterrey to Guadalajara or Ciudad de Mexico is subtantially more exposed to crime relative to shipping to Ciudad Juarez in the North. At the same time, route alternatives matter. If a firm ships from Tijuana, as in panel d), it will not have many alternatives to ship to Ciudad Juarez or Monterrey as it has available only single major roads. As a result, this firm will be heavily exposed to changes in crime that occur along its only way.

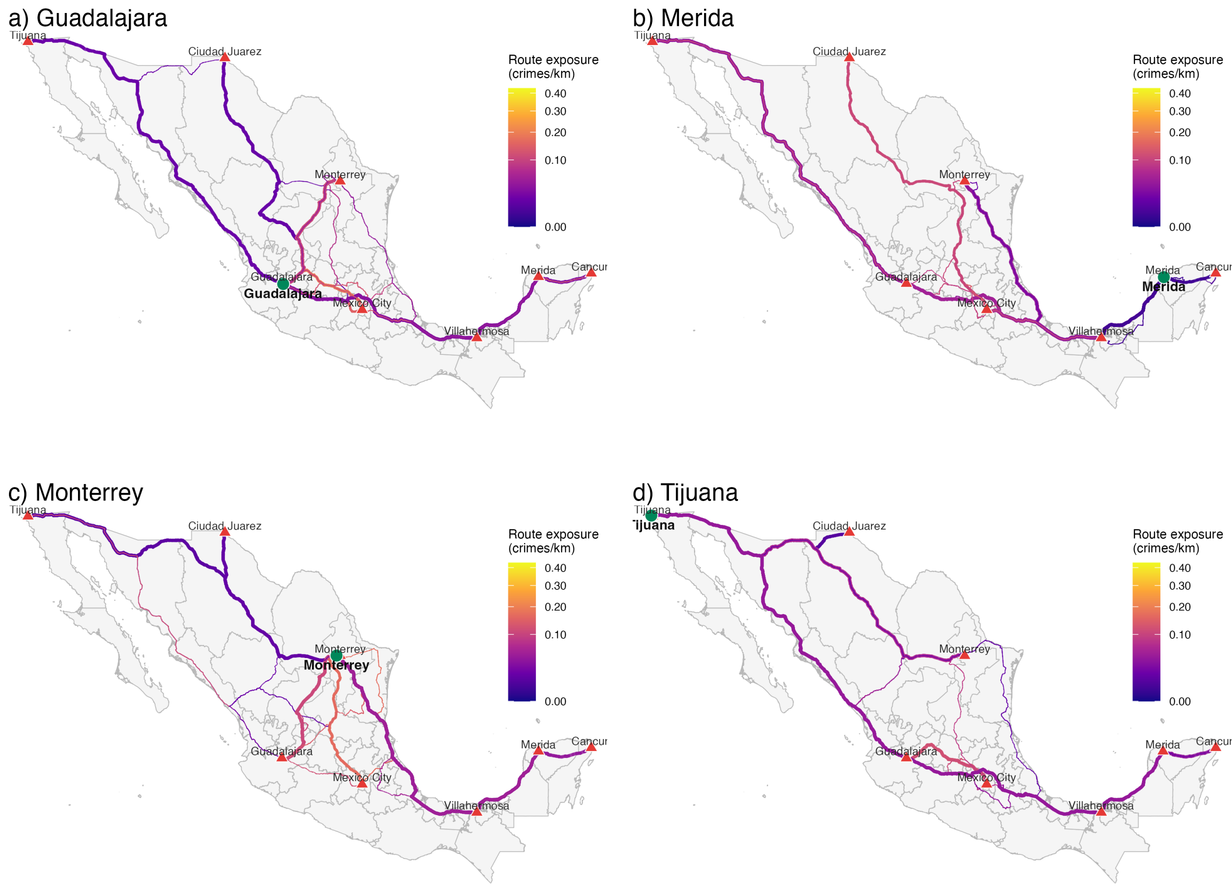


Figure 3.7: Route alternatives and crime exposure for selected origins

Notes: Each panel displays road-network route alternatives from the origin municipality named in the panel to eight wholesale-market destinations: Tijuana, Ciudad Juarez, Mexico City, Guadalajara, Monterrey, Cancun, Villahermosa, and Merida. Routes are colored by average crime exposure over 2011–2025, measured as the three-crime bundle per kilometer along the route. The color scale is capped at the 95th percentile across these example routes. Line width increases with inverse-time route share; triangles mark destination markets and filled circles mark origins.

From route-level crime to OD exposure. Most origin-destination pairs have several feasible routes. Because realized route choice is not observed directly in all of our datasets, we approximate the relative importance of each feasible corridor using pre-determined route weights. In the baseline, ω_{odr} is the inverse-travel-time share of route r within the OD-specific route set. For each route we compute a baseline-relative crime shift,

$$g_{odrt} = \frac{\kappa_{odrt} - \bar{\kappa}_{odr,2011\text{–}2013}}{\bar{\kappa}_{odr,2011\text{–}2013}}, \tag{3.22}$$

and aggregate across feasible routes:

$$\kappa_{od,t} = \sum_{r \in \mathcal{R}_{od}} \omega_{odr} \cdot g_{odrt}. \tag{3.23}$$

This OD-level exposure index captures both where crime is located and how relevant each corridor is for moving goods between that origin and destination.

Identification and the shift-share exposure design. Our route-crime measure has the structure of a shift-share exposure design. In the language of Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2025), the object in equation (3.23) is a weighted average of crime shifts, where the shares ω_{odr} are predetermined route weights and the shifts g_{odrt} are changes in crime exposure along each feasible corridor. The shift-share literature emphasizes that the credibility of such designs must come from a clear argument about the shares, the shifts, or both. We interpret our identifying variation primarily through the share-exogeneity logic emphasized by this literature.

The shares are not chosen as a function of contemporaneous crime, prices, or firm outcomes; they are engineering features of the road network, determined by feasible route paths and inverse driving time. Moreover, they are tailored to the treatment: they measure precisely the extent to which an origin–destination pair is mechanically exposed to crime while goods are in transit, rather than a generic measure of market size, distance, or local economic activity. The identifying comparison is therefore between otherwise similar shipment relationships whose predetermined road geography exposes them differentially to subsequent changes in crime along transport corridors.

The shifts are less naturally viewed as fully as-good-as-random municipality-level shocks. Crime may respond to broader local conditions and to the value of economic activity on some corridors. For this reason, we construct shifts as changes relative to the 2011–2013 baseline, a period in which the relevant road-crime measures were comparatively low and stable. This removes persistent differences in crime levels across corridors and focuses the variation on the sharp post-baseline increase in insecurity. Under our preferred interpretation, the exclusion restriction is that, conditional on the fixed effects and controls in the estimating equation, origin–destination pairs, firms, or destination markets with different predetermined route shares would not have experienced systematically different changes in costs or prices absent the subsequent crime increases along those routes.

3.5.2 Effects of Crime on Transport Firms

Using the model as a guide. In the model, τ_{odt} is a reduced-form ad valorem shipping wedge. It summarizes the per-unit frictions faced when serving an origin–destination pair, including distance, fuel, labor, maintenance, security costs, delay, and crime risk. Because route crime raises this wedge, the model predicts that firms serving more exposed OD pairs should experience changes in outcomes that move with shipping frictions. We therefore begin with the firm data, where the route-cost mechanism is most direct.

Firm-level exposure. EAT reports each firm's main origin-destination markets together with their revenue shares. We use those shares to aggregate OD-level crime exposure into a firm-level Bartik-style index:

$$B_{f,t} = \frac{\sum\limits_{od \in \mathcal{M}_{f,t}} w^{rev}_{f,od,t} \cdot \kappa_{od,t}}{\sum\limits_{od \in \mathcal{M}_{f,t}} w^{rev}_{f,od,t}}, \tag{3.24}$$

where $\mathcal{M}_{f,t}$ is the subset of the firm's reported OD pairs that matches to the route-level crime files. This exposure measure asks whether firms whose commercial geography places more weight on crime-exposed corridors experience different outcomes than otherwise similar firms. It is the firm-level analogue of the model's route-crime wedge: the same OD exposure $\kappa_{od,t}$ that shifts τ_{odt} is averaged over the firm's reported shipment portfolio.

Firm-level specification. Our baseline firm-level specification is

$$\ln Y_{f,t} = \alpha_f + \alpha_t + \beta \cdot B_{f,t} + \varepsilon_{f,t}, \tag{3.25}$$

where α_f are firm fixed effects that control for time-invariant characteristics of the firm, and α_t are year fixed effects that control for economy-wide conditions. We estimate this equation for total operating costs, fuel expenditure, insurance expenditure, total trips, total distance, employment, revenue, and shipped value.

We do not treat the EAT regression as a full structural estimation of the model. Instead, it is the first reduced-form test of the mechanism emphasized by the theory. For cost-side outcomes, β is the empirical counterpart of ϕ: it measures how strongly observed firm costs respond to the same route-crime variation that shifts τ_{odt} in the model. If crime operates through the route-cost channel, then β should be positive for outcomes such as total costs, fuel, and distance. For activity outcomes such as revenue or shipped value, the sign is theoretically more ambiguous because firms may pass costs through, reroute, or shrink activity.

3.5.3 Effects of Crime on Wholesale Prices

The second implication of the model is that higher shipping costs should not remain confined to the balance sheet of transport firms. In the origin–destination version of the model, the delivered wholesale price of product k shipped from origin o to destination d in year t is

$$P^k_{od,t} = \mu^k_I c^k_{ot} \tau_{od,t}, \tag{3.26}$$

where c^k_{ot} is the origin–product–time farm-gate cost and $\tau_{od,t}$ is the bilateral transport wedge generated by the route network. Taking logs gives

$$\ln P^k_{od,t} = \ln \mu^k_I + \ln c^k_{ot} + \ln \tau_{od,t}. \tag{3.27}$$

Thus, the model says that delivered wholesale prices vary because of origin-specific production costs and because of bilateral transport costs. The empirical price regression must therefore control for origin–product–time production costs and isolate the component of the bilateral transport wedge that varies with crime exposure along the routes connecting o and d.

Mapping route crime into the transport wedge. The model's routing problem implies that the OD transport wedge is an aggregate of route-specific wedges. Let κ_{odrt} denote crime exposure along route r between origin o and destination d in year t. A first-order approximation to the route aggregator around baseline route shares gives

$$\Delta \ln \tau_{od,t} \simeq \phi \sum_{r \in \mathcal{R}_{od}} \omega_{odr,0} \, \Delta\kappa_{odrt}, \tag{3.28}$$

where $\omega_{odr,0}$ is the baseline importance of route r in the OD transport wedge and ϕ is the semi-elasticity of route costs with respect to crime exposure.

Equation (3.28) is a change equation. Combining it with Eq. (3.27) gives

$$\Delta \ln P^k_{od,t} \simeq \Delta \ln c^k_{ot} + \phi \sum_{r \in \mathcal{R}_{od}} \omega_{odr,0} \, \Delta\kappa_{odrt}. \tag{3.29}$$

Define the scalar OD crime exposure bundle as

$$\kappa_{od,t} \equiv \sum_{r \in \mathcal{R}_{od}} \omega_{odr,0} \, \Delta\kappa_{odrt}. \tag{3.30}$$

Then the model implies that changes in delivered wholesale prices are proportional to changes in route-share-weighted crime exposure, after accounting for changes in origin–product production costs. The same approximation can be written in levels once the exposure measure is normalized relative to a baseline period. In particular,

$$\ln \tau_{od,t} \simeq \lambda_{od} + \phi\kappa_{od,t}, \tag{3.31}$$

where λ_{od} absorbs the baseline OD transport wedge and all time-invariant route characteristics, including distance, road quality, and geography. Substituting Eq. (3.31) into Eq. (3.27) yields the levels counterpart of the model-implied price equation:

$$\ln P^k_{od,t} \simeq \ln c^k_{ot} + \lambda_{od} + \phi\kappa_{od,t} + \text{constant}. \tag{3.32}$$

Price regression. Guided by Eq. (3.32), our the model-implied wholesale-price specification is

$$\ln P^k_{od,t} = \theta\kappa_{od,t} + \alpha^k_{ot} + \psi^k_{dt} + \lambda_{od} + \varepsilon^k_{od,t}, \tag{3.33}$$

where $P^k_{od,t}$ is the observed SNIIM wholesale price for good k shipped from origin o to destination market d, and $\kappa_{od,t}$ is the scalar route-share-weighted crime exposure bundle for that OD pair. The coefficient θ is the price pass-through of route crime exposure into delivered wholesale prices.

The fixed effects in Eq. (3.33) come directly from the price equation. The origin–product–year fixed effects α^k_{ot} absorb the farm-gate cost term $\ln c^k_{ot}$, including origin-specific productivity, wages, land rents, harvest shocks, and source-side crime that affects production. The destination–product–year fixed effects ψ_{dkt} absorb product-specific wholesale-market shocks at the destination, such as local demand, storage and handling conditions. The origin–destination–product fixed effects λ_{od} absorb permanent bilateral heterogeneity, including distance and road quality. Identification therefore comes from changes over time in route-level crime exposure within the same OD–product cell, after comparing only variation that is not explained by origin–product–year or destination–product–year shocks.

An alternative specification that we run is

$$\ln P^k_{od,t} = \theta\kappa_{od,t} + \alpha^k_{o} + \psi^k_{d} + \lambda^k_{t} + \varepsilon^k_{od,t}, \tag{3.34}$$

where variation comes from relative prices from the same good across different destinations, rather than within an origin-destination pair. This specification effectively compares, for example, the price of an avocado from Michoacán delivered at Mexico City relative to Guadalajara at a given point in time, controling for origin-crop and destination-crop time-invariant characteristics.

3.6 Results

This section presents three sets of evidence. We first use the EAT firm panel to test whether route-crime exposure appears in transport firms' operating costs and activity. We then use SNIIM wholesale prices to ask whether the same exposure is reflected in delivered agricultural prices. Finally, we use the matched EAT–ENVE sample to describe the direct burden of victimization and prevention spending among transport firms.

3.6.1 Firm-level evidence from EAT

The firm-level regressions estimate Eq. (3.25): log firm outcomes are regressed on firm-level route-crime exposure, with firm and year fixed effects. The preferred sample is EAT Serie 2013. As discussed in Section 3.4, this is the cleanest subpanel for a within-firm design because it follows a stable set of firms over 2017–2021. Appendix 3.E reports analogous results using the pooled EAT sample.

Table 3.3: Firm-level EAT regressions: Series 2013, cost outcomes

	Log total costs				Log fuel expense				Log insurance expense			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Transporter-robbery exposure	-0.00005***				-0.00004***				-0.00000			
	(0.00001)				(0.00001)				(0.00002)			
Extortion exposure		0.00439***				0.00659***				0.00096		
		(0.00062)				(0.00101)				(0.00197)		
Homicide exposure			0.09660**				0.06475				-0.08916	
			(0.04552)				(0.05947)				(0.18326)	
Crime exposure				0.02855**				0.02400*				0.00265
				(0.01178)				(0.01429)				(0.03774)
Observations	775	800	805	805	775	800	805	805	696	715	720	720
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm IDs	303	314	314	314	303	314	314	314	278	287	287	287
Years	4	4	4	4	4	4	4	4	4	4	4	4

Notes: Each column is a separate firm-level regression with firm and year fixed effects; only the row for the exposure included in that column is populated. The displayed regressors are firm-year crime exposure measures built in the INEGI lab by merging each firm's reported top origin-destination revenue shares to OD-year route-crime exposure, then averaging over matched OD pairs and renormalizing over the matched share. OD-year route exposure is based on municipal crimes normalized by road kilometers along the relevant routes and expressed as a percent-change-from-base shift. The single-crime rows report transporter robbery, extortion, and homicide exposures; the bundled row reports the crime-bundle exposure available in the frozen lab return. Standard errors in parentheses are clustered at the firm-series level. All numeric coefficient and standard-error entries are reported to five decimals. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3.3 reports the cost-side results. The clearest and most stable pattern comes from extortion exposure. A one-unit increase in extortion exposure–that is, a 100 percent increase relative to the baseline exposure index–is associated with a 0.00439 increase in log total costs and a 0.00659 increase in log fuel expenses. Both coefficients are statistically significant at the one percent level. The implied magnitudes are modest, about 0.4 and 0.7 percent, respectively, but they are precisely estimated and lie on the margins most directly connected to the model's route-shipping channel. By contrast, the insurance coefficient is small and not statistically significant, suggesting that the observed adjustment is not mainly an insurance-premium response.

The other crime measures point in the same broad direction although unevenly. Homicide exposure is positive and statistically significant for total costs, with a coefficient of 0.09660, but it is not significant for fuel or insurance. The bundled crime exposure is positive for total costs and fuel expenses, with coefficients of 0.02855 and 0.02400, significant at the five and ten percent levels, respectively; the insurance estimate is again small and insignificant. Thus, the cost table supports the central claim of the paper: crime exposure raises transport operating costs. The evidence is driven most cleanly by extortion and, secondarily, by the bundled measure, rather than by every crime category.

A separate caveat applies to transporter-robbery exposure. In 2018, transporter robbery was recoded from a local to a federal crime category. In our municipal series, this institutional change appears to generate an artificial decline in municipal transporter-robbery counts after 2018. Because our exposure measure relies on municipality-year crime counts, the transporter-robbery coefficients are contaminated by this reporting break. We therefore report them for transparency, but we do not interpret them as evidence for the main mechanism. This is especially important because the significant cost coefficients are economically tiny and have the opposite sign from the route-cost channel.

Table 3.4: Firm-level EAT regressions: Series 2013, shipment and activity outcomes

	Log shipped value				Log total distance				Log total trips				Log total tons			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Transporter-robbery exposure	-0.00001				-0.00015***				-0.00001				-0.00003			
	(0.00002)				(0.00001)				(0.00001)				(0.00002)			
Extortion exposure		-0.01565**				0.00637***				0.00579***				-0.01152***		
		(0.00645)				(0.00193)				(0.00164)				(0.00346)		
Homicide exposure			0.00256				0.09133				-0.07391				0.19021*	
			(0.32725)				(0.08879)				(0.08060)				(0.09923)	
Crime exposure				-0.07496				0.01358				0.03003				0.08424**
				(0.08265)				(0.02772)				(0.02988)				(0.03365)
Observations	505	525	529	529	775	800	805	805	775	800	805	805	775	800	805	805
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm IDs	236	244	245	245	303	314	314	314	303	314	314	314	303	314	314	314
Years	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

Notes: Each column is a separate firm-level regression with firm and year fixed effects; only the row for the exposure included in that column is populated. The displayed regressors are firm-year crime exposure measures built in the INEGI lab by merging each firm's reported top origin-destination revenue shares to OD-year route-crime exposure, then averaging over matched OD pairs and renormalizing over the matched share. OD-year route exposure is based on municipal crimes normalized by road kilometers along the relevant routes and expressed as a percent-change-from-base shift. The single-crime rows report transporter robbery, extortion, and homicide exposures; the bundled row reports the crime-bundle exposure available in the frozen lab return. Standard errors in parentheses are clustered at the firm-series level. All numeric coefficient and standard-error entries are reported to five decimals. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3.4 shows how firms adjust on shipment and operating margins. Extortion again provides the most coherent pattern. A 100 percent increase in extortion exposure is associated with lower shipped value, higher total distance, more trips, and lower tons shipped. These estimates are consistent with the interpretation that extortion raises the cost of serving exposed shipment portfolios: firms travel farther and complete more trips, while the value and tonnage moved through those portfolios fall. This combination is consistent with a story of rerouting, fragmentation of shipments, delays, or reduced service on exposed corridors.

The remaining activity estimates are less systematic. Homicide exposure is imprecise for shipped value, distance, and trips, and positive for total tons at the ten percent level. The bundled crime exposure is positive and statistically significant only for total tons; its effects on shipped value, distance, and trips are not statistically significant. Transporter robbery again delivers very small coefficients, with a significant negative estimate only for total distance, and should be interpreted cautiously for the reporting reasons described above. Taken together, Tables 3.3 and 3.4 indicate that the firm-level evidence is not a generic "crime increases everything" result. It is a cost-side and operations-side pattern, with extortion as the crime category that most clearly maps into the transport-cost mechanism.

Table 3.5: Yearly crop-level price regressions: crime exposure

	Log price							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Transporter-robbery exposure	0.00001				0.00016***			
	(0.00001)				(0.00005)			
Extortion exposure		0.00060				0.00278		
		(0.00185)				(0.00357)		
Homicide exposure			0.01127**				0.00703	
			(0.00524)				(0.00889)	
Crime exposure				0.01186**				0.00596
				(0.00490)				(0.00732)
Observations	20,419	21,503	21,708	21,708	20,419	21,503	21,708	21,708
Outcome mean (log price)	2.488	2.483	2.487	2.487	2.488	2.483	2.487	2.487
Crop x year FE	✓	✓	✓	✓				
Origin x crop FE	✓	✓	✓	✓				
Destination x crop FE	✓	✓	✓	✓				
Origin x crop x year FE					✓	✓	✓	✓
Destination x crop x year FE					✓	✓	✓	✓
OD pair FE					✓	✓	✓	✓
Two-way cluster SE	O,D	O,D	O,D	O,D	O,D	O,D	O,D	O,D

Notes: The dependent variable is the yearly mean of monthly log SNIIM modal wholesale prices, collapsed to crop-origin-destination-year cells by first averaging monthly log prices at the product-origin-destination-year level and then averaging across product varieties within crop-origin-destination-year cells. Exposure variables are yearly percent-change flownet route exposures relative to the 2011–2012 annual route-level base. At the OD-year level, monthly SESNSP crimes are summed to municipality-year cells, normalized by municipal road kilometers, averaged over the interior municipalities crossed by each route, and aggregated across alternative routes with inverse travel-time weights. The crop-state-origin bridge uses SIAP production shares to average municipality-origin OD exposure to SNIIM crop-origin-state cells. Each column is a separate OLS regression; only the row for the exposure included in that column is populated. Standard errors in parentheses are clustered two-way by origin state and destination market. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

3.6.2 Wholesale price evidence from SNIIM

We next ask whether route-level crime exposure is reflected in wholesale agricultural prices. The SNIIM regressions use a crop–origin–destination–year panel constructed by averaging monthly log modal prices and then averaging across product varieties within crop cells. The table reports two fixed-effect designs. Columns (1)–(4) include crop-year, origin-crop, and destination-crop fixed effects. This specification exploits cross-destination price variation in SNIIM: intuitively, it compares the price of the same crop from a given producing origin across different destination markets that are reached through corridors with different crime exposure, while absorbing common crop-year shocks and permanent origin-crop and destination-crop differences. Columns (5)–(8) are more demanding. They add origin-crop-year fixed effects, destination-crop-year fixed effects, and OD-pair fixed effects, so identification comes from changes in route exposure within OD pairs after absorbing origin-side supply shocks, destination-side demand shocks, and permanent corridor characteristics.

Table 3.5 shows that the price evidence is positive but noisier than the firm-level evidence. In the cross-destination specification, homicide exposure and the bundled crime exposure are positive and statistically significant at the five percent level, with coefficients of 0.01127 and 0.01186. These magnitudes imply that a 100 percent increase in exposure is associated with roughly 1.1–1.2 percent higher wholesale prices. Extortion exposure is small and statistically insignificant in this price table, and transporter-robbery exposure is essentially zero in the same fixed-effect design.

The saturated specifications in columns (5)–(8) point in the same direction but with less precision. The bundled coefficient falls to 0.00596 and is not statistically significant; homicide remains positive but insignificant; and extortion remains positive but imprecise. Transporter-robbery exposure is positive and statistically significant in column (5), but the coefficient is only 0.00016. Given both the small magnitude and the reporting break in transporter robbery after 2018, this estimate is not central to our interpretation. We therefore read the SNIIM evidence as suggestive pass-through from route crime exposure to delivered prices, strongest in the across-destination variation emphasized by the data. Given the estimates, we conclude that there is a noisy crime-to-price pass-through with an elasticity between the range 0.6% and 1.2%.

3.6.3 Matched EAT–ENVE evidence

In addition to the reduced-form analysis, we provide complementary descriptive evidence of how crime affects firms. We do so by analyzing a subset of transportation firms that appear in both the transport EAT and victimization EAT surveys. While the EAT reveals how exposure is associated with operating outcomes, ENVE records whether the firm suffered any type of crime and the direct economic burden of victimization and prevention. The matched EAT-ENVE sample contains 237 firm-year observations corresponding to 154 firms. Of those firms, 84 report at least one targeted crime and 70 do not, so 54.5 percent of matched firms are victimized.

Table 3.6 reports the ENVE descriptive statistics for the matched sample. The available ratio summaries are reported for victimized matched firms, which makes the table especially useful for quantifying the intensive margin of crime costs. The distributions are highly right-skewed. Among victimized firms, average targeted losses are 2.82 million, average prevention spending is 1.03 million, and average total crime cost reaches 3.90 million. Medians are much smaller, which indicates that the burden is concentrated in the upper tail rather than spread evenly across firms.

Table 3.6: ENVE matched sample: victimization, losses, and prevention

Metric	Firms	Mean	Median	Min	Max	P25	P75
Panel A. Victimization and incidents							
Victim extortion (%)	82	16	0	0	100	0	0
Victim goods in transit (%)	84	53	100	0	100	0	100
Victim vehicle theft (%)	82	84	100	0	100	100	100
Number of extortion incidents	14	3	1	1	12	1	3
Number of goods-in-transit incidents	49	2	1	1	19	1	2
Number of vehicle theft incidents	68	3	2	1	24	1	3
Panel B. Economic losses							
Conservative targeted losses (MXN)	60	3,216,695	1,800,000	10,000	22,000,000	600,000	3,000,000
Total targeted losses (MXN)	69	2,816,202	1,000,000	0	22,000,000	277,500	2,850,000
Goods-transit proxy losses (MXN)	47	8,155	0	0	295,000	0	0
Vehicle theft losses (MXN)	60	3,216,695	1,800,000	10,000	22,000,000	600,000	3,000,000
Losses / total costs (%)	69	113	1	0	9,231	0	2
Losses / total revenue (%)	69	58	1	0	4,800	0	1
Panel C. Prevention spending							
Prevention spending (MXN)	67	1,028,642	250,000	10,000	30,000,000	80,000	500,001
Prevention / total costs (%)	67	34	0	0	2,769	0	1
Prevention / total revenue (%)	67	18	0	0	1,440	0	0
Panel D. Total crime burden							
Total crime cost (losses + prevention, MXN)	56	3,901,903	2,000,000	18,000	30,085,000	591,250	4,196,250
Total crime cost / total costs (%)	56	174	1	0	12,000	0	3
Total crime cost / total revenue (%)	56	90	1	0	6,240	0	3

Notes: The table summarizes firms observed in both the EAT transport survey and the ENVE victimization survey, separately for firms reporting a crime and firms not reporting a crime. Rows are grouped into victimization and incident counts, economic losses, prevention spending, and total crime burden. Columns report the number of firms represented in each row and the mean, median, minimum, maximum, 25th percentile, and 75th percentile of the row variable. Monetary variables are in Mexican pesos. Victimization indicators and ratios to total costs or total revenue are reported as percentages. Total crime burden is the sum of economic losses and prevention spending.

From Table 3.6 we can establish four main descriptive facts for victimized matched firms. In a given year, extortion, cargo robbery, and vehicle theft are common; incident counts can be large, with maximum counts of 12, 19, and 24 incidents, respectively; yearly direct economic losses represent 113% of total costs and 58% of total revenue; and crime-prevention spending represents 34% of total costs and 18% of total revenue. This evidence shows that the burden of crime for affected transport firms can be substantial, potentially an order of magnitude larger than regular taxes.

3.7 Calibration

This section describes how we bring the model in Section 3.3 to the data. The exercise is intentionally static. The time subscripts in the model are useful for motivating the empirical specifications, but the counterfactuals compare two steady environments: the observed economy with crime as a route-level friction and a counterfactual economy in which that friction is removed.

3.7.1 Geography and Route Choice

Origins, destinations, and goods. The quantitative geography is built from Mexican municipalities. Let $\mathcal{J}$ denote the set of municipalities with positive employment and value added in the Economic Census. Agricultural origins $\mathcal{O}$ are the municipalities that produce goods in the routed SNIIM–SIAP geography. Destination nodes $\mathcal{D}$ are the SNIIM wholesale markets. We interpret SNIIM destination wholesale markets as regional hubs, so nearby locations are assigned corresponding price indices. Formally, each municipality $j \in \mathcal{J}$ is assigned to a hub

$$h(j) = \arg\min_{d \in \mathcal{D}} \text{dist}(j, d), \tag{3.35}$$

where distance is computed between the municipality centroid and the wholesale-market coordinate. Goods are the agricultural products for which we can connect SNIIM prices to production shares.

Routes. For every origin–destination pair, we construct a set of feasible driving routes on the road network. We keep the four lowest-cost alternatives for each pair. These alternatives capture the fact that shipments are not tied to a single shortest path. When one corridor becomes more expensive because of crime, intermediaries may shift some traffic toward other feasible corridors.

The non-crime component of route costs is measured by the distance of the route in kilometers and enters in logs. Thus the route-level transport wedge used in the quantitative exercise is

$$\tau^r_{odr} = \exp(\delta \log(\mathrm{km}_{odr}) + \phi\, \kappa_{odr}) \,, \tag{3.36}$$

where km_{odr} is route distance in kilometers and κ_{odr} is the normalized route crime index described below. We set $\delta = 0.25$. This value treats distance as a trade-cost elasticity rather than as a raw gravity coefficient. It lies above the clean route-based estimate in Donaldson (2018), who recovers trade costs from price gaps and route-level effective distance, and close to the freight-cost distance elasticities around 0.3 summarized by Anderson and van Wincoop (2004). The important normalization is that δ multiplies log kilometers. If instead distance entered in kilometers, the comparable coefficient would be a small semi-elasticity per kilometer rather than 0.25.

We set the raw route-quality shifter equal to one for every active route. The CES weights η_{odr} are therefore equal-normalized within each origin–destination pair, so that the route aggregator does not mechanically make pairs with more alternatives cheaper simply because more routes are available.

3.7.2 Crime Exposure

Municipality crime intensity. The crime object entering Eq. (3.36) is designed to keep the same route-based logic as the empirical Bartik construction, without importing the full dynamic Bartik object into the static model. The model treats crime as a scalar route friction. On the data side, we construct this scalar as total annual municipal crime counts across the transport-relevant categories used in the paper. Because municipalities differ sharply in the amount of road infrastructure they contain, we normalize annual counts by municipal road kilometers. This gives a crime intensity measured as incidents per road kilometer.

Route aggregation. For each route, we identify the municipalities crossed by the route and exclude the origin and destination municipalities. The exclusion keeps the route measure focused on crime encountered along the corridor rather than crime at the shipment endpoints. The raw route crime measure is the mean of average annual crime intensity across the interior municipalities crossed by the route:

$$\widetilde{\kappa}_{odr} = \frac{1}{|\mathcal{M}^I_{odr}|} \sum_{m \in \mathcal{M}^I_{odr}} \overline{\left(\frac{\text{crime}_m}{\text{road km}_m}\right)}, \tag{3.37}$$

where $\mathcal{M}^I_{odr}$ is the set of interior municipalities crossed by route r, and the overbar denotes the average over available years.

The model-facing crime object is a unit-free index. We divide the raw route measure by its mean among routes with positive exposure:

$$\kappa_{odr} = \frac{\widetilde{\kappa}_{odr}}{\overline{\widetilde{\kappa}}_{+}}, \qquad \overline{\widetilde{\kappa}}_{+} = \frac{1}{|\mathcal{R}_{+}|} \sum_{(o,d,r) \in \mathcal{R}_{+}} \widetilde{\kappa}_{odr}. \tag{3.38}$$

Thus $\kappa_{odr} = 1$ corresponds to an average positively exposed route, while $\kappa_{odr} = 0$ corresponds to no measured interior-route exposure. Removing crime in the counterfactual means setting this route crime index to zero, leaving the non-crime distance component and the route set fixed.

Calibrating the route-crime slope. The EAT cost regressions discipline ϕ. Their crime regressors are percent-change Bartik exposures in proportional units: a one-unit increase is a 100 percent increase relative to the base exposure, not a one percentage point increase. We set $\phi = 0.02855$, the coefficient on the total-crime exposure in the total-cost specification. This choice maps the firm-side evidence most directly to the transport-cost channel in Eq. (3.36). It also keeps the quantitative model agnostic about the composition of crime: the counterfactual removes the scalar route-crime friction, not a separate set of category-level frictions.

3.7.3 Economic Fundamentals

Labor and population weights. Municipality labor size comes from the 2024 Economic Census. We use total employed workers to construct the fixed population weights $\bar{T}_o$. In the shipping-only benchmark the mobility parameter is set to zero, so these weights pin down the spatial distribution of labor. They are not merely starting values for an iterative migration process; with fixed population, they are the population allocation used in the equilibrium.

Productivity and land. The agricultural productivity terms $\bar{A}_o^k$ are calibrated from municipality-level census output together with crop production shares. The Economic Census gives total gross output and employment at the municipality level, while the agricultural data give crop-specific production shares. We allocate municipality output and labor across crops using those crop shares and then invert the Cobb–Douglas production function in Eq. (3.1). This inversion uses a calibration labor share of 0.75. In the benchmark equilibrium, the production block sets $\mu = 1$ so that land does not introduce an additional fixed-factor loop. The same crop shares also provide the fixed agricultural support terms $\bar{H}_o^k$ and the baseline origin weights in destination demand. Appendix 3.A.1 gives the exact inversion and normalizations.

This inversion lets high-output, high-labor municipalities enter the model with stronger fundamentals, while preserving crop-specific heterogeneity from the agricultural production data. Because crop-specific census labor and crop-specific value added are not observed, municipality-level census moments are allocated across crops using SIAP production shares.

Demand and expenditure. Goods weights are based on aggregate crop production, and origin weights within each good are based on crop production shares. Destination expenditure is closed inside the model using local income. The outside good has price normalized to one and stands in for the non-modeled part of local activity.

Let I_j^X denote exogenous outside-sector income in municipality j, calibrated from value added in the Economic Census. Hub outside income aggregates the municipalities assigned to that hub:

$$I_d^X = \sum_{j:h(j)=d} I_j^X. \tag{3.39}$$

Let χ_{do} indicate the hub assignment of producing origin o, so that $\chi_{do} = 1$ when $h(o) = d$ and zero otherwise. Hub income is

$$I_d = I_d^X + \sum_o \chi_{do} w_o L_o, \tag{3.40}$$

and expenditure on the modeled food block is

$$E_d^T = \alpha I_d. \tag{3.41}$$

We set $\alpha = 0.38$, using the ENIGH 2024 share of food, beverages, and tobacco in monthly expenditure. The remaining 62 percent is spent on the outside good. With outside-good price normalized to one, municipality welfare is real income. For municipality j assigned to hub $h(j)$,

$$V_j = \frac{I_j/L_j}{(P_{h(j)}^T)^\alpha}, \qquad I_j = I_j^X + \mathbb{1}\{j \in \mathcal{O}\} w_{o(j)} L_{o(j)}. \tag{3.42}$$

This closure keeps expenditure tied to local income rather than to quote counts or equal market weights, while allowing the empirical destination nodes to remain the wholesale hubs observed in the price data. Appendix 3.A.2 describes the construction of the outside-income layer and the numeraire normalization.

Table 3.7: Structural parameters

Parameter	Value	Role and calibration
$R_{od}^{\max}$	4	Geography: Route alternatives kept for each origin–destination pair. The route-choice build keeps four alternatives.
σ	5.0	Demand: Origin-variety elasticity within an agricultural good. Baseline Armington value.
ρ	3.0	Demand: Elasticity across agricultural goods. Baseline value.
ζ	4.0	Intermediation: Elasticity across alternative route services. Baseline value for rerouting.
α	0.38	Preferences: Expenditure share on modeled food goods. Set to 0.38 from ENIGH 2024: food, beverages, and tobacco account for 38 percent of monthly expenditure.
μ	1.0	Production: Agricultural labor share. Set to one; productivity inversion uses a 0.75 calibration share.
ξ	0.3	Production: Fixed-factor share in non-tradables. Inactive in the outside-good benchmark.
δ	0.25	Transport costs: Route-cost elasticity with respect to log kilometers. Set to 0.25 from the distance-elasticity literature.
ϕ	0.02855	Transport costs: Route-cost semi-elasticity with respect to the unit-free route crime index. Set to 0.02855 from the EAT total-cost coefficient on the total-crime percent-change Bartik.
ν	0.0	Population: Population mobility. Set to zero; population is fixed at Censo labor weights.

Note: *Notes:* Parameter values used in the quantitative exercise. The benchmark lets route crime affect transport costs, holds population fixed, and uses an outside good with price normalized to one.

3.7.4 Parameters

Table 3.7 reports the parameter values used in the quantitative exercise. The benchmark isolates the route-shipping channel: crime affects route-level transport costs through ϕ, distance affects route costs through δ, and intermediaries can substitute across feasible routes according to ζ.

Population is fixed by setting $\nu = 0$. This benchmark is the channel most directly disciplined by the firm-side cost evidence and the route-exposure design.

3.8 Counterfactuals

Welfare cost of crime. We use the calibrated model to ask how much route crime matters through the shipping channel alone. The benchmark economy is the calibrated Mexican equilibrium with observed route crime entering transport costs through

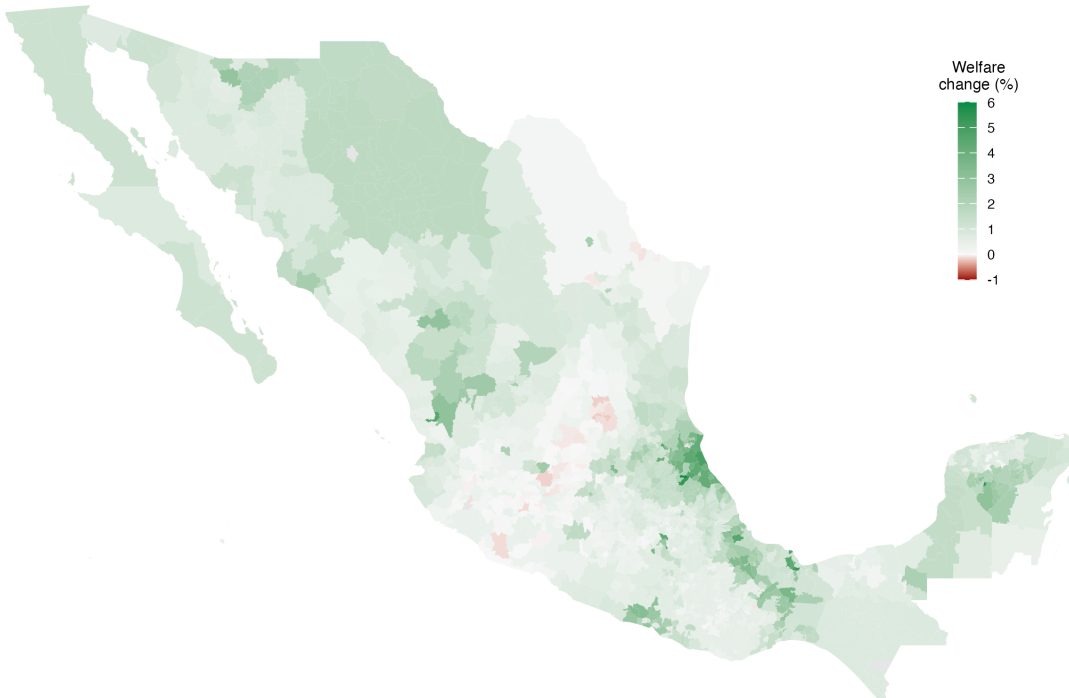
$$\tau_{odr} = \exp(\delta \log(\text{km}_{odr}) + \phi \kappa_{odr}) . \tag{3.43}$$

Table 3.8: Shipping-crime counterfactual

	Change (percent)
Panel A. Welfare	
Population-weighted municipality welfare, V_j	+0.63
P10 municipality welfare	+0.16
P90 municipality welfare	+1.69
Panel B. Trade costs and prices	
Mean route-level trade cost, τ_{odr}	-2.90
Mean destination tradable price index, P_d^T	-2.17
Panel C. Nominal OD expenditure	
Mean nominal OD expenditure change	+1.68

Notes: Percent changes compare the calibrated benchmark with a no-crime counterfactual that sets $\kappa_{odr} = 0$ on every route, holding routes, distance, fundamentals, and hub assignments fixed. Non-shipping crime channels are inactive. Modeled food expenditure is $\alpha = 0.38$ of hub income, and municipality welfare is $V_j = (I_j/L_j)/(P_{h(j)}^T)^\alpha$. The nominal OD expenditure row averages origin-destination pairs with positive baseline expenditure.

The no-crime counterfactual sets $\kappa_{odr} = 0$ for every route. We keep the route network, distance, route-quality weights, productivities, demand weights, and municipality-to-hub price-index assignment fixed. In this exercise, crime does not directly affect farm productivity, non-tradable productivity, amenities, or demand shifters. Those channels are part of the general model, but we leave them inactive because the available evidence does not discipline their crime semi-elasticities. The counterfactual should therefore be read as the welfare cost of route crime operating through



Notes: The map reports percent changes in municipality welfare, $V_j = (I_j/L_j)/(P^T_{h(j)})^\alpha$, for Censo municipalities in the calibrated shipping-only benchmark. Each municipality is assigned the tradable price index of its nearest SNIIM wholesale hub. Municipalities outside the mapped model-location universe are shown in grey.

Figure 3.8: Municipality Welfare Changes from Removing Route Crime

shipping costs.

Table 3.8 reports the resulting equilibrium changes. Welfare is the real-income object from the model. With the outside good as numeraire, municipality welfare is $V_j = (I_j/L_j)/(P^T_{h(j)})^\alpha$, where $h(j)$ is the wholesale hub assigned to municipality j. Population-weighted welfare rises by 0.63 percent when route crime is removed. The gains are not concentrated only in the upper tail: the 10th percentile municipality gains 0.16 percent, while the 90th percentile gains 1.69 percent. These welfare gains are smaller than the changes in modeled-goods prices because the outside good absorbs the part of expenditure outside the broad food category.

The mechanism is visible in the price and trade-cost rows. Removing route crime lowers mean active route-level costs by 2.90 percent. Destination tradable price indices fall by 2.17 percent. These changes are larger than the aggregate welfare effect because welfare is computed over the full consumption basket, while the modeled route-cost channel applies to the food block governed by α.

The nominal OD expenditure row should be interpreted as a reallocation statistic rather than as a real aggregate-trade quantity. It averages percentage changes in nominal origin–destination expenditure across pairs with positive baseline expenditure, and those flows are measured in the model numeraire. This statistic can move differently from welfare because the counterfactual lowers delivered prices, changes sourcing shares, and lets equilibrium wages and local expenditure adjust. The welfare object is real income, while this row summarizes how nominal spending is redistributed across OD links.

Spatial effects. Figure 3.8 maps the municipality-level welfare changes underlying Panel A and shows that welfare gains are not spatially uniform. They are largest in parts of Veracruz, Puebla, Tabasco, and Yucatán in the east and southeast; in parts of Guerrero and Oaxaca in the south; and in the western and northwestern states of Durango and Nayarit. Interestingly, some of these states, most notably Yucatán, are places where crime is not a central concern. In fact, Yucatán is one of the safest states in the country. The spatial effects thus show that the benefits need not occur in the ex-ante dangerous locations, but rather in places that benefit substantially from improved connectivity to other locations.

A small number of municipalities have slightly negative welfare changes. This is not because removing route crime makes local goods more expensive; prices also fall in those places. Rather, the losses come from the income side of the adjustment. As trade costs fall, spending reallocates across producing origins, and a few municipalities see local income fall slightly more than prices. These losses are small and reflect spatial redistribution through equilibrium wages.

3.8.1 Discussion

We find an average welfare gain of 0.63 percent from counterfactually removing crime from route-level shipping costs. We view this estimate as a likely lower bound on the total welfare cost of crime. The benchmark counterfactual isolates the mechanism most directly disciplined by the transport-firm and price evidence: crime raises the cost of moving goods along routes. It does not allow crime to lower farm-gate productivity, reduce non-tradable productivity in warehousing, distribution, and local services, reduce local amenities, induce migration, or make buyers avoid exposed origins for reliability or reputational reasons. Appendix 3.B shows how these channels can be embedded in the same model through crime-dependent agricultural productivity, non-tradable productivity, amenities, and Armington demand shifters. We leave those channels inactive because the available evidence does not estimate their semi-elasticities, but each would plausibly add to the welfare cost measured here.

The ENVE accounting estimate in Section 3.2 provides a useful benchmark. Direct losses and prevention spending by economic units amounted to 0.51 percent of GDP in 2023 (Instituto Nacional de Estadistica y Geografia, 2024). Our 0.63 percent welfare gain is roughly 20 percent larger than that direct accounting burden, even though the counterfactual includes only route-crime shipping costs rather than the full set of victimization and prevention costs. This difference is the reason to compute welfare in equilibrium: removing crime from shipping routes changes delivered prices, sourcing, and market access, so the welfare effect can exceed the direct resources recorded in firm surveys.

Relative to the broader domestic-trade-frictions literature, the magnitude is modest but economically meaningful. It is smaller than the gains from major infrastructure or market-access changes. Donaldson (2018) estimates that railroads in colonial India raised real agricultural income by about 16 percent in connected districts, with social savings equal to 11.2 percent of agricultural income. Asturias et al. (2019) find real-income gains of 2.7 percent from India's Golden Quadrilateral highway project, with allocative-efficiency improvements accounting for 7.4 percent of those gains. Sotelo (2020) finds that paving roads in Peru raises aggregate agricultural productivity by 4.9 percent and median farmer welfare by 2.7 percent, while harming more than 20 percent of farmers through increased competition. Our 0.63 percent welfare gain is therefore about one quarter of the Golden Quadrilateral and Peru median-welfare magnitudes, and much smaller than Donaldson's district-level railroad estimate. It is closer in spirit to friction-specific or network-efficiency counterfactuals: Graff (2024) estimates that reorganizing Africa's road network would raise welfare by about 1.3

percent, roughly twice our benchmark effect. The comparison is not one-for-one because those papers study infrastructure construction or network reallocation, while our exercise removes one institutional wedge from the existing road network. The relevant conclusion is therefore not that route-crime removal has the same scale as a national road or rail program, but that even a narrow shipping-only crime channel generates welfare changes in the range that the domestic trade literature treats as first-order.

This interpretation also connects the counterfactual to evidence on extortion and corruption as trade-cost wedges. Olken and Barron (2009) find that illegal payments on trucking routes in Aceh averaged about 13 percent of trip costs. Sequeira and Djankov (2014) show that firms facing coercive port corruption reroute substantially, traveling an additional 319 kilometers on average to avoid the more corrupt port, and that the corrupt corridor carries a 70 percent transport-price premium. Brown et al. (2025) find that gang collusion in El Salvador raised extortion payments by about 20 percent and retail pharmaceutical prices by 7.6 percent. These magnitudes are larger than our aggregate welfare number because they are local cost or price effects, not economy-wide real-income changes. Our contribution is to map an analogous crime-to-shipping-cost margin into equilibrium prices and real income, while keeping the non-shipping channels transparent for future calibration.

3.9 Conclusion

This paper studies how route crime—extortion, robbery, and homicide—affects firm outcomes and agricultural goods prices in Mexico. Exploiting detailed confidential transportation-firm microdata, origin–destination wholesale agricultural prices, and administrative crime records, we show that doubling our crime exposure i) increases transportation firms' operating costs, with estimates ranging from 0.4% to 9.6%, depending on the type of crime; and ii) increases agricultural prices between 0.6% and 1.1%, although estimates are noisy.

The contribution of the paper is to demonstrate empirically a crime-to-cost and crime-to-price channel, and to quantify the aggregate welfare effects using a spatial trade model with transport intermediaries. We show that counterfactually removing the route-crime friction lowers mean route-level trade costs by about 2.9%, lowers destination tradable price indices by about 2.2%, and raises population-weighted municipality welfare by about 0.63% in a benchmark where modeled food goods account for 38% of expenditure.

The immediate policy implication of the paper is that road security should be treated as a form of trade facilitation. Policy could prioritize economically central shipping routes, particularly corridors with high exposure to crime and limited route substitutes, rather than focusing exclusively on crime at origins or destinations. In this sense, reducing predation on transport networks can operate like an improvement in domestic infrastructure by lowering the effective cost of connecting markets.

A natural next step is to discipline the broader channels that are present in the model but inactive in the benchmark counterfactual. Confidential CPI microdata, together with land and housing price data, would allow us to study how crime affects non-tradable and service prices, local costs of living, and the spatial allocation of economic activity. This would move the quantification from the shipping-only exercise toward a fuller measure of the real aggregate effect of crime, taking into account trade, local productivity, and prices across tradables, non-tradables, and services.

Future work should also use the route structure to evaluate targeted policy experiments. Natural extensions are to remove crime only on the highest-exposure corridors, or compare route-based crime reductions to origin-based reductions. These exercises would help distinguish policies that reduce the measured friction where it is most severe from policies that generate the largest general-equilibrium gains.

Appendix

3.A Calibration Details

3.A.1 Productivity Inversion

This appendix describes the construction of agricultural productivities $\bar{A}_o^k$. The objective is to let high-output, high-labor municipalities enter the quantitative model with stronger fundamentals, while using crop production data to allocate municipality-level census moments across goods.

Let Y_o denote total gross output in municipality o from the Economic Census, and let L_o denote total employed workers in the same municipality. The census data are available at the municipality level, not at the crop-by-municipality level. We therefore use SIAP crop volumes to allocate municipality output and labor across modeled goods. Let q_o^k be SIAP volume for crop k in municipality o. Define the within-municipality crop weight

$$s_o^k = \frac{q_o^k}{\sum_\ell q_o^\ell}, \tag{3.44}$$

where the denominator sums over modeled crops with SIAP support in municipality o. The calibration then allocates

$$Y_o^k = s_o^k Y_o, \qquad L_o^k = s_o^k L_o. \tag{3.45}$$

Because the production data also give crop-specific support in the form of SIAP production shares, we set the fixed-factor shifter to

$$\bar{H}_o^k = \max\{\gamma_o^k, \varepsilon\}, \qquad \varepsilon = 10^{-6}, \tag{3.46}$$

where γ_o^k is the SIAP crop share used elsewhere in the calibration.

Given a calibration labor share $\mu_{cal} = 0.75$, the raw productivity inversion is

$$\bar{A}^k_{o,raw} = \frac{Y^k_o}{(L^k_o)^{\mu_{cal}}(\bar{H}^k_o)^{1-\mu_{cal}}}. \tag{3.47}$$

The model uses productivities normalized relative to the median active crop-origin cell:

$$\bar{A}^k_o = \frac{\bar{A}^k_{o,raw}}{\text{median}_{(o,k):\gamma^k_o>0}\,\bar{A}^k_{o,raw}}. \tag{3.48}$$

Crop-origin cells with no crop support are assigned the small value ε. The same SIAP crop shares enter the demand side as origin weights within each good.

Two normalizations are worth keeping explicit. First, the inversion uses Censo gross output rather than value added; value added is used separately to discipline the outside-income layer in the expenditure closure. Second, the productivity inversion allows a fixed factor through $\mu_{cal} = 0.75$, but the benchmark equilibrium sets the production labor share to $\mu = 1$. Thus $\bar{H}^k_o$ is recorded in the fundamentals, but the fixed-factor cost loop is inactive in the benchmark counterfactual.

3.A.2 Outside Income and Numeraire

The quantitative model does not treat the modeled agricultural goods as the entire economy. We therefore add an outside-good income layer that disciplines local expenditure while preserving the route-level shipping channel as the object of interest. The outside good has price normalized to one,

$$P^X = 1, \tag{3.49}$$

so all nominal objects in the quantitative model are measured in outside-good units.

Let VA_j denote Censo value added in municipality j, and let N_j denote Censo employment. We convert value added into model units using two reference quantities: the median positive Censo value added per worker,

$$\widetilde{y} = \text{median}_{j:VA_j>0}\,\frac{VA_j}{N_j}, \tag{3.50}$$

and the average labor force among routed producing origins,

$$\widetilde{L} = \frac{1}{|\mathcal{O}|} \sum_{o \in \mathcal{O}} N_o. \tag{3.51}$$

The outside-income and labor objects used by the model are then

$$I_j^X = \frac{VA_j}{\widetilde{y}\,\widetilde{L}}, \qquad L_j = \frac{N_j}{\widetilde{L}}. \tag{3.52}$$

This normalization implies that a municipality with median value added per worker has outside income per worker equal to one in model units. Model wages and incomes are solved in the same units.

We use value added rather than remunerations because the income object is intended to measure local expenditure capacity from non-modeled activity. Value added is closer to total local income generated by production, including payroll, mixed income, and operating surplus. Remunerations measure formal wage payments and are therefore a narrower object. A remuneration-based income layer is a natural robustness exercise if the counterfactual is interpreted as applying only to formal labor income.

SNIIM wholesale markets are treated as regional hubs. If $h(j)$ is the hub assigned to municipality j, hub outside income is

$$I_d^X = \sum_{j:h(j)=d} I_j^X. \tag{3.53}$$

Total hub income adds endogenous labor income from producing origins assigned to the hub:

$$I_d = I_d^X + \sum_o \chi_{do} w_o L_o, \tag{3.54}$$

where $\chi_{do} = 1$ if origin o is assigned to hub d and zero otherwise. Modeled food expenditure is

$$E_d^T = \alpha I_d, \qquad \alpha = 0.38. \tag{3.55}$$

The residual share $1 - \alpha$ is spent on the outside good. Therefore the municipality cost of living is

$$P_j = (P_{h(j)}^T)^\alpha (P^X)^{1-\alpha} = (P_{h(j)}^T)^\alpha, \tag{3.56}$$

and municipality welfare is

$$V_j = \frac{I_j/L_j}{(P_{h(j)}^T)^\alpha}. \tag{3.57}$$

3.B Model extensions

The main text studies crime as a route-level shipping friction. This appendix records additional margins that can be embedded in the same equilibrium environment but are not used in the quantitative exercise. These extensions are useful for organizing future versions of the model, where crime may affect local production conditions, amenities, or sourcing preferences in addition to transport costs.

Farm-gate productivity. Crime in producing origins may reduce agricultural productivity by disrupting planting, harvesting, storage, and local input markets. One way to introduce this channel is

$$A_{ot}^k = \bar{A}_{ot}^k \exp(-\phi_A \kappa_{ot}), \tag{3.58}$$

where κ_{ot} is crime at the source and $\phi_A > 0$ is the productivity semi-elasticity with respect to local crime.

Non-tradable productivity. Crime may also lower productivity in local services, warehousing, distribution, and other non-tradables. This can be written as

$$A_{ot}^{NT} = \bar{A}_{ot}^{NT} \exp(-\phi_{NT} \kappa_{ot}) \,, \tag{3.59}$$

with $\phi_{NT} > 0$. This channel would raise local non-tradable prices and feed back into the cost of living.

Amenities. Crime may make locations less desirable to workers and firms through fear, harassment, or a deterioration in public space. A reduced-form amenity channel is

$$B_{ot} = \bar{B}_{ot} \exp(-\phi_B \kappa_{ot}) \,, \tag{3.60}$$

where $\phi_B > 0$. In a model with mobile labor, lower amenities would affect population and wages through the mobility condition in Eq. (3.18).

Demand avoidance. Buyers may avoid sourcing from crime-exposed supply chains beyond the direct cost increase, for example because reliability falls or because certain origins become reputationally risky. This can be captured through the Armington demand shifter:

$$\beta_{od}^{k} = \bar{\beta}_{od}^{k} \exp(-\psi \, \kappa_{odt}) \,, \tag{3.61}$$

where κ_{odt} is a bilateral supply-chain crime exposure index and $\psi > 0$ governs the strength of buyer avoidance.

3.C Balanced-Trade Closure and Wage Determination

This appendix derives the expenditure closure used in the quantitative model. The key point is that the destination expenditure object is not an independent primitive in a closed general-equilibrium trade model. It is local income earned by residents of the destination location.

3.C.1 Expenditure Shares

For each destination location d and good k, the CES price index across origins is

$$P_{dt}^{k} = \left[\sum_{o} \beta_{od}^{k} \left(p_{odt}^{k} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} . \tag{3.62}$$

The expenditure share of destination d on origin o for good k is therefore

$$S_{odt}^{k} = \beta_{od}^{k} \left(\frac{p_{odt}^{k}}{P_{dt}^{k}} \right)^{1-\sigma} . \tag{3.63}$$

Across goods, the tradable price index is

$$P_{dt}^{T} = \left[\sum_{k} \alpha_{d}^{k} \left(P_{dt}^{k} \right)^{1-\rho} \right]^{\frac{1}{1-\rho}} , \tag{3.64}$$

and the expenditure share of destination d on good k is

$$s_{dt}^{k} = \alpha_{d}^{k} \left(\frac{P_{dt}^{k}}{P_{dt}^{T}} \right)^{1-\rho} . \tag{3.65}$$

If total tradable expenditure in destination d is E_{dt}^{T}, then nominal expenditure on origin o and good k is

$$X_{odt}^{k} = S_{odt}^{k} s_{dt}^{k} E_{dt}^{T}. \tag{3.66}$$

This is the gravity block. It determines expenditure shares conditional on prices and total destination expenditure. It does not by itself determine E_{dt}^{T}.

3.C.2 Local Income and Balanced Trade

Let I_{dt} denote total income available to residents of destination location d. With Cobb–Douglas preferences over modeled tradables and the residual outside good, modeled tradable expenditure is

$$E_{dt}^{T} = \alpha I_{dt}. \tag{3.67}$$

Balanced trade closes the model by setting destination expenditure equal to income rather than treating it as an external demand shifter.

In the shipping-only benchmark used in the quantification, income may include exogenous outside-sector income. Let I_{dt}^{X} denote this outside income and let χ_{do} map income from modeled producing locations to destination expenditure nodes. Then

$$I_{dt} = I_{dt}^{X} + \sum_{o} \chi_{do} w_{ot} L_{ot}, \qquad E_{dt}^{T} = \alpha I_{dt}. \tag{3.68}$$

If the destination and origin sets are the same, χ_{do} is the identity matrix. If outside income is shut down as well, this reduces to the standard expression $E_{dt}^{T} = \alpha w_{dt} L_{dt}$.

When land is active in agricultural production, total income also includes payments to the fixed factor. Under constant returns and perfect competition, crop revenue in origin o and good k is

$$R_{ot}^{k} = \sum_{d} X_{odt}^{k}. \tag{3.69}$$

Labor receives share μ of crop revenue and land receives share $1-\mu$:

$$w_{ot} L_{ot}^{k} = \mu R_{ot}^{k}, \qquad r_{ot}^{k} \bar{H}_{o}^{k} = (1-\mu) R_{ot}^{k}. \tag{3.70}$$

If local residents own the fixed factor locally, total local agricultural income is the sum of labor and land income:

$$I_{ot}^{A} = \sum_{k} R_{ot}^{k}. \tag{3.71}$$

If land rents are rebated nationally or owned outside the location, the income closure must specify that ownership rule. The shipping-only benchmark avoids this additional incidence choice by setting $\mu = 1$.

3.C.3 Wage Fixed Point

The wage fixed point follows from combining expenditure shares with labor-market clearing. In the labor-only benchmark with outside income, origin revenue is

$$R_{ot} = \sum_{d,k} S^k_{odt} s^k_{dt} \alpha \left(I^X_{dt} + \sum_j \chi_{dj} w_{jt} L_{jt} \right). \tag{3.72}$$

Labor-market clearing requires

$$w_{ot} L_{ot} = R_{ot}. \tag{3.73}$$

Equations (3.72)–(3.73) define the wage system:

$$w_{ot} = \frac{1}{L_{ot}} \sum_{d,k} S^k_{odt}(\mathbf{w}) s^k_{dt}(\mathbf{w}) \alpha \left(I^X_{dt} + \sum_j \chi_{dj} w_{jt} L_{jt} \right), \tag{3.74}$$

where the shares depend on wages through farm-gate costs and delivered prices. In the no-outside-income version only relative wages matter, so one wage can be chosen as the numeraire. In the outside-good benchmark, the outside-good price is the numeraire and the Censo income layer pins the nominal wage scale.

With land active, labor-market clearing instead uses labor's revenue share:

$$w_{ot} L^T_{ot} = \mu \sum_{d,k} X^k_{odt}, \tag{3.75}$$

and the farm-gate cost depends on the equilibrium labor-to-land ratio,

$$c^k_{ot} = \frac{w_{ot}}{\mu A^k_{ot}} \left(\frac{L^k_{ot}}{\bar{H}^k_o} \right)^{1-\mu}. \tag{3.76}$$

This is why the richer model requires a subsidiary fixed point over farm-gate costs. The benchmark counterfactual sets $\mu = 1$, so this loop is shut down and the balanced-trade closure reduces to Eq. (3.74).

3.D Additional figures

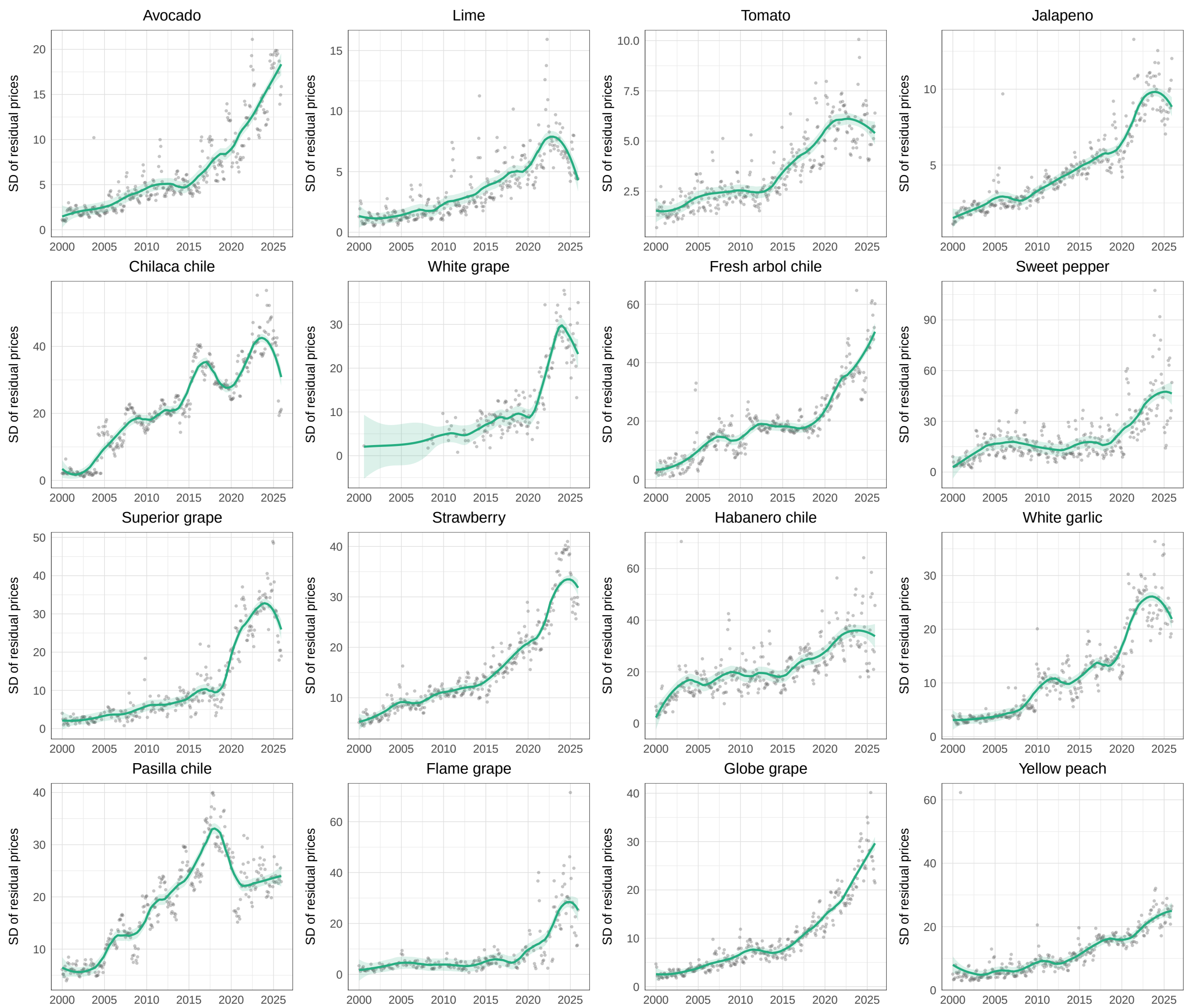


Figure 3.9: Cross-destination residual price dispersion over time

Notes: Each point is the product-month standard deviation of residual frequent wholesale prices across destination markets for selected products with rising dispersion over time. Residual prices remove product-specific calendar-month and year means and add back the product mean. Product-months require at least three destination markets and ten price observations; lines are local-polynomial smooths with pointwise confidence bands. Prices are nominal Mexican pesos per kilogram.

3.E Additional EAT Regression Tables

This appendix reports the firm-level EAT regression tables that are not shown in the main text. The first table keeps Serie 2013 but switches to additional firm-scale outcomes. The remaining tables use the pooled series sample, combining series 2008, 2013, and 2018, for the cost, activity, and additional outcome blocks.

Table 3.9: Firm-level EAT regressions: Series 2013, additional outcomes

	Log employment				Log total vehicles			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Transporter-robbery exposure	0.00000				-0.00004***			
	(0.00001)				(0.00001)			
Extortion exposure		0.00189				0.00658**		
		(0.00236)				(0.00270)		
Homicide exposure			-0.00478				0.00782	
			(0.05804)				(0.04413)	
Crime exposure				0.03041				0.01268
				(0.03809)				(0.02371)
Observations	735	759	764	764	774	799	804	804
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Firm IDs	294	305	305	305	303	314	314	314
Years	4	4	4	4	3	3	3	3

Notes: Each column is a separate firm-level regression with firm and year fixed effects; only the row for the exposure included in that column is populated. The displayed regressors are firm-year crime exposure measures built in the INEGI lab by merging each firm's reported top origin-destination revenue shares to OD-year route-crime exposure, then averaging over matched OD pairs and renormalizing over the matched share. OD-year route exposure is based on municipal crimes normalized by road kilometers along the relevant routes and expressed as a percent-change-from-base shift. The single-crime rows report transporter robbery, extortion, and homicide exposures; the bundled row reports the crime-bundle exposure available in the frozen lab return. Standard errors in parentheses are clustered at the firm-series level. All numeric coefficient and standard-error entries are reported to five decimals. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3.10: Firm-level EAT regressions: Pooled series 2008, 2013, and 2018, cost outcomes

	Log total costs				Log fuel expense				Log insurance expense			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Transporter-robbery exposure	-0.00003***				-0.00003***				-0.00003***			
	(0.00001)				(0.00001)				(0.00001)			
Extortion exposure		0.00309**				0.00609***				-0.00099		
		(0.00147)				(0.00101)				(0.00378)		
Homicide exposure			0.03537				0.07164*				-0.00158	
			(0.03060)				(0.03700)				(0.09432)	
Crime exposure				0.01099*				0.01034				-0.02625
				(0.00665)				(0.00864)				(0.01750)
Observations	2,272	2,360	2,368	2,368	1,985	2,059	2,066	2,066	2,048	2,122	2,130	2,130
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm IDs	988	1,024	1,025	1,025	801	829	830	830	917	949	950	950
Years	8	8	8	8	7	7	7	7	8	8	8	8

Notes: Each column is a separate firm-level regression with firm and year fixed effects; only the row for the exposure included in that column is populated. The displayed regressors are firm-year crime exposure measures built in the INEGI lab by merging each firm's reported top origin-destination revenue shares to OD-year route-crime exposure, then averaging over matched OD pairs and renormalizing over the matched share. OD-year route exposure is based on municipal crimes normalized by road kilometers along the relevant routes and expressed as a percent-change-from-base shift. The single-crime rows report transporter robbery, extortion, and homicide exposures; the bundled row reports the crime-bundle exposure available in the frozen lab return. Standard errors in parentheses are clustered at the firm-series level. All numeric coefficient and standard-error entries are reported to five decimals. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3.11: Firm-level EAT regressions: Pooled series 2008, 2013, and 2018, shipment and activity outcomes

	Log shipped value				Log total distance				Log total trips				Log total tons			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Transporter-robbery exposure	-0.00004				-0.00006**				-0.00002				-0.00001			
	(0.00017)				(0.00003)				(0.00003)				(0.00004)			
Extortion exposure		-0.01643**				0.00647***				0.00427				-0.01770***		
		(0.00824)				(0.00246)				(0.00267)				(0.00538)		
Homicide exposure			-0.41002				0.06225				-0.00043				0.03074	
			(0.26857)				(0.05800)				(0.06206)				(0.09736)	
Crime exposure				-0.07878				0.01374				0.02361				0.04172**
				(0.05062)				(0.01249)				(0.01555)				(0.01918)
Observations	1,314	1,363	1,369	1,369	2,272	2,360	2,368	2,368	2,272	2,360	2,368	2,368	2,272	2,360	2,368	2,368
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm IDs	631	652	654	654	988	1,024	1,025	1,025	988	1,024	1,025	1,025	988	1,024	1,025	1,025
Years	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8

Notes: Each column is a separate firm-level regression with firm and year fixed effects; only the row for the exposure included in that column is populated. The displayed regressors are firm-year crime exposure measures built in the INEGI lab by merging each firm's reported top origin-destination revenue shares to OD-year route-crime exposure, then averaging over matched OD pairs and renormalizing over the matched share. OD-year route exposure is based on municipal crimes normalized by road kilometers along the relevant routes and expressed as a percent-change-from-base shift. The single-crime rows report transporter robbery, extortion, and homicide exposures; the bundled row reports the crime-bundle exposure available in the frozen lab return. Standard errors in parentheses are clustered at the firm-series level. All numeric coefficient and standard-error entries are reported to five decimals. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3.12: Firm-level EAT regressions: Pooled series 2008, 2013, and 2018, additional outcomes

	Log employment				Log total vehicles			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Transporter-robbery exposure	-0.00001				0.00006***			
	(0.00000)				(0.00002)			
Extortion exposure		0.00093				0.00468**		
		(0.00193)				(0.00216)		
Homicide exposure			-0.02872				-0.04355	
			(0.04942)				(0.03959)	
Crime exposure				0.01120				-0.00319
				(0.01327)				(0.01008)
Observations	2,138	2,222	2,230	2,230	2,271	2,359	2,367	2,367
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Firm IDs	939	974	975	975	988	1,024	1,025	1,025
Years	8	8	8	8	8	8	8	8

Notes: Each column is a separate firm-level regression with firm and year fixed effects; only the row for the exposure included in that column is populated. The displayed regressors are firm-year crime exposure measures built in the INEGI lab by merging each firm's reported top origin-destination revenue shares to OD-year route-crime exposure, then averaging over matched OD pairs and renormalizing over the matched share. OD-year route exposure is based on municipal crimes normalized by road kilometers along the relevant routes and expressed as a percent-change-from-base shift. The single-crime rows report transporter robbery, extortion, and homicide exposures; the bundled row reports the crime-bundle exposure available in the frozen lab return. Standard errors in parentheses are clustered at the firm-series level. All numeric coefficient and standard-error entries are reported to five decimals. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Bibliography

Allen, Treb, David Atkin, Santiago Cantillo Cleves, and Carlos Eduardo Hernandez, "The Traveling Trucker Problem," *AEA Papers and Proceedings*, 2024, *114*, 334–339.

Anderson, James E. and Eric van Wincoop, "Trade Costs," *Journal of Economic Literature*, 2004, *42* (3), 691–751.

Asturias, Jose, Manuel Garcia-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2019, *17* (6), 1881–1940.

Atkin, David and Amit K. Khandelwal, "How Distortions Alter the Impacts of International Trade in Developing Countries," *Annual Review of Economics*, 2020, *12* (1), 213–238.

_ **and Dave Donaldson**, "The Role of Trade in Economic Development," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics*, Vol. 5, Elsevier, 2022, pp. 1–59.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "A Practical Guide to Shift-Share Instruments," *Journal of Economic Perspectives*, 2025, *39* (1), 181–204.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou, "Geography, Transportation, and Endogenous Trade Costs," *Econometrica*, 2020, *88* (2), 657–691.

Brown, Zach Y., Eduardo Montero, Carlos Schmidt-Padilla, and Maria Micaela Sviatschi, "Market Structure and Extortion: Evidence from 50,000 Extortion Payments," *Review of Economic Studies*, 2025, *92* (3), 1595–1624.

Dell, Melissa, "Trafficking Networks and the Mexican Drug War," *American Economic Review*, 2015, *105* (6), 1738–1779.

_ **, Benjamin Feigenberg, and Kensuke Teshima**, "The Violent Consequences of Trade-Induced Worker Displacement in Mexico," *American Economic Review: Insights*, 2019, *1* (1), 43–58.

Dix-Carneiro, Rafael, Rodrigo R. Soares, and Gabriel Ulyssea, "Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization," *American Economic Journal: Applied Economics*, 2018, *10* (4), 158–195.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Gollin, Douglas and Richard Rogerson, "Productivity, Transport Costs and Subsistence Agriculture," *Journal of Development Economics*, 2014, *107*, 38–48.

Graff, Tilman, "Spatial Inefficiencies in Africa's Trade Network," *Journal of Development Economics*, 2024, *171*, 103319.

Instituto Nacional de Estadistica y Geografia, "Encuesta Nacional de Victimizacion de Empresas 2024," 2024. ENVE 2024.

Khanna, Gaurav, Carlos Alberto Medina-Durango, Anant Nyshadham, Daniel Ramos-Menchelli, Jorge Andres Tamayo-Castano, and Audrey Tiew, "Spatial Mobility, Economic Opportunity, and Crime," 2025. Working paper.

Olken, Benjamin A. and Patrick Barron, "The Simple Economics of Extortion: Evidence from Trucking in Aceh," *Journal of Political Economy*, 2009, *117* (3), 417–452.

Sequeira, Sandra and Simeon Djankov, "Corruption and Firm Behavior: Evidence from African Ports," *Journal of International Economics*, 2014, *94* (2), 277–294.

Sotelo, Sebastian, "Domestic Trade Frictions and Agriculture," *Journal of Political Economy*, 2020, *128* (7), 2690–2738.

Chapter Acknowledgement. Chapter 3, in full, is coauthored with Alfonso Cebreros and Daniel Ramos-Menchelli. The dissertation author was the primary author of this chapter.