Optimal Transit Networks in Developing Countries*

Leticia Juarez† IDB
Jordán Mosqueda‡ UCSD
Román Zárate§ UCSD

March 2026

Abstract

A key allocation problem for budget-constrained governments in developing cities is where and how much to build public mass transit infrastructure, relative to allowing privatized and often informal transit operators to operate. Mass transit technologies are fixed-cost intensive but involve little marginal cost, while minibuses have higher marginal costs, entail externalities, but require little upfront investment. In this paper we study the optimal transit network problem in general spatial equilibrium, where a planner can choose between technologies to connect locations within a city. We show that the optimal public-private transit mix on any given edge of the network depends crucially on the trade-off between relative marginal and fixed costs across technologies. Leveraging a unique dataset of public and private transit networks in Mexico City and a battery of wages, land use, and commuting flows microdata, we quantitatively study the gains from budget-feasible expansions of the transit system. We find that increasing the infrastructure budget by 50% raises welfare by 4.4% under public-only expansion and by 5.2% when both public and private transit can expand optimally, with the additional 18% welfare gain driven by private transit. In both cases, optimal investments primarily improve connections from peripheral areas to productive outer nodes and towards the existing network. These results suggest that developing cities could benefit substantially by jointly designing public and private networks, rather than public mass transit in isolation.

*We are deeply grateful to Santiago Veleda for outstanding research assistance. We also acknowledge and appreciate help from Claude Code and Codex.

†Inter-American Development Bank, email: leticiaj@umich.edu
‡University of California San Diego, email: jmosqued@ucsd.edu
§University of California San Diego, email: rzaratevasquez@ucsd.edu

1 INTRODUCTION

Urban transit systems in developing cities face a fundamental allocation problem: how much and where should scarce public resources be allocated to publicly-provided mass transit relative to privatized and 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 the city structure, welfare and commuting patterns by shaping congestion, environmental externalities, and the overall cost to commute 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 should a planner 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. On the other hand, private transit has higher marginal operating costs, 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—precluding 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 the actual private routes is scarce throughout the developing world. In the majority of cities service is informal, therefore trips geographies' and service characteristics are rarely centrally collected.

In this paper we address these challenges by 1) providing a general spatial equilibrium framework that delivers the planner's optimal public and private networks, and by 2) quantifying it 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 should a planner 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.

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 done 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.

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) that was collected by a private firm "Where Is My Transport". This datasets provides a unified

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

representation of the city's full transit system, including both formal public mass transit — metro, BRT, and suburban rail — and private and 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. These sources 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 operation cost parameters to the equilibrium where the population distribution and the network are fixed to the observed one. We identify higher fixed investment costs to build public transit relative to private, 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 current 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 4.4% welfare gains, but the welfare gains under joint public-private planning are approximately 18% larger, delivering 5.2% of welfare gains. 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 in the outskirts of the city. Also, we show that new investments attempt to connect peripheral nodes between them and towards the existing transit infrastructure, 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 by optimal planner choices and how its interaction with public transit shapes the city 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 Model

We build on Fajgelbaum and Schaal (2020) and extend their optimal road network and trade between locations framework to a city setting with commuting and transit networks. The key addition is that an edge in the graph has a characterization of public and/or private, s_{jk}. That is, given an allocation of residents, workers, and flows across locations, the planner decides whether to build a public edge (e.g., subway), build a link where private operators will operate, or both.

Environment. The economy consists of a discrete set of locations $\mathcal{J} = \{1, ..., J\}$ 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, by virtue of commuting. The mass of workers living and working in a given location j is denoted by L_j^R and L_j^L, respectively. Commuting of workers between locations generates a commuting flow L_{jk} across any node pair jk. 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 $k' \notin \mathcal{N}(j)$, they must transit through a sequence of connected locations.

Transport technology. In Fajgelbaum and Schaal (2020), each edge is thought of as a road. The fundamental difference in our framework is that edges can be of different types or technologies: public, private, or both. We define s_{jk} to be the share of public infrastructure—relative to private—in link jk. Let I_{jk}^{pub} and I_{jk}^{pri} denote investment in public and private infrastructure on link jk, respectively. The technology share is thus

$$s_{jk} = \frac{I_{jk}^{\text{pub}}}{I_{jk}^{\text{pub}} + I_{jk}^{\text{pri}}}.$$

Transporting a worker from j to k requires τ_{jk} units of labor, so $(1 + \tau_{jk})$ corresponds to the iceberg cost of transportation, which captures the fact that commuting implies a real resource loss as captured by time, fatigue, mortality, or productivity loss. We assume that the cost τ_{jk} takes a tractable log-linear form

$$\tau_{jk} = \delta_{jk}^{\tau} \left[(1 - s_{jk})\, \theta_{\text{pri}} + s_{jk}\, \theta_{\text{pub}} \right] L_{jk}^{\beta}\, I_{jk}^{-\gamma}, \tag{1}$$

where L_{jk} represents commuting flows, $I_{jk} \equiv I_{jk}^{\text{pub}} + I_{jk}^{\text{pri}}$ is total infrastructure, δ_{jk}^{τ} captures geographic frictions that affect per-unit transport costs and θ_{pri} and θ_{pub} are technology-specific shift operating costs that capture , with $\theta_{\text{pub}} < \theta_{\text{pri}}$ reflecting the lower marginal cost of public transit. Transportation costs decline with investment and increase with flows, reflecting congestion externalities. The parameter β is the elasticity of congestion, and controls how costs increase with more flows; while the parameter γ controls how costs decrease with investment.

We assume that building a link requires investment in infrastructure: usage of a resource K that is in fixed aggregate supply in the economy. The cost of building infrastructure varies across links and technologies, reflecting geographic heterogeneity. Building public infrastructure I_{jk}^{pub} on link jk requires $\delta_{jk}^{I,\text{pub}}\, I_{jk}^{\text{pub}}$ units of K, while building private infrastructure I_{jk}^{pri} requires $\delta_{jk}^{I,\text{pri}}\, I_{jk}^{\text{pri}}$ 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:

$$\sum_{(j,k) \in \mathcal{E}} \left[\delta_{jk}^{I,\text{pub}}\, I_{jk}^{\text{pub}} + \delta_{jk}^{I,\text{pri}}\, I_{jk}^{\text{pri}} \right] \leq K \tag{2}$$

Flow constraint. At each location j, effective labor usage must equal effective labor availability:

$$\underbrace{L_j^L}_{\text{Local workers}} + \underbrace{\sum_{k \in N(j)} (1+\tau_{jk}) L_{jk}}_{\text{Exports of workers}} \leq \underbrace{L_j^R}_{\text{Endowment of workers}} + \underbrace{\sum_{i \in N(j)} L_{ij}}_{\text{Imports of workers}} . \tag{3}$$

The right-hand side of the inequality is the total amount of agents available in a given location, aggregating across local residents and imported workers, i.e., inflowing commuters from neighboring locations. The left-hand side comprises local workers plus workers that are exported to neighboring locations, i.e., outflowing commuters, inclusive of the commuting cost as commuting is subject to iceberg losses.

Preferences. Workers derive utility from consumption of a homogenous 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 produced the final good with a Cobb-Douglas technology combining productivity, labor, and commercial land. That is,

$$C_j = A_j {L_j^L}^{\psi} H_j^{c\,1-\psi}$$

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.

2.1 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

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

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—equals demand—local workers plus outgoing commuters net of commuting costs:

$$\underbrace{L_j^L}_{\text{Local workers}} + \underbrace{\sum_{k\in N(j)} (1+\tau_{jk})\, L_{jk}}_{\text{Exports of workers}} = \underbrace{L_j^R}_{\text{Endowment of workers}} + \underbrace{\sum_{i\in N(j)} L_{ij}}_{\text{Imports of workers}} .$$

Commuting costs. The iceberg cost on edge (j,k) depends on the modal composition of infrastructure, commuter volume, and total capacity. Letting $I_{jk} \equiv I_{jk}^{\text{pub}} + I_{jk}^{\text{pri}}$ and $s_{jk} \equiv I_{jk}^{\text{pub}}/I_{jk}$ denote total investment and the public share, respectively,

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

Market clearing. Aggregate population is fixed at $\bar{L}$, final-good consumption cannot exceed output, and housing consumption at each location is bounded by the local land endowment:

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

Production technology. Output at each location is linear in effective labor:

$$C_j = A_j\, L_j^L.$$

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

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

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

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

$$c_j,\ h_j,\ L_j^R,\ L_j^L \ge 0 \quad \forall j \in \mathcal{J}, \qquad L_{jk} \ge 0 \quad \forall (j,k) \in \mathcal{E}.$$

2.2 Optimal commuting flows

The problem of the planner 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. Analogous to (Fajgelbaum and Schaal, 2020), this is an optimal transport problem that involves choosing how workers move from residence to workplace following the least-cost route. The solution of the problem consists on finding a Lagrange multiplier for each location pair and then expressing flows as a function of the difference between the multipliers across locations. From the first-order conditions of the problem, as shown in Appendix B.1, we can show that the optimal flows in each link are given by

$$L_{jk} = \left[\frac{1}{1+\beta} \frac{I_{jk}^{\gamma}}{\delta_{jk}^{\tau} \left[(1-s_{jk})\, \theta_{\text{pri}} + s_{jk}\, \theta_{\text{pub}} \right]} \frac{\max\{P_k^F - P_j^F, 0\}}{P_j^F + P^L} \right]^{\frac{1}{\beta}}, \tag{4}$$

which mimics the result found by Fajgelbaum and Schaal (2020), where commuting flows depend positively on infrastructure, negatively on congestion β and commuting costs, and are a function of the Lagrange multipliers P. In particular, they are a function of the balanced-flows constraint multiplier difference between two neighboring nodes, $P_k^F - P_j^F$. This suggests that the allocation of flows in a given link should reflect society's relative valuation of allocating workers to some node relative to another, relative to the effective shadow value of a marginal resident in node j, $P_j^F + P^L$. Further, we can only obtain a positive flow L_{jk} when the "destination potential" is high enough relative to the origin after accounting for commuting costs. In other words, the multiplier difference is the marginal welfare gain—expressed in terms of the planner's numeraire—from shifting an additional commuter from j towards workplace k, holding everything else fixed.

2.3 Optimal transit network

Given consumption and flows allocation, the planner then chooses the optimal investment of both public and private infrastructure in each link. If P_K is the multiplier of the network-building constraint (3), the first-order condition with respect to public infrastructure is

$$[I_{jk}^{\text{pub}}] \qquad \underbrace{(P_j^F + P^L)\, L_{jk} \left(-\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pub}}} \right)}_{\text{Marginal benefit of public infr.}} = \underbrace{P^K\, \delta_{jk}^{I,\text{pub}}}_{\text{Marginal cost of building}} \qquad \forall (j,k).^{2} \tag{5}$$

This condition states that the planner invests in public infrastructure until the marginal benefit (left-hand side) equals the marginal cost (right-hand side). The marginal benefit consists of the marginal reduction in commuting costs when building more public infrastructure, and the marginal cost consists of the marginal public-specific building cost.

By replacing the value of $\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pub}}}$ we can further decompose the marginal benefit into two com-

[2] For exposition purposes, this equation corresponds to an interior solution so the complementary slackness multipliers are zero.

ponents:

$$-\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pub}}} = \frac{\tau_{jk}}{I_{jk}} \left[\underbrace{\gamma}_{\text{Direct effect}} - \underbrace{\frac{(\theta_{\text{pub}} - \theta_{\text{pri}})\,(1 - s_{jk})}{\Theta(s_{jk})}}_{\text{Composition effect}} \right], \tag{6}$$

a (i) direct infrastructure effect that captures the reduction of congestion due to more total capacity, and a (ii) composition effect, which captures that shifting toward public technology reduces operating costs whenever the marginal cost of operating public transit is lower than that of the private technology, i.e., $\theta_{\text{pub}} < \theta_{\text{pri}}$.

The analogous insights hold with respect to private transit investment. The planner will choose to build private transit infrastructure until the marginal cost equals the benefit:

$$[I_{jk}^{\text{pri}}] \qquad \underbrace{(P_j^F + P^L)\, L_{jk} \left(-\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pri}}} \right)}_{\text{Marginal benefit of private infr.}} \quad = \quad \underbrace{P^K\, \delta_{jk}^{I,\text{pri}}}_{\text{Marginal cost of building}} \qquad \forall (j,k), \tag{7}$$

where the marginal benefit can be decomposed into a direct and a composition effect, via the impact of building infrastructure on commuting costs:

$$-\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pri}}} = \frac{\tau_{jk}}{I_{jk}} \left[\underbrace{\gamma}_{\text{Direct effect}} + \underbrace{\frac{(\theta_{\text{pub}} - \theta_{\text{pri}})\, s_{jk}}{\Theta(s_{jk})}}_{\text{Composition effect}} \right]. \tag{8}$$

Note, however, that whenever the marginal cost of operating public transit is lower than that of the private technology, i.e., $\theta_{\text{pub}} < \theta_{\text{pri}}$, the composition effect now works against the infrastructure effect: adding private infrastructure increases total capacity but shifts the technology mix toward the higher-cost mode.

Optimal technology mix. We now characterize the optimal mix of infrastructure across public and private technologies. Combining the first-order conditions for public and private infrastructure, we summarize the optimal mix of public-private investments along with some implications in the following proposition.

Proposition 1 (Optimal Public Transit Share). Consider the planner's problem defined in Section 2.1. Consider a link $(j,k) \in \mathcal{E}$ with strictly positive investment in both technologies ($I_{jk}^{\text{pub}} > 0$ and $I_{jk}^{\text{pri}} > 0$). Let $\Delta\theta \equiv \theta_{\text{pub}} - \theta_{\text{pri}}$ and define the *relative infrastructure cost* on link jk as:

$$\rho_{jk} \;\equiv\; \frac{\delta_{jk}^{I,\text{pub}}}{\delta_{jk}^{I,\text{pri}}}.$$

Then, the optimal public transit share in link jk is given by:

$$s_{jk}^* = \frac{\Delta\theta - \gamma\,\theta_{\text{pri}}(1-\rho_{jk})}{\Delta\theta\,(\gamma+1)(1-\rho_{jk})}. \tag{9}$$

Furthermore, the following relationships hold:

- $\partial s_{jk}^*/\partial \rho_{jk} > 0$: Higher relative cost of public infrastructure *increases* the optimal public share.
- $\partial s_{jk}^*/\partial(\theta_{\text{pri}} - \theta_{\text{pub}}) > 0$ when $\rho_{jk} > 1$: A larger operating cost advantage of public transit increases the optimal public share whenever public infrastructure is more expensive to build, which is the empirically relevant case.
- $\partial s_{jk}^*/\partial \gamma > 0$ when $\rho_{jk} < 1$: When public infrastructure is relatively cheaper to build, higher infrastructure elasticity favors public transit. The opposite holds when $\rho_{jk} > 1$.[3]

Proof. See Appendix B.2 for the proof. □

Proposition 1 shows that the optimal share of public infrastructure s_{jk}^* depends exclusively on relative technology-specific cost and construction parameters, and the elasticity of commuting costs to infrastructure investments—all parameters. Importantly, it shows that in links where public infrastructure is relatively more expensive to build (higher ρ_{jk}) will optimally have a higher public transit share. Intuitively, when public infrastructure becomes more expensive to build, total investment falls, making public's operating cost advantage relatively more valuable. This generates spatial variation in the technology mix: corridors with favorable geology for tunneling, for example, will see more metro, while areas where surface construction is relatively easier will rely more on private transit. Furthermore, it shows that whenever the marginal cost of operating public transit is smaller than that of the private technology, the optimal share of public transit will be higher, reflecting, for example, that the cost per passenger of an electric metro wagon is substantially smaller than the per-passenger cost of a fossil-fuel-based minibus.

Note that the analytical formula of the optimal mix characterizes interior solutions. Corner solutions ($s = 0$ or $s = 1$) occur when flows are sufficiently low or high, respectively. The threshold flows depend on the shadow prices and can be characterized by the complementary slackness conditions.

Optimal infrastructure investment. Having characterized the optimal mix of technologies in each link, we can recover the optimal level of total infrastructure, and technology-specific investments. Plugging the optimal public-private share from equation (9) into either FOCs (18) or (21) and rearranging, we obtain that total investment in link jk is given by

$$I_{jk}^* = \left[\frac{(P_j^F + P^L)\,\delta_{jk}^{\tau}\,L_{jk}^{\beta+1}}{P^K} \left(\frac{\theta_{\text{pub}} - \theta_{\text{pri}}}{\delta_{jk}^{I,\text{pri}} - \delta_{jk}^{I,\text{pub}}} \right) \right]^{\frac{1}{\gamma+1}}. \tag{10}$$

[3]Derivations can be found in Appendix B.2.

Technology-specific investments are thus

$$I_{jk}^{\text{pub}\,*} = s_{jk}^{*}\, I_{jk}^{*}, \qquad I_{jk}^{\text{pri}\,*} = (1 - s_{jk}^{*})\, I_{jk}^{*}. \tag{11}$$

The optimal infrastructure therefore depends on the commuting flow, congestion and infrastructure elasticities, geography, cost and investment parameters, and constraint Lagrange multipliers. In particular, relative to the optimal infrastructure in Fajgelbaum and Schaal (2020), our characterization adds the explicit trade-off of technologies as captured by the second multiplicative term in equation (10): the optimal investment in each link depends on the trade-off between relative fixed and marginal costs across technologies.

3 Illustrative example on a stylized city

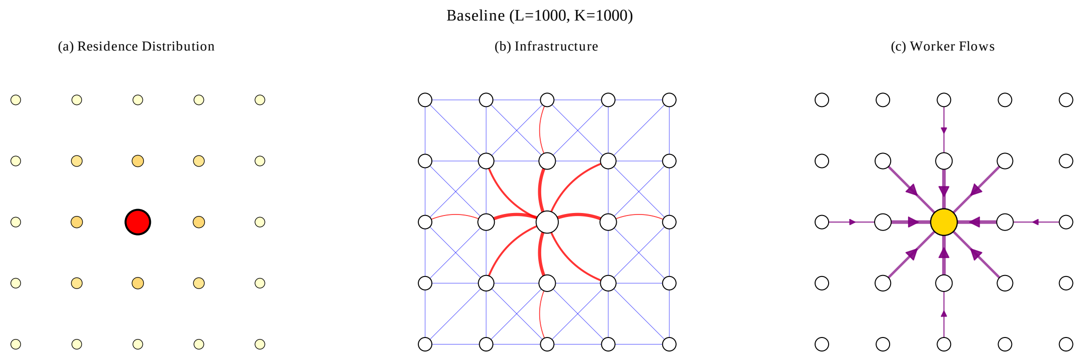
To build intuition and understand the key insights from the model, consider a stylized city that consists of a simple grid of 5×5 locations arranged on a 2-dimensional plane. In this city all nodes are equally productive but the center node, which is twice as productive. All locations have the same endowment of the non-tradable good—land. The size of the city is of $\bar{L} = 1{,}000$ labor units. Locations are connected via a network where each node has up to 8 neighbors (horizontal, vertical, and diagonal connections). The planner allocates a budget K across edges, choosing both the level of infrastructure investment and the share devoted to public versus private transit on each link. The baseline parametrization in this economy is

- Elasticities: congestion $\beta = 0.8$ and infrastructure $\gamma = 0.4$
- Marginal cost parameters: $\theta_{\text{pub}} = 0.1$, $\theta_{\text{pri}} = 0.5$
- Fixed/investment cost parameters: $\delta_{jk}^{I,\text{pub}} = 0.5\,\delta_{jk}$, $\quad \delta_{jk}^{I,\text{pri}} = 0.1\delta_{jk}$
- Budget: $K = 1{,}000$
- Commuting cost shifter: $\delta_{jk}^{\tau} = \delta_{jk}$
- Consumption share: $\alpha = 0.7$

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

Figure 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 in and around the center, as shown in panel (a), reflecting the trade-off between commuting costs and access to higher productivity. The planner responds endogenously by directing infrastructure investment toward links that connect peripheral locations to the center (panel (b)), effectively reducing commuting frictions along the most heavily used corridors. As a result, commuting flows exhibit a pronounced radial pattern toward the central node (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 1. EQUILIBRIUM OUTPUT: BASELINE EXAMPLE



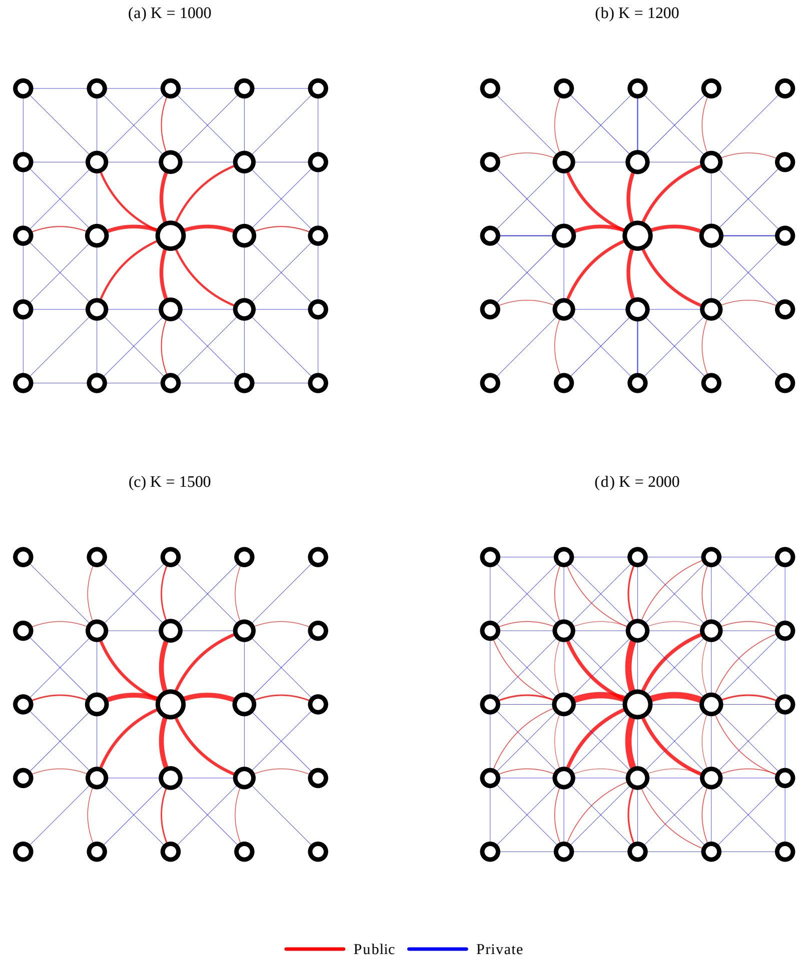
Note: The figure illustrates the baseline equilibrium in a stylized 5×5 spatial economy. Nodes represent locations and edges denote commuting links between neighboring locations (horizontal, vertical, and diagonal). In panel (a), node size and color intensity capture the equilibrium residential distribution, 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 $K = 1000$

3.1 Comparative Statistics

In this subsection, we are interested in understanding how the network is designed whenever i) we increase the budget of the planner, ii) we change the marginal cost parameters or the iii) 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 budget of the planner? In particular, we multiply the baseline budget by 1.2, 1.5, and 2 times the initial budget. Figure 2 shows the optimal transit network for different values of K. The baseline case is shown in panel (a). First, private infrastructure (blue lines) forms a thin uniform network across the entire city and provides baseline connectivity but does not intensify with additional resources. In the baseline case with fewer resources, nodes are connected diagonally exclusively, incentivizing commuting from the outskirts to the center but not across peripheral locations. Second, public infrastructure (red curves) is concentrated on edges directly connecting to the CBD, reflecting the high commuting demand toward the productive center. Third, as K increases (panels (b), (c), (d)), public investment expands in the intensive margin, with thicker lines on existing CBD-adjacent edges indicating greater capacity. The planner prioritizes public transit near the CBD because its lower operating cost ($\theta_{\text{pub}} < \theta_{\text{pri}}$) makes it more efficient for high-volume corridors, despite its higher construction cost ($\delta^I_{\text{pub}} > \delta^I_{\text{pri}}$). Only the highest-demand edges justify the fixed cost, and as resources grow the planner focuses on greater infrastructure rather than expanding the network—the private technology has very low investment but enough to connect the rest of the city.

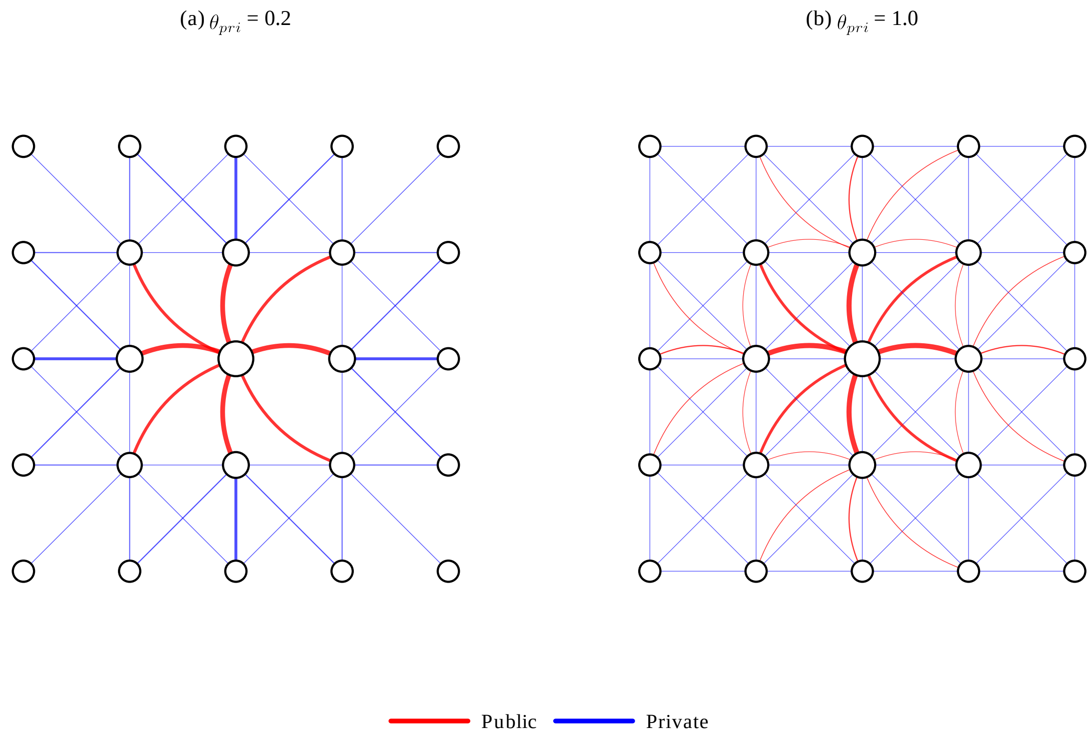
FIGURE 2. OPTIMAL NETWORK FOR DIFFERENT BUDGETS K



Note: The figure shows the optimal infrastructure allocation for a 5×5 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), $\theta_{\text{pub}} = 0.1$, $\theta_{\text{pri}} = 0.5$, $\delta^I_{\text{pub}} = 0.5\delta$, $\delta^I_{\text{pri}} = 0.1\delta$.

Increasing marginal costs θ. How does the network change when we make the private technology relatively cheaper to operate as compared to relatively more expensive? Relative to the baseline scenario where $\theta_{\text{pri}} = 5$, here we consider a case where it goes from 2 to 10. Figure 3 shows the networks under these parameters in panel (a) and (b), respectively. When private transit is cheap to operate ($\theta_{\text{pri}} = 2$, panel a), the planner concentrates private infrastructure on the high-demand edges connecting directly to the CBD—these thick blue lines serve the heaviest commuting corridors where private provision is cost-effective. Public transit, meanwhile, spreads across the entire network as a thin web connecting all locations. The logic reverses when private transit becomes expensive ($\theta_{\text{pri}} = 10$, panel b): private infrastructure thins out into a uniform, low-capacity network covering all edges, while public transit increases slightly to serve the CBD-adjacent corridors where its lower operating cost ($\theta_{\text{pub}} = 1$) justifies the much higher construction expense—since it is five times more expensive than private infrastructure. In both cases, the planner exploits comparative advantage: the technology with lower operating cost handles the high-volume routes, while the other provides baseline connectivity across the city. This illustrates a key insight: the optimal technology mix depends not on absolute costs but on the relative efficiency of public versus private provision.

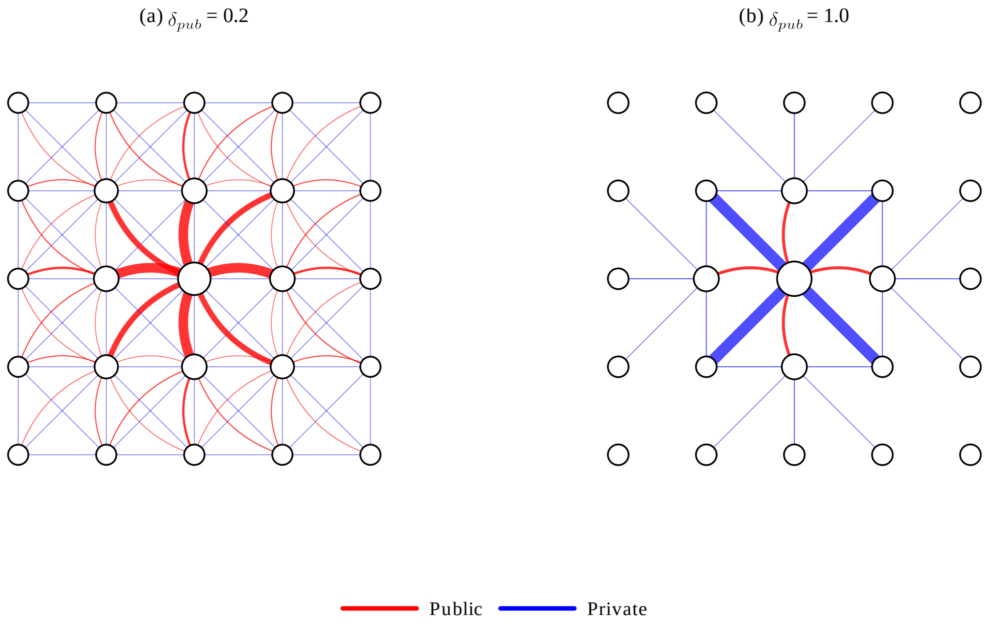
FIGURE 3. OPTIMAL NETWORK FOR DIFFERENT MARGINAL COSTS



Note: The figure shows the optimal infrastructure allocation for a 5×5 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), $\delta^I_{\text{pub}} = 0.5\delta$, $\delta^I_{\text{pri}} = 0.1\delta$.

Increasing building costs δ. How does the network change when we make the public technology relatively cheaper to build as compared to relatively more expensive? Relative to the baseline scenario where $\delta^I_{\text{pub}} = 5\delta$, here we consider cases where it ranges from 2δ to 10δ, holding $\delta^I_{\text{pri}} = \delta$ fixed. Figure 4 shows the networks under these parameters in panels (a) and (b), respectively. When public infrastructure is cheap to build ($\delta_{\text{pub}} = 2$, panel a), the planner invests heavily in public transit on the CBD-adjacent edges—thick red curves connecting the center to its immediate neighbors. Private infrastructure forms a thin, uniform network across the entire city, providing baseline connectivity. When public infrastructure becomes expensive ($\delta_{\text{pub}} = 10$, panel b), the pattern shifts dramatically: public transit now spreads as a thin web across all edges (since its lower operating cost $\theta_{\text{pub}} < \theta_{\text{pri}}$ still justifies some presence everywhere), while private infrastructure concentrates on the high-demand CBD corridors with thick blue lines. The key insight is that construction costs (δ) and operating costs (θ) interact to determine the optimal technology mix: the technology that is cheap to build gets deployed on high-capacity routes where fixed costs can be amortized over heavy traffic, while the technology that is cheap to operate spreads across low-volume edges where per-trip efficiency matters more than upfront investment.

FIGURE 4. OPTIMAL NETWORK FOR DIFFERENT BUILDING COSTS



Note: The figure shows the optimal infrastructure allocation for a 5×5 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), $\theta_{\text{pub}} = 0.1$, $\theta_{\text{pri}} = 0.5$.

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.

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 coexistence 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 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 6, we describe the data sources that underlie our analysis.

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.

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 data at the tract, block, and residential/commercial unit levels.

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.

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

TABLE 1—COMPARISON BETWEEN PRIVATE AND PUBLIC TRANSPORTATION IN USAGE AND KM COVERED

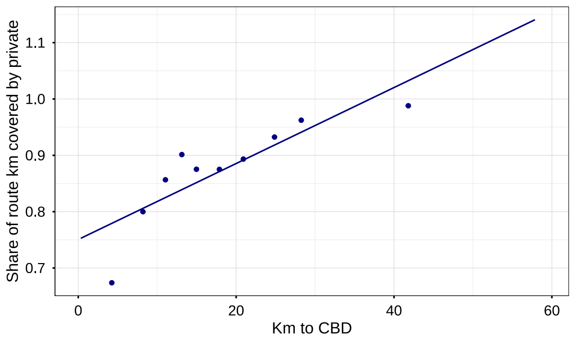


FIGURE 5. DISTANCE TO THE CBD AND SHARE OF KM COVERED BY PRIVATE TRANSPORTATION

Table 1 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 (Appendix Figure 11). This coverage is spatially uneven: Appendix Figure 9 shows that formal public infrastructure—Metro, BRT, and commuter rail—is concentrated within about 15 km of the CBD, while private operators cover nearly all route-kilometers in peripheral districts, with many at or above 90 percent private coverage. 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 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. This gradient is confirmed in Appendix Figure 10, which breaks down total route-kilometers by distance band: private transit accounts for 68 percent of route-kilometers in the innermost ring (0–5 km), rising steadily to 100 percent beyond 35 km. This pattern is consistent with the model's prediction that high fixed costs of public transit construction lead planners to concentrate public investment in high-density central corridors, leaving peripheral areas to private provision (Mosqueda, 2026).

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.

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

TABLE 2—COMPLEMENTARITY OF PUBLIC AND PRIVATE TRANSIT

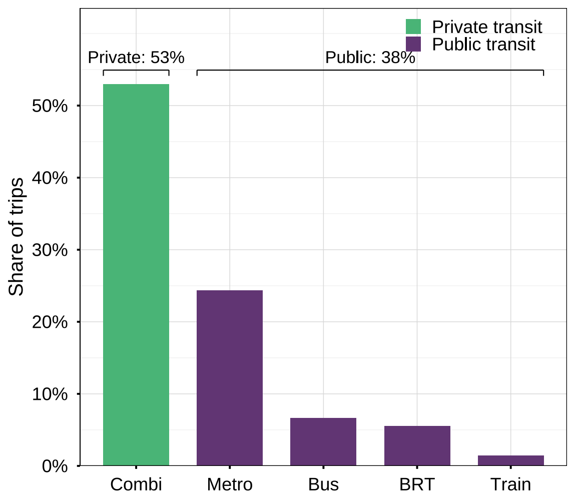


FIGURE 6. SHARE OF TRIPS BY TRANSPORTATION MODE

Table 2 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 (see also Appendix Figure 12, which shows the full breakdown of commutes into private-only, mixed, and public-only trips). 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. The structure of these multimodal trips is informative: Appendix Figure 13 breaks each journey into access, trunk, and egress stages, showing that the most common pattern involves private access, public trunk, and private egress—consistent with combis playing a first-and-last-mile role. This is also visible at the mode-to-mode level: Appendix Figure 14 shows that combi-to-metro transitions account for 49.6 percent of all consecutive mode pairs within multimodal trips, confirming that private transit mainly serves as a feeder into public trunk infrastructure.

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. Appendix Figure 16 confirms that even on trips using public infrastructure, private transit takes up a large share of total travel time. Mixed trips are also the longest and most complex—averaging 96 minutes and 3.5 segments—while private-only trips are shorter (57 minutes) but the most expensive per trip (20.9 MXN), reflecting higher per-kilometer costs of combi service (Appendix Figure 17). These patterns are consistent with commuters combining modes to take advantage of the speed of public trunk services while relying on private transit for spatial coverage.

This first-and-last-mile role of private transit is also confirmed at the infrastructure level. Some 81 percent of metro stations have at least one combi stop within 100 meters, and 64 percent have two or more, with an average of 2.44 combi stops per metro station nearby (Appendix Figure 15). This co-location reflects how private feeder services have grown around public infrastructure. Figure 6 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.

These patterns have two important implications for the model. First, they support the assumption that commuting costs along any given edge depend on the mix of technologies present, since individual journeys routinely involve both. 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 4.2) and plays a structural first-and-last-mile role in multimodal commuting (Fact 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.

5 Calibration

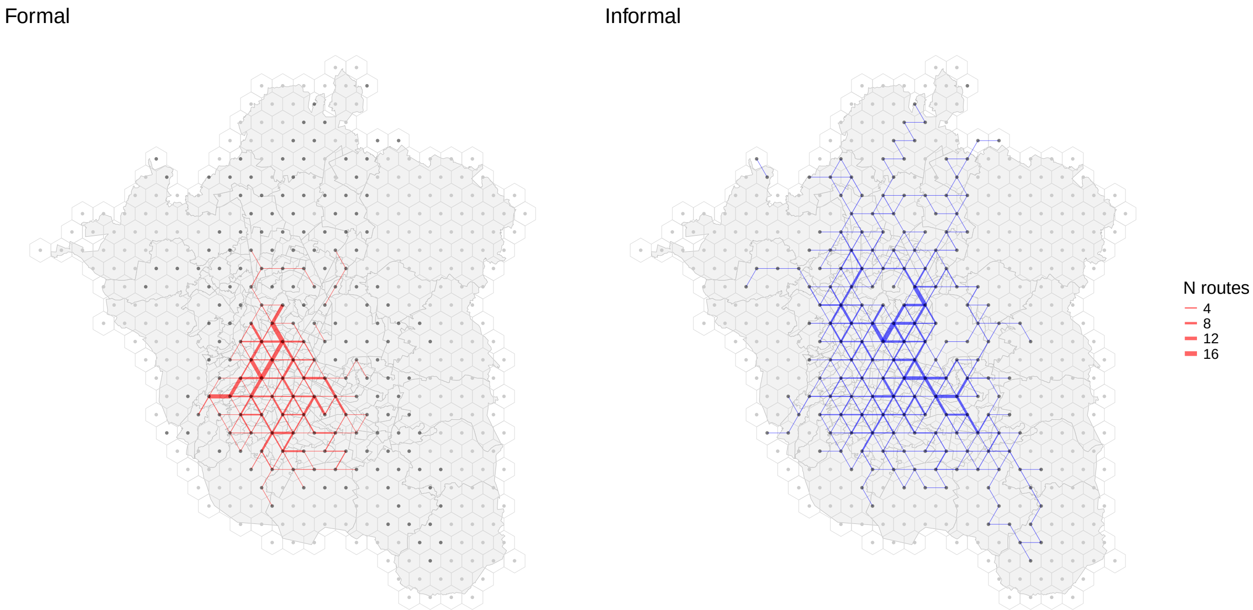
5.1 Building a simplified network

Transit systems are more complex to represent than standard road networks. GTFS data contain many overlapping routes for multimodal 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 7 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 7. 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. Neighbor nodes are defined to be connected to a given node if there is overlap of a transit line passing between them. Thickness of the lines represents the number of transit routes that connect two given nodes. Network data comes from WhereIsMyTransport's GTFS file.

5.2 Parameter calibration

Marginal costs of operation θ. To calibrate the marginal costs 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 their 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 of speed of public and private lines. We normalize $\theta_{\text{pub}} = 1$ given that 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 $\theta_{\text{pub}} = 1.64$. This suggests that minibuses are 64% more costly to operate relative to 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) 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 links 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 between δ_{pub} and δ_{priv} matter, we normal-

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

ize $\delta_{\text{pub}} = 1$ 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 one.

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 such elasticity maps directly to our interpretation of private transit infrastructure generating congestion.

Commuting costs shifters. We calibrate δ^{τ} to match the variance of commuting flows observed from Encuesta Origen-Destino. We focus on this moment because commuting flows are unavailable at the level of granularity in our simplified hexagonal network. The geography of the origin-destination districts in EOD is much more aggregated so typically it is not possible to observe the paths that workers going from a residence district to a workplace district take.[5] Therefore, we set as target observed moment the variance of the commuting flows from district-to-district commuting in the EOD, and set the analogous model moment to be the variance of the hex-to-hex commuting flows resulting from the baseline model. Intuitively, this calibration strategy attempts to recover what would be the level of the model's commuting costs that matches one moment—spread—of the distribution of observed commuting flows. Due to this spatially-granular commuting data availability limitation, we further assume that the shifters are constant across links, so there is only a single scalar δ^{τ} that governs the overall level of commuting costs. We set δ^{τ} = 2e-6.

Land and productivity. To obtain a measure of H and H^c we rely on land use data that shows 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. Then we aggregate into our hexagon geographical units, and map average wages directly to productivity.

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 population is fixed to the observed one, and the public and private networks are the observed ones. Given our calibration, we are assuming that the current network was built with a K budget, which we increase by 50% in the

[5] This is not an issue in developed countries, e.g. the U.S., where commuting flow data is available at the tract-to-tract level

counterfactuals.

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, 1) where the planner only considers public transit infrastructure investments, and 2) where the planner considers both public and private infrastructure. By comparing these two scenarios, it allows us to assess what are the potential additional welfare gains of investing in private transit relative to public alone.

Counterfactual	Δ% vs Baseline (%)	Δ vs (B1)
(B1) Expand public only	4.4	—
(B2) Expand public and private	5.2	0.8

TABLE 3—WELFARE GAINS OF TRANSIT EXPANSIONS

Note: Table shows the welfare changes of two counterfactual scenarios: expanding infrastructure using public transit only and using both public and private transit investments. Baseline is the equilibrium with observed population and current network. Counterfactuals are computed by increasing the base budget by 50%. The third column shows the percentage point difference between counterfactual (b) relative to (a).

Welfare gains from network expansions. Table 3, shows the welfare gains of increasing the baseline budget by 50% and optimally expanding the current network. The first row shows a 4.4% gain of expanding the network when the planner only implements public transit investments, while the second row shows a 5.2% gain when the planner can invest in both technologies. By comparing these numbers, it suggests that private transit can deliver an extra marginal welfare gain of 0.8 percentage points, or 18% additional welfare gain. This magnitude is quantitatively important: governments in developing cities could benefit substantially by considering the joint design of the network, rather than the common recipe of expanding and evaluating public mass transit in isolation.

Spatial effects. In which part of the network should the investments be made? Figure 8 shows the spatial distribution of counterfactual investments. The left panel shows the new public investments when the planner considers only public infrastructure. The middle and right panels show new public and private investments, respectively, when the planner considers both technologies. In each panel, we show the current public or private network in grey lines.

First, focusing on the left panel, most of the investment is concentrated in the Eastern and Western peripheral parts of the city. In particular, most investment is strongly concentrated around two specific nodes, which correspond to two of the most productive locations in the city.[6] These nodes correspond to Mexico City's International Airport (to the East) and Santa Fe (to the West), which is an important hub for firms' headquarters.[7] The remaining investments are located on the outskirts of the city and primarily connect peripheral locations to these

[6] In Appendix A.3 we provide an additional figure showing the productivity of locations.

[7] Intuitively, these results make sense. Readers familiar with the Mexican setting and context may know that getting to these two places by transit is particularly difficult, as service availability is limited.

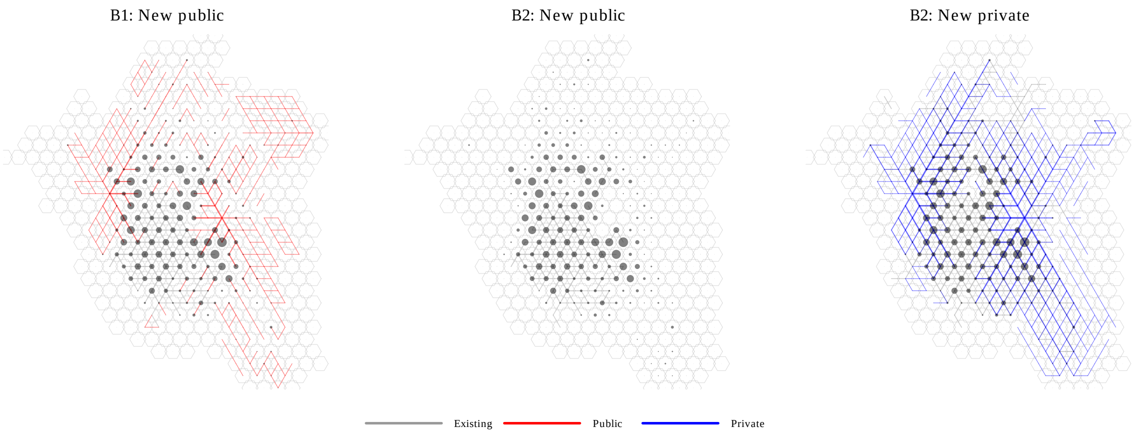


FIGURE 8. OPTIMAL TRANSIT EXPANSIONS

Note: Figure shows two counterfactual scenarios of expanding transit infrastructure relative to the current network. We consider a 50% increase of the K budget. The B1 scenario shows the optimal public network expansion when the planner uses public transit exclusively. The B2 scenario shows the optimal expansions of both public and private transit: the middle panel shows public infrastructure expansions, and the right panel shows private expansions. The size of the nodes represents the distribution of population by residence, with larger nodes indicating higher population density. Grey lines represent existing infrastructure.

productive nodes and to nodes that are already connected to the existing transit infrastructure.

Second, comparing the left and middle panels, we see that when the planner considers both technologies, most investment is allocated to the private technology. This mainly reflects that, in our calibration, private investment is an order of magnitude cheaper than public investment, and that the additional marginal costs and congestion generated by the private technology are outweighed by large budget savings. The location of the investments is similar to the previous scenario: most of the private investment aims to connect the majority of peripheral locations, particularly those linked to productive nodes and nodes that serve as connectors to the existing network.

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 showed 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.

REFERENCES

ADAMOPOULOS, T., L. BRANDT, J. LEIGHT, AND D. RESTUCCIA (2022): "Misallocation, selection, and productivity: A quantitative analysis with panel data from China," *Econometrica*, 90, 1261–1282.

ALDER, S. (2025): "Chinese roads in India: The effect of transport infrastructure on economic development," *Journal of International Economics*, 104140.

ALLEN, T. AND C. ARKOLAKIS (2022): "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 89, 2911–2957.

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

ASTURIAS, J., M. G. SANTANA, AND R. RAMOS (2014): *Misallocation, internal trade, and the role of transportation infrastructure*, CEMFI (Centro de Estudios Monetarios y Financieros) Madrid, Spain.

BAUM-SNOW, N. (2007): "Did highways cause suburbanization?" *The quarterly journal of economics*, 122, 775–805.

BERNOT, M., V. CASELLES, AND J.-M. MOREL (2009): *Optimal transportation networks: models and theory*, Springer.

BJÖRKEGREN, D., A. DUHAUT, G. NAGPAL, AND N. TSIVANIDIS (2025): "Public and Private Transit: Evidence from Lagos," Tech. rep., National Bureau of Economic Research.

CHANDRA, A. AND E. THOMPSON (2000): "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional science and urban economics*, 30, 457–490.

CONWELL, L. (2024): "Privatized Provision of Public Transit," .

DESMET, K. AND E. ROSSI-HANSBERG (2013): "Urban accounting and welfare," *American Economic Review*, 103, 2296–2327.

DONALDSON, D. (2018): "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 108, 899–934.

DURANTON, G., P. M. MORROW, AND M. A. TURNER (2014): "Roads and Trade: Evidence from the US," *Review of Economic Studies*, 81, 681–724.

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

FAJGELBAUM, P. D., E. MORALES, J. C. SUÁREZ SERRATO, AND O. ZIDAR (2019): "State taxes and spatial misallocation," *The Review of Economic Studies*, 86, 333–376.

FAJGELBAUM, P. D. AND E. SCHAAL (2020): "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 88, 1411–1452.

FELBERMAYR, G. J. AND A. TARASOV (2022): "Trade and the spatial distribution of transport infrastructure," *Journal of Urban Economics*, 130, 103473.

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

HSIEH, C.-T. AND P. J. KLENOW (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124, 1403–1448.

HSIEH, C.-T. AND E. MORETTI (2019): "Housing constraints and spatial misallocation," *American economic journal: macroeconomics*, 11, 1–39.

KHANNA, G., C. MEDINA, A. NYSHADHAM, J. RAMOS, J. TAMAYO, AND Z. H. TIEW (2024): "Spatial Mobility, Economic Opportunity, and Crime," Working paper; add link when available.

KREINDLER, G., A. GADUH, T. GRAFF, R. HANNA, AND B. A. OLKEN (2023): "Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta," Tech. rep., National Bureau of Economic Research.

MOSQUEDA, J. (2026): "Equilibrium Commuting Costs: The Role of Private and Public Transit," Job Market Paper; working paper.

ROGERSON, R. AND D. RESTUCCIA (2004): "Policy distortions and aggregate productivity with heterogeneous plants," in *2004 Meeting Papers*, Society for Economic Dynamics, 69.

TARASOV, A. AND G. FELBERMAYR (2014): "Trade and the Spatial Distribution of Transport Infrastructure," .

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

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

Appendix

Appendix Contents

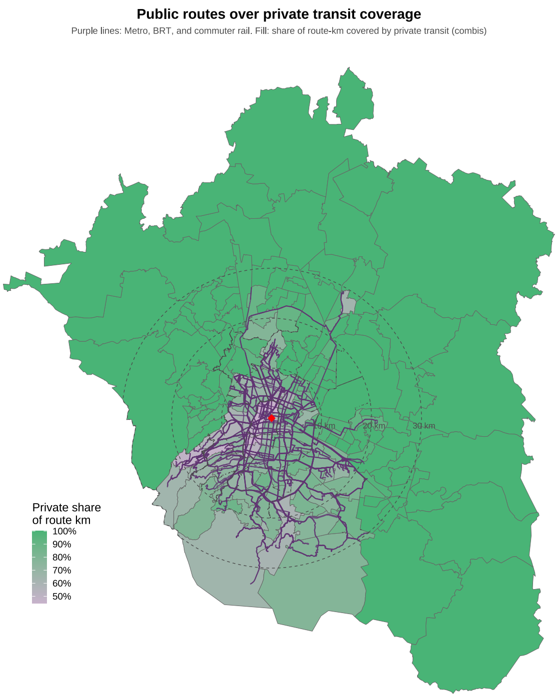
A Transit network in Mexico

A.1 Private Transit and Network Coverage

This section documents the dominance of private transit operators (*combis*) in the ZMVM's transit network. We measure network coverage using total route-kilometers by mode and district.

Figure 9 displays a map of the ZMVM in which each district is shaded according to the share of route-kilometers covered by private transit (combis). Formal public transit routes—Metro, BRT (Metrobús), and commuter rail (Tren Suburbano)—are overlaid as purple lines, with concentric rings indicating distances of 10, 20, and 30 km from the CBD. Two patterns stand out. First, formal public transit infrastructure is concentrated within approximately 15 km of the CBD. Second, in the vast majority of districts—particularly those in the urban periphery—private operators account for nearly all route-kilometers, with many districts at or above 90% private coverage.

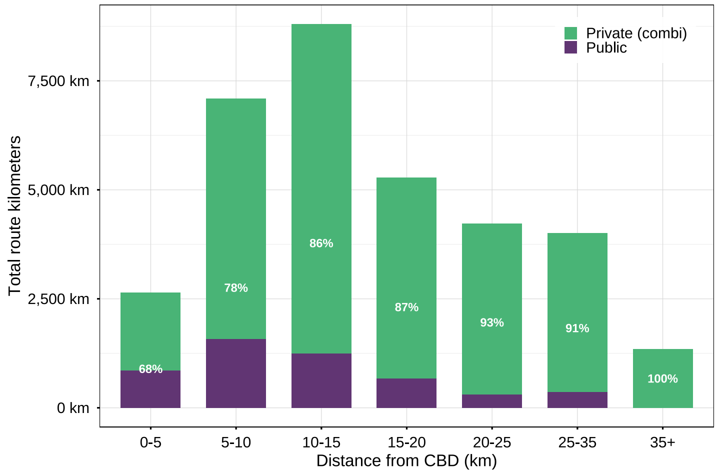
FIGURE 9. SPATIAL DISTRIBUTION OF PRIVATE TRANSIT COVERAGE



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 10 decomposes total route-kilometers by distance band from the CBD. Even in the innermost ring (0–5 km), private transit accounts for 68% of route-kilometers. This share rises monotonically with distance: from 78% in the 5–10 km band to 100% beyond 35 km. The gradient reflects the fact that formal public transit networks were built to serve central areas, leaving peripheral districts almost entirely dependent on informal private operators.

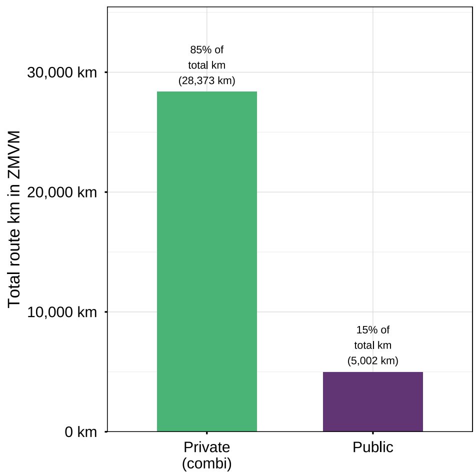
FIGURE 10. ROUTE-KILOMETERS BY DISTANCE FROM CBD



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.

Figure 11 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.

FIGURE 11. AGGREGATE NETWORK COVERAGE: PRIVATE VS. PUBLIC TRANSIT



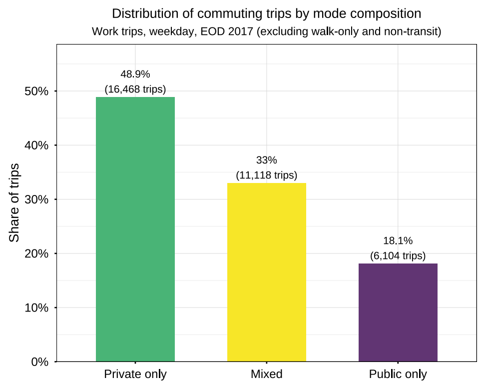
Notes: The figure shows total route-kilometers in the ZMVM by operator type. Private transit (combis) accounts for 28,373 km (85%) and formal public transit for 5,002 km (15%).

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 4.3. We decompose commuting trips by mode composition, examine the spatial co-location of public and private stops, and characterize how trip attributes vary across trip types.

Trip composition. Figure 12 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.

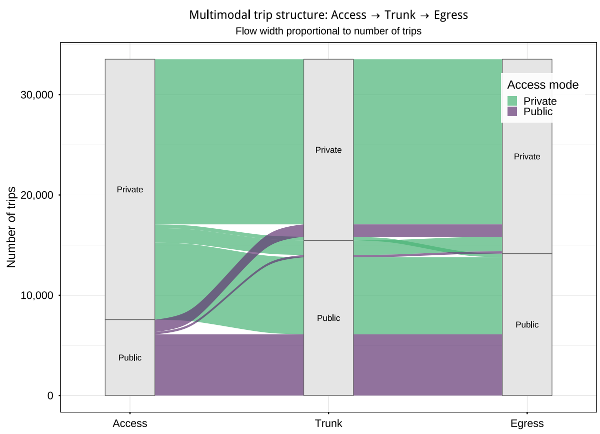
FIGURE 12. DISTRIBUTION OF COMMUTING TRIPS BY MODE COMPOSITION



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 13 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.

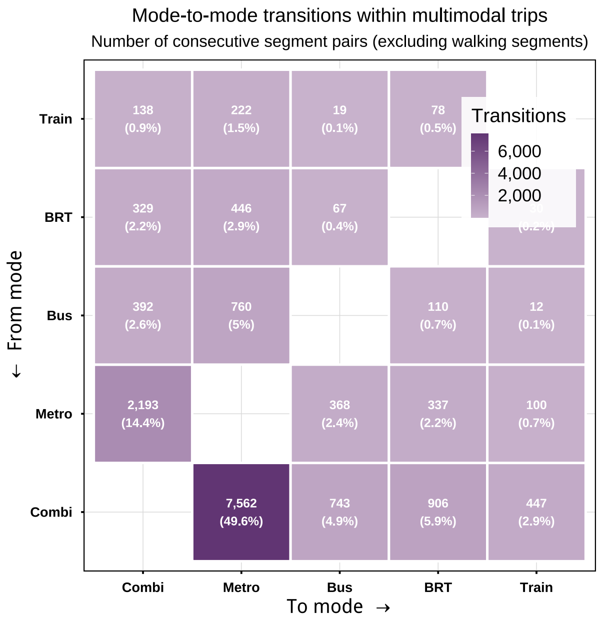
FIGURE 13. MULTIMODAL TRIP STRUCTURE: ACCESS, TRUNK, AND EGRESS



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.

Mode-to-mode transitions. Figure 14 displays a transition matrix of consecutive mode pairs within multimodal trips, excluding walking segments. The single most frequent transition is from combi to metro, accounting for 49.6 percent of all observed transitions (7,562 pairs). The next largest combi transitions are combi-to-BRT (5.9 percent) and combi-to-bus (4.9 percent). Metro-to-combi is the dominant reverse transition (14.4 percent). These patterns confirm that private transit overwhelmingly serves as the feeder mode into public trunk infrastructure, with the combi–metro interface constituting the backbone of multimodal commuting in the ZMVM.

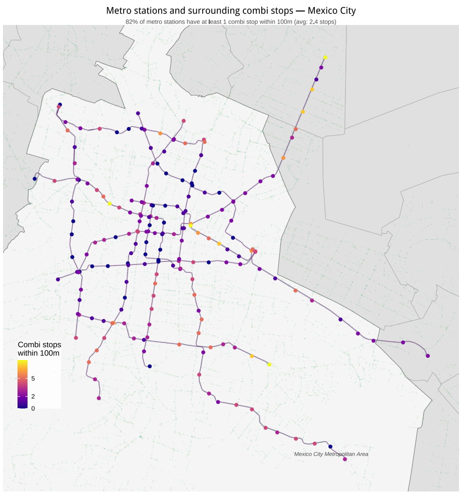
FIGURE 14. MODE-TO-MODE TRANSITIONS WITHIN MULTIMODAL TRIPS



Notes: Each cell reports the number and share of consecutive segment pairs transitioning from the row mode to the column mode, excluding walking segments. Darker shading indicates higher frequency.

Spatial co-location. Figure 15 maps every metro station in Mexico City, colored by the number of combi stops located within 100 meters. Eighty-two percent of metro stations have at least one combi stop in their immediate vicinity, with an average of 2.4 combi stops per station. This physical co-location reflects the organic emergence of private feeder services around public infrastructure and provides infrastructure-level confirmation of the trip-level complementarity documented above.

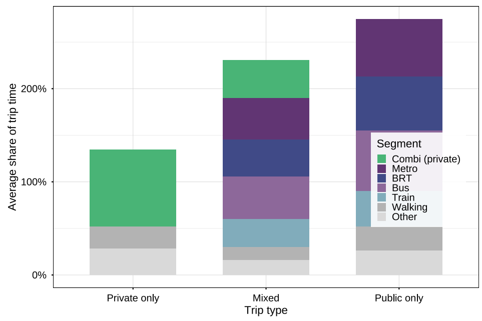
FIGURE 15. METRO STATIONS AND SURROUNDING COMBI STOPS



Notes: Each dot represents a metro station, colored by the number of combi stops within 100 meters. Eighty-two percent of stations have at least one combi stop nearby, with an average of 2.4 stops per station.

Time decomposition and trip characteristics. Figure 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.

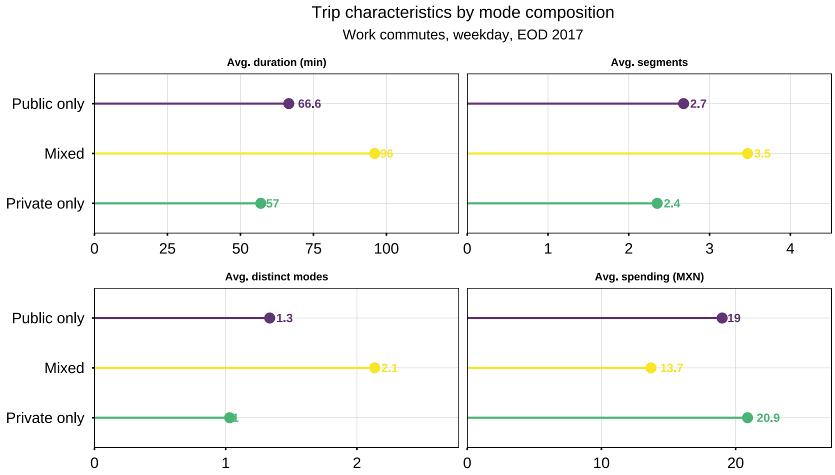
FIGURE 16. AVERAGE SHARE OF TRIP TIME BY MODE SEGMENT



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 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.

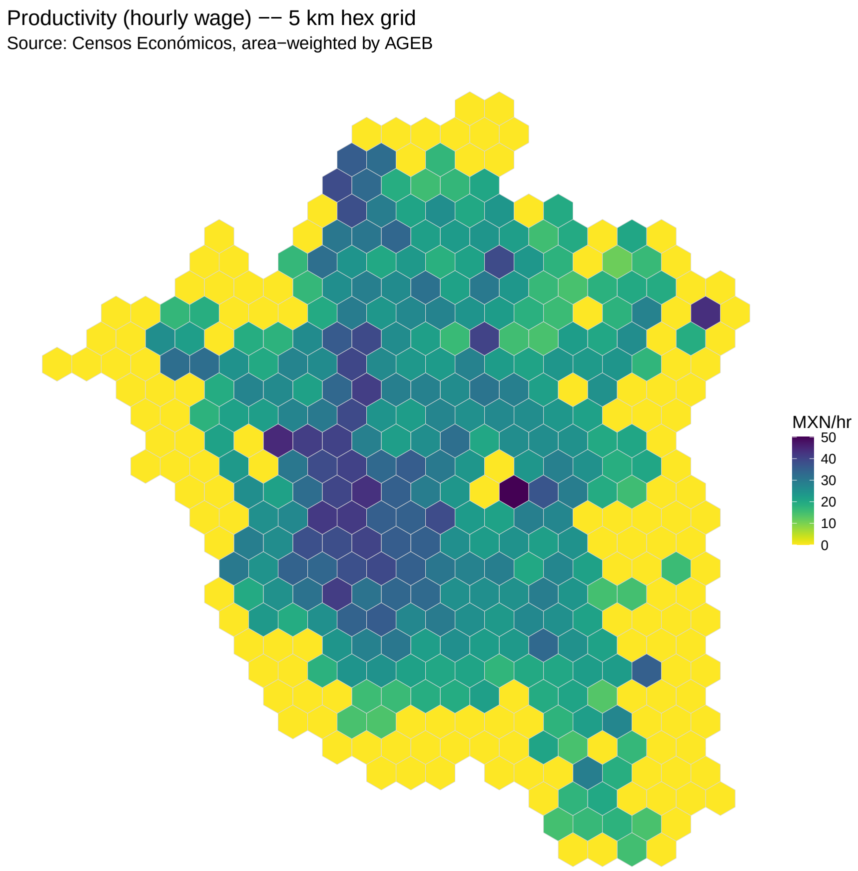
FIGURE 17. TRIP CHARACTERISTICS BY MODE COMPOSITION



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

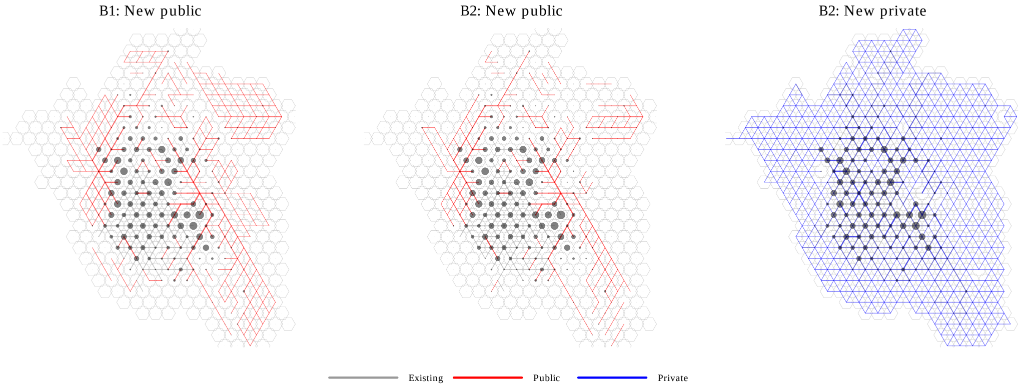
A.3 Productivity map

FIGURE 18. SPATIAL DISTRIBUTION OF PRODUCTIVITY IN MEXICO CITY



A.4 Sensitivity analysis: Increasing marginal costs of private transit

FIGURE 19. THETA RATIO = 10



B Model derivations

B.1 First order conditions of the planner's problem

The Lagrangian of the planner's problem is

$$\begin{aligned}
\mathcal{L} = u &+ \sum_{j\in\mathcal{J}} \omega_j \, L_j^R \big(U(c_j, h_j) - u\big) \\
&+ \sum_{j\in\mathcal{J}} P_j^F \left(L_j^R + \sum_{i\in N(j)} L_{ij} - L_j^L - \sum_{k\in N(j)} (1+\tau_{jk}) L_{jk} \right) \\
&+ P^R \Big(\bar{L} - \sum_{j\in\mathcal{J}} L_j^R \Big) \\
&+ P^L \left(\bar{L} - \sum_{k\in\mathcal{J}} L_k^L - \sum_{k\in\mathcal{J}} \sum_{j\in N(k)} \tau_{jk} L_{jk} \right) \\
&+ P^C \Big(\sum_{j\in\mathcal{J}} A_j L_j^L - \sum_{j\in\mathcal{J}} c_j L_j^R \Big) + \sum_{j\in\mathcal{J}} P_j^H \big(H_j - h_j L_j^R \big) \\
&+ P^K \left(K - \sum_{(j,k)\in\mathcal{E}} \big[\delta_{jk}^{I,\text{pub}} \, I_{jk}^{\text{pub}} + \delta_{jk}^{I,\text{pri}} \, I_{jk}^{\text{pri}} \big] \right) \\
&+ \sum_{j\in\mathcal{J}} \mu_j^c \, c_j + \sum_{j\in\mathcal{J}} \mu_j^h \, h_j + \sum_{j\in\mathcal{J}} \mu_j^R \, L_j^R + \sum_{j\in\mathcal{J}} \mu_j^L \, L_j^L + \sum_{(j,k)\in\mathcal{E}} \nu_{jk} \, L_{jk} \\
&+ \sum_{(j,k)\in\mathcal{E}} \underline{\mu}_{jk}^{I,\text{pub}} \, (I_{jk}^{\text{pub}} - \underline{I}_{jk}^{\text{pub}}) + \sum_{(j,k)\in\mathcal{E}} \overline{\mu}_{jk}^{I,\text{pub}} \, (\overline{I}_{jk}^{\text{pub}} - I_{jk}^{\text{pub}}) \\
&+ \sum_{(j,k)\in\mathcal{E}} \underline{\mu}_{jk}^{I,\text{pri}} \, (I_{jk}^{\text{pri}} - \underline{I}_{jk}^{\text{pri}}) + \sum_{(j,k)\in\mathcal{E}} \overline{\mu}_{jk}^{I,\text{pri}} \, (\overline{I}_{jk}^{\text{pri}} - I_{jk}^{\text{pri}}).
\end{aligned} \tag{12}$$

The first-order conditions with respect to consumption and labor allocations are standard:

$$[c_j] \qquad \omega_j L_j^R \, U_c(c_j, h_j) \;+\; \mu_j^c \;=\; P^C \, L_j^R, \qquad \forall j. \tag{13}$$

$$[h_j] \qquad \omega_j L_j^R \, U_h(c_j, h_j) \;+\; \mu_j^h \;=\; P_j^H \, L_j^R, \qquad \forall j. \tag{14}$$

$$[L_j^L] \qquad -\; P_j^F \;-\; P^L \;+\; P^C A_j \;+\; \mu_j^L \;=\; 0, \qquad \forall j. \tag{15}$$

$$[L_j^R] \qquad \omega_j \big(U(c_j, h_j) - u \big) \;+\; P_j^F \;-\; P^R \;-\; P^C c_j \;-\; P_j^H h_j \;+\; \mu_j^R \;=\; 0, \qquad \forall j. \tag{16}$$

The first-order condition with respect to L_{jk} is:

$$[L_{jk}] \qquad P_k^F \;-\; P_j^F\Big[(1+\tau_{jk}) + L_{jk}\,\frac{\partial \tau_{jk}}{\partial L_{jk}}\Big] \;-\; P^L\Big[\tau_{jk} + L_{jk}\,\frac{\partial \tau_{jk}}{\partial L_{jk}}\Big] \;+\; \nu_{jk} \;=\; 0, \qquad \forall (j,k).$$

To compute the derivative, note that τ_{jk} depends on L_{jk} both directly through the L_{jk}^{β} term and indirectly through s_{jk}. However, since $s_{jk} = I_{jk}^{\text{pub}}/(I_{jk}^{\text{pub}} + I_{jk}^{\text{pri}})$ depends only on infrastructure investments, we have $\partial s_{jk}/\partial L_{jk} = 0$. Thus:

$$\frac{\partial \tau_{jk}}{\partial L_{jk}} = \frac{\beta}{L_{jk}}\,\tau_{jk}.$$

Substituting and simplifying:

$$[L_{jk}] \;:\; P_k^F - P_j^F\Big[1 + (1+\beta)\tau_{jk}\Big] - P^L\,(1+\beta)\tau_{jk} + \nu_{jk} = 0, \qquad \forall (j,k). \tag{17}$$

The first-order condition with respect to public infrastructure is:

$$[I_{jk}^{\text{pub}}] \qquad -\,(P_j^F + P^L)\,L_{jk}\,\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pub}}} \;-\; P^K\,\delta_{jk}^{I,\text{pub}} \;+\; \underline{\mu}_{jk}^{I,\text{pub}} - \overline{\mu}_{jk}^{I,\text{pub}} \;=\; 0, \qquad \forall (j,k). \tag{18}$$

To compute $\partial \tau_{jk}/\partial I_{jk}^{\text{pub}}$, we use the chain rule. Recall that $I_{jk} = I_{jk}^{\text{pub}} + I_{jk}^{\text{pri}}$ and $s_{jk} = I_{jk}^{\text{pub}}/I_{jk}$. Define $\Theta(s_{jk}) \equiv (1-s_{jk})\theta_{\text{pri}} + s_{jk}\theta_{\text{pub}}$, so that $\tau_{jk} = \delta_{jk}^{\tau}\Theta(s_{jk})L_{jk}^{\beta}I_{jk}^{-\gamma}$. Then:

$$\frac{\partial I_{jk}}{\partial I_{jk}^{\text{pub}}} = 1,$$

$$\frac{\partial s_{jk}}{\partial I_{jk}^{\text{pub}}} = \frac{I_{jk} - I_{jk}^{\text{pub}}}{I_{jk}^2} = \frac{I_{jk}^{\text{pri}}}{I_{jk}^2} = \frac{1-s_{jk}}{I_{jk}}.$$

Since $\Theta'(s_{jk}) = \theta_{\text{pub}} - \theta_{\text{pri}} \equiv \Delta\theta$ (negative under the maintained assumption $\theta_{\text{pub}} < \theta_{\text{pri}}$), we have:

$$\begin{aligned}\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pub}}} &= \tau_{jk}\left[\frac{\Delta\theta}{\Theta(s_{jk})}\cdot\frac{1-s_{jk}}{I_{jk}} - \frac{\gamma}{I_{jk}}\right] \\ &= \frac{\tau_{jk}}{I_{jk}}\left[\frac{\Delta\theta\,(1-s_{jk})}{\Theta(s_{jk})} - \gamma\right].\end{aligned} \tag{19}$$

Substituting into equation (18) and assuming interior solutions ($\underline{\mu}_{jk}^{I,\text{pub}} = \overline{\mu}_{jk}^{I,\text{pub}} = 0$):

$$[I_{jk}^{\text{pub}}] \qquad (P_j^F + P^L)\,L_{jk}\,\frac{\tau_{jk}}{I_{jk}}\left[\gamma - \frac{\Delta\theta\,(1-s_{jk})}{\Theta(s_{jk})}\right] \;=\; P^K\,\delta_{jk}^{I,\text{pub}}. \tag{20}$$

This condition states that the planner invests in public infrastructure until the marginal benefit (left-hand side) equals the marginal cost (right-hand side). The marginal benefit has two components: (i) the direct infrastructure effect—more total capacity reduces congestion costs (the γ term), and (ii) the composition effect—shifting toward public technology reduces operating costs when $\theta_{\text{pub}} < \theta_{\text{pri}}$ (the term involving $\Delta\theta$, which is negative, making the subtraction a net addition to benefits).

Analogously, the first-order condition with respect to private infrastructure is:

$$[I_{jk}^{\text{pri}}] \qquad -\,(P_j^F + P^L)\, L_{jk}\, \frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pri}}} \;-\; P^K\, \delta_{jk}^{I,\text{pri}} \;+\; \underline{\mu}_{jk}^{I,\text{pri}} - \overline{\mu}_{jk}^{I,\text{pri}} \;=\; 0, \qquad \forall (j,k). \tag{21}$$

Following similar steps:

$$\frac{\partial s_{jk}}{\partial I_{jk}^{\text{pri}}} = -\frac{I_{jk}^{\text{pub}}}{I_{jk}^2} = -\frac{s_{jk}}{I_{jk}},$$

so that:

$$\frac{\partial \tau_{jk}}{\partial I_{jk}^{\text{pri}}} = \frac{\tau_{jk}}{I_{jk}} \left[-\frac{\Delta\theta\, s_{jk}}{\Theta(s_{jk})} - \gamma \right]. \tag{22}$$

Substituting into equation (21) and assuming interior solutions:

$$[I_{jk}^{\text{pri}}] \qquad (P_j^F + P^L)\, L_{jk}\, \frac{\tau_{jk}}{I_{jk}} \left[\gamma + \frac{\Delta\theta\, s_{jk}}{\Theta(s_{jk})} \right] \;=\; P^K\, \delta_{jk}^{I,\text{pri}}. \tag{23}$$

B.2 Proof of Proposition 1

Proof. When both infrastructure types are strictly positive, the bound constraints are slack and $\underline{\mu}_{jk}^{I,\text{pub}} = \overline{\mu}_{jk}^{I,\text{pub}} = \underline{\mu}_{jk}^{I,\text{pri}} = \overline{\mu}_{jk}^{I,\text{pri}} = 0$. Taking the ratio of equations (20) and (23):

$$\frac{\gamma - \dfrac{\Delta\theta\,(1 - s_{jk})}{\Theta(s_{jk})}}{\gamma + \dfrac{\Delta\theta\, s_{jk}}{\Theta(s_{jk})}} \;=\; \frac{\delta_{jk}^{I,\text{pub}}}{\delta_{jk}^{I,\text{pri}}} \;=\; \rho_{jk}. \tag{24}$$

Multiplying numerator and denominator of the left-hand side by $\Theta(s_{jk})$:

$$\frac{\gamma\,\Theta(s_{jk}) - \Delta\theta\,(1 - s_{jk})}{\gamma\,\Theta(s_{jk}) + \Delta\theta\, s_{jk}} \;=\; \rho_{jk}.$$

Cross-multiplying:

$$\gamma\,\Theta(s_{jk}) - \Delta\theta\,(1 - s_{jk}) \;=\; \rho_{jk} \left[\gamma\,\Theta(s_{jk}) + \Delta\theta\, s_{jk} \right].$$

Expanding and collecting terms involving $\Theta(s_{jk})$:

$$\gamma\,\Theta(s_{jk})(1-\rho_{jk}) \;=\; \Delta\theta\,(1-s_{jk}) + \rho_{jk}\,\Delta\theta\,s_{jk} \;=\; \Delta\theta\,[1-s_{jk}(1-\rho_{jk})].$$

Substituting $\Theta(s_{jk}) = \theta_{\text{pri}} + s_{jk}\,\Delta\theta$:

$$\gamma\,[\theta_{\text{pri}} + s_{jk}\,\Delta\theta](1-\rho_{jk}) \;=\; \Delta\theta\,[1-s_{jk}(1-\rho_{jk})].$$

Expanding the left-hand side:

$$\gamma\,\theta_{\text{pri}}(1-\rho_{jk}) + \gamma\,s_{jk}\,\Delta\theta\,(1-\rho_{jk}) \;=\; \Delta\theta - \Delta\theta\,s_{jk}(1-\rho_{jk}).$$

Collecting terms in s_{jk}:

$$s_{jk}\,\Delta\theta\,(1-\rho_{jk})(\gamma+1) \;=\; \Delta\theta - \gamma\,\theta_{\text{pri}}(1-\rho_{jk}).$$

Solving for s_{jk}:

$$s^*_{jk} \;=\; \frac{\Delta\theta - \gamma\,\theta_{\text{pri}}(1-\rho_{jk})}{\Delta\theta\,(\gamma+1)(1-\rho_{jk})}.$$

Multiplying numerator and denominator by -1 yields the expression in the proposition:

$$s^*_{jk} \;=\; \frac{\gamma\,\theta_{\text{pri}}(1-\rho_{jk}) - \Delta\theta}{-\Delta\theta\,(\gamma+1)(1-\rho_{jk})}. \tag{25}$$

□

Comparative statics. Define $N \equiv \gamma\,\theta_{\text{pri}}(1-\rho_{jk}) - \Delta\theta$ and $D \equiv -\Delta\theta\,(\gamma+1)(1-\rho_{jk})$, so that $s^*_{jk} = N/D$.

Effect of relative infrastructure cost ρ_{jk}: Differentiating:

$$\frac{\partial N}{\partial \rho_{jk}} = -\gamma\,\theta_{\text{pri}},$$
$$\frac{\partial D}{\partial \rho_{jk}} = \Delta\theta\,(\gamma+1).$$

Then:

$$\begin{aligned}
\frac{\partial s_{jk}^*}{\partial \rho_{jk}} &= \frac{1}{D^2}\left[\frac{\partial N}{\partial \rho_{jk}} D - N\frac{\partial D}{\partial \rho_{jk}}\right] \\
&= \frac{1}{D^2}\left[(-\gamma\,\theta_{\mathrm{pri}})(-\Delta\theta(\gamma+1)(1-\rho_{jk})) - N\cdot\Delta\theta(\gamma+1)\right] \\
&= \frac{\Delta\theta(\gamma+1)}{D^2}\left[\gamma\,\theta_{\mathrm{pri}}(1-\rho_{jk}) - N\right] \\
&= \frac{\Delta\theta(\gamma+1)}{D^2}\left[\gamma\,\theta_{\mathrm{pri}}(1-\rho_{jk}) - \gamma\,\theta_{\mathrm{pri}}(1-\rho_{jk}) + \Delta\theta\right] \\
&= \frac{\Delta\theta^2(\gamma+1)}{D^2}.
\end{aligned}$$

Since $\Delta\theta^2 > 0$, $(\gamma+1) > 0$, and $D^2 > 0$, we have:

$$\frac{\partial s_{jk}^*}{\partial \rho_{jk}} > 0.$$

Effect of operating cost advantage $a \equiv \theta_{pri} - \theta_{pub} = -\Delta\theta$: Note that $\partial/\partial a = -\partial/\partial\Delta\theta$. We have:

$$\begin{aligned}
\frac{\partial N}{\partial \Delta\theta} &= -1, \\
\frac{\partial D}{\partial \Delta\theta} &= -(\gamma+1)(1-\rho_{jk}).
\end{aligned}$$

Then:

$$\begin{aligned}
\frac{\partial s_{jk}^*}{\partial \Delta\theta} &= \frac{-D - N\cdot[-(\gamma+1)(1-\rho_{jk})]}{D^2} \\
&= \frac{-D + N(\gamma+1)(1-\rho_{jk})}{D^2}.
\end{aligned}$$

Substituting $D = -\Delta\theta(\gamma+1)(1-\rho_{jk})$:

$$\begin{aligned}
\frac{\partial s_{jk}^*}{\partial \Delta\theta} &= \frac{\Delta\theta(\gamma+1)(1-\rho_{jk}) + N(\gamma+1)(1-\rho_{jk})}{D^2} \\
&= \frac{(\gamma+1)(1-\rho_{jk})[\Delta\theta + N]}{D^2} \\
&= \frac{(\gamma+1)(1-\rho_{jk})\cdot\gamma\,\theta_{\mathrm{pri}}(1-\rho_{jk})}{D^2} \\
&= \frac{\gamma\,\theta_{\mathrm{pri}}(\gamma+1)(1-\rho_{jk})^2}{D^2}.
\end{aligned}$$

Therefore:

$$\frac{\partial s_{jk}^*}{\partial a} = -\frac{\partial s_{jk}^*}{\partial \Delta\theta} = -\frac{\gamma\,\theta_{\mathrm{pri}}(\gamma+1)(1-\rho_{jk})^2}{D^2}. \tag{26}$$

Since $D^2 > 0$ and $\gamma\,\theta_{\mathrm{pri}}(\gamma+1) > 0$, the sign depends on $(1-\rho_{jk})^2 > 0$ always, but the overall expression has a negative sign in front. However, we must be careful: $D = -\Delta\theta(\gamma+1)(1-$

ρ_{jk}), so $D^2 = \Delta\theta^2(\gamma+1)^2(1-\rho_{jk})^2$. Thus:

$$\frac{\partial s^*_{jk}}{\partial a} = -\frac{\gamma\,\theta_{\text{pri}}(\gamma+1)(1-\rho_{jk})^2}{\Delta\theta^2(\gamma+1)^2(1-\rho_{jk})^2} = -\frac{\gamma\,\theta_{\text{pri}}}{\Delta\theta^2(\gamma+1)}.$$

Since $\Delta\theta < 0$ implies $\Delta\theta^2 > 0$, and $\gamma, \theta_{\text{pri}}, (\gamma+1) > 0$, we have:

$$\frac{\partial s^*_{jk}}{\partial(\theta_{\text{pri}} - \theta_{\text{pub}})} = \frac{\gamma\,\theta_{\text{pri}}}{\Delta\theta^2(\gamma+1)} > 0. \tag{27}$$

A larger operating cost advantage of public transit (higher $\theta_{\text{pri}} - \theta_{\text{pub}}$) increases the optimal public share, as expected.

Effect of infrastructure elasticity γ: Differentiating N and D:

$$\begin{aligned}\frac{\partial N}{\partial \gamma} &= \theta_{\text{pri}}(1-\rho_{jk}),\\ \frac{\partial D}{\partial \gamma} &= -\Delta\theta(1-\rho_{jk}).\end{aligned}$$

Then:

$$\begin{aligned}\frac{\partial s^*_{jk}}{\partial \gamma} &= \frac{\theta_{\text{pri}}(1-\rho_{jk})\cdot D - N\cdot(-\Delta\theta)(1-\rho_{jk})}{D^2}\\ &= \frac{(1-\rho_{jk})[\theta_{\text{pri}}D + N\Delta\theta]}{D^2}.\end{aligned}$$

Substituting $D = -\Delta\theta(\gamma+1)(1-\rho_{jk})$ and $N = \gamma\theta_{\text{pri}}(1-\rho_{jk}) - \Delta\theta$:

$$\begin{aligned}\theta_{\text{pri}}D + N\Delta\theta &= -\theta_{\text{pri}}\Delta\theta(\gamma+1)(1-\rho_{jk}) + [\gamma\theta_{\text{pri}}(1-\rho_{jk}) - \Delta\theta]\Delta\theta\\ &= -\theta_{\text{pri}}\Delta\theta(\gamma+1)(1-\rho_{jk}) + \gamma\theta_{\text{pri}}\Delta\theta(1-\rho_{jk}) - \Delta\theta^2\\ &= \theta_{\text{pri}}\Delta\theta(1-\rho_{jk})[-(\gamma+1)+\gamma] - \Delta\theta^2\\ &= -\theta_{\text{pri}}\Delta\theta(1-\rho_{jk}) - \Delta\theta^2\\ &= -\Delta\theta[\theta_{\text{pri}}(1-\rho_{jk}) + \Delta\theta].\end{aligned}$$

Note that $\theta_{\text{pri}}(1-\rho_{jk}) + \Delta\theta = \theta_{\text{pri}} - \rho_{jk}\theta_{\text{pri}} + \theta_{\text{pub}} - \theta_{\text{pri}} = \theta_{\text{pub}} - \rho_{jk}\theta_{\text{pri}}$.

Therefore:

$$\frac{\partial s^*_{jk}}{\partial \gamma} = \frac{(1-\rho_{jk})(-\Delta\theta)[\theta_{\text{pub}} - \rho_{jk}\theta_{\text{pri}}]}{D^2}. \tag{28}$$

Since $-\Delta\theta > 0$ (as $\theta_{\text{pub}} < \theta_{\text{pri}}$) and $D^2 > 0$, the sign depends on:

- $(1-\rho_{jk})$: positive if $\rho_{jk} < 1$, negative if $\rho_{jk} > 1$.
- $[\theta_{\text{pub}} - \rho_{jk}\theta_{\text{pri}}]$: positive if $\rho_{jk} < \theta_{\text{pub}}/\theta_{\text{pri}} < 1$, negative otherwise.

For the empirically relevant case where $\rho_{jk} > 1$ (public infrastructure more expensive to build) and $\theta_{\text{pub}}/\theta_{\text{pri}} < 1$ (public cheaper to operate), both factors are negative, so:

$$\frac{\partial s^*_{jk}}{\partial \gamma} > 0 \quad \text{when } \rho_{jk} > 1. \tag{29}$$

Higher infrastructure elasticity increases the optimal public share when public is more expensive to build, because the direct capacity effect (which benefits both technologies equally) becomes more important relative to the composition effect.

B.3 Derivation of optimal investment

Substituting commuting costs into the private-infrastructure FOC (21) yields

$$(P^F_j + P^L)\, L_{jk}\, \delta^{\tau}_{jk}\, \Theta(s_{jk})\, L^{\beta}_{jk}\, I^{-(\gamma+1)}_{jk} \left(\gamma + \frac{\Delta\theta\, s_{jk}}{\Theta(s_{jk})}\right) = P^K\, \delta^{I,\text{pri}}_{jk}.$$

Rearranging gives

$$I^{\gamma+1}_{jk} = \frac{(P^F_j + P^L)\, \delta^{\tau}_{jk}\, L^{\beta+1}_{jk}}{P^K\, \delta^{I,\text{pri}}_{jk}}\, \Theta(s_{jk}) \left(\gamma + \frac{\Delta\theta\, s_{jk}}{\Theta(s_{jk})}\right). \tag{30}$$

Noting that the ratio of FOCs implies

$$\frac{\gamma - \dfrac{\Delta\theta\,(1 - s^*_{jk})}{\Theta(s^*_{jk})}}{\gamma + \dfrac{\Delta\theta\, s^*_{jk}}{\Theta(s^*_{jk})}} = \rho_{jk},$$

evaluating at the optimal mix s^*_{jk}, and using the identities

$$\gamma + \frac{\Delta\theta\, s}{\Theta(s)} = \gamma + 1 - \frac{\theta_{\text{pri}}}{\Theta(s)}, \qquad \gamma - \frac{\Delta\theta\,(1-s)}{\Theta(s)} = \gamma + 1 - \frac{\theta_{\text{pub}}}{\Theta(s)},$$

the ratio condition implies

$$(\gamma+1)\,(1-\rho_{jk}) = \frac{\theta_{\text{pub}} - \rho_{jk}\theta_{\text{pri}}}{\Theta(s^*_{jk})}, \qquad \Rightarrow \qquad \Theta(s^*_{jk}) = \frac{\theta_{\text{pub}} - \rho_{jk}\theta_{\text{pri}}}{(\gamma+1)(1-\rho_{jk})}.$$

Plugging this into $\Theta(s^*_{jk})\big(\gamma + \frac{\Delta\theta s^*_{jk}}{\Theta(s^*_{jk})}\big) = \Theta(s^*_{jk})\big((\gamma+1) - \frac{\theta_{\text{pri}}}{\Theta(s^*_{jk})}\big)$ gives

$$\Theta(s^*_{jk}) \left(\gamma + \frac{\Delta\theta\, s^*_{jk}}{\Theta(s^*_{jk})}\right) = (\gamma+1)\Theta(s^*_{jk}) - \theta_{\text{pri}} = \frac{\Delta\theta}{1-\rho_{jk}}.$$

Substituting this expression into (30) yields (10) (expanding $\Delta\theta$ and $(1-\rho_{jk})$).