The Centralizing Effects of Tokyo's Train System

Mark Bamba*

January 3, 2026
Latest draft: **link**

Abstract

We evaluate the impact of Tokyo's train system on where people live and work and the resulting welfare. We do so by calibrating a class of urban commuting models to Tokyo and using them to quantify how welfare and the spatial distribution of residence and employment respond to counterfactual changes in Tokyo's rail network and in the surrounding transport environment and urban form.

Three findings emerge. First, Tokyo's rail system centralizes activity and generates substantial net benefits: relative to a no-rail benchmark, the existing network raises the center's share of population and employment, and household welfare gains are between 9–16% in Greater Tokyo, corresponding to 1.8–5 times the present value of construction and maintenance costs. Second, network design matters. Core-focused subway lines account for most of the centralizing effect and a disproportionate share of welfare gains relative to construction expenditure. Third, the gains are sensitive to urban density and mobility options. Increasing automobile travel speeds toward typical U.S. levels reduces welfare gains from rail by about 40% and makes the city less centralized. If, in addition, Tokyo's population density is reduced to New York–like levels, the rail system remains socially beneficial; at Los Angeles–like density, net returns become negative, highlighting high population density and the absence of alternative modes that encourage dispersion as important factors for successful rail investment.

*This project would not have been possible without Eduardo Morales, Steve Redding, and David Weinstein's continued advice and support. Atsushi Yamagishi, and Kentaro Nakajima provided invaluable guidance throughout. This project has benefitted from conversations with Richard Rogerson, Owen Zidar, Derek Wenning, John Sturm, Michael Jenuwine, Pietro Buri, Mariko Nakagawa, Shota Fujishima, Chihiro Shimizu, Yuyang Jiang among others. Participants from the Princeton Student Trade Workshop and Hitotsubashi's Urban Seminar are thanked for their constructive comments. This project is the result of a collaboration with the Center for Spatial Information Science, the University of Tokyo (Project No. 1296) and used the 1991 and 1996 *Establishment and Enterprise Census Statistical Data by Small Areas*, and the 1990 and 1995 *Population Census Tabulation for Small Areas*, all of which were provided by Sinfonica.

1 Introduction

Metropolitan areas today are characterized by dense spatial concentrations of economic activity. For example, New York City and Tokyo's special wards have about 8 million and 10 million residents, respectively, and population densities several times those of surrounding suburbs and peripheries. These metropolitan areas are served by urban rail systems that carry millions of passenger trips each day, both within central districts and between centers and outlying areas. These rail systems likely play a role in the concentration of economic activity and have large welfare consequences. However, the magnitudes of these effects are less well known. Moreover, rail systems entail large construction and operating costs, raising a central question: do the benefits exceed these costs?[1]

This paper takes up these questions in the context of Tokyo's rail network. We ask whether the welfare gains from Tokyo's train system exceed its construction and maintenance costs, and to what extent the system centralizes population and employment in the city.

To evaluate the impact of Tokyo's rail network, we adopt a class of urban commuting models in which households choose where to live and work and incur commuting costs. We estimate these models using detailed spatial data on population, employment, and land prices, together with commuter flows and travel times between locations across multiple travel modes, and we combine the model with estimates of construction costs, land costs, and operating expenditures for Tokyo's rail network.

We then use the estimated models to quantify how welfare and the spatial distribution of residence and employment respond to changes in the train network. As a baseline, we compare equilibrium outcomes in Tokyo with and without its rail network, holding travel times by non-rail modes fixed at their observed levels. We then study how the effects of rail vary with (i) the placement of lines—dense core subway lines versus longer radial commuter lines, (ii) the overall extent of the rail network, and (iii) citywide population density and automobile travel times. These density and car-speed experiments speak to the external validity of our results for lower-density, car-oriented U.S. cities. To interpret these counterfactuals, we consider a simple model-based statistic that summarizes the direct effect of changes in commuting costs on commuting access, and use it to characterize how changes to Tokyo's rail network affect where people live and work.

Three findings emerge. First, Tokyo's rail network centralizes both population and employment and generates large net benefits. Removing the train system reduces the center's share of Tokyo's population and employment, and household welfare gains from the full system are greater than the present value of construction and maintenance costs. Second, network design matters at least as much as network scale. The core subway network, though only a small share of total track and expenditure, accounts for most of the system's centralizing effect on population and employment and a disproportionate share of its welfare gains relative to construction costs. Along Tokyo's historical expansion path, marginal benefits generally rise as the network grows, but typical investments in core subway lines yield returns comparable to, or higher than, even the most productive phases of

[1] For example, New York City added three new subway stations in 2017 at a cost of $4.5 billion for stations with roughly 200,000 riders per day.

overall network expansion, underscoring that where new capacity is added is at least as important for welfare as how much is built. Third, the gains are sensitive to automobile travel speeds and population density. Raising car speeds to typical U.S. levels decentralizes Tokyo and substantially reduces the welfare gains from the network. If, in addition, Tokyo's population density is exogenously reduced to New York–like levels while the costs of construction and maintenance are held fixed, the system remains socially beneficial. At Los Angeles–like density, net returns turn negative, highlighting high population density and the absence of alternative modes that encourage dispersion as important factors for successful rail investment.

Tokyo provides an especially stark environment for studying the impact of rail systems. The broader metropolitan area we study is home to roughly 38 million residents, making it one of the largest urban areas in the world, and is served by an extensive, heavily utilized rail network. In our data, rail accounts for slightly more than half of commute trips across the metropolitan area and roughly 70% of commute trips for residents in central Tokyo. The overall rail network comprises about 3,900 km of track, of which the subway system accounts for around one tenth, concentrated within roughly 10 km of the Imperial Palace. Despite its limited share of total track length, the subway carries about 9.3 million passenger trips on a typical weekday (almost a quarter of total train-system ridership[2]), making it one of the highest-ridership subway systems in the world.[3] These features make Tokyo a canonical example of a dense, rail-oriented metropolitan area and offer lessons for policymakers and urban planners.

We study the train system using a class of quantitative urban commuting models introduced by Ahlfeldt et al. (2015), in which households choose where to live and work and incur commuting costs. In our baseline specification, households trade off wages, residential amenities, housing costs, and commuting costs when selecting residential and workplace locations, and commuting costs depend on travel times and access to different modes, including rail, car, bus, cycling, and walking. Residential amenities vary by location and may depend on local population, generating residential agglomeration forces. Firms in each location combine labor and floorspace to produce a common final good that is freely traded across locations, with productivity that varies by location and may depend on local employment, generating production agglomeration forces. Floorspace is shared efficiently between residential and commercial uses and is supplied by a competitive construction sector with a constant elasticity of supply.

We estimate the model using detailed spatial data on commuting flows and travel times by origin, destination, and mode, as well as the spatial distributions of population, employment, and land prices. Commuting flows and travel times pin down workers' sensitivity to travel times and their preferences over residence, workplace, and mode. Combined with information on floorspace expenditure shares and supply elasticities, these data allow us to recover location-specific amenities,

[2]According to the government's Metropolitan Transportation Census, the sum of weekday ridership across all train operators in the metropolitan area was approximately 40 million in 2015. This figure overstates ridership due to double counting passengers who used multiple operators on a single trip.

[3]For comparison, average weekday ridership on the New York City subway was about 5.5 million in 2019, and the London Underground carried roughly 3.5–4 million passengers per weekday over a similar period.

productivities, quality-adjusted land, and location's commuter access, and to estimate agglomeration forces using variation from the expansion of Tokyo's rail network between 1995 and 2020.

Equipped with these estimated objects, we simulate counterfactual equilibria under alternative transport systems and evaluate their implications for welfare and the spatial distribution of population and employment.

The first set of counterfactuals addresses net benefits and centralization. We compare the observed equilibrium in Tokyo to a counterfactual in which the rail network is removed, holding the road network, non-rail travel times, and total population fixed. In our preferred specification with flexible floorspace shared efficiently between commercial and residential uses, this experiment reduces the center's shares of total population and employment by about 9 and 10 percentage points (-22% and -15%), respectively, and substantially lowers the log population and employment density gradients. Household welfare gains from the existing rail network are 9–16% in Greater Tokyo, corresponding to 1.8–5 times the present value of construction and maintenance costs at discount rates between 2–6%. When we allow Tokyo's population to migrate to the rest of Japan under the assumption that the system is financed nationally, total population in the metropolitan area falls by nearly one third. Together, these results imply that Tokyo's train system contributes significantly to the spatial concentration of economic activity both by shifting population and employment toward the core and by raising the overall population of the metropolitan area.

To interpret this counterfactual, we use simple model-based measures of the *direct effects* of rail-induced commuting-cost changes on residents and firms. These two measures summarize, for each location, how the rail network changes commuting costs for residents commuting from that location and for workers commuting to firms in that location, respectively, in the baseline equilibrium. The two measures explain 70–98% of the cross-location variation in the total model-implied population and employment responses to removing Tokyo's train system and show that commuting-cost improvements are disproportionately concentrated in the core, particularly for trips within central Tokyo. These patterns help account for the system's centralizing effect.

Motivated by the disproportionately large commuting-cost shocks for commutes within central Tokyo, the second set of counterfactuals examines how network design shapes welfare and centralization. We focus on the contrast between core-focused subway lines and radial commuter lines in the suburbs. We first remove individual subway segments in the reverse order of historical construction and measure the associated welfare loss relative to the segment's construction and operating expenditures. For each subway segment, we then construct a counterfactual in which the subway segment remains in place and we instead remove radial segments of similar length, deleting stations from the outskirts inward until reaching the 20 km core, again in reverse construction order. These experiments show that subway segments generate substantially larger welfare gains and stronger centralization of population and employment than similar track lengths of radial segments. Although the subway network accounts for about 10% of total track length and roughly 14% of estimated construction and related expenditures, it delivers about 37% of the welfare gain from the entire system (around 60% when construction costs are excluded), and the benefit–cost

ratio of the subway network is roughly six times that of a similar track length of radial lines. In terms of spatial patterns, removing the subway network reduces the share of Tokyo's population and employment within 20 km of the center by 5 and 5.7 percentage points (−12% and −10%), implying that the subway accounts for a majority of the train system's total centralization effects on population and employment.

We then relate these segment-level findings to the evolution of the network along Tokyo's historical expansion path. Marginal benefit–cost ratios for 20-year construction periods tend to rise as the network expands and are above one for most periods, peaking around 15. By comparison, segment-level benefit–cost ratios for subway investments average about 16, whereas radial investments yield average ratios similar to or below the historical mean. Thus, while network scale matters and marginal returns rise as the network expands over much of the twentieth century, the composition of investment between core subway lines and peripheral radial extensions is at least as important for welfare as further increases in total track length.

A third set of counterfactuals examines how these conclusions depend on population density and car travel conditions. We first increase car travel speeds toward typical U.S. levels while holding the rail network fixed. This change reduces the shares of residents and employment within 20 km of the center by 9 percentage points (−19 %) and 3 percentage points (−6 %), respectively, and lowers welfare gains from rail by about 40%, as a more dispersed city makes less use of trains. We then vary Tokyo's population density by scaling total population while holding the rail network and construction and operating costs fixed, so that a smaller population must cover the same fixed costs of building and maintaining the system. At New York–like density, roughly half of Tokyo's density, the rail system remains socially beneficial and passes a cost–benefit test at our baseline discount rate. At Los Angeles–like density, roughly a tenth of Tokyo's, net returns become negative, with welfare gains falling well below the present value of construction and operating costs.

Taken together, these results imply large net benefits from Tokyo's rail system, with returns that depend on network placement, the level of automobile infrastructure, and citywide population density.

Related Literature

This paper contributes to a number of literatures. First, this paper relates to the broad literature on the economic impact of transportation infrastructure, particularly commuting infrastructure. These include papers from the quantitative spatial literature that have developed and applied methods to analyze the economic impact of transportation infrastructure including Ahlfeldt et al. (2015), Tsivanidis (2023), Fajgelbaum and Schaal (2020), Allen and Arkolakis (2022), Monte et al. (2018), Redding and Sturm (2008), Donaldson (2018), Donaldson and Hornbeck (2016), Fajgelbaum et al. (2023), Bordeu (2024), Severen (2023), Gorback (2024), Miyauchi et al. (2025), Donald et al. (2025), Conwell et al. (2023), Zárate (2022), among others. They also include empirical papers that have used quasi-experimental methods to study the impact of transportation infrastructure on economic outcomes, such as Duranton et al. (2020), Baum-Snow (2007), Baum-Snow et al.

(2017), Asahi et al. (2023), Berger and Enflo (2017), Banerjee et al. (2020), and Faber (2014), and papers such as Brooks and Liscow (2021) that study the costs of infrastructure. Studies on Japan include Bernard et al. (2019), which studies how high-speed passenger rail improved firms' buyer-seller linkages, and Kazunobu et al. (2021), which studies the impact of high-speed rail on the distribution of economic activity across Japan. Miyauchi et al. (2025) provide empirical evidence for Tokyo using smartphone data and theory to study how trip chains (trips with intermediate stops) generate consumption externalities that imply strong agglomeration forces for economic activity in the city center. The authors find that trip chains offer an additional rationale for transportation improvements toward the city center, providing evidence that trip chains disproportionately occur on lines that connect central locations to other central locations.[4]

Relative to this literature, the main contribution of this paper is to apply a class of quantitative commuting models (those originating from Ahlfeldt et al. (2015)) to evaluate a large urban rail system and quantify its welfare impacts, its centralizing effects on population and employment, and the role of network design (core versus radial lines), density, and car travel conditions.

Second, this paper is related to the literature that studies the impact of transportation infrastructure (particularly commuting infrastructure) on patterns of spatial centralization of economic activity. Previous studies have mostly emphasized the decentralizing impact of new commuting infrastructure. Glaeser and Kahn (2001) document that American cities first began to decentralize in the late 19th century as the first commuter trains and streetcars allowed workers to move their residence to the suburbs, and that decentralization accelerated in the 20th century with the proliferation of the automobile, first for population, and then in the latter half of the century, for employment.[5] Baum-Snow (2007) uses the construction of the interstate highway system as a natural experiment and finds large decentralization in population from central cities. Baum-Snow et al. (2017) find large decentralizing impacts of radial and ring-road highways on population across cities in China. Gorback (2024) finds that entry of UberX across U.S. cities led to a dispersion in non-tradable amenities within cities.

The extent to which commuter rail in urban areas centralizes population and employment has been less clear-cut. Heblich et al. (2020) find that the advent of the steam railway in the 19th century facilitated the large-scale shift in residential population to the suburbs and concentration of employment in the city center in absolute numbers.[6] At the same time, they find that the highest percentage growth rates in both employment and population occurred in the suburbs. Gonzalez-Navarro and Turner (2018) find little evidence of urban population growth from subways, while Baum-Snow et al. (2017) find little evidence of population shift to the suburbs from radial railroads.

Relative to this literature, the contribution of this paper is to provide model-based evidence of commuting infrastructure that centralizes both population and employment, driven by a dense

[4]See chapters by Donaldson (2025), Redding and Turner (2015), Redding (2025), Redding and Rossi-Hansberg (2017) for reviews and Fogel (1970) and Meyer (1965) for classical texts.

[5]See also *Crabgrass Frontier* by Kenneth Jackson for a thorough historical account.

[6]The authors also find similar patterns of shifting population from downtown to the suburbs in other large metropolitan areas following the transport improvements of the 19th century, including Berlin, Paris, Boston, Chicago, New York, and Philadelphia.

subway network in the core of the city.

Third, this paper relates to the literature on the welfare impact of urban rail. Beginning with the classic texts of Meyer et al. (1959) and Meyer (1965), a large literature has documented that since the advent and proliferation of the automobile (beginning with Henry Ford's Model T in 1913) and the road infrastructure that complements it, there has been a decline in population density in urban areas (especially in the U.S.), threatening the economic viability of urban train systems, which require large fixed costs of construction. More recent works by Baum-Snow et al. (2005), Gonzalez-Navarro and Turner (2018), Glaeser and Kahn (2001), Glaeser and Kahn (2004), and Glaeser et al. (2008) have also argued in a similar vein. Severen (2023) study Los Angeles's Metro Rail and find that the commuting gains are between 2–64% of the annualized cost by the year 2000 for their baseline estimates. Gupta et al. (2020) evaluate the expansion of New York City's subway network (the most utilized subway system in the U.S.) and find that the value of the expansion, as measured by the increase in property values, is between 53–177% of construction cost.[7] Thus, recent studies of urban rail in the U.S. have found mixed returns on investment. In contrast, Heblich et al. (2020) who study the impact of London's train system in the 19th and early 20th centuries, a period prior to the proliferation of automobiles, find welfare gains from urban rail that substantially exceed their construction costs.

Relative to this literature, we provide evidence from a setting where rail has high utilization, quantify welfare gains that substantially exceed construction and operating costs, and assess how far such gains could be replicated in lower-density, car-oriented U.S. cities.

Fourth, this paper is related to theoretical work that studies comparative statics for gravity-based spatial models. Allen et al. (2020) provide first-order comparative static expressions and Neumann series expansions for a wide class of constant-elasticity trade models (e.g., Arkolakis et al. (2012)), which the authors refer to as "universal gravity" models. Allen and Arkolakis (2025) relate the relevant parameters necessary for these comparative statics to reduced-form "local" and "global" elasticities. Kleinman et al. (2024) show that first-order comparative statics for productivity shocks in a constant-elasticity economic geography model can be represented using a friend-enemy matrix that summarizes each location's exposure to productivity shocks in all locations and analyze the comparative statics's Neumann representation. They also analyze the first-order approximation error of such models.[8]

Relative to this literature, the contribution of this paper is to derive theoretical bounds on the general-equilibrium remainder in a canonical commuting model, clarifying when model-based direct-effect proxies provide reliable guidance for comparative-static inference without solving the full general-equilibrium system. We note that in a companion paper, Bamba (2025) show how the methodology can be applied to models in the family of constant-elasticity spatial models, such as models considered in Arkolakis et al. (2012), Allen et al. (2020), and whose equilibrium existence and uniqueness properties are studied in Allen et al. (2024).

[7] They also note that the government captures only 31% of this value through higher property-tax revenue.

[8] See also the working paper version, Kleinman et al. (2020).

The remaining sections are as follows. Section 2 discusses the historical background, Section 3 discusses the data, Section 4 presents the commuting models, Section 5 discusses the estimation and calibration of the model, Section 6 presents the counterfactual simulations of removing Tokyo's train system, Section 7 analyzes the direct effects of the train system and constructs proxies and theoretical bounds on the comparative statics, Section 8 analyzes the role of Tokyo's subway system and the extent to which our findings can be translated to New York City or Los Angeles, and Section 9 concludes.

2 Background

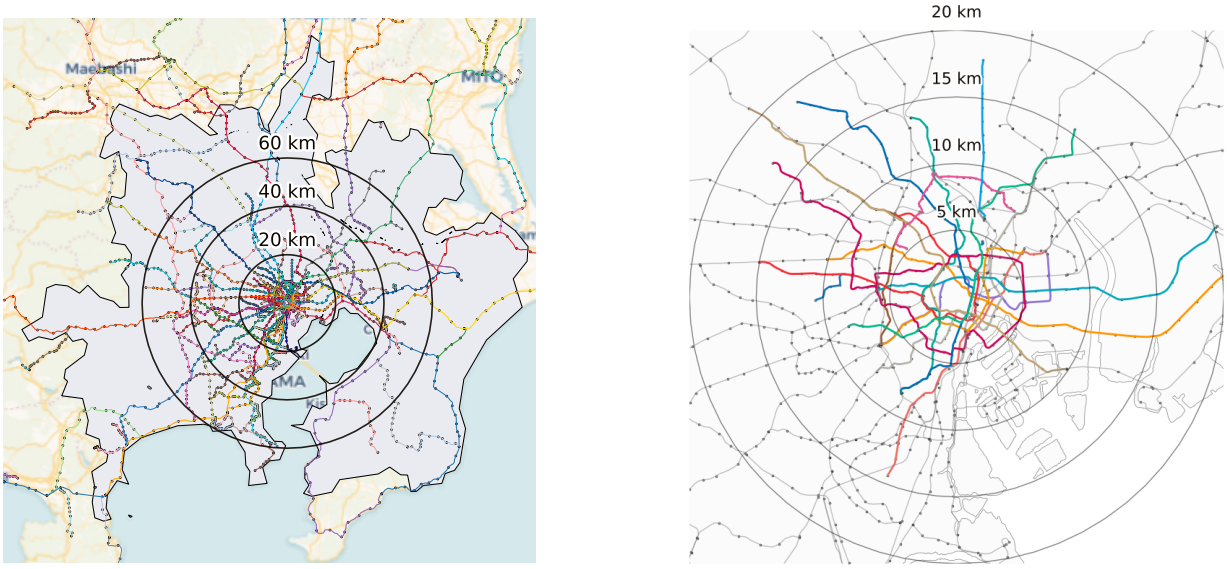
This section summarizes four features of Tokyo that are relevant for our evaluations of welfare and centralization of Tokyo's train system: the spatial boundary we study, the evolution of population concentration, the geography and timing of rail construction, and the institutions that govern and finance the system.

Spatial boundary and city rings. We study the Greater Tokyo Area using the Japanese government's commuting-based definition of the metropolitan area, which extends roughly 80 km from the Imperial Palace (Figure 1). This definition includes Tokyo prefecture and surrounding municipalities in Kanagawa, Saitama, and Chiba that are integrated with the core through commuting and school-going flows.[9] To organize spatial comparisons, we define *central Tokyo* as locations within 20 km of the Imperial Palace, the *suburb* as 20–40 km, and the *periphery* as beyond 40 km. These rings align with the official definition of the core: the inner 20 km covers Tokyo's 23 special wards, the "central city" in the commuting-based metropolitan definition (Figure 1).

Population growth and the evolution of centralization. Greater Tokyo is unusually large and dense among metropolitan areas: it has roughly 38 million residents in 2020 and a population density around 2,900 people per square kilometer. High population density raises the potential surplus from fixed-cost transit infrastructure by spreading construction and operating costs over many riders. At the same time, Tokyo's spatial concentration has evolved substantially over the last century. Figure 2b plots population density by distance to the center in selected years. The gradient flattens between the prewar and postwar periods, reflecting outward population shifts that coincide with rising incomes and automobile usage. Because these forces move together with transport investment, reduced-form correlations between rail expansion and decentralization are difficult to interpret, motivating our counterfactual analysis.

The timing and geography of rail construction. Tokyo's modern rail system is the cumulative outcome of construction that spans both the prewar and postwar eras. Figure 2a overlays

[9] Since the 1975 national census, the "metropolitan area" definition combines a "central city" (Tokyo's special wards and designated cities) with surrounding municipalities that are contiguous and have a commuting-and-school-going share (age 15+) into the central city of at least 1.5% of resident population, plus municipalities enclosed by these areas.



(a) Greater Tokyo and Train Network

(b) Subway network in Central Tokyo

Figure 1: Tokyo's Train Network

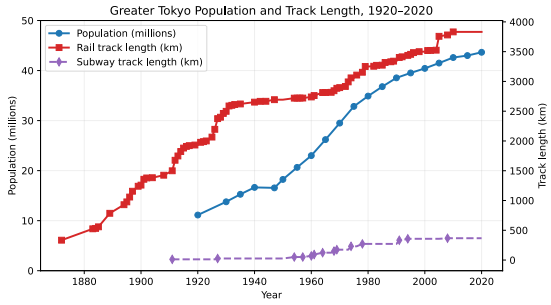
Notes: Numbers indicate the distance of the corresponding ring from the Imperial Palace. Colored and black lines are train lines and black dots are stations. Shaded region is the Greater Tokyo Area. Data is from the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT)

Greater Tokyo's population (from 1920 onward) with cumulative route-kilometers of the *current* rail network, dated by construction years.[10] Figure 2a shows that many of today's rail corridors were established before the postwar population boom: by 1950, cumulative corridor route-kilometers already account for a substantial share of the modern network, while the metropolitan population remains far below its current level. This timing suggests that rail did not simply expand in response to postwar population growth and highlights the need to separate the contribution of rail from contemporaneous highway construction and automobile adoption. Within the core, the composition of rail service also changed over time: surface tram service dominated early in the twentieth century, while subway construction and expansion accelerated in the postwar period as road congestion increased and demand for high-capacity rapid transit grew. Together, these patterns imply that much of the network footprint was built prewar, while the postwar period disproportionately builds transit in central Tokyo, shaping access within the core.

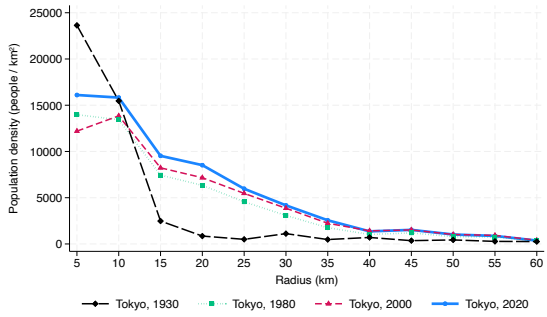
Tokyo's network is distinctive in its core–radial structure (Figure 1). Employment in central Tokyo is organized around multiple centers within the inner 10–15 km rather than a single point. Radial commuter rail lines bring commuters from suburban corridors into a small set of gateway stations in central Tokyo, while the dense subway grid and the *Yamanote Line* distribute travelers across these employment centers. Central Tokyo is also densely residential: the inner 20 km contains 39% of Greater Tokyo's residents (and 53% of its employment). Reflecting heavy use by both

[10]This series dates a line extension to the initial segment's opening year, so it summarizes the timing of corridor formation rather than the exact historical route-kilometers in each year.

inbound and within-core trips, the subway accounts for roughly one-quarter of rail trips despite comprising about one-tenth of track length. This distinction has implications for our counterfactuals: changes to the core network can affect access across central employment centers and the commuting options of residents located in the core, while changes outside the core primarily affect long-haul access from suburban residential areas to central Tokyo workplaces.



(a) Historical Track Length and Population of Greater Tokyo



(b) Population density by distance to CBD

Note: Panel (a) plots historical population of the Greater Tokyo Area overlayed with the historical track length of the overall train system and subway. Panel (b) plots population density by ring radius to the Imperial Palace calculated using municipality-level data. For example, the figure for 10 km represents the population density for municipalities with centroids within 5 km and 10 km from the Imperial Palace. See Section C.8.2 for more information.

Figure 2: Population Patterns and Train Construction Over Time

Institutional organization and project finance. Rail in Greater Tokyo is provided through a public–private system in which multiple operators build and run lines, while national and local governments approve projects, set key regulations, and coordinate across operators in an integrated network. Projects rely on a mix of public support (loans, subsidies, and public investment), retained earnings from fare revenue, and station-area development mechanisms often described as *land-value capture*. These land-value-capture mechanisms include coordinated land development around stations, regulatory changes that permit denser construction, and revenues from real-estate and commercial businesses operated by rail companies in station areas. These denser developments permit higher population and employment around train stations, leading to potentially higher welfare impacts of the train system.

3 Data

Our baseline counterfactuals and travel-time calculations use a unit of analysis known as *oaza* in Japanese that is roughly the same size as a US census tract (and hence we will use the term "census tract"), with an average employment of approximately 2000 residents. There are approximately 8700 census tracts in the Greater Tokyo area. We use this level of aggregation because most counterfactuals of interest (e.g. removing segments of the train system) would have less concerns regarding the measurement of commuting costs for commutes within a census tract and to census

tracts close-by. Because our railroad data contains 2286 train stations in the Greater Tokyo area in 2020, this geographic definition allows us to typically contain at most one train station in a geographic unit, and thus accurately measure how commuters connect to the the train network.

The general strategy for calculating commuting costs is to calculate the travel time on a particular mode between the centroids of each region.[11] If commuting costs were measured at a higher level of aggregation (e.g. between centroids of municipalities, for which there are 254), there would be a challenge in measuring current commuting costs and predicting counterfactual commute costs since households would use multiple train stations and determining the average commute cost associated with commutes from such a location requires taking a weighted average of commuting times across the multiple stations that are used. Thus, even if the counterfactual analysis were done at a higher level of aggregation, accurate counterfactual predictions of changes to commuting costs would require travel time information using geographical units comparable to the catchment area of a train station (e.g. with each region ideally containing at most one station). Summary statistics in Table 2 show that over 75 percent of municipalities contain more than one train station while less than five percent of census-tracts contain more than one train station.

Another reason for this unit of analysis is that it is likely large enough that the model's assumption that agglomeration externalities do not spill over to neighboring districts is reasonable. Ahlfeldt et al. (2015) who use block-level data to estimate the strength of spillovers to neighboring blocks find rapid decay in spillovers that amount to a half-life of roughly two minutes for production externalities and one minute for residential externalities. The average travel time between neighboring census tracts in Tokyo is approximately 6 minutes by foot and therefore agglomeration spillover to neighboring tracts are likely small.

There are four main datasets that are used to calibrate the quantitative model, in addition to several supplementary data sources that are used for estimation, calibration, and validation of the model. We describe each data set in turn and include additional details in the appendix.

First, population data by census tract comes from the Japanese Population Census. The census has been conducted almost every five years since 1920 in order to provide statistics on the state of Japan's population and households. This data is available at the block level from 1995, which is aggregated to the census tract level for our analysis.

Second, employment by region comes from two sources – the Economic Census for Business Activity, which has surveyed all establishments and companies every five years since 2012, and the Establishment and Enterprise Census, which occurred every five years before it was replaced by the Economic Census for Business Activity. These datasets, which are available at the block level from 1996, are also aggregated to the census-tract level for our analysis.[12]

Third, the primary source of land price data is the Official Land Price Announcements, which is published annually by the Japanese Ministry of Land, Infrastructure, Transport and Tourism

[11] If a centroid is in an inaccessible location (e.g. in the mountains), then we alter the location to a nearby road that is accessible).

[12] The 1995 Establishment and Enterprise Census block-level data was digitized and provided to us through the University of Tokyo's Center for Spatial Information Science.

(MLIT). Land prices across Japan are collected by real estate appraisers hired by the national and local government to provide policymakers with reliable land values that are important for calibrating national and local property tax levies. This data is supplemented by prefecture-level surveys that are conducted similarly by local governments. Merging both datasets results in a panel of 11288 land plots across the Greater Tokyo Area in 2020, which we spatially interpolate to infer land prices at the census-tract level. Additional details can be found in subsection C.5. The full dataset is from the years 1983-2020 while we utilize the time variation between 1995 and 2020.

Table 2 reports summary statistics for population, employment, and land prices at both the municipality and census-tract levels.

Fourth, we use Tokyo's Personal Trip Survey for data on commuter flows and travel times between origin and destination districts by mode. The survey is of roughly 300,000 households (about 700,000 individuals) in the Greater Tokyo Area. The data provides commuter flows and travel times by mode, but commuters often take multiple modes to reach their destination. The survey therefore uses a hierarchy of modes to determine the "primary mode of transport" for each trip. The hierarchy is as follows: train, bus, car, bicycle, and then foot. Thus, if a trip involves taking a train and a bus, the trip is classified as a train trip for the *primary mode*.[13] We use these definitions of modes as given from the travel survey in our gravity estimation and model calibration. Based on these definitions, we present statistics on commuting in Tokyo in Table 1. The survey finds that 54 percent of all commutes to work are by train, 23 percent by car, 10 percent by bicycle, 7 percent by foot, 2 percent by bus, and 3 percent by other modes. For households living in central Tokyo, 70% of work commuters take the train while 5% take the car. Thus, the train is utilized significantly more for households in the center, than in the periphery.

The average commute time by train is 63 minutes, about double the time for car (28 minutes), and quadruple the times for walking (15 minutes) and cycling (17 minutes). That trains have both a higher commuting share and longer commuting times than other modes suggests distinct demand shifters systematically raise the attractiveness of rail relative to other modes and motivates our methodology to estimate heterogeneous preferences for travel across the different travel modes.

Mode	% of Commutes		Average Commute Time (min)	
	Overall	Central Tokyo	Overall	Central Tokyo
Train	54%	70%	63	59
Car	23%	5%	28	36
Cycle	10%	12%	17	17
Foot	6%	8%	16	15
Bus	2%	3%	38	38

Table 1: Commuting in Tokyo

[13]For trips involving train, the survey found that travel between the origin or destination to the train station occurs 80 percent by foot, 8 percent by bus, 6 percent by bicycle, and 2 percent by car.

Travel Times

Travel times reported in the survey are those reported by commuters as their commuting times between their home and workplace. We believe these times are a more accurate measure of workers' commute times than those that we compute using GIS tools (discussed next) and subject to less systematic bias. Our general strategy for determining travel times in estimation and counterfactuals is therefore to use the survey travel times wherever possible to calibrate our GIS tools such that our calculated times are consistent with what is reported in the travel survey. To calculate these travel times using GIS tools, we take the following steps. First, we calculate travel times for cars, cycling, and walking between the centroids of origin-destination pairs using OpenRoute Services(ORS), an open-source tool that utilizes the street network provided by OpenStreetMap to calculate the shortest routes between locations and their associated travel times.

Next, we utilize data on the train network provided by MLIT. The data includes GIS shapefiles for the train lines and train stations, in addition to the beginning and ending dates of operation of segments of train lines from the 1800s to the present. We exclude high-speed rail lines (shinkansen) from our analysis, as they account for a negligible share of commuting.[14] We use the GIS shapefiles to construct a network on NetworkX, a network analysis library in Python, and construct minimum travel times between the centroids of origin-destination pairs for trains.

We then adjust the travel times from our GIS tools to match the travel times reported in the survey. Additional information can be found in Section C.

Train Cost Data

To calculate the costs associated with having a train system, we seek information about construction costs (inclusive of land), maintenance costs, and operating costs. To be conservative in our estimates, we calculate an upper bound on costs whenever we are uncertain about the cost calculation.

We use data provided by Transit Costs Project from the NYU's Marron Institute. This data provides the construction costs from recent train projects, gathered from trade publications, reports, and confirmed with media sources. Table C.2 lists the transit projects and their respective construction costs. The average above-ground project cost 300 million US dollars per kilometer at 2023 prices (PPP adjusted), while projects that involved substantial tunneling cost 471 million US dollars per kilometer. These figures are roughly around the median of construction costs across countries that the Transit Project has considered at the time of this writing, the highest being New Zealand at about 1 billion US dollars per kilometer and the lowest being Portugal at about 100 million US dollars per kilometer. We assume that construction costs (excluding land costs) scale linearly with track length. In counterfactual exercises where subways are removed, we use the average cost per kilometer from projects that required tunneling (471 mm in 2023 USD), while for non-subway lines we use the cost estimates for projects without tunneling (300 mm in 2023 USD).

[14] See Bernard et al. (2019) for a study on the impact of high-speed rail on firms' buyer-seller linkages and Kazunobu et al. (2021) for a study on the impact of high-speed rail on the distribution of economic activity across Japan.

In theory, the construction cost figures should be inclusive of land-acquisition costs. However, there is the possibility that little land was acquired because the government owned the land already. To account for the opportunity cost of land, we therefore add estimates of the value of land that is used for the train system. To do so, we determine the amount of land used by the train system across locations, using data provided by the MLIT. The Geospatial Information Authority of Japan uses satellite images (in conjunction with preexisting maps) to categorizes each 50m-50m square of land in Japan as rice field, other farming, forest, waste land, building, road, train, river or lake, shore, ocean, golf course, or other, and provides a 1km-1km mesh map of Japan with the area of each of these categories. We use this data to determine the land area used by the train system in each region that is then multiplied with the (previously discussed) land price to determine land values.[15] The total value of land associated with Tokyo's train system is 14t yen 140B USD using PPP rate of 94 yen to the dollar, or an average of 2.7b yen (27m USD) per kilometer of track. This land-value amount exceeds the value of construction costs, consistent with our suspicion that construction-cost data does not fully account for land costs.

In theory, if passengers are paying fares to train operators that pay for operating costs, then operating costs should be captured in our model-based measure of commuting costs. However, Japan has a hidden subsidy on train commuting: employer-paid commuting costs (e.g., train passes) are excluded from an employee's taxable income—up to a monthly cap that is approximately five times higher for public-transit use than automobile use. As a result of this system, most employers in Tokyo opt to pay for their employees public-transit commutes (with limitations), and therefore many households may not internalize the monetary cost of public transit in their location choices. Given the apparent subsidy on public-transit use and to be conservative with our cost estimates, we include the operating costs of train operators into our calculation.

Operating and maintenance costs for Tokyo's main subway operator (Tokyo Metro), five other subway lines, and two above-ground train lines were taken from the operators' financial statements. To extrapolate to the entire train system, we run a regression of these costs on the number of stations and the track length, and then extrapolate to the entire train system, which has 5200km of track and 2,286 stations. Since most of the data in the regression are subway lines, which have higher traffic than peripheral lines per unit of track or station than the typical non-subway line and hence higher costs, we view this extrapolation as a conservative estimate of costs (e.g. likely higher than the actual amount). Annualized costs as a percentage of Tokyo's GDP are shown in Table 3. Additional information on data sources and cost calculations can be found in the appendix Section C.

[15]This train category also includes track that is dedicated to freight. However, freight-only tracks accounts for about 1% of track in the Tokyo area, so we believe this measurement error would not contribute significantly to our measure of land value.

4 Urban Commuting Models

Commuting Model à la Allen and Arkolakis (2022)

We begin with a canonical model of commuting that maintains the core features of those of the remaining models considered in this study.

Locations in a city are indexed by $i \in \mathcal{N} \equiv \{1, \ldots, N\}$ and differ in their primitives by their exogenous components of amenities $\bar{U}_i$, productivities $\bar{A}_i$, and exogenous commuting costs τ_{ijm} experienced by workers commuting from residence location i to workplace location j using a particular mode $m \in \mathcal{M}$. There are five modes: $\mathcal{M} = \{\text{trains, cars, walking, buses, cycling}\}$.

A continuum of households with mass $\bar{L}$ who are perfectly mobile choose where to live and work and what travel mode to commute between their residence and workplace. They earn a wage by supplying their labor inelastically in their workplace j and have a preference for a freely traded numéraire good. Their utility is augmented by the amenities of their residence location, which are endogenous, and diminished by commuting costs τ_{ijm}.

Setting the price of the final good to one, the indirect utility for an individual living in location i, working in location j and commuting on mode m is given by:

$$V_{ijm}(\omega) = \frac{U_i w_j \epsilon_{ijm}(\omega)}{\tau_{ijm}} \tag{1}$$

where $\epsilon_{ijm}(\omega)$ is an idiosyncratic preference shock for worker ω commuting between residence i and workplace j using mode m, U_i is the level of amenities enjoyed in residence i, and w_j are the wages earned in workplace j. In the baseline model, we assume ϵ_{ijm} is distributed i.i.d. Fréchet with unit scale and a shape parameter θ. In our baseline specification, commuting costs are parametrized to follow an exponential function:

$$\tau_{ijm} = \exp(\kappa^O_{im} + \kappa^D_{jm} + \kappa_m t_{ijm}) \tag{2}$$

where t_{ijm} is the travel time for the commute from i to j on commuting mode m; κ_m governs the elasticity of commuting costs with respect to time for mode m and is expressed in units of 100 percent (or log) of wages per unit of time; κ^O_{im} and κ^D_{jm} are residence-mode-specific and workplace-mode-specific constants, respectively, and represent additional costs associated with a particular mode in a particular location that is not captured in travel times, such as the quality of highways or train stations in a particular location or the cost of car parking in a particular location.

The Fréchet distribution implies that commuter flows—the number of commuters between residence i and workplace j—are

$$L_{ij} = \bar{L} \frac{U_i^\theta w_j^\theta \tau_{ij}^{-\theta}}{\sum_i \sum_j U_i^\theta w_j^\theta \tau_{ij}^{-\theta}}, \tag{3}$$

where $\bar{L}$ is total employment in the economy and $\tau_{ij} = [\sum_m \tau_{ijm}^{-\theta}]^{-1/\theta}$ is the Fréchet-implied ag-

gregate commuting costs across the five modes for commutes between i and j. In our baseline specification, we assume that $\bar{L}$ is exogenous.

Each location j produces the homogeneous and costlessly traded final good, Y_j linearly with labor, augmented by a potentially endogenous productivity A_j: $Y_j = A_j L_{Fj}$.

Amenities and productivities exhibit spillovers that are local to a location. Amenities are a function of the exogenous component $\bar{U}_i$ and local residential population L_{Ri}, while productivity is similarly a function of the exogenous component $\bar{A}_j$ and local employment L_{Fj}.

$$U_i = \bar{U}_i L_{Ri}^{\phi_U}, \tag{4}$$

$$A_j = \bar{A}_j L_{Fj}^{\phi_A} \tag{5}$$

Agents take the level of amenities and productivities as given.

Commuter market clearing requires that residential population in each location equals the sum of outbound commuter flows, and that employment in each location equals the sum of inbound commuter flows:

$$\sum_i L_{ij} = L_{Fj}, \tag{6}$$

$$\sum_j L_{ij} = L_{Ri}. \tag{7}$$

The Fréchet distribution implies that the average utility of the household, $\bar{V}$ is given by:[16]

$$\bar{V} = \left[\sum_i \sum_j U_i^\theta w_j^\theta \tau_{ij}^{-\theta} \right]^{1/\theta}. \tag{8}$$

Substituting equations (8) and (3) into equations (6) and (7) imply that the supply of residents and employment in each location can be expressed as:

$$L_{Ri} = \frac{\bar{L}}{\bar{V}^\theta} U_i^\theta \underbrace{\sum_j \tau_{ij}^{-\theta} w_j^\theta}_{\substack{\text{Residential} \\ \text{Commuter Access}}} \tag{9}$$

$$L_{Fi} = \frac{\bar{L}}{\bar{V}^\theta} w_j^\theta \underbrace{\sum_j \tau_{ji}^{-\theta} U_j^\theta}_{\substack{\text{Firm} \\ \text{Commuter Access}}} \tag{10}$$

We refer to $\sum_j \tau_{ij}^{-\theta} w_j^\theta$ and $\sum_j \tau_{ji}^{-\theta} U_j^\theta$ as *Residential Commuter Access (RCA)* and *Firm Com-*

[16] Technically, the average utility for a Fréchet model includes a constant multiplier that is a function of the heterogeneity parameter θ. We omit this multiplier because the constant does not change in counterfactuals and it does not affect our welfare calculations.

muter Access (FCA), respectively, following the international trade and spatial literature (e.g. Ahlfeldt et al. (2015), Tsivanidis (2023), Redding and Venables (2004)). Residential population is increasing in RCA, which summarizes a residential location's accessibility to work locations with high wages, while employment is increasing in FCA, which summarizes a workplace's accessibility to residential locations with high amenities. These market access terms will be important for the direct-effect measures that we discuss in Section 7.

One can then define the following equilibrium:

Definition. Given $\left\{\bar{U}_i, \bar{A}_i\right\}_{i\in\mathcal{N}}$, $\{\tau_{ij}\}_{i,j\in\mathcal{N}}$ where $\tau_{ij} = \left[\sum_m \tau_{ijm}^{-\theta}\right]^{-1/\theta}$, total population $\bar{L}$, and parameters $\{\theta, \phi_U, \phi_A\}$ equilibrium variables $\bar{V}, \{L_{Ri}, L_{Fi}\}_{i\in\mathcal{N}}$ solve equations (8) (population mobility), (9) (supply of residents), (10) (supply of workers), (5), (4) (spillovers), and firm optimality: $w_j = A_j$.

Allen et al. (2024) study the existence and uniqueness properties for models of this type. In this case, one can show that when $\left\{\bar{U}_i, \bar{A}_i\right\}_{i\in\mathcal{N}}$, and $\{\tau_{ij}\}_{i,j\in\mathcal{N}}$ are positive and finite and in addition, if $|\frac{\phi_U\theta\phi_A\theta}{(1-\phi_U\theta)(1-\phi_A\theta)}| < 1$, then there exists an equilibrium that is unique.

Commuting Model with Shared Floorspace à la Ahlfeldt et al. (2015)

We augment the model above to include floorspace that is shared between commercial and residential uses, in the spirit of Ahlfeldt et al. (2015). Allowing floorspace to be efficiently shared allows for the possibility that transportation improvements induce specialization of locations as either residences or workplaces, as has been documented in numerous studies.

In this augmented model, locations continue to differ in their primitives by their exogenous levels of amenities $\bar{U}_i$, productivities $\bar{A}_i$, and exogenous commuting costs τ_{ijm}, and in addition, locations contain an exogenous amount of quality-adjusted land, D_i.

Instead of consuming only the final good, households now have Cobb-Douglas preferences over the final good and housing (i.e. floorspace). Final-good producers, too, now use floorspace: they combine labor and floorspace using a Cobb-Douglas technology, whose Hicks-neutral productivity is again endogenous to the employment in the region. Floorspace is allocated efficiently between residential and commercial land uses such that the floorspace prices are equalized between the two uses of land. Floorspace in each location is produced with a Cobb-Douglas technology using land, in fixed supply, and the final good. Rents from land are rebated to households proportionally to their wages.

Indirect utility for an individual living in location i, working in location j and commuting on mode m is given by:

$$V_{ijm}(\omega) = \frac{U_i(w_j + \pi_j)r_{Ri}^{\beta-1}\epsilon_{ijm}(\omega)}{\tau_{ijm}} \tag{11}$$

where r_{Ri} is the residential price of floorspace and β is the expenditure on the final good, π_j are the land-rents rebated to households, and $U_i, w_j, \epsilon_{ijm}, \tau_{ijm}$ are amenities, wages, preference shocks,

and commuting costs, as in the previous section. The price of the costlessly traded final good has been normalized to one.

Price-taking firms in each location produce the freely traded final good with a Cobb-Douglas technology $Y_j = A_j L_{Fj}^{\alpha} H_{Fj}^{1-\alpha}$, where H_{Fj} is commercial floor space and L_{Fj} and A_j are labor and productivity as in the previous section.

Floorspace is produced in each location by a construction sector using the final good and land with a Cobb-Douglas technology: $H_i = K_i^{\mu} D_i^{1-\mu}$, where K_i is the final good and D_i is exogenous quality-adjusted land. Floorspace is shared such that floorspace supply is equal to the sum of commercial and residential floorspace demand, and the no-arbitrage condition requires that residential and commercial floor prices be equalized if both land-uses are used. Otherwise, land is used entirely by the use (commercial or residential) that has the higher bid and equilibrium floor prices are equal to this higher bid. Rent from land is rebated to households proportionally to their income, such that $\pi_j = k w_j$, where k is a proportionality factor such that all land payments are returned to households.

Productivities and amenities exhibit spillovers as before

$$A_j = \bar{A}_j L_{Fj}^{\mu_A}, \tag{12}$$

$$U_i = \bar{U}_i L_{Ri}^{\mu_U}, \tag{13}$$

where μ_A and μ_U are the elasticities of productivity with respect to employment and amenities with respect to residential population, respectively (analogous to ϕ_A and ϕ_U in the previous section). Households and firms take A_j and U_i as given.

As in the previous section, we require that commuter flows sum to residential population at each origin (7) and sum to employment at each destination (6), while average utility is now given by:

$$\bar{V} = (1+k) \left[\sum_j \sum_i \sum_m \left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ijm}} \right)^{\theta} \right]^{1/\theta} \tag{14}$$

.

Analogous to (10) and (9), the commuter flows implied by the indirect utility (14) and the Fréchet assumption on $\epsilon_{ijm}(\omega)$, together with the commuter market clearing assumptions (7) and (6), imply that the supply of residents and workers in each location can be expressed as:

$$L_{Ri} = \frac{\bar{L}(1+k)^{\theta}}{\bar{V}^{\theta}} U_i^{\theta} r_{Ri}^{\theta(\beta-1)} \underbrace{\sum_j \tau_{ij}^{-\theta} w_j^{\theta}}_{\substack{\text{Residential} \\ \text{Commuter Access}}} \tag{15}$$

$$L_{Fi} = \frac{\bar{L}(1+k)^{\theta}}{\bar{V}^{\theta}} w_i^{\theta} \underbrace{\sum_j \tau_{ji}^{-\theta} U_j^{\theta} r_{Rj}^{\theta(\beta-1)}}_{\substack{\text{Firm} \\ \text{Commuter Access}}} \tag{16}$$

where RCA has the same expression as before while FCA now summarizes a workplace's access to locations that are attractive to live in, taking into account both amenities *and floor prices*.

Given $\left\{\bar{U}_i, \bar{A}_i, D_i\right\}_{i\in\mathcal{N}}$, $\{\tau_{ij}\}_{i,j\in\mathcal{N}}$ where $\tau_{ij} = \left[\sum_m \tau_{ijm}^{-\theta}\right]^{-1/\theta}$, total population $\bar{L}$, and parameters $\{\theta, \alpha, \beta, \mu_U, \mu_A, \mu\}$, an equilibrium can be defined as equilibrium variables $\bar{V}, k, \{L_{Ri}, L_{Fi}, w_i, r_i\}_{i\in\mathcal{N}}$ such that population mobility (14) is satisfied, and in each location households select locations optimally (15), (16), residents optimally choose floorspace and final good consumption, final-good producers optimally choose labor and floorspace, the construction sector optimally choose land and the final good (the intermediate input) as inputs, and land, floorspace, and labor markets clear. In addition, the final good market clears for the city as a whole.

Section A.1 provides additional details and formally defines the equilibrium. Ahlfeldt et al. (2015) prove that there exists a unique equilibrium whenever $\mu_U = \mu_A = 0$.

Intuition

The common intuition of a transportation improvement that is put forward goes as follows. Suppose a new commuting mode (the train system) is introduced and suppose that the city is closed. The introduction of a new mode reduces commuting costs between all locations as households now have a new choice for commuting. If in addition, suppose trains disproportionately benefits travel to faraway locations (possibly due to lower κ_m or fast travel speeds), then workers can commute for longer distances, and thus workers will reallocate to workplaces with higher wages and residence locations with higher amenities. In the model without floorspace, equilibrium congestion forces will then counteract these forces. Overall, such a model emphasizes the important of locations' *absolute advantage* in productivity and amenities: locations with absolute advantages in productivity and amenities gain employment and residents, respectively.

In contrast, the model with shared floorspace emphasizes the importance of locations' *comparative advantages*. A transportation-induced shock to employment (due to an absolute advantage in production) in a location will bid up rents, displacing residents as firms outbid households for space. Similarly, a shock to population (due to an absolute advantage in amenities) in a location will bid up rents, displacing employment as residents outbid firms for space. These two competing forces, together with a fixed total population, imply that the relative sizes of productivity versus amenities are important for the predictions of the effects of transportation improvements on popu-

lation and employment, and it is therefore common for improved transportation improvements to lead to either increased population or increased employment according to that location's comparative advantage but not both. As a result, it is typical for transportation improvements to increase specialization of locations as either workplaces or residences, based on the locations' comparative advantage.

While this intuition has been put forward and validated both theoretically and empirically (see e.g. Redding (2025) and Heblich et al. (2020)), the theory does not preclude other comparative static possibilities, as we will see in the remaining sections.

Extensions

Appendix Section B introduce a number of extensions to check the sensitivity of our results to models with additional margins that may be of importance. Section B.1 allows travel times by car to be endogenous to the vehicle traffic on the route taken between residence and workplace, following Allen and Arkolakis (2022). When trains are removed and automobiles are the main alternative, an increase in traffic congestion may raise travel times on automobiles leading to further welfare losses and additional adjustment in population and employment, which are incorporated in this section. Section B.3 relaxes the assumption of a closed economy and allows for migration into and out of Tokyo from the rest of Japan to evaluate how the overall population and land prices respond to changes in the transportation infrastructure. We also consider an alternate parametrization of commuting costs where $\tau_{ijm} = \exp(\kappa_{0m} + \kappa_{1m} t_{ijm})$, with κ_{0m} and κ_{1m} as parameters. This specification focuses on travel times as the only source of variation in commuting costs across locations. While such a specification would inevitably have a worse fit than the baseline specification, by focusing on travel times, it is less susceptible to the concern that its parameters (κ_{0m} and κ_{1m}) may be endogenous to the infrastructure that is available.

5 Estimation and Calibration

Given the parameter θ, the matrix of commuting costs, $\{\tau_{ij}\}_{i,j\in\mathcal{N}}$ and employment and population in each location, we can identify wages and rent-adjusted amenities in each location, up to a normalization.[17] Without floor-space the rent-adjusted amenities correspond to simply the amenities of a location $\{U_i\}_{i\in\mathcal{N}}$ while in the shared-floorspace model, we identify the composite: $\left\{U_i r_i^{\beta-1}\right\}_{i\in\mathcal{N}}$. The identification is similar to that of a Rosen-Roback model that is augmented with a notion of commuting access in each location. For each workplace location, conditional on the commuting costs to each residence location and the population in each of those locations, a higher number of workers (i.e. employment) in a location implies higher wages, representing the attractiveness of a workplace to work. Similarly, for each residence location, conditional on the commuting costs to each workplace location and the employment in each of those locations, a higher number of

[17] Under the assumption that land rents are rebated to households in proportion to wages, the identified wages coincide with total household income up to a constant factor.

residents (i.e. population) in a location implies higher rent-adjusted amenities, representing the attractiveness of a residence location to live in. In the model with shared floorspace, floor prices are recovered using the cost-function of the construction sector, a monotonic function of land prices. This data, together with parameter β, then recovers amenities $\{U_i\}_{i\in\mathcal{N}}$.

In the model without floorspace, the recovered wages are equal to productivity, given that production is linear in labor. When floorspace is used in production, the final-good producer's cost function recovers productivity as a function of floor prices and wages, which have been identified previously.

Given the spillover parameters ($\{\mu_U, \mu_A\}$ or $\{\phi_U, \phi_A\}$) and population and employment, one can then recover fundamental productivities and amenities $\left\{\bar{U}_i, \bar{A}_i\right\}_{i\in\mathcal{N}}$. Demand functions for floorspace from firms and residents, together with floor prices, recover equilibrium floorspace, which in turn recovers quality-adjusted land through the construction sector's conditional input demand function(as a function of land prices). Details of the identification can be found in Section D.

Thus, given parameters, we require data on the vector of employment and population, the matrix of commuting costs, and in addition, for the model with floorspace, the vector of land prices. Employment and population numbers are obtained directly from Census data. Land prices are obtained from the Official Land Price Announcements.

What remains then is to recover commuting costs and to estimate parameters of the model.

Travel times are calculated using a combination of the street map (using Openroute Services), the geography of the train network (taken from MLIT), and reported travel times from the Tokyo Person-Trip Survey. Details of the travel time calculations are discussed in Section (C.1).

We then proceed as follows in order to estimate the necessary parameters of the model. First, we use the gravity structure of the model to estimate the semi-elasticity of commuting flows with respect to travel time for each mode, corresponding to the product, $\theta\kappa_m$, as well as location-mode specific fixed effects $\theta\kappa^O_{im}$ and $\theta\kappa^D_{jm}$. Second, we use these parameters together with employment and population data, and bilateral travel times for each mode to recover transformed wages, w_i^θ, which are recoverable because they are proportional to commuter flows, conditional on commuting costs and multilateral resistance (RCA). An identifying assumption is that $\sum_m \kappa^O_{im} = \sum_m \kappa^D_{im} = 0$ for all $i \in \mathcal{N}$: in other words, the overall preference for a destination (across all five modes) is summarized by the wage of that destination. Similarly, the overall preference for residential location (across all five modes) is summarized by the amenity of that location (or rent-adjusted amenities in the model with shared floorspace). We then match the model's prediction of the dispersion of log earnings across locations with that of Tokyo to separately recover θ and κ_m, and thus wages. Intuitively, a higher θ implies less heterogeneity in preference shocks, higher sensitivity to wages, and thus lower dispersion in wages conditional on observed commuter flows. Next, we use the wages, together with floor-space prices, residential population, and travel times to back out productivity and amenity levels for every location. Last, we use the expansion of the train network between 1995-2020 and employ a difference-in-difference strategy to estimate the spillover parameters μ_U and μ_A that correspond to the elasticity of amenities with respect to residential

population and the elasticity of productivity with respect to employment, respectively.

5.1 Parameters from the Urban literature

Of the parameters in our model $\{\alpha, \beta, \mu, \theta, \kappa, \mu_U, \mu_A\}$, a subset will be taken from previous estimates in the literature. We set the share of consumer expenditure on residential floor space $(1-\beta)$ equal to 0.25, consistent with the estimates from Davis and Ortalo-Magné (2011). The parameter that governs the share of firm expenditure on commercial floor space $(1-\alpha)$ is set to 0.2, consistent with findings from Valentinyi and Herrendorf (2008). Last, the share of land in construction costs, $(1-\mu)$ is set to 0.5 to match the floorspace supply elasticity found in Nakajima et al. (2018).

5.2 Estimating parameters governing commuting costs

First we use the commuter flow data from the Person Trip survey to estimate a gravity equation of the following form, implied by equation (3) (and analogous expression for the model with shared floorspace, as derived in Appendix A):

$$\ln \pi_{ijm} = \gamma_{im} + \vartheta_{jm} - \theta \kappa_m t_{ijm} \tag{17}$$

where $\pi_{ijm} = \frac{L_{ijm}}{L}$ is the share of workers that have selected residence in location i and workplace in location j and mode m, $\gamma_{im} = \theta(\ln U_i + \kappa^O_{im})$, and $\vartheta_{jm} = \theta(\ln w_j + \kappa^O_{im})$.[18] Following the gravity-estimation literature, we assume that there is measurement error in the shares, leading to the estimating equation:

$$\ln \pi_{ijm} = \gamma_{im} + \vartheta_{jm} - \theta \kappa_m t_{ijm} + \varepsilon_{ijm} \tag{18}$$

We estimate this model using Poisson-Pseudo Maximum Likelihood (PPML) using commuting flows and travel times from the Person-Trip Survey data (where missing travel times have been imputed using GIS software, as discussed in Section 3). The commuter-flows data is at the commuting-district level, whose population is about 7-8 times larger than that of a census tract on average. Translating our estimates at the commuting-district level to the census-tract level assumes that we can approximate the average of the logarithm of commuting probabilities with the logarithm of the average, as discussed in Section D.1 and Ahlfeldt et al. (2015). In practice, Ahlfeldt et al. (2015) find that the approximation is close in the sense that it leads to a negligible difference in parameter estimates. We therefore follow the literature and employ the same procedure.

Given estimates of $\{\theta\kappa_m\}_{m\in\mathcal{M}}$, $\{\gamma_{im}, \vartheta_{jm}\}_{i\in\mathcal{N}, j\in\mathcal{M}}$, and travel times $\{t_{ijm}\}_{i,j\in\mathcal{N}, m\in\mathcal{M}}$, data on employment and population (and floor prices for the floorspace model) together with the identifying assumption that $\sum_m \kappa^O_{im} = \sum_m \kappa^D_{im} = 0$ are then used to recover transformed wages: w_i^θ for the N locations using these N equations (up to a choice of normalization), as shown in Section D.[19]

[18] In the model with shared floorspace $\gamma_{im} = \theta(\ln U_i r_i^{\beta-1} + \kappa^O_{im})$

[19] When the unit of analysis for the commuter-flows regression is the same as the rest of the analysis, an alternative approach is to utilize the destination fixed effects from the commuter-flows gravity regression, as these fixed effects are equal to log transformed wages.

We then calibrate the heterogeneity parameter θ to match the heterogeneity in log earnings by residence location across the municipalities of Tokyo. Because the Fréchet parameter applies a power function to wages, adjusting the inequality in wages across locations, we use data on the dispersion of log earnings across districts, which are a weighted average of workplace wages, to calibrate θ.[20]

5.3 Estimation of Spillovers

We estimate the agglomeration parameters for the shared-floorspace model, which is our preferred specification for welfare analysis. Once productivities and amenities are recovered, (13) and (12) relate these location characteristics to fundamentals and to the levels of population and employment:

$$\ln A_j = \mu_A \ln L_{Fj} + \ln \bar{A}_j, \qquad \ln U_i = \mu_U \ln L_{Ri} + \ln \bar{U}_i$$

where $\bar{A}_j$ and $\bar{U}_i$ denote fundamental productivity and amenity terms.

A simple OLS regression of $\ln A_j$ or $\ln U_i$ on the corresponding log employment or population is likely to suffer from confounding. A central concern is reverse causality: in the model, higher fundamental productivities and amenities raise local employment and population, so treating $\ln L_{Fj}$ or $\ln L_{Ri}$ as exogenous will generally bias the estimate of μ_A or μ_U. To address this issue, we exploit the large expansion of Tokyo's train network between 1995 and 2020 and estimate the spillover parameters using a difference-in-differences–type design.

Between 1995 and 2020, approximately 350km of new track and 263 new stations were added to Tokyo's railway network, corresponding to roughly a 7 percent and 16 percent increase in total track length and station count, respectively. Figure 10 maps this expansion. We calibrate the model separately for 1995 and 2020, recover $\{A_{j,1995}, A_{j,2020}\}$ and $\{U_{i,1995}, U_{i,2020}\}$, and work with first differences:

$$\Delta \ln A_j = \mu_A \Delta \ln L_{Fj} + \Delta \ln \bar{A}_j, \qquad \Delta \ln U_i = \mu_U \Delta \ln L_{Ri} + \Delta \ln \bar{U}_i$$

where $\Delta \ln X_j \equiv \ln X_{j,2020} - \ln X_{j,1995}$. First differencing absorbs all time-invariant unobserved heterogeneity in fundamentals at the census-tract level but leaves open the possibility that changes in fundamentals, $\Delta \ln \bar{A}_j$ and $\Delta \ln \bar{U}_i$, are correlated with changes in local employment and population. Our first IV strategy uses increases in the local number of stations as a source of variation in $\Delta \ln L_{Fj}$ and $\Delta \ln L_{Ri}$. For each census tract j and radius $R \in \{1, 3, 5\}$km, we compute the number of stations within Rkm in 1995 and 2020, denoted $S_{j,1995}(R)$ and $S_{j,2020}(R)$. We then define three indicator variables

$$Z_{1j} = \mathbf{1}\{S_{j,2020}(1 \text{ km}) > S_{j,1995}(1 \text{ km})\}, \quad Z_{3j} = \mathbf{1}\{S_{j,2020}(3 \text{ km}) > S_{j,1995}(3 \text{ km})\},$$

$$Z_{5j} = \mathbf{1}\{S_{j,2020}(5 \text{ km}) > S_{j,1995}(5 \text{ km})\}$$

[20] We then perform an over-identification check, as discussed in Section 5.6.

which equal one if the number of stations within 1km, 3km, or 5km of tract j increases between 1995 and 2020, and zero otherwise. These indicators capture whether the local station network around tract j expands over the period (allowing for the possibility that some stations close) and shift the change in employment and population implied by the model. We use Z_{1j}, Z_{3j}, and Z_{5j} as excluded instruments for $\Delta \ln L_{Fj}$ and $\Delta \ln L_{Ri}$ in two-stage least squares regressions of the form

$$\Delta \ln L_{Fj} = \pi_1 Z_{1j} + \pi_2 Z_{3j} + \pi_3 Z_{5j} + \gamma_k + \varepsilon_j$$

$$\Delta \ln A_j = \mu_A \widehat{\Delta \ln L_{Fj}} + \gamma_k + \eta_j$$

and analogously for $\Delta \ln U_i$ with population on the left-hand side and $\Delta \ln L_{Ri}$ as the endogenous regressor. Here γ_k denotes municipality fixed effects, so identification comes from comparing tracts within the same municipality that experience different increases in the number of nearby stations.

Our second IV strategy exploits variation in tracts that gain access to their *first* nearby station over the sample period. For each census tract, we construct three indicator variables that equal one if the tract receives its first train station within 1km, 3km, or 5km from the tract, respectively, between 1995 and 2020, and zero otherwise. Tracts for which all three indicators are zero either already had at least one station within 5km in 1995 or never receive a first station within 5km between 1995 and 2020; these locations form the control (omitted) group in this specification. These "first-station" indicators isolate changes in employment and population driven by newly created access to the train network, rather than marginal improvements in already well-served locations. We re-estimate the 2SLS specifications above using these indicators as excluded instruments in place of the station-increase dummies.

Intuitively, both IV strategies are similar to a difference-in-differences setup in which "treated" tracts are those differentially exposed to the network expansion and "control" tracts are less exposed tracts in the same municipality. In the first strategy, treatment intensity is measured by whether the number of nearby stations increases, so tracts with larger expansions in local station counts experience larger instrumented changes in employment and population than tracts with no such expansion. In the second strategy, treatment is measured by gaining a *first* nearby station, so treated tracts are those that switch from having no nearby station to having at least one within 1–5km, while control tracts either already had a nearby station in 1995 or never gain one over the period. In both cases, the first stage uses differential exposure to station expansions to predict differential growth in local employment and population, and the second stage relates differential changes in recovered productivities and amenities to these predicted changes in population and employment. Under the exclusion restriction that, conditional on municipality fixed effects, exposure to station expansions affects $\Delta \ln A_j$ and $\Delta \ln U_i$ only through its effect on $\Delta \ln L_{Fj}$ and $\Delta \ln L_{Ri}$ (i.e., is uncorrelated with tract-specific shocks to $\Delta \ln \bar{A}_j$ and $\Delta \ln \bar{U}_i$), the IV coefficients μ_A and μ_U recover the causal agglomeration elasticities in the shared-floorspace model.

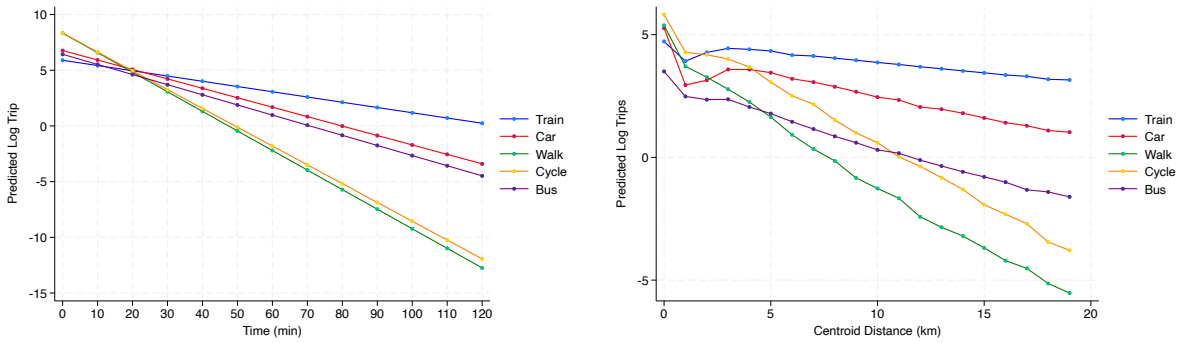
Section D.3 discusses the calibration approach for the road congestion model.

5.4 Results

Gravity results using OLS and PPML are presented in Table 4. Predicted commuter flows implied from PPML specification with mode-origin and mode-destination specific constants and mode-specific slopes are shown in Figure 3. Given these estimates, we separately identify $\kappa_{im}^O, \kappa_{im}^D, \kappa_m$ from θ by matching the model's predictions of the variance of log earnings with that of the data. We estimate the heterogeneity parameter as $\theta = 2.3$ and the implied commuting-cost parameters are shown in Table 5. The results show substantial heterogeneity both in the value of time across different modes and in fixed mode-specific costs, independent of travel time. Trains have the cheapest marginal costs with respect to a unit of travel time whereas cars cost about 76 percent more than trains per unit of time. The difference may reflect not only differences in amenity value (e.g. a commuter may choose to read on the train but will have to settle for an audiobook if they choose to drive) but also monetary costs (e.g. of fuel). A ten-minute increase in one-way commuting time on a train is equivalent to a 21 percent decrease in wages while a ten-minute increase in one-way commuting time on a car is equivalent to a 37 percent decrease in wages. On the other hand, trains have higher fixed costs than walking, cycling or driving, on average, leading to fewer predicted trips than other modes at the intercept (i.e. $t = 0$). Figure 3b shows the implied predicted probabilities by straight-line distance rather than by time. The figure shows that commuters' higher willingness to spend time on trains translate to higher willingness to travel further distances as well. Thus, introducing trains gives commuters better access to locations further away, allowing households to better separate their workplace and residence. However, we emphasize that the figure does not show the geographical distribution of where these trips take place. We note that the x-axis range is until 20 km, which is the radius that we defined for central Tokyo. Therefore, even as trains may allow better separation between workplace and residence, if trains are concentrated in central Tokyo (as discussed in Section 2), then trains may still lead to increases in both population and employment.

The results in Table 5 and our estimate of $\theta = 2.3$ are broadly consistent with previous estimates from the literature. The estimated commuting cost parameters κ_m for trains are comparable to those found in Ahlfeldt et al. (2015) and Tsivanidis (2023) of roughly 0.01. However, we estimate separate parameters for different modes and find substantial heterogeneity. The heterogeneity helps to explain why commuting times are significantly larger for train commuters than car-commuters and also for train-centered cities like Tokyo (about 60 minutes) than in car-centered cities like Los Angeles (about 30 min). The heterogeneity also has implications for welfare and the spatial impact of train removal, as residents will switch to shorter commutes leading to less separation between workplace and residence locations.

Our estimate of the Fréchet parameter is similar to those estimated in Tsivanidis (2023) of between 2.7 and 3.3 and the preferred estimate in Severen (2023) of 2.2 and lower than the estimate of 6.5 to 6.8 in Ahlfeldt et al. (2015) which use a similar estimation strategy of matching the variance of log wages. The difference may be because incorporating more than one mode requires



(a) By Time (b) By Distance Between Zones

Figure 3: Predicted Commuter Flows from Gravity Model

Predicted commuter flows from the "PPML" "Modes" specification from Table 4. The left figure shows one-way travel times on the x-axis varied from 0-120 minutes for each mode and the predicted log commuter flows are computed for each observation (origin-destination-mode) using the estimated coefficients and origin-mode and destination-mode fixed effects. The figure then reports the average predicted log commuter flows for each mode across these observations. The right figure calculates the straight-line distance between the centroids of origin-destination zones and calculates the average predicted log commuter flows for each mode for each 1-km distance bin between centroids. Predicted flows were computed using Stata's "margins" and "marginsplot" commands after estimation with the "ppmlhdfe" command.

more preference heterogeneity to rationalize the data.[21]

Table F.5 summarizes the instrumental-variable estimates for the productivity and amenity spillovers for the two IV strategies. Other results such as OLS, first-stage, and reduced-form estimates can be found in Tables F.3, F.4, F.5, and F.6. Our two IV strategies yield amenity spillover estimates between 0.35 and 0.37 that are estimated precisely (standard error of 0.01). The production spillovers are estimated between 0.28 and 0.31 (standard error of 0.01). These estimates are largely consistent with previous studies. Tsivanidis (2023) find amenity spillovers for low- and high-skilled workers of 0.42 and 0.58, while Ahlfeldt et al. (2015) find an estimate of 0.15. Tsivanidis (2023) find production spillover estimates of 0.21, while Ahlfeldt et al. (2015) estimate 0.07. The discrepancy between our parameter estimates and those in Ahlfeldt et al. (2015) is largely due to differences in the estimated preference heterogeneity parameter θ. If we constrain θ to their estimate of $\theta = 6.8$, then we instead estimate production and residential spillovers between 0.107 and 0.119 and between 0.139 and 0.144, respectively, comparable to their spillover estimates.

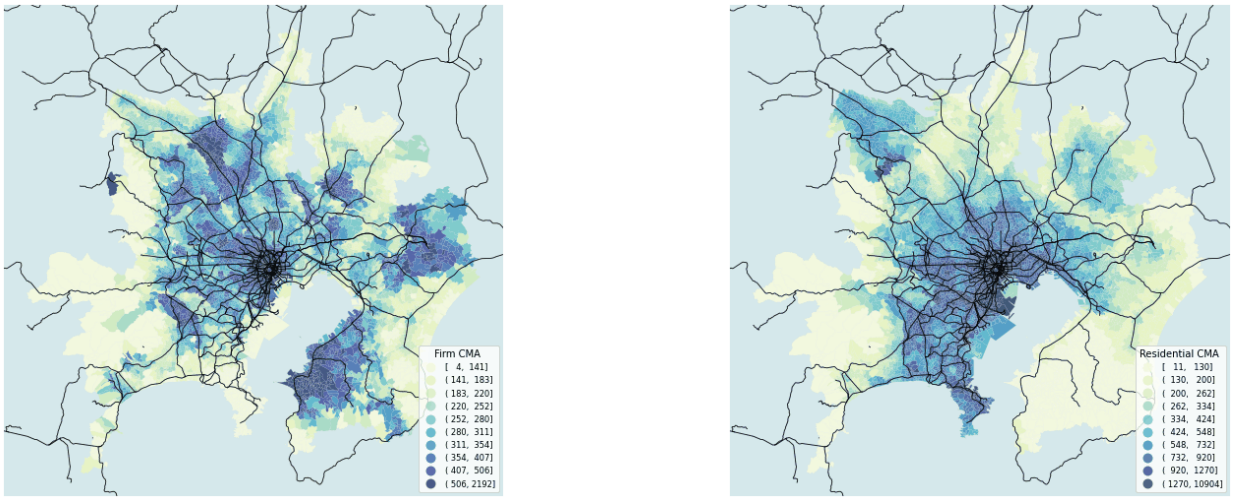
We do not have access to geographically linked tract-level census data prior to 1995, and therefore we cannot test for pre-trends with our sample. We are currently implementing a pre-trend analysis for land prices, for which we have a longer time series. However, threats to the parallel-trends assumption would require differing trends in amenities or productivities between census tracts that are in close proximity within municipalities, which are not large (approximately 50 sq. km, or an approximate radius of 3.7 km, on average). Analysis of the first-stage and reduced-form regressions in Tables F.3, F.4, F.5, and F.6 shows that the estimated coefficients are larger for tracts that are closer to the new train stations, as expected. For robustness, we run a specification

[21] Another reason may be due to the difference in moment conditions. Ahlfeldt et al. (2015) match the variance in log wages across employment locations while the present study matches the variance in log earnings across residential locations. Because wage data at the workplace is not easily available for Tokyo, we are unable to verify whether this difference may be the reason for our different estimates of θ.

in which we restrict the sample to tracts that are within 3 km of a new train station. The treatment group consists of tracts within 1 km of a new station, while the control group consists of census tracts between 1 and 3 km from a new station within the same municipality. Thus, the treatment and control groups are more comparable locations. The results, shown in Table F.7, are similar to our earlier results.

5.5 2020 Equilibrium

In the 2020 equilibrium, about 39% of the residential population, 53% of employment, and 85% of total land value is contained within 20km of the Imperial Palace (the "center"), a region that contains 9% of the metro's land area, while 19%, 16%, and 3%, respectively, are contained in the region beyond 40km from this point (the "periphery"), a region that contains 67% of the land area. Model-implied wages are on average 2.5 times higher in the center than in the periphery partly due to larger fundamental productivity (2.1 times higher), more capital deepening (40% higher floorspace) and agglomeration (contributing an additional 17% to the difference in productivity). Amenities on the other hand are similar between central and peripheral regions, while fundamental amenities are 19% *lower* in the center. Relative to the periphery, the center therefore has an absolute advantage in production, a small absolute disadvantage in residence (as measured by fundamental amenities), and a comparative advantage in production. What sustains the high population level in the center if it does not have an advantage (either absolute or comparatively) in amenities? The answer lies in the regions' commuter access metrics. Firm commuter access is approximately 15% higher in the center than the periphery while residential commuter access is nine times larger. The center's close proximity to the high density of firms makes the center an attractive location to live, sustaining a large population in the center.



(a) Firm Commuter Access (b) Residential Commuter Access

Figure 4: Commuter Access in 2020

5.6 Model Fit

We perform a number of exercises to assess the fit of our model to relevant moments from the data. Figure 12 plots the cumulative distribution of travel times for the five travel modes between the model and the commuter flow data. We find a close fit for all five modes. Figure 11 plots the commuter shares for cars and train for each municipality from the model against the commuter-flow data. The correlation between model versus the data's car mode shares is 0.94 while that of train is 0.9. The regression coefficients of the population-weighted OLS are 0.72 and 0.84 respectively. The overall mode share between the data and model can be found in Table 7 and are similar to each other. The correlation in average income across municipalities between the data and the model is 0.58, and the ratio in average income between residents in the center and the periphery is 1.75 in the model and 1.5 in the data. Overall, we assess the model to have reasonable fit with important moments of the data.

6 Removing The Train System

Equipped with parameters that have either been estimated or taken from the literature and structural residuals that have been calculated using model-inversion, we then perform model simulations of removing the train system for our class of models. Specifically, we remove the train system while keeping travel times for other modes of transportation fixed using three specifications: the model without floorspace, the model with floorspace but with spillovers turned off ($\mu_U = \mu_A = 0$), and the model with floorspace and with spillovers. Parameters from the literature that are used for the quantification are shown in Table 8.

Specifically, we remove all choices involving train from agents' choice sets such that commuters can now only choose from walking, cycling, driving, or taking the bus. The new commuting cost between two locations i and j, denoted as τ'_{ij} are then given by:

$$\tau'_{ij} = \left[\sum_{m \setminus \text{train}} \tau_{ijm} \right]^{-1/\theta}$$

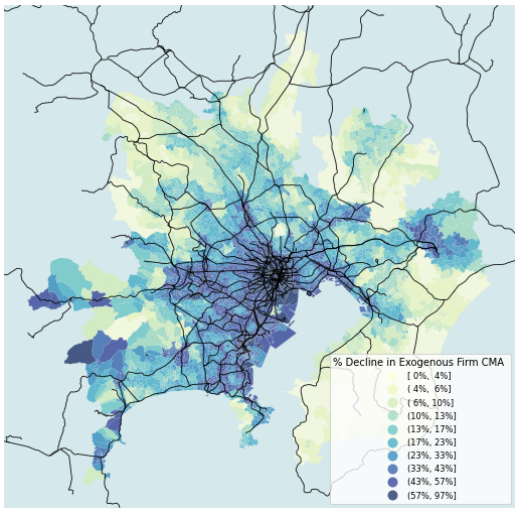
where $\sum_{m \setminus \text{train}}$ denotes that the sum is taken over all modes except for the train.

We represent the exogenous shock to commuting access in Figure 5. These are calculated as the change in commuter access (expressed in equations 9, 10, 15, and 16) while holding constant endogenous variables (wages, amenities, and floor prices) at the 2020 equilibrium.

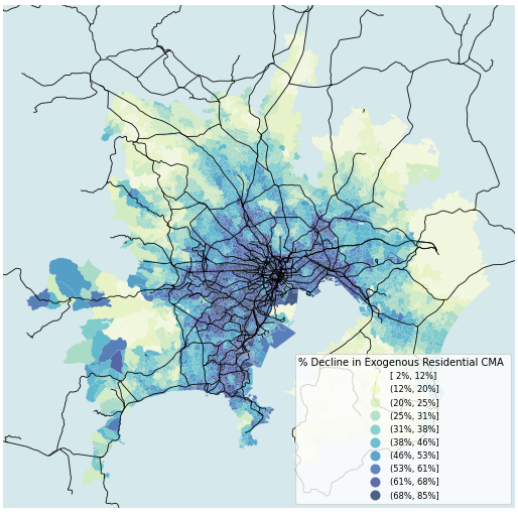
The direct impact (holding constant endogenous variables) of removing the train is that all regions lose commuting access, with the median census tract declining by 17% in firm access and 38% in residential access. However, the decline in central Tokyo's firm access is over four times more than that of peripheries while the decline in residential access is more than twice that of the peripheries. These discrepancies lead to shift in population and employment from the city center to its peripheries, as will be discussed further when we seek to understand the reason for

our centralizing results, in Section 7.

The results in Table 9 presents three types of results for the three specifications. First, we present results regarding centralization patterns: we study the share of total population and employment within a 10 km and 20 km from the Imperial palace, the gradient of the log population and employment densities with respect to distance from the Imperial Palace, and the ratio of employment to population in the center (20 km radius). We then present three types of welfare results based on an assumption of a closed economy. First, we present the change in household's average utility excluding the cost of construction and related expenses. Second, we annualize construction costs using a number of discount rates, compare them to Tokyo's GDP, and use this figure to calculate a consumption-equivalent measure of welfare of the train system for households of the Greater Tokyo Area. Last, we calculate benefit-to-cost ratios by comparing the gains (the first welfare measure) to the costs (as a percentage of Tokyo's GDP). Underneath the welfare measures for a closed economy, we present additional statistics for the version of the model that allows Tokyo's population to depopulate in response to the loss of train system. Figure 13 plots the employment and population densities by distance to the Imperial Palace for the closed- and open-economy model assumptions.



(a) Exogenous Change to Firm CA (% Change)



(b) Exogenous Change to Residential CA (% Change)

Figure 5: Shock to Commuting Access

Notes: Panels plot the percent change in firm (left) and residential (right) commuter access when the rail network is removed in Section 6. The darkest declines concentrate within the 20 km of the center, where the magnitude of the direct impacts on employment and population are four times and two times more than those of peripheries, respectively.

6.1 Discussion

We find across our three specifications that the train system has a centralizing effect on both population and employment, that the welfare gains from the train system exceed the costs (based on the discount rates between 2-6%), and that if Tokyo's total population were allowed to adjust in response to the removal of the train system, there would be substantial depopulation of Tokyo. In

our preferred "Floorspace + Spillovers" specification the erosion of Tokyo's core is stark: the share of residents living within 10 km of the Imperial Palace falls from 23.6 % to 13.8 % (a 41.4 % decline) when rail is removed, and the within-10 km employment share drops from 41.6 % to 33.2 % (a 20.3 % decline). Focusing on the broader 20 km core, the population share falls by 10.4 percentage points (a 22.6 % decline) and employment by 8.9 percentage points (a 15.6 % decline), while the gradients of log population and employment density flatten by 29.4 % and 14.8 %, respectively. When we allow Tokyo's overall population to shrink in response to rail removal, central locations depopulate further, as Tokyo's overall population declines by 28.9 %.

The findings that Tokyo's train system leads to centralization in both population and employment may at first glance seem inconsistent with previous studies, which have mostly emphasized the decentralizing impact of new transportation infrastructure. Glaeser and Kahn (2001) document that American cities first began to decentralize in the late 19th century as the first commuter trains and streetcars allowed workers to move their residence to the suburbs. During the 20th century, the advent of automobiles first led to further decentralization in population and then in the latter half of the 20th century also led to the decentralization of firms. Baum-Snow (2007) use the construction of the interstate highway system as a natural experiment and find large decentralization in population from central cities. Baum-Snow et al. (2017) find large decentralizing impacts of radial and ring-road highways on population across cities in China.

The extent to which commuter rail in urban areas centralizes population and employment has been less clear cut. Heblich et al. (2020) find that the advent of the steam railway in the 19th century facilitated the large-scale shift in residential population to the suburbs and concentration of employment in the city center in absolute numbers. The authors also find similar patterns of shifting population from downtown to the suburbs in other large metropolitan areas following the transport improvements of the 19th century, including Berlin, Paris, Boston, Chicago, New York, and Philadelphia. At the same time, the highest *percentage* growth rates in both employment and population occurred in the suburbs. Studying more recent expansions in urban rail using a panel of the world's largest cities, Gonzalez-Navarro and Turner (2018) found little evidence of urban population growth from subways, while Baum-Snow et al. (2017) found little evidence of population shift to the suburbs from radial railroads.

Yet, our results align with the narrative suggested in the introduction: Tokyo's train system plays an important role in facilitating the city's high population density and concentration of economic activity in the core, leading to higher overall city population and substantially sharper population and employment density gradients.

The welfare impact on Tokyo's household is similarly stark. Household consumption-equivalent utility falls by 20.6 % when trains disappear (and without rebating households for the construction costs), and the implied benefit–cost ratios remain comfortably above one once we annualize construction costs (4.9, 3.4, 1.78at discount rates of 2%, 4%, and 6%, respectively). These magnitudes are large when set against recent U.S. evidence: Severen (2023) estimate that the commuting benefits of Los Angeles's Metro Rail equal 2–64% of annualized costs by 2000, while Gupta et al. (2020)

find that the value of New York City's subway expansion amounts to 53–177% of construction cost. Other works on rail in the U.S., such as Glaeser and Kahn (2004), Glaeser et al. (2008), and Baum-Snow et al. (2005), have documented the declining ridership and thus limited economic viability of urban rail given the high fixed costs. Our benefit–cost ratios resemble instead the "London-style" returns documented by Heblich et al. (2020) for the pre-automobile era.

Since the construction costs of Tokyo's train system is similar to those of other countries at least in dollar terms (as discussed in Section 3), the discrepancy in the welfare gains for Tokyo relative to other developed cities is likely due to conditions in Tokyo being apt for higher benefits. We explore these possible conditions in Section 8.

6.2 Robustness and Extensions

Congestion Model

In Table F.2, we report the results for the model in Section B.1 where travel times by cars are endogenized to allow for changes in traffic congestion following train removal to have an additional equilibrium effect. After the train removal, the adjustment of car-travel times induces further decentralization of households who are replaced by firms. Relative to the equilibrium where the train is removed but car-travel times are held fixed, population in the inner 10 km and inner 20 km decline by an additional 2.3 and 0.3 percentage points, respectively, while the decline in employment in the inner 10 km and inner 20 km is diminished by 1.6 and 0.9 percentage points, respectively. The welfare gains associated with the train system increase by 2.1 percentage points.

Sensitivity of Centralization Results

We explore the robustness of our centralizing results to the specifications of the model. In particular, we make the following modifications. First, we increase the speed of all train travel by a factor of two. Second, we decrease the speed of all train travel by factor of one half. Third, we reduce the value-of-time parameter of cars such that it matches that of trains (this change is equivalent to increasing the speed of all car trips). Fourth, we remove automobiles as a choice set. Last, we reduce the speed of train travel by a factor of one-half for commutes within the center (residence and workplace are in the center).

For each case, we first simulate a counterfactual where the above conditions are met. We then remove the train system as before to see how the results change. The results can be found in Table 10 and Table 11. We find that with the exception of the last exercise, these modifications affect the magnitude of the centralizing effects, but they do not change the qualitative pattern of a centralizing effect of the train system. Whenever train travel speeds are reduced for commutes between the central regions however, the centralizing patterns are reversed for the case of population–the train system has a centralizing effect on employment but has a decentralizing effect on population. Thus, the centralization results are robust to various adjustments of the model. However, our results are sensitive to the magnitude of the commuting-cost shock for commutes within the center.

7 The Direct Effect on Population and Employment in the Core and Peripheries

To understand the reason for our centralization results we study the exogenous change in commuting access, introduced in Section 6. As discussed in Tsivanidis (2023) and Redding (2025), commuting access in both models are equivalent (up to normalization), because they can be expressed as equivelent functions of the matrix of commuting costs, and the vectors of population and employment, which are the same in the two models: referring to Φ_{Ri} as residential commuter access for location i and Φ_{Fi} as firm commuter access for location i, Tsivanidis (2023) show that (given $\{\tau_{ij}\}_{i,j\in\mathcal{N}}, \{L_{Ri}, L_{Fi}\}_{i\in\mathcal{N}}$) these variables are the unique (up to scale) solution to the following equations:

$$\Phi_{Ri} = \sum_j \tau_{ij}^{-\theta} \frac{L_{Fj}}{\Phi_{Fj}} \qquad\qquad \Phi_{Fj} = \sum_i \tau_{ij}^{-\theta} \frac{L_{Fi}}{\Phi_{Ri}}$$

As discussed in Section 6, given a change in commute costs from an initial matrix $\{\tau_{ij}\}_{i,j\in\mathcal{N}}$ to an alternative matrix $\left\{\tau'_{ij}\right\}_{i,j\in\mathcal{N}}$ and initial vectors $\{L_{Ri}, L_{Fi}\}_{i\in\mathcal{N}}$, we define the exogenous change of RCA and FCA as the change in Φ_{Ri} and Φ_{Fj}, holding constant endogenous variables (L_{Ri} and L_{Fi}, or equivalently w_i, U_i and if floorspace exists, then r_i too) in all locations, and we denote these variables as $\hat{\Phi}_{Ri}^{DE}$ and $\hat{\Phi}_{Fi}^{DE}$, respectively:

$$\hat{\Phi}_{Ri}^{DE} \equiv \frac{\sum_j \tau_{ij}'^{-\theta} \frac{L_{Fj}}{\Phi_{Fj}}}{\sum_j \tau_{ij}^{-\theta} \frac{L_{Fj}}{\Phi_{Fj}}} \qquad\qquad \hat{\Phi}_{Fj}^{DE} \equiv \frac{\sum_i \tau_{ij}'^{-\theta} \frac{L_{Fi}}{\Phi_{Ri}}}{\sum_i \tau_{ij}^{-\theta} \frac{L_{Fi}}{\Phi_{Ri}}} \tag{19}$$

As in Dekle et al. (2008), these can be expressed as a commuter-share weighted average of changes to accessibility to other locations:

$$\hat{\Phi}_{Ri}^{DE} = \sum_j \frac{L_{ij}}{L_{Ri}} \hat{\tau}_{ij}^{-\theta} \qquad\qquad \hat{\Phi}_{Fj}^{DE} = \sum_i \frac{L_{ij}}{L_{Fj}} \hat{\tau}_{ij}^{-\theta} \tag{CA}$$

where $\hat{\tau}_{ij}$ are the change in commuting costs and L_{ij} are commuter flows from i to j. Thus exogenous change in commuter access are a weighted average of changes to $\tau_{ij}^{-\theta}$, which we will refer to as *accessibility*. Returning to our gravity specification in Section 5.2, accessibility can be thought of as the commuter-flow-relevant measure of commuting-costs: the partial elasticity of commuter flows with respect to accessibility is one, holding constant wages in workplaces and amenities (and floor prices, depending on the model) for residences. Weighting the commuter flows by their commuting shares then, the exogenous change in commuter access then can be viewed as a measure (at least in terms of its units) as the percent change in population (in the case of RCA) or employment (in the case of FCA) in partial equilibrium, and this intuition is reflected in the expressions in the commuter supply equations (9), (10), (15), (16), where if all other endogenous variables were held constant, then these variables precisely measure the percent change in population and employment.

We therefore refer to $\hat{\Phi}_{Ri}^{DE}$ and $\hat{\Phi}_{Fj}^{DE}$ as the *direct effect* on population and employment, respectively.

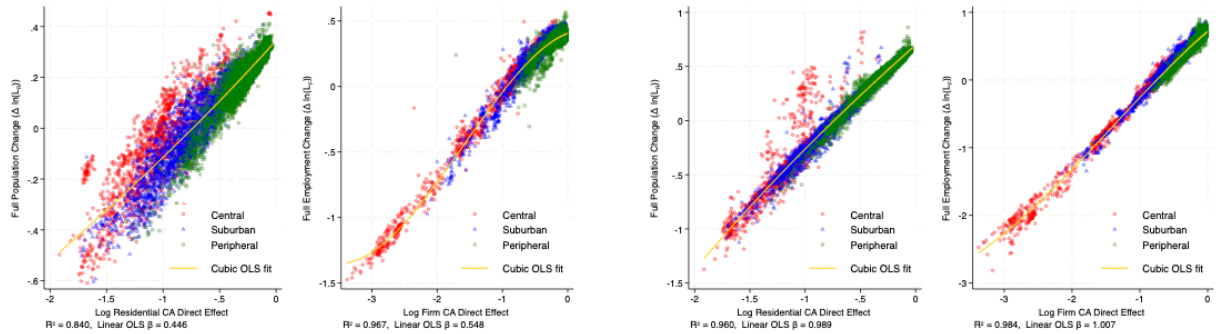
We use these direct-effect measures to study whether the direct effect alone explains the centralizing patterns that were observed in the previous section. The appeal is twofold. First, because the direct effects are easy to measure (they require commuting shares and accessibility) and analyze (since they are weighted averages), we can study them further to investigate the reason for the centralization *if they have good explanatory power over the full general-equilibrium effect*. Second, if they have good explanatory power, then the direct-effect measures may offer a back-of-the-envelope metric to predict the general equilibrium response of transportation improvements more generally.

In this section, we use the direct-effect measures, $\hat{\Phi}_{Ri}^{DE}$ and $\hat{\Phi}_{Fi}^{DE}$ to analyze the reason for the centralization patterns observed in the previous section. We find that the direct effects have strong explanatory power over the general equilibrium response and there is a positive relationship. Given that Tokyo's train system had a much larger direct effect to the core than to the periphery (see Section 6) for both residents and firms, our findings that Tokyo's train system centralizes both population and employment is therefore consistent with the disparities in these direct effects between core and peripheries driving (or at least explaining statistically) the centralization patterns.

Decomposing $\hat{\Phi}_{Ri}^{DE}$ into its components shows that 88% of the center's direct impact derives from improved access to other central regions while 12% was from improved access to suburban and peripheral regions. This disparity suggests that it is the accessibility improvements to connections in the core that are the culprit, and we then explore the role of the transportation network in the city's core for the centralization patterns.

Given their good in-sample explanatory power, can the direct effects be used to make numerical predictions on the change in population and employment, more generally for other transportation improvements and other empirical settings? We show how a researcher can construct proxies, which we refer to as heuristics, for the general equilibrium effect using the direct effects, and also construct bounds on their error that are applicable for small transportation improvements (i.e. when first-order approximations are valid). The advantage of this approach, relative to full general-equilibrium analysis, are twofold. First, the data requirements are smaller than for traditional GE analysis: the proxy approach requires information on commuting shares and commuting-cost changes only for the locations that directly experienced the transportation improvement. In contrast, traditional GE analysis requires all commuting shares and changes to commuting costs. Second, because both the proxies and bounds can be computed in closed-form, the mapping from observables and assumptions to counterfactuals is transparent. In contrast, closed-form solutions are rarely available in traditional GE analysis and counterfactual equilibria are solved on computers. The disadvantage of the proxy approach is that depending on the empirical setting or the strength of spillovers (governed by the parameters), the proxies may be inaccurate and the bounds too large for either the proxy or the bounds to be useful. Fortunately, the size of the bounds (and therefore the potential error in our proxies) scales by a factor that is a function of only parameters, θ, ϕ_U and ϕ_A. Therefore, given these parameters, the researcher has a way to evaluate the potential size of the bounds before proceeding to analyze the direct effects.

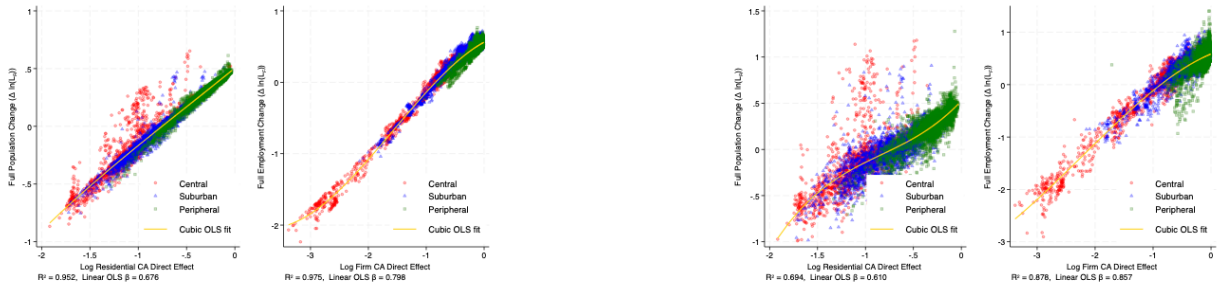
7.1 Explanatory Power of Direct Effects



(a) No Floorspace (b) Shared Floorspace

Figure 6: Exact Counterfactual Analysis Versus Direct Effects

Notes: Each panel scatters the exact changes in population (blue) and employment (orange) implied by the train-removal counterfactual against the direct-effect statistic from Section 7. The cubic fits summarize the monotone relationship emphasized in the text: central tracts (red) suffer the largest negative direct impacts and, correspondingly, the largest general-equilibrium losses, while peripheral tracts (green) are far closer to zero. Slopes below one illustrate the GE dampening discussed above.



(a) No Spillovers (b) Spillovers + Congestion

Figure 7: Exact Counterfactual Analysis Versus Direct Effects

Notes: The additional specifications—no spillovers (left) and the congestion model (right)—show that the direct-effect statistic continues to line up with full model outcomes even when GE mechanisms differ. As Section 7 explains, flatter slopes for the congestion case reflect stronger attenuation from automobile feedbacks, whereas the spillover-free model nearly lies on the 45° line.

Figures 6 and 7 shows scatter plots and cubic-polynomial fits of the regression of the exact counterfactual variables on the direct-effect variables for the counterfactual of removing Tokyo's train system. Figure 6 shows the relationship for the two models discussed in Section 4 while Figure 7 shows the relationship for the shared-floorspace model without spillovers (where $\mu_U = \mu_A = 0$) and the model with endogenous travel times for automobiles. Underneath the plots indicate the R-squared of the cubic-polynomial fit and the slope coefficient of a (univariate) linear OLS regression. The direct effects explain a large share of the variation across our specifications, with R-squared ranging from 0.70 (population for the model with road congestion) to 0.98 (employment for the model with shared floorspace). In addition, the relationships appears to be monotonic–a larger direct impact (in magnitude) is associated with a larger decrease in the variable of interest. As a result, the direct-effects perform well in predicting the ordering of endogenous variables, as measured by the correlation of the ranks (i.e. Spearman's correlation), which ranges from 0.71 to 0.98. The

shape of the relationship, however, depends on the model, with some linear OLS coefficients being close to one (model with shared floorspace) while others are as low as 0.45 (population in model without floorspace). A flatter slope is consistent with congestion forces playing a larger role in dampening the direct effects. In the model with shared floorspace, a fixed supply of land acts as a congestion force while higher agglomeration forces (μ_A, μ_A) and floorspace supply elasticity $\frac{\mu}{1-\mu}$ counteract these forces. In the model without floorspace, the congestion forces act through the parameters ϕ_U and ϕ_A which are both negative.

The plots also show the difference in the size of the direct effect across locations. Central tracts (with a red marker) experience a direct-effect on residences and firms of -61% and -51% respectively while peripheral locations (in green) experience -25% and -12%, respectively. Thus, the train system disproportionately affects commuting costs for centrally located firms and households, and according to the correlations, these disparities are associated with the train removal leading to larger declines in population and employment in the center.

When we decompose $\hat{\Phi}_{Ri}^{DE}$ into its components, we find that 88% of the center's direct impact on households derives from improved access to other central regions while 12% was from improved access to suburban and peripheral regions. This disparity suggests that it is the accessibility improvements to connections in the core that drive the centralization patterns. In Section 8, we test this hypothesis and explore the role of the transportation network in the city's core for the centralization patterns.

While the exact functional relationship between direct effects and the true general equilibrium comparative static likely depends on the model (especially the congestion parameters) and empirical setting, the strong explanatory power and monotonic relationship offers promise that the direct effect may be useful as a back-of-the-envelope metric of how population and employment will shift across locations, given a change in transportation infrastructure. The simple closed-form expression–of a weighted average of accessibility improvements–is easy to calculate and requires significantly less data and resources than traditional general equilibrium analysis. Thus, given that the direct effects correlates well with the true general-equilibrium effect, can the direct-effect be used to make numerical predictions about the general equilibrium effect? We study this question in the following section.

7.2 Proxies for the General Equilibrium Effect

The previous section shows that the direct effect measures have a tendency to move monotonically with the full general equilibrium effect (statistically). However, the scatter plots varied in their functional relationship, highlighting the fact that model assumptions will alter the functional relationship. In addition, we have no guarantee that these functional relationships can be translated to other empirical settings.

Can we derive a metric for the general equilibrium effect that would be a good proxy for the general equilibrium effect, in general, not only directionally, but also numerically? In this section, we show how to construct two such proxies and derive bounds on the full general equilibrium

counterfactual, given a particular model (and parameters). The bounds are applicable for small changes to the transportation infrastructure (i.e. first-order effects). We focus on the model without floorspace due to its simplicity. In a related paper, Bamba (2025) show that the same approach can be applied to any model in the constant-elasticity spatial family studied by Allen et al. (2024), which includes the models of Allen et al. (2020) and Arkolakis et al. (2012).

Computing the proxies and its bounds have a similar flavor to computing the direct effects. The first proxy, which we refer to as the *direct-effect proxy*, is simply a scaled version of the direct-effect and the bounds require information about the largest proxy across all locations for both population ($\hat{\Phi}_{Ri}^{DE}$) and employment ($\hat{\Phi}_{Fi}^{DE}$). In addition to this data, these proxies and bounds require the parameters from the model, because the parameters mediate the strength of general equilibrium forces that create the deviation between the direct effects and the full comparative static. The second proxy, which we call the *augmented proxy*, improves on the first one by making a guess for the counterfactual values of $\frac{L_{Fj}}{\Phi_{Fj}}$ and $\frac{L_{Fi}}{\Phi_{Ri}}$ in (19) based on the direct effects (to proxy the change in access, in the denominator) and direct-effect proxy (to proxy the change population or employment, in the numerator). Thus the second proxy requires more information: in addition to the changes in commuting costs and commuting shares for a particular location, one also needs the direct effects for all locations. The bounds once again require information about the largest possible augmented proxy across locations, and both the proxy and the bounds require the parameters from the model. The upside is that the bounds are strictly smaller than those of the direct-effect proxy.

For both proxies and bounds, the data requirements are smaller than for traditional analysis: we require information on commuting shares and commuting-cost changes only for the locations that directly experienced the transportation improvement. If all locations experience a transportation improvement however, then the data requirements for computing the bounds are, strictly speaking, the same as in traditional analysis (one needs both matrices of commuter shares and commuting-cost improvements), because computing the largest of the proxies requires knowledge of all commuting shares and changes to accessibility. However, a researcher or practitioner may be able to speculate ex-ante about which locations will experience the largest direct impacts: they are likely the locations close to where the transportation improvement occurs (e.g. close to a new train station). Thus, a practitioner can make an "educated guess" and compute the direct-effect bounds with information on this location. If, in addition, the practitioner has knowledge of the direct-effects in other locations, then one can make improvements on the direct-effect proxy and construct the augmented proxy based on the information that is available, and also potentially construct improved bounds.

Perhaps, the biggest strength of the heuristic approach is that because the proxies and bounds are expressed in closed-form, the researcher can make assumptions as necessary based on the available data and resources and provide a prediction and bounds based on the data and resources that are available.

When we apply our proxies and bounds to Tokyo's train system, the bounds for the augmented proxy account for 13% of the range and 78% of one standard deviation of the first-order changes in population. As a result, the bounds are tight enough to make sharp predictions regarding

how population and employment shifts as a result of the train system, while requiring potentially less data and less computation than traditional general equilibrium analysis. For our purposes, the bounds are able to guarantee that the first-order effect of Tokyo's train system (e.g. if Tokyo's train system made a uniform marginal improvement to its speeds) would lead to the average municipality in the center to gain population relative to the average municipality in the periphery.

In contrast, the bound for the direct-effect proxy are 36% larger than the range of the data, and hence are too large to guarantee a particular shift in population in one direction or another.

Two Proxies for General Equilibrium Effects

We return to the model without floorspace from Section 4. Substitutions of equations (4) and (5) into equations (9) and (10) and considering once again using exact-hat notation, the change in population and employment can be written as:

$$\hat{L}_{Ri}^{1-\tilde{\phi}_U} = \frac{\hat{\bar{L}}}{\hat{\bar{V}}^{\theta}} \hat{\Phi}_{Ri}, \qquad \hat{L}_{Fi}^{1-\tilde{\phi}_A} = \frac{\hat{\bar{L}}}{\hat{\bar{V}}^{\theta}} \hat{\Phi}_{Fi}$$

$$\Rightarrow \hat{L}_{Ri} = \left[\frac{\hat{\bar{L}}}{\hat{\bar{V}}^{\theta}}\right]^{\frac{1}{1-\tilde{\phi}_U}} \hat{\Phi}_{Ri}^{\frac{1}{1-\tilde{\phi}_U}}, \qquad \Rightarrow \hat{L}_{Fi} = \left[\frac{\hat{\bar{L}}}{\hat{\bar{V}}^{\theta}}\right]^{\frac{1}{1-\tilde{\phi}_A}} \hat{\Phi}_{Fi}^{\frac{1}{1-\tilde{\phi}_A}}$$

where we define $\tilde{\phi}_U = \theta\phi_U$ and $\tilde{\phi}_A = \theta\phi_A$. The relative change in population and employment between two locations k and l are then independent of the change in total population or average utility:

$$\frac{\hat{L}_{Rk}}{\hat{L}_{Rl}} = \left[\frac{\hat{\Phi}_{Rk}}{\hat{\Phi}_{Rl}}\right]^{\frac{1}{1-\tilde{\phi}_U}}, \qquad \frac{\hat{L}_{Fk}}{\hat{L}_{Fl}} = \left[\frac{\hat{\Phi}_{Fk}}{\hat{\Phi}_{Fl}}\right]^{\frac{1}{1-\tilde{\phi}_A}} \tag{20}$$

That these expressions are independent of total population and average utility is useful because it means a researcher does not need to make assumptions about population mobility in and out of the city if they are merely interested in the relative changes in population and employment between locations, given the changes in commuter access for the two locations. To abstract from population-mobility assumptions, we therefore focus proxies for the *relative* changes in population and employment across locations.

Our first proxy, termed the direct-effect proxy, replaces $\hat{\Phi}_R$ and $\hat{\Phi}_F$, with our direct-effects, $\hat{\Phi}_R^{DE}$ and $\hat{\Phi}_F^{DE}$. We refer to $\hat{\tilde{L}}_{Ri} \equiv \hat{L}_{Ri}\left[\frac{\hat{\bar{V}}^{\theta}}{\hat{\bar{L}}}\right]^{\frac{1}{1-\tilde{\phi}_A}}$ and $\hat{\tilde{L}}_{Fi} \equiv \hat{L}_{Fi}\left[\frac{\hat{\bar{V}}^{\theta}}{\hat{\bar{L}}}\right]^{\frac{1}{1-\tilde{\phi}_A}}$ as the *mobility-adjusted* change in population and employment, respectively. We define the direct effect proxy as follows.

Definition (Direct-Effect Proxy)**.** The direct-effect proxies are:

$$\hat{\tilde{L}}_{Ri}^{DE} \equiv \left[\hat{\Phi}_{Ri}^{DE}\right]^{\frac{1}{1-\tilde{\phi}_U}}, \qquad \hat{\tilde{L}}_{Fi}^{DE} \equiv \left[\hat{\Phi}_{Fi}^{DE}\right]^{\frac{1}{1-\tilde{\phi}_A}} \tag{21}$$

Given the direct impact on commuter access in each location, the direct-effect proxy allows

for each location's local congestion forces to equilibrate: for residents the amenities in a location respond to the population changes, while for firms wages respond to the employment change. In both situations, these local equilibrium forces act as a congestion force (when ϕ_U and ϕ_A are negative). Beyond these equilibrium forces, no further equilibrium response is permitted under the direct-effect proxy. For example, the direct-effect proxy does not allow for the population and employment to respond to the changes in wages and amenities in other locations. The next proxy will consider these forces. In contrast to the direct-effect on commuter access where we assume that endogenous variables stay constant, in the augmented proxy, we proxy the change in endogenous variables using the direct-effect and direct-effect proxies. Instead of holding constant the endogenous variables in (19) we proxy $\hat{L}_{Ri}$ and $\hat{L}_{Fi}$ with $\hat{\tilde{L}}^{DE}_{Ri}$ and $\hat{\tilde{L}}^{DE}_{Fi}$, respectively, and proxy $\hat{\Phi}_{Ri}$ and $\hat{\Phi}_{Fi}$ with $\hat{\Phi}^{DE}_{Ri}$ and $\hat{\Phi}^{DE}_{Fi}$, respectively. [22] Alternatively, one can view this exercise as proxying changes in wages, $\hat{w}_i = \hat{L}^{\phi_A}_{Fi}$, in (9) with $\left[\hat{\tilde{L}}^{DE}_{Fi}\right]^{\phi_A}$ and proxying changes in amenities, $\hat{U}_i = \hat{L}^{\phi_U}_{Ri}$, in (10) with $\left[\hat{\tilde{L}}^{DE}_{Ri}\right]^{\phi_U}$. Then the expressions in (19) are replaced with:

$$\hat{\Phi}^{AP}_{Ri} \equiv \frac{\sum_j \tau'^{-\theta}_{ij} \frac{L_{Fj}}{\Phi_{Fj}} \hat{\tilde{L}}^{DE}_{Fj}}{\sum_j \tau^{-\theta}_{ij} \frac{L_{Fj}}{\Phi_{Fj}}}, \qquad \hat{\Phi}^{AP}_{Fj} \equiv \frac{\sum_i \tau'^{-\theta}_{ij} \frac{L_{Fi}}{\Phi_{Ri}} \hat{\tilde{L}}^{DE}_{Fi}}{\sum_i \tau^{-\theta}_{ij} \frac{L_{Fi}}{\Phi_{Ri}}} \tag{22}$$

which can be expressed using commuting shares such that we can then define our augmented proxy as follows:

Definition (Augmented Proxy)**.** The augmented proxies are:

$$\hat{\tilde{L}}^{AP}_{Ri} \equiv \left[\sum_j \frac{L_{ij}}{L_{Ri}} \hat{\tau}^{-\theta}_{ij} \left(\hat{\Phi}^{DE}_{Fj}\right)^{\frac{\tilde{\phi}_A}{1-\tilde{\phi}_A}}\right]^{\frac{1}{1-\tilde{\phi}_U}}, \qquad \hat{\tilde{L}}^{AP}_{Fj} \equiv \left[\sum_i \frac{L_{ij}}{L_{Fj}} \hat{\tau}^{-\theta}_{ij} \left(\hat{\Phi}^{DE}_{Ri}\right)^{\frac{\tilde{\phi}_U}{1-\tilde{\phi}_U}}\right]^{\frac{1}{1-\tilde{\phi}_A}} \tag{23}$$

The augmented proxy captures the change in population and employment to a change in the transportation network as if local equilibrium forces (wages and amenities) responded to the direct effect in each location, and in addition, *if population and employment responded to those changes in wages and amenities in other locations.* Thus, one can view the two proxies as capturing the first two effects of the propagation of the commuting-cost shock through the network.[23]

How do these proxies perform in predicting the true general equilibrium effects in the no-floorspace model? We first provide theoretical bounds on the two proxies, such that a researcher can know the performance of their proxies with similar data that was used to compute the proxies.

Log-linearizing equations (CA) and (21) with respect to $d \ln \tau_{ij}$ (holding constant endogenous variables) results in:

[22] Because our focus is on relative changes across locations, differences in scalars do not affect the results.

[23] See Allen et al. (2020) for an extensive discussion on the network-propagation interpretation in gravity-based trade models.

$$\Delta \ln \Phi_{Rl}^{DE} = \theta \sum_j \frac{L_{lj}}{L_{Rl}} \Delta \ln \tau_{lj}, \qquad \Delta \ln \Phi_{Fl}^{DE} = \theta \sum_j \frac{L_{il}}{L_{Fl}} \Delta \ln \tau_{il} \tag{24}$$

$$\Delta^{DE} \ln \tilde{L}_{Rl} = \frac{1}{1-\tilde{\phi}_U} \Delta \ln \Phi_{Rl}^{DE}, \qquad \Delta^{DE} \ln \tilde{L}_{Fl} = \frac{1}{1-\tilde{\phi}_A} \Delta \ln \Phi_{Fl}^{DE} \tag{25}$$

Similarly log-linearizing the definition of the augmented proxy, (23) results in:

$$\Delta^{AP} \ln \tilde{L}_{Rl} = \frac{1}{1-\tilde{\phi}_U} \left\{ \Delta \ln \Phi_{Rl}^{DE} + \frac{\tilde{\phi}_A}{1-\tilde{\phi}_A} \sum_j \frac{L_{lj}}{L_{Rl}} \Delta \ln \Phi_{Fj}^{DE} \right\} \tag{26}$$

$$\Delta^{AP} \ln \tilde{L}_{Fl} = \frac{1}{1-\tilde{\phi}_A} \left\{ \Delta \ln \Phi_{Fl}^{DE} + \frac{\tilde{\phi}_U}{1-\tilde{\phi}_U} \sum_i \frac{L_{il}}{L_{Fl}} \Delta \ln \Phi_{Ri}^{DE} \right\} \tag{27}$$

Then one can place bounds on the error using the following proposition.

Proposition 1 (First-Order Bounds)**.** Consider a series of first-order comparative statics:

$$\Delta \ln \tilde{L}_{Rl} = \sum_i \sum_j \frac{\partial \ln L_{Rl}}{\partial \ln \tau_{ij}} \Delta \ln \tau_{ij}, \qquad d \ln L_{Fl} = \sum_i \sum_j \frac{\partial \ln L_{Fl}}{\partial \ln \tau_{ij}} \Delta \ln \tau_{ij}$$

where $\frac{\partial \ln \tilde{L}_{Rl}}{\partial \ln \tau_{ij}}$ and $\frac{\partial \ln \tilde{L}_{Fl}}{\partial \ln \tau_{ij}}$ are the *total* first-order comparative-static derivatives (accounting for changes in endogenous variables), and $\{\Delta \ln \tau_{ij}\}_{i,j \in \mathcal{N}}$ are a small changes to commuting costs. Then whenever $|\frac{\tilde{\phi}_U \tilde{\phi}_A}{(1-\tilde{\phi}_U)(1-\tilde{\phi}_A)}| < 1$:

1. The difference between the direct-effect proxy and the comparative static is bounded by the following :

$$|\Delta \ln \tilde{L}_{Rl} - \Delta^{DE} \ln \tilde{L}_{Rl}| \leq \max \left\{ \left| \frac{1}{1-\tilde{\phi}_U} \right|, \left| \frac{1}{1-\tilde{\phi}_A} \right| \right\} \max \left\{ \Delta \ln \Phi^{DE} \right\} \frac{\zeta^{DE}}{1-\zeta^{DE}} \tag{28}$$

$$|\Delta \ln \tilde{L}_{Fl} - \Delta^{DE} \ln \tilde{L}_{Fl}| \leq \max \left\{ \left| \frac{1}{1-\tilde{\phi}_U} \right|, \left| \frac{1}{1-\tilde{\phi}_A} \right| \right\} \max \left\{ \Delta \ln \Phi^{DE} \right\} \frac{\zeta^{DE}}{1-\zeta^{DE}} \tag{29}$$

   where $\zeta^{DE} \equiv \max \left\{ \left| \frac{\tilde{\phi}_A}{1-\tilde{\phi}_A} \right|, \left| \frac{\tilde{\phi}_u}{1-\tilde{\phi}_u} \right| \right\}$ and

$$\max \left\{ \Delta \ln \Phi^{DE} \right\} \equiv \max \left\{ \max_i \left\{ \Delta \ln \Phi_{Ri}^{DE} \right\}, \max_j \left\{ \Delta \ln \Phi_{Fj}^{DE} \right\} \right\}$$

2. The difference between the augmented proxy and the comparative static is bounded by the following:

$$|\Delta \ln \tilde{L}_{Rl} - \Delta^{AP} \ln \tilde{L}_{Rl}| \leq \max_i \left\{ \Delta \ln \Phi^{AP}_{Ri} \right\} \frac{\zeta^{AP}}{1-\zeta^{AP}} \tag{30}$$

$$|\Delta \ln \tilde{L}_{Fl} - \Delta^{AP} \ln \tilde{L}_{Fl}| \leq \max_j \left\{ \Delta \ln \Phi^{AP}_{Fj} \right\} \frac{\zeta^{AP}}{1-\zeta^{AP}} \tag{31}$$

where $\zeta^{AP} = |\frac{\tilde{\phi}_U \tilde{\phi}_A}{(1-\tilde{\phi}_U)(1-\tilde{\phi}_A)}|$

Proof. See Appendix Section E.1 □

The proof derives the first-order comparative statics using the implicit function theorem and shows that under the assumptions $|\frac{\tilde{\phi}_U \tilde{\phi}_A}{(1-\tilde{\phi}_U)(1-\tilde{\phi}_A)}| < 1$, then the direct-effect proxy and augmented proxy can be expressed as the leading terms of a decaying geometric series of matrices (i.e. a Neumann series). Properties of decaying geometric series can then be used to bound the remainder.

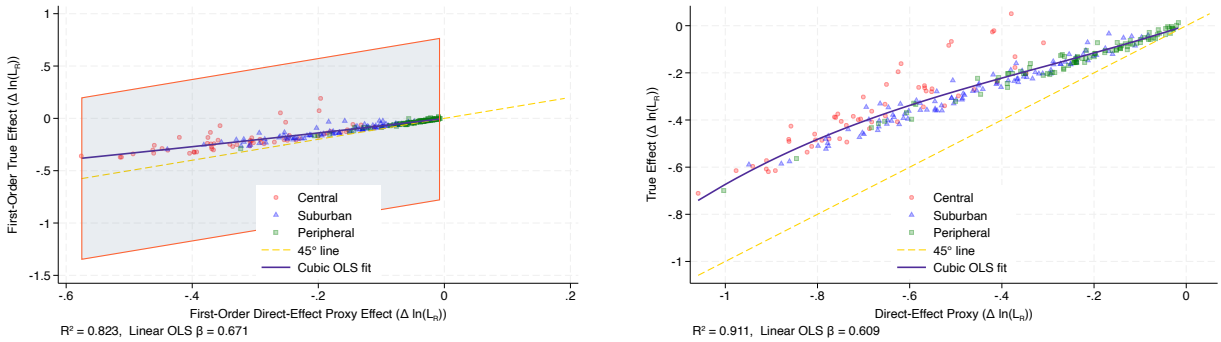
The bounds for both the direct-effect proxy and the augmented proxy require a measure of the *worst-case possible proxy*. In the case of the direct-effect proxy, this measure is the product of the largest direct effect and the larger of $\left|\frac{1}{1-\tilde{\phi}_U}\right|$ and $\left|\frac{1}{1-\tilde{\phi}_A}\right|$. For the augmented proxy, this measure is simply the largest augmented proxy across locations. These worst-case measures are then multiplied with a geometric factor, $\frac{\zeta^{DE}}{1-\zeta^{DE}}$ or $\frac{\zeta^{AP}}{1-\zeta^{AP}}$, which is a function of parameters. This geometric factor governs the size of general equilibrium effects that lead to the deviation between these proxies and the true effect. Whenever spillovers are zero $\phi_U = \phi_A = 0$, then amenities and wages are exogenous, and the geometric factors are both zero.[24] Thus, the direct effect, the direct-effect proxy, and the augmented proxy all correspond to the true comparative static. In other cases, one can measure the importance of general-equilibrium effects (relative to the size of the direct effects) using the geometric factors.

The key benefit of using the augmented proxy is that the geometric factor is smaller, leading to a smaller error. Inputting our estimates $\phi_A = -0.82, \phi_U = -0.68$, the geometric factor for the direct-effect proxy is $\frac{\zeta^{DE}}{1-\zeta^{DE}} \approx 0.82$ while for the augmented proxy, the geometric factor is $\frac{\zeta^{DE}}{1-\zeta^{DE}} \approx 0.22$. As a result, the error using the augmented proxy improves on the direct-effect proxy significantly.

Figure 8 plots true counterfactuals with the direct effect proxy for population. The left figure scales the transportation improvement to be small such that first-order effects dominate, and hence our theoretical bounds can be applied. On the right, we plot the full counterfactual of train removal with the direct-effect proxy. The direct-effect proxy is closer to the 45-degree line than the direct effect on RCA alone–the OLS coefficient increases from 0.45 to 0.61-0.67. However, there continues to be a bias, amounting roughly to 40% of the direct proxy. These highlight that while the direct effect may have strong predictive properties, general equilibrium effects continue to play a prominent role. As a result of the strength of general equilibrium effects, the bounds are larger than the variation in comparative statics, making their utility limited.

[24] Note that in the model with shared floorspace, floor-price-adjusted amenities and wages are not exogenous even when spillovers are zero: $\mu_U = \mu_A = 0$.

Figure 9 plots true counterfactuals with the augmented proxy for population. The OLS coefficient is now between 1.1 and 1.075, closer to unity. The R-squared is now close to around 0.996-0.997. Thus, by incorporating a proxy for the change in endogenous variables in each location, the augmented proxy comes remarkably close to the true counterfactual. With this information, the bounds are also smaller and informative. As a result, one can guarantee for example, that between certain central locations (located in the bottom left) and peripheral locations (located in the top right) the first-order effect of the train removal would shift population from certain central municipalities to the periphery. Indeed the average central municipality's augmented proxy is 0.18 log points larger in magnitude than that of the average peripheral municipality, while the height of the error bar is 0.15 log points. Therefore, one can guarantee (without computing the general equilibrium comparative static) that the first-order effect of the train removal would shift population on average (in geometric averages) from the center to the periphery.[25] In the next section, we verify these findings using exact counterfactuals and in addition, analyze the welfare implications.



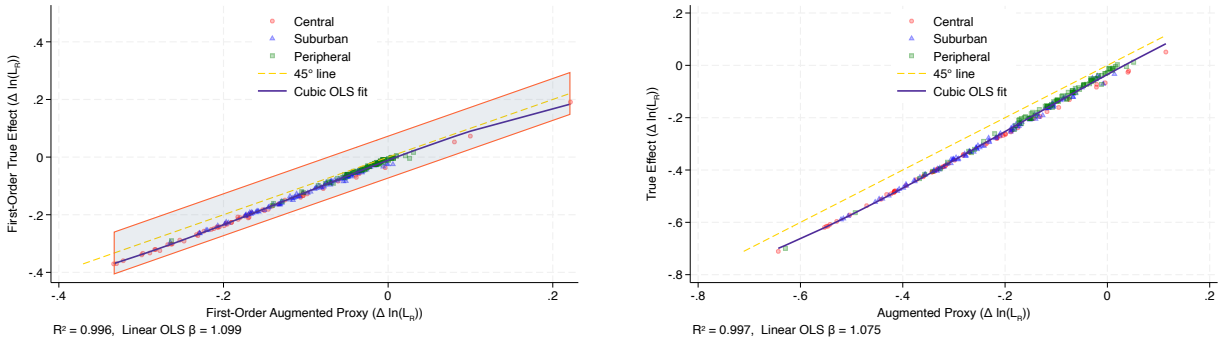
(a) Small Shock (First-order)

(b) Removing The Entire Train System

Figure 8: Direct-Effect Proxies and Their Bounds for Population

Notes: Each point compares the direct-effect proxy in Definition 21 against the true change in population for municipalities. Panel (a) uses a scaled (first-order) shock, panel (b) removes the entire rail system. Error bars depict the bound from Proposition 1. Consistent with the discussion following equation (E.4), the proxy captures the ranking of municipalities but tends to understate the magnitude of central losses, leaving wide error bars when spillovers are strong.

[25] Assumptions on population mobility in and out of the city can then be used to calculate the absolute change in population.



(a) Small Shock (First-order)

(b) Removing The Entire Train System

Figure 9: Augmented Proxies and Their Bounds for Population

Notes: The augmented proxy in Definition 23, which feeds direct effects back through local GE responses, closely matches the true counterfactual in both the linear approximation (left) and the full train-removal experiment (right). The tighter error bars mirror the smaller geometric factor ζ^{AP} in Proposition 1, allowing us to sign population shifts between core and periphery without solving the entire model.

8 Network Design, Scale, and Urban Environment

This section asks which features of Tokyo's rail network account for its large welfare and centralizing effects. We focus on three margins. First, we compare the welfare gains from core subway segments to expenditure-equivalent bundles of radial commuter segments. Second, we study how marginal benefit–cost ratios evolve as the network expands along Tokyo's historical construction path and relate these patterns to the segment-level results. Third, we evaluate how the value of a network of Tokyo's scale depends on car travel speeds and population density by adjusting these parameters to match those of large U.S. metropolitan areas.

8.1 Subway segments versus expenditure-equivalent radial segments

We first remove segments of the subway in the reverse order in which they were constructed. Then, for each subway segment that we remove, we perform an alternative counterfactual in which the subway segment remains and we instead remove an expenditure-equivalent bundle of radial train lines. Similar to the subway removal, we remove radial lines in the reverse order in which they were built, but we stop once the present value of construction, operating, and maintenance costs matches that of the corresponding subway segment. For each counterfactual, we compute the welfare loss relative to the observed 2020 equilibrium and compare this welfare loss to the present value of the segment's construction, operating, and maintenance costs, using the same cost assumptions as in Section 6 based on recent Japanese rail projects and land prices around the network. We are thus able to make a comparison for two different transportation networks where the expenditures are equivalent.

Figure 15 reports benefit–cost ratios for subway segments and for expenditure-equivalent bundles of radial segments. For each subway construction period on the horizontal axis, we remove the subway segments built in that period from the full 2020 network and compute the ratio of the implied welfare loss to the present value of construction, operating, and maintenance costs for

those segments. We do the same for the corresponding radial bundles. At our baseline 4% discount rate, subway benefit–cost ratios range between about 8 and 33 across periods and average about 16, whereas radial ratios range between roughly 0.05 and 6 and average about 2.7. Subway investments therefore yield several times higher returns than expenditure-equivalent radial investments in every period. Although the subway network accounts for about 10% of total track length and roughly 14% of estimated construction and related expenditures, it delivers about 37% of the welfare gain from the entire system (around 60% when construction costs are excluded), and the overall gains from the subway network are roughly six times those of an expenditure-equivalent bundle of radial lines.

Figure 16 shows the corresponding effects on the spatial distribution of population and employment. Removing subway segments reduces the share of both population and employment located within 20 km of the center, whereas removing expenditure-equivalent radial bundles slightly increases these shares. Without the subway network, the share of Tokyo's population within 20 km falls by 5 percentage points (12 percent), accounting for about 54 percent of the train system's total centralization effect on population. The corresponding decline for employment is 5.7 percentage points (10 percent), accounting for about 80 percent of the system's centralization effect on employment. These patterns show that core subway segments play an important role in both the welfare gains and the centralization effects of the rail system, while radial segments contribute relatively little and tend to decentralize activity.

The subway and radial segments also have different implications for specialization of the core as a workplace. Because the subway's centralizing effect is somewhat stronger for population than for employment, the subway network reduces the ratio of employment to population within 20 km of the center, whereas expenditure-equivalent radial bundles slightly increase this ratio. Thus, improvements in core subway infrastructure that generate large welfare gains and centralize activity do not necessarily increase the degree to which the core is specialized as a place to work.

To assess the role of complementarities across lines, we repeat the subway-removal counterfactuals after first disallowing transfers to and from subway stations. Figure 17 shows that removing transfers reduces the total welfare gain from the subway network by about 43 percent, indicating sizable complementarities across lines. For most construction periods, the marginal benefit–cost ratios are much lower when transfers are disallowed. For the most recent subway additions, by contrast, marginal gains are similar with and without transfers and remain high. This pattern is consistent with network effects playing an important role for earlier parts of the subway network, while the high returns to the most recent segments depend more on the placement of lines in already dense areas than on additional transfer opportunities.

8.2 Network scale along the historical expansion path

The subway–radial comparison above focuses on the composition of investment, conditional on the 2020 network. We now turn to how the welfare performance of the network evolves as it expands along Tokyo's historical construction path. Panel (a) of Figure 19 reports benefit–cost ratios for

the marginal 20-year construction period of all rail segments. For each period, we compare welfare in the network that includes all segments built up to that period with welfare in the network that includes only segments built up to the previous period, and divide the implied welfare gain by the present value of construction, operating, and maintenance costs for that period.

Marginal benefit–cost ratios are modest for the earliest period and then rise markedly as the network expands, remaining comfortably above one for all periods from 1900 through 1980.

Over the twentieth century as a whole, marginal benefit–cost ratios for 20-year periods average about 3.4 and peak around 15 during the 1980–2000 period, indicating that successive expansions of the network became more, rather than less, valuable at the margin over most of Tokyo's history. The marginal benefit–cost ratio then falls sharply for the 2000–2020 period, when new investment consists almost entirely of radial or peripheral segments and no new core subway lines are added.

Comparing these historical patterns to the segment-level results in Figure 15 shows that subway segments have average benefit–cost ratios of about 16, exceeding even the peak marginal benefit–cost ratio along the historical expansion path, whereas expenditure-equivalent radial investments average about 2.7, below the historical mean. The typical subway segment therefore delivers a higher benefit–cost ratio than even the historically most successful 20-year expansion of the network as a whole, while radial segments perform similarly to or worse than the historical average. Thus, the composition of investment between core subway lines and peripheral radial extensions has been least as important for welfare as further increases in total track length.

8.3 Urban environment: car speeds and density

We next explore how sensitive our results are to features of Tokyo's environment by simulating counterfactuals that resemble large U.S. metropolitan areas. We first increase the speed of car travel in Tokyo to match typical car accessibility in large U.S. cities, following Conwell et al. (2023).[26] This procedure implies that car speeds in Tokyo must increase by 43 percent. In this "fast-car" environment, population and employment decentralize substantially: the shares of residents and jobs within 20 km of the center fall by 10 and 18 percent, respectively. The welfare gains from the entire rail system fall to about 8.3 percent, roughly 42 percent lower than under Tokyo's current car speeds (30 percent lower excluding construction costs), as a more dispersed economy makes less intensive use of rail. Both subways and radial lines experience similar proportional declines in benefits, as shown in Figure 18.

We then investigate the sensitivity of our findings to total population by reducing Tokyo's population to match the densities of two large U.S. metropolitan areas: Greater New York and Greater Los Angeles. Greater Tokyo's population density is roughly twice that of Greater New York and about ten times that of Greater Los Angeles, so in the latter environments there are fewer residents to share the fixed costs of construction. Mechanically, lowering population lowers

[26] We use the TravelTime GIS tool to compute car isochrone areas from a point near the Imperial Palace for various travel times, and increase the effective car speed until we match the 45-minute isochrone area for the average large U.S. city.

the city's total income while construction and related costs of the train system fixed. Figure 19 reports cumulative benefit–cost ratios for the network as it expands along the historical path under four environments: baseline Tokyo, Tokyo with fast cars, and fast-car Tokyo with population scaled down to New York-like and Los-Angeles-like densities. At New York-like densities, the cumulative benefit–cost ratio of the network remains positive at our baseline 4 percent discount rate (about 1.2, corresponding to a welfare gain of 1.3 percent), but falls below one at a 6 percent rate (around 0.64, implying welfare losses of 1.5 percent). At Los-Angeles-like densities, the cumulative benefit–cost ratio is about 0.26 at a 4 percent discount rate, corresponding to a welfare loss of roughly 9 percent. While these exercises do not measure whether Tokyo's network would be socially beneficial if directly transplanted into these cities, the magnitudes are broadly consistent with the mixed empirical evidence on rail investments in New York and Los Angeles. For example, Severen (2023) estimate that gains from Los Angeles's Metro Rail were between 2 and 64 percent of its annualized cost by 2000, while Gupta et al. (2020) find that the expansion of New York City's subway network raised property values by between 53 and 177 percent of construction costs.

Taken together, the results in this section highlight three facets of Tokyo's rail system that are important for understanding its success. First, core-focused subway investment contributes disproportionately to the welfare gains of the system and to the centralization of population and employment, reflecting both the favorable placement of subway lines in areas with many commuters and complementarities across lines. Second, expanding the network along Tokyo's historical path generated high marginal benefit–cost ratios that have tended to grow for most construction periods, with lower returns only for recent peripheral extensions, but the segment-level comparisons show that average benefit–cost ratios for subway segments exceed even the peak marginal ratios along the historical expansion path, whereas radial segments perform similarly to or worse than the historical average. Thus, these comparisons suggest that, conditional on a network of Tokyo's scale, the composition of investment between core subway lines and peripheral radial extensions is at least as important for welfare as further increases in total track length. Third, the value of a network of Tokyo's scale depends critically on urban form and mobility options: faster car travel and lower population density substantially reduce the benefit–cost ratio of the system and can render it unprofitable in environments that resemble car-oriented, low-density U.S. metropolitan areas.

9 Conclusion

This paper has evaluated the economic impact of Tokyo's rail network on the spatial distribution of population and employment and on welfare. Using a quantitative commuting model disciplined by detailed data on population, employment, land prices, commuting flows, travel times across modes, and construction and operating costs, we quantified the systems' impacts on the location of economic activity and welfare relative to a no-rail benchmark, how the system's effects depend on the composition of the network between core subway lines and radial extensions, the scale of the system, and car travel conditions and population density.

Three main conclusions emerge. First, Tokyo's rail network substantially centralizes both population and employment and generates large net benefits. Relative to a no-rail benchmark, the existing system raises the center's share of residents and jobs and delivers welfare gains that exceed the present value of construction and maintenance costs by a comfortable margin. Second, network design matters at least as much as network scale. Core subway lines, though only a small share of total track length and expenditure, account for a disproportionate share of the network's welfare and centralization effects, and typical subway segments yield benefit–cost ratios that meet or exceed those of the most productive periods of overall network expansion. Third, the gains from rail are sensitive to the broader urban environment. Faster car travel and lower population density both weaken the role of rail: when car access approaches that of large U.S. cities or when density is reduced toward Los Angeles–like levels, the net returns to a Tokyo-style network fall sharply and can become negative.

The analysis has several implications for urban transport policy. For dense, rail-oriented cities like Tokyo, the results suggest that well-designed rail networks can deliver sizable net benefits and that concentrating investment in core connectivity, rather than extending radial lines ever farther into the periphery, can be particularly productive. At the same time, the sensitivity of our findings to car speeds and density cautions against directly extrapolating Tokyo's experience to lower-density, car-oriented cities: in such environments, comparable investments in rail are unlikely to be justified unless they are accompanied by land-use and transport policies that support dense, transit-oriented development. More broadly, understanding when large rail systems can replicate Tokyo's performance will require further evidence from other cities that differ in their land-use patterns, transport options, and institutional settings.

Additional Figures and Tables

Summary Statistics, Background, and Train Costs

Level	Variable	N	Mean	Median	Min	Max	p_5	p_{25}	p_{75}	p_{95}	SD
Municipality	Approx. radius (km)	254	3.73	3.26	1.28	10.8	1.8	2.45	4.59	7.22	1.75
Municipality	Area (sq km)	254	53.2	33.5	5.16	369.1	10.2	18.9	66.1	163.7	56.2
Municipality	Employment	252	77,424	37,002	357	1,211,738	3,523	14,315	81,504	228,099	138,819
Municipality	Land Price	252	258,598	123,686	11,158	5,607,025	17,691	43,128	235,932	758,157	566,941
Municipality	Population	252	150,979	107,324	987	943,664	8,833	37,937	203,767	470,741	155,468
Municipality	Station count	254	8.04	4	0	65	0	2	9	31.3	10.3
Oaza	Approx. radius (km)	8,778	0.58	0.51	0.02	5.93	0.17	0.34	0.71	1.22	0.39
Oaza	Area (sq km)	8,778	1.53	0.82	0	110.6	0.09	0.36	1.6	4.69	3.36
Oaza	Employment	8,778	2,223	692	0	301,230	20	191	2,072	7,649	7,167
Oaza	Land Price	8,778	275,715	107,183	5,142	31,052,916	18,235	41,825	242,923	881,167	937,017
Oaza	Population	8,778	4,334	1,987	0	62,708	87	546	5,725	16,480	5,884
Oaza	Station count	8,778	0.23	0	0	11	0	0	0	1	0.71

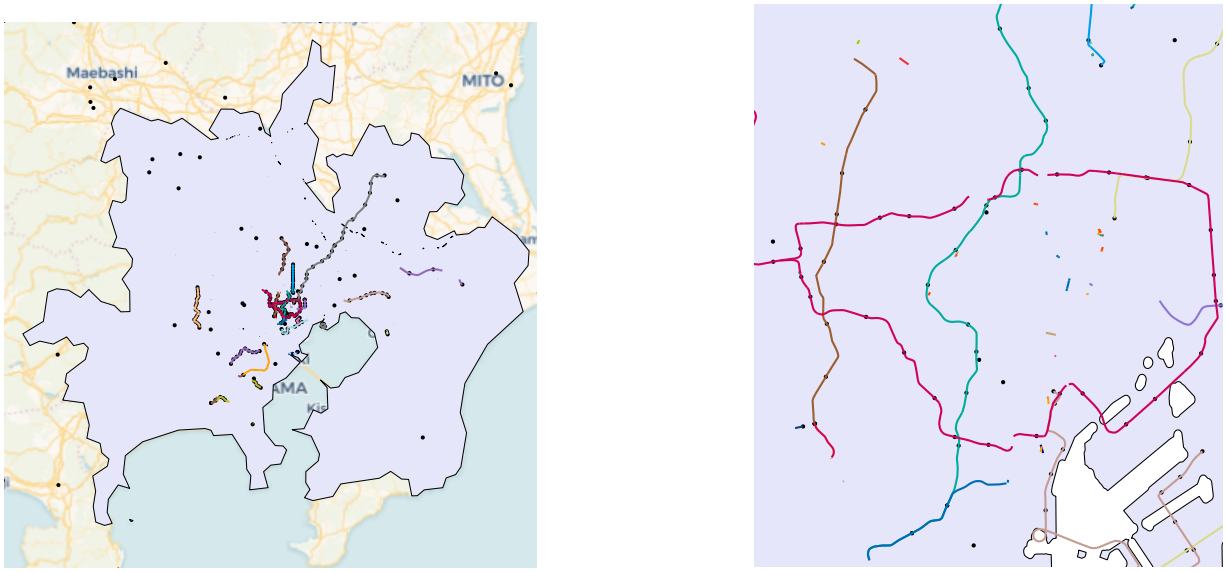
Table 2: Summary statistics for observed variables and station access

Notes: Land prices come from the MLIT Official Land Price Announcements and are reported in yen per square meter. Station counts are the number of stations contained in a region. Approx radius refers to the radius of a circle with the same area as the region.

Discount rate	Construction costs	Land cost	Operating & Maintenance	Total cost
0.5%	0.51%	0.57%	1.86%	2.94%
1.0%	1.03%	1.13%	1.86%	4.03%
2.0%	2.06%	2.27%	1.86%	6.19%
3.0%	3.09%	3.40%	1.86%	8.35%
4.0%	4.12%	4.53%	1.86%	10.51%
5.0%	5.15%	5.67%	1.86%	12.68%
6.0%	6.18%	6.80%	1.86%	14.84%
7.0%	7.20%	7.93%	1.86%	17.00%
8.0%	8.23%	9.07%	1.86%	19.16%
9.0%	9.26%	10.20%	1.86%	21.33%

Table 3: Annualized Costs as Percent of Tokyo's GDP

The table reports the annualized costs of Tokyo's train system as a percent of Tokyo's GDP for different discount rates. The construction costs and land costs are annualized using a standard annuity formula, while the operating and maintenance costs are assumed to be constant each year. The total cost is the sum of these three components. Construction costs are derived from estimates from NYU Marron's Transit Costs Project, scaled by the track length of Tokyo's train system. Land costs are calculated by combining geospatial data on the location of the train network and land prices. Operating and maintenance costs are derived from operators' financial statements, scaled by the station count and track length of Tokyo's system. See Section 3 for additional information.



(a) Greater Tokyo Area

(b) Central Tokyo

Figure 10: Changes to Train Lines from 1995-2020

The maps show the new train lines and stations that were added to the system between 1995 and 2020. Colored lines are train lines while black dots are stations. Shaded region is the Greater Tokyo Area. Data is from the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT)

	OLS Modes	PPML Modes	OLS Same Slope	PPML Same Slope	OLS Modes	PPML Modes
Time			-0.013	-0.066		
			(0.000)	(0.000)		
Mode # Time						
bus	-0.010	-0.054			-0.011	-0.085
	(0.001)	(0.001)			(0.001)	(0.002)
car	-0.014	-0.090			-0.015	-0.085
	(0.000)	(0.001)			(0.000)	(0.001)
cycle	-0.029	-0.145			-0.034	-0.169
	(0.002)	(0.003)			(0.002)	(0.003)
train	-0.012	-0.054			-0.011	-0.047
	(0.000)	(0.000)			(0.000)	(0.000)
walk	-0.034	-0.070			-0.052	-0.172
	(0.002)	(0.002)			(0.003)	(0.006)
Mode						
car	0.241	2.784	0.030	1.326		
	(0.042)	(0.085)	(0.016)	(0.046)		
cycle	0.827	3.617	0.251	0.995		
	(0.061)	(0.112)	(0.021)	(0.053)		
train	0.477	2.805	0.406	3.085		
	(0.040)	(0.074)	(0.016)	(0.044)		
walk	1.171	0.891	0.458	0.273		
	(0.071)	(0.117)	(0.028)	(0.070)		
N	47456	1722825	47456	1722825	46790	1506798
Origin FE	Yes	Yes	Yes	Yes		
Destination FE	Yes	Yes	Yes	Yes		
Origin-Mode FE					Yes	Yes
Destination-Mode FE					Yes	Yes

Table 4: Gravity Estimates

OLS and PPML estimates of commuter flows on travel times. Travel times refer to average one-way travel times in minutes. "Stacked" estimates estimate one slope coefficient for all modes. "Modes" specification estimate separate slopes for each mode. "Same Slope" allows for mode-specific constants but one slope across all modes. Slope coefficients are interpreted as the log-point change in commuter flows associated with a one-minute increase in commuting time for particular mode, holding constant origin-mode and destination-mode fixed effects (multilateral resistance). The unit of observation is the basic zone, which contains on average 50 thousand residents. Data on commuter trips and reported times are obtained by Tokyo's Person-Trip Survey. Reported travel times are only available for observations with trips, so travel times for observations with no trips are imputed using predictions from travel times that are computed using the street map and the train network. Details are found in section C.4 of the appendix. Robust standard errors in parentheses. OLS estimates were estimated with Stata's "reghdfe" package while PPML estimates were estimated with Stata's "ppmlhdfe" command. See Correia et al. (2020) and Correia (2023).

Mode	Slope (κ_m)
Train	0.021
Car	0.037
Walk	0.076
Cycle	0.074
Bus	0.037

Table 5: Value of Time By Mode

This table reports value of time for each mode in units of log wages per minute of one-way travel time. These are calculated by dividing the gravity estimates in Table 4 by the heterogeneity parameter reported in Table 8.

	New Station		First Station	
	Prod.	Amen.	Prod.	Amen.
Log Change in Employment	0.277		0.305	
	(0.013)		(0.010)	
Log Change in Population		0.367		0.351
		(0.012)		(0.014)
N	8719	8594	8719	8594
K–P rk F-Stat	15.0	15.3	17.7	12.2
Munic FE	Yes	Yes	Yes	Yes
# Stations <1km	835	835	229	229
# Stations <3km	2683	2683	174	174
# Stations <5km	3977	3977	107	107

Table 6: Summary of Agglomeration IV Results

This table summarizes the main IV results from Tables F.3, F.4, F.5, and F.6. The first two columns correspond to the results using the "New Station" instruments (i.e. whether or not a tract experienced an increase in the number of train stations within 1km, 3km, or 5km), while the last two columns correspond to the results using the "First Station" instruments (i.e. whether or not a tract experienced its first train station within 1km, 3km, or 5km). The "# Stations <X km" rows report the number of census tracts that experienced an increase from 0 to at least 1 station within X km (for the First Station instrument) or the number of census tracts that experienced an increase in the number of stations within X km (for the New Station instrument). All specifications include municipality fixed effects. Standard errors are clustered at the municipality level. K–P rk F-Stat refers to the Kleibergen–Paap rk F statistic for weak instruments.

Model Validation

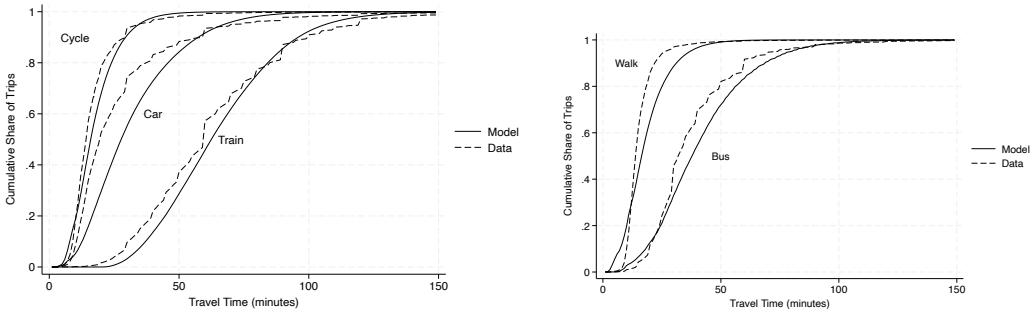
	Model	Survey
mode		
bus	3.42%	2.42%
car	19.32%	22.89%
cycle	13.61%	10.38%
train	54.10%	54.47%
walk	9.55%	6.63%

Table 7: Mode Shares for Greater Tokyo (Data vs Model)

(a) $corr = 0.9$ $\beta = 0.74, \beta_{\text{weighted}} = 0.84$ (b) $corr = 0.94$ $\beta = 0.61, \beta_{\text{weighted}} = 0.72$

Figure 11: Car and Train Mode Shares by Municipality (Data vs Model)

The figures compare the mode shares by municipality for the data and the model (for the estimation discussed in Section 5. The left figure compares the mode shares for trains and the right figure compares the mode shares for cars.
The deviations from a perfect fit arise for two reasons. First, there are differences in geographical aggregation between the commuter flow data and the other spatial units (municipality and census tract). Second, there are census tracts for which not all five mode-location fixed effects were identified in the PPML estimation for a particular region due to there not being enough commuter flows for a mode-location (e.g. train trips for a location with no train stations in its proximity). For these census tracts, we performed a spatial interpolation of the mode-location fixed effects, which leads to a deviation in the corresponding municipality's mode share relative to the data. This missing data largely occurred in the peripheries, because there may be fewer trips from these regions on particular modes, such as trains and buses). Hence, we believe these discrepancies lead to larger deviations from the model to the data for peripheral regions, which is what we find.



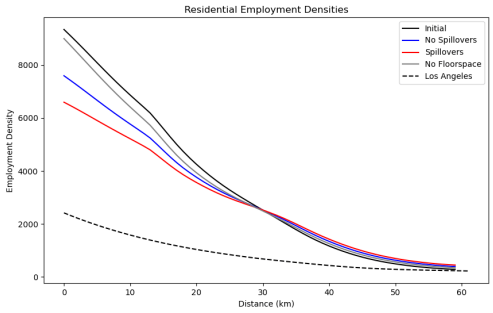
(a) Train/Car/Cycle (b) Bus/Walk

Figure 12: Travel Time Distribution by Mode (Data vs Model)

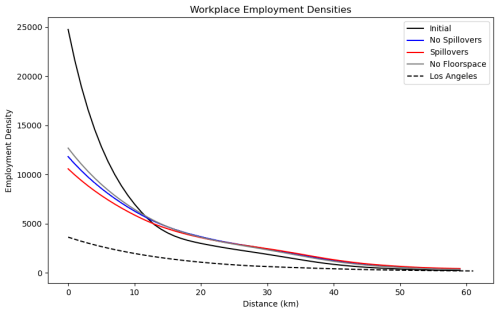
Removing the Train System

Parameter	Value	Source
Heterogeneity Parameter (θ)	2.3	Estimated
Cost share of Labor (α)	0.8	Valentinyi and Herrendorf (2008)
Expenditure share of final good (β)	0.75	Davis and Ortalo-Magné (2011)
Construction sector cost share of physical capital (μ)	0.5	Nakajima et al. (2018)
Production spillover (μ_A)	0.28	Estimated
Amenity spillovers (μ_U)	0.37	Estimated
First-Stage Migration Parameter (η)	2	Head and Mayer (2021), Bordeu (2024)

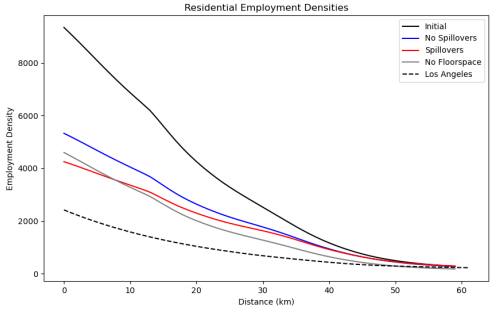
Table 8: Parameters Used in Counterfactual



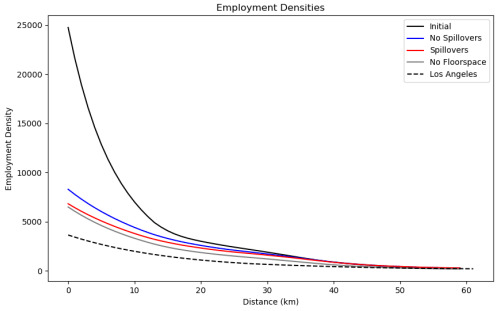
(a) Population Density (Closed City)



(b) Employment Density (Closed City)



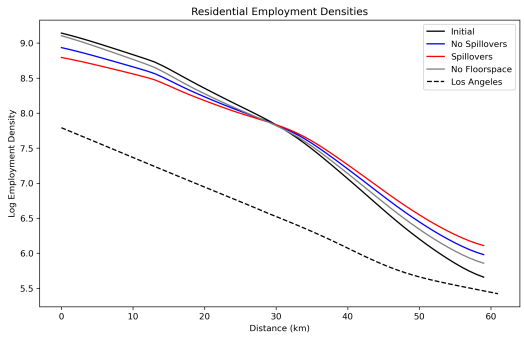
(c) Population Density (Open City)



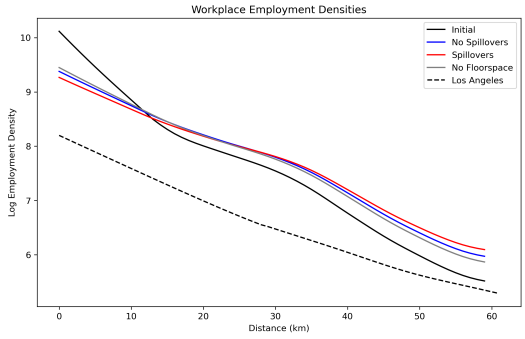
(d) Employment Density (Open City)

Figure 13: Population and Employment Density by Distance to City Center

The figure plots the population and employment density by distance to the Imperial Palace for Tokyo for the counterfactual of removing Tokyo's train system while holding the travel times of other travel modes fixed. "Initial" refers to the 2020 equilibrium. "No Spillover" refers to the model of shared floorspace with $\mu_U = \mu_A = 0$. "Spillovers" sets μ_U and μ_A according to Table 8. "No Floorspace" refers to the model without floorspace. "Los Angeles" refers to the population and employment density by distance to Downtown for Los Angeles.



(a) Population Density



(b) Employment Density

Figure 14: Log Population and Employment Density by Distance to City Center (Closed City)

The figure plots the log population and log employment density by distance to the Imperial Palace for Tokyo for the counterfactual of removing Tokyo's train system while holding the travel times of other travel modes fixed. "Initial" refers to the 2020 equilibrium. "No Spillover" refers to the model of shared floorspace with $\mu_U = \mu_A = 0$. "Spillovers" sets μ_U and μ_A according to Table 8. "No Floorspace" refers to the model without floorspace. "Los Angeles" refers to the population and employment density by distance to Downtown for Los Angeles.

	With Train	Remove Train					
		No Floorspace		Shared Floorspace		Floorspace + Spillovers	
	Level	Level	Pct. Chg	Level	Pct. Chg	Level	Pct. Chg
Centralization							
<10km Population Share	0.236	0.187	-20.9%	0.162	-31.4%	0.138	-41.4%
<10km Employment Share	0.416	0.366	-12.0%	0.348	-16.4%	0.332	-20.3%
Central: Population Share	0.46	0.413	-10.2%	0.385	-16.4%	0.356	-22.6%
Central: Employment Share	0.576	0.529	-8.2%	0.507	-12.0%	0.487	-15.6%
Log Pop. Density Gradient	0.075	0.065	-13.3%	0.059	-21.0%	0.053	-29.4%
Log Emp. Density Gradient	0.094	0.087	-7.7%	0.083	-11.2%	0.08	-14.8%
Central $\frac{\text{Employment}}{\text{Population}}$	1.252	1.28	2.2%	1.318	5.2%	1.366	9.1%
Closed Economy: Welfare							
Welfare (excl. construction)	1	0.763	-23.7%	0.781	-21.9%	0.794	-20.6%
Welfare (2% discount rate)	–	19.6%	–	17.7%	–	16.4%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (2% discount rate)	–	5.669	–	5.219	–	4.917	–
Welfare (4% discount rate)	–	17.7%	–	15.8%	–	14.5%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (4% discount rate)	–	3.898	–	3.589	–	3.381	–
Welfare (6% discount rate)	–	12.2%	–	10.3%	–	9%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (6% discount rate)	–	2.052	–	1.889	–	1.78	–
With Migration							
Total Population	1	0.815	-18.5%	0.761	-23.9%	0.711	-28.9%
Total Land Value	1	1	0.0%	0.737	-26.3%	0.663	-33.7%

Table 9: Impact of Train Removal on Centralization and Welfare

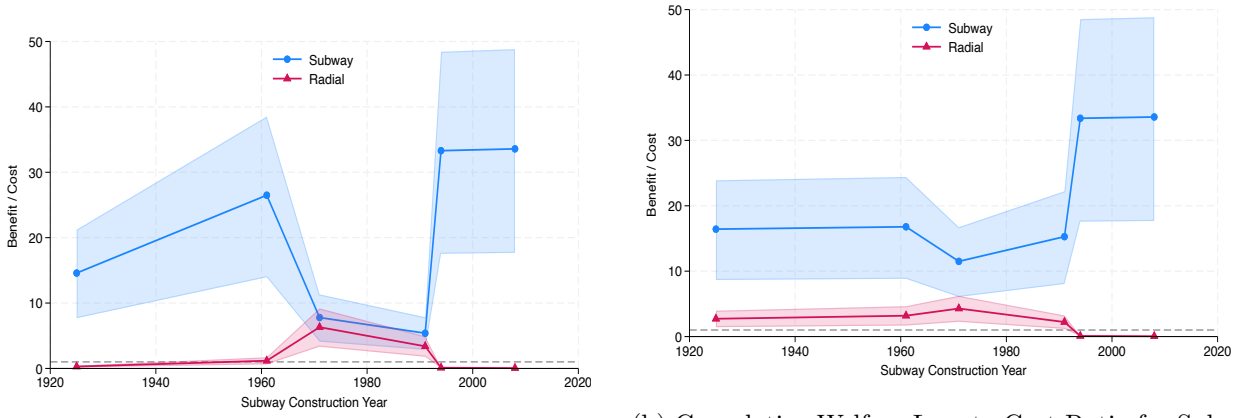
Notes: "With Train" reports the 2020 equilibrium, while the remaining columns remove all rail segments under the three model variants described in Section 6. "No Floorspace" refers to the model that omits floorspace from the model but amenities and productivities exhibit congestion spillovers. "Shared Floorspace" includes floorspace that is efficiently shared between residential and commercial uses, and where amenities and productivities are exogenous. "Floorspace + Spillover" allows amenities and productivities to be functions of the level of population and employment, respectively. Percentage changes are relative to the baseline level in each row. "Central" refer to regions within 20 km of the Imperial Palace. "Log Pop. Density Gradient" and "Log Emp. Density Gradient" reports the slope of the log population and employment density with respect to distance (in kilometers) from the center. "Welfare (excl. construction)" reports the consumption equivalent change in welfare for the household for a closed city if the construction, maintenance, and operating costs are not considered. Underneath, welfare rows report closed-city consumption equivalents (accounting for construction, operating, and maintenance costs) and associated benefit–cost ratios for discount rates of 2%, 4%, and 6% when annualizing construction and land costs. Migration rows allow Tokyo's total population and land value to adjust, holding constant Japan's population and assuming construction costs are paid by the national government.

	Train Slow			Train Fast			Train Slow in Center		
	No Train	With Train	% Chg.	No Train	With Train	% Chg.	No Train	With Train	% Chg.
Central Population Absolute	0.40	0.42	4.11	0.40	0.43	6.80	0.40	0.35	-14.40
Central Employment Absolute	0.42	0.44	5.23	0.42	0.67	59.02	0.42	0.50	18.57
Log Population Gradient	0.06	0.06	3.62	0.06	0.07	13.31	0.06	0.06	-4.07
Log Employment Gradient	0.06	0.06	4.23	0.06	0.09	53.56	0.06	0.07	17.91
Central Employment vs Population Ratio	1.04	1.05	1.08	1.04	1.55	48.89	1.04	1.45	38.51
Central vs Periph. RCA Ratio	1.65	1.90	15.52	1.65	2.20	33.60	1.65	1.47	-10.80
Central vs Periph. FCA Ratio	0.69	0.74	8.14	0.69	2.63	283.80	0.69	0.98	43.31

Table 10: Robustness to Train Speeds

	Baseline			Same VoT			No Cars		
	No Train	With Train	% Chg.	No Train	With Train	% Chg.	No Train	With Train	% Chg.
Central Population Absolute	0.40	0.46	13.92	0.33	0.39	16.23	0.49	0.51	4.76
Central Employment Absolute	0.42	0.58	36.68	0.39	0.49	24.30	0.50	0.65	31.13
Log Population Gradient	0.06	0.07	16.17	0.05	0.06	15.74	0.07	0.08	8.15
Log Employment Gradient	0.06	0.08	30.74	0.06	0.07	19.66	0.07	0.09	27.80
Central Employment vs Population Ratio	1.04	1.25	19.98	1.19	1.27	6.94	1.02	1.28	25.17
Central vs Periph. RCA Ratio	1.65	2.62	58.79	0.98	1.50	52.91	3.16	4.05	28.30
Central vs Periph. FCA Ratio	0.69	1.37	99.84	0.54	0.80	47.36	1.44	3.13	117.85

Table 11: Robustness to Change in Value of Time and Removing Cars

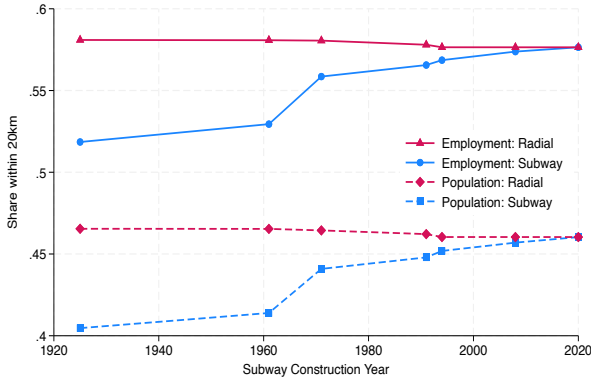


(a) Benefit/Cost of Marginal Segments

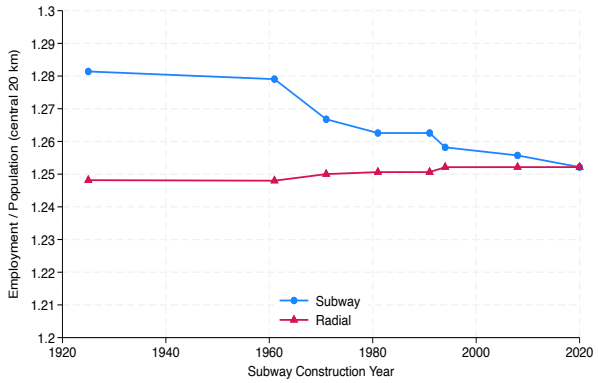
(b) Cumulative Welfare Loss to Cost Ratio for Subway and Radial Lines Relative to 2020 Equilibrium

Figure 15: Benefit-Cost Ratios for Subways and Radial Lines

Notes: The left series plots the Benefit/Cost from removing expenditure-equivalent segments of subways and radial lines. The right series display the cumulative Loss/Cost ratio, relative to the 2020 equilibrium with the train network. The "Subway" plot removes subway segments according to the subway segments' construction year (removing the most recent subway lines first). The "Radial" plot removes the expenditure-equivalent amount of the most recently constructed radial lines.
Subway investments consistently deliver ratios well above one, whereas recent radial-line additions exhibit far smaller payoffs, illustrating that late peripheral expansions generate limited incremental value.



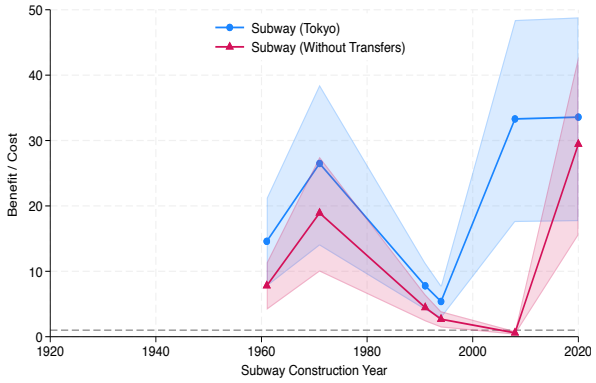
(a) Share of Total Population and Employment in the Center



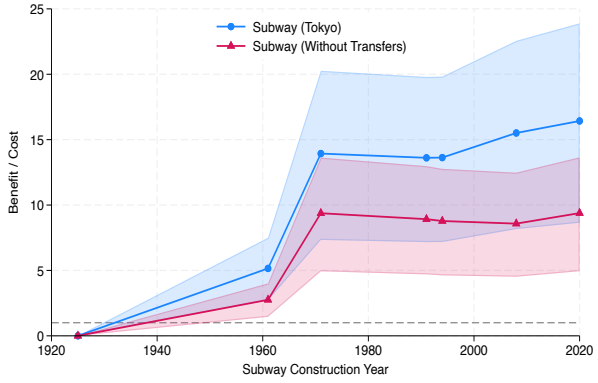
(b) Employment vs Population in the Center

Figure 16: The Impact of Train Segments on Centralization and Specialization Patterns

Notes: The series displays metrics from removing expenditure-equivalent segments of subways and radial lines, relative to the 2020 equilibrium. The "Subway" plot removes subway segments according to the subway segments' construction year (removing the most recent subway lines first). The "Radial" plot removes the expenditure-equivalent amount of the most recently constructed radial lines. Panel (a) shows that incremental subway expansions raise both the population and employment shares in the core, as discussed in Section 8. Panel (b) shows that population responds more strongly than employment, implying that subway expansions reduce core specialization (employment-to-population ratios fall). In contrast, radial lines have a decentralizing effect and contributes to specialization. Metrics are computed for municipalities within 20 km of the Imperial Palace.



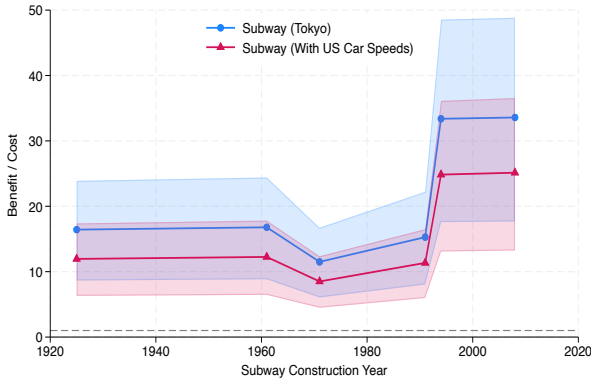
(a) Benefit/Cost of Marginal Segment



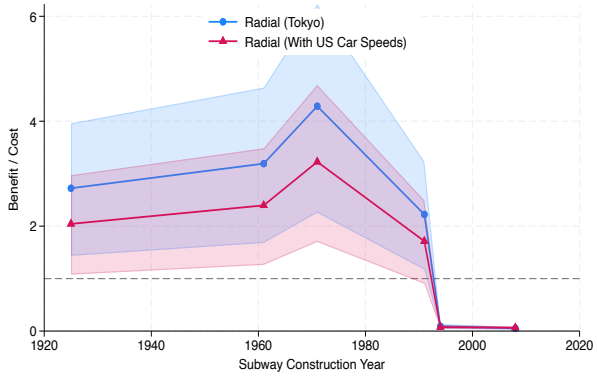
(b) Cumulative Benefit/Cost from 1925 to Construction Year

Figure 17: Welfare Gains to Cost Ratio for Subway Segments When Transfers Between Train Lines Are Not Permitted

Notes: The right series display the cumulative Benefit/Cost ratio from adding every segment of the subway, relative to a counterfactual where the subway does not exist. The left series display the Benefit/Cost ratio from the marginal segment of the subway. "Subway (Tokyo)" refers to our baseline counterfactual of removing subway segments. "Subway (Without Transfers)" refers to counterfactuals of the train network where commuters are not permitted to make transfers to and from subway stations. Disallowing transfers reduces the benefit–cost profile of every subway segment, underscoring the network effects highlighted in Section 8: a large share of the value of new subway segments derive from how they expand the train network.



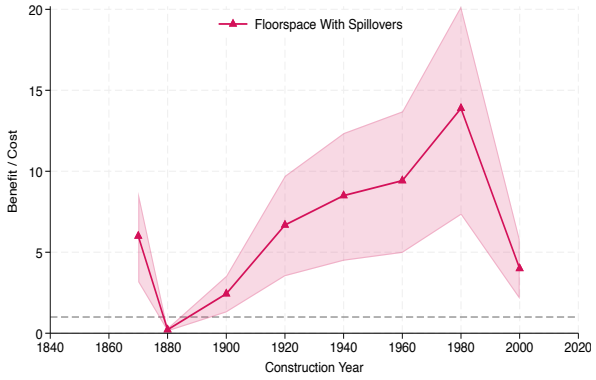
(a) Subway



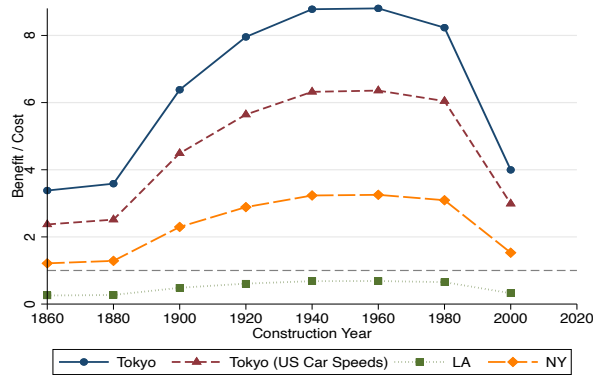
(b) Radial Train Lines

Figure 18: Cumulative Welfare Gains to Cost Ratio for Subways and Radial Lines when Car Speeds Match Typical US City

Notes: The series display the cumulative Benefit/Cost ratio from removing expenditure-equivalent segments of subways and radial lines, relative to the 2020 equilibrium. The "Subway" plot removes subway segments according to the subway segments' construction year (removing the most recent subway lines first). The "Radial" plot removes the expenditure-equivalent amount of the most recently constructed radial lines. The "Tokyo" plots shows our baseline results (also found in Figure 15) while the "With US Car Speeds" shows the Benefit/Cost ratios in an alternative environment where car speeds match those of the average U.S. city. Benefits from both subways and radial lines are reduced by similar proportions.



(a) Benefit/Cost of Marginal Segment



(b) Cumulative Benefit/Cost with Population Density Matched to New York and Los Angeles

Figure 19: Historical Welfare Gains to Cost Ratio

Notes: Panel (a) displays the Benefit/Cost ratios of each of the train-network segments that were constructed for each 20-year period. The Benefit/Cost ratio of marginal segments increased between 1880 and 1999, but dropped significantly for the period from 2000-2020. We suspect that the drop is largely due to the fact that the lines constructed from 2000-2020 were radial or were otherwise placed outside the city's core.

Panel (b) shows the cumulative Benefit/Cost ratios relative to the 2020 equilibrium for different environments. The series for "Tokyo" are the baseline results. "LA" and "NY" first raises car speeds in Tokyo to those of a typical US city (as in Figure 18). The total population of Tokyo is then exogenously reduced to those of the Greater Los Angeles and Greater New York, while the costs of construction, maintenance, and operations are held fixed. The train network continues to be socially beneficial at New York-like population densities but turn negative at Los Angeles-like densities. Thus, the results on the benefits of Tokyo's train system are sensitive to the accessibility of alternative transportation modes and of the city's population density.

	With Train (2020)	Remove Subway			Remove Radial	
	Level	Level	vs Base	vs No Transfers	Level	vs Base
Centralization						
<10km Population Share	0.236	0.177	-25.0%		0.239	1.1%
<10km Employment Share	0.416	0.349	-16.3%		0.42	0.8%
Central: Population Share	0.46	0.405	-12.1%		0.465	1.1%
Central: Employment Share	0.576	0.519	-10.0%		0.581	0.8%
Log Pop. Density Gradient	0.075	0.065	-13.7%		0.075	0.7%
Log Emp. Density Gradient	0.094	0.086	-8.4%		0.094	0.6%
Central $\frac{\text{Employment}}{\text{Population}}$	1.252	1.281	2.3%		1.248	-0.3%
Closed Economy: Welfare						
Welfare (excl. construction)	100%	93.6%	-6.4%	–	98.9%	-1.1%
Welfare (2% discount rate)		6.2%	–	3.4%	0.8%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (2% discount rate)		23.891	–	13.648	3.959	–
Welfare (4% discount rate)		6.1%	–	3.3%	0.7%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (4% discount rate)		16.429	–	9.385	2.723	–
Welfare (6% discount rate)		5.7%	–	2.9%	0.3%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (6% discount rate)		8.647	–	4.94	1.433	–
With Migration						
Total Population	1	0.91	-9.0%		0.985	-1.5%
Total Land Value	1	0.861	-13.9%		0.988	-1.2%

Table 12: Impact of Removing Subway and Expenditure-Equivalent Length of Radial Trains

Notes: "With Train" reports the 2020 equilibrium, while the remaining columns removes segments of the train system. "Remove Subway" removes all subways located in the core of the city. "vs Base" reports percentage changes relative to the "With Train (2020)" equilibrium. "vs No Transfers" reports the welfare figures relative to an equilibrium where subways are present but transfers to and from subways are not permitted. Thus, the welfare figures represent welfare from the subway excluding the value from connecting to the train network. "Remove Radial" removes the expenditure-equivalent amount of recently constructed radial train lines. We find that the value from connecting to the network is almost as large as the individual subway lines themselves, underscoring strong network effects from the train network. In addition, we find (i.) subways account for a disproportionate share of the welfare gains; (ii.) subways are a centralizing force for both employment and population while radial lines are a decentralizing force; (iii.) subways centralize population more than it does employment, leading to reduced specialization of the core as a workplace, as measured by the ratio of employment to population.
"Central" refer to regions within 20 km of the Imperial Palace. "Log Pop. Density Gradient" and "Log Emp. Density Gradient" reports the slope of the log population and employment density with respect to distance (in kilometers) from the center. "Welfare (excl. construction)" reports the consumption equivalent change in welfare for the household for a closed city if the construction, maintenance, and operating costs are not considered. Underneath, welfare rows report closed-city consumption equivalents (accounting for construction, operating, and maintenance costs) and associated benefit–cost ratios for discount rates of 2%, 4%, and 6% when annualizing construction and land costs. Migration rows allow Tokyo's total population and land value to adjust, holding constant Japan's population and assuming construction costs are paid by the national government.

	Baseline			US Car Speeds			
	Train	No Train		Train		No Train	
	Level	Level	Pct. Chg	Level	Pct. Chg	Level	Pct. Chg
Centralization							
<10km Population Share	0.236	0.138	-41.4%	0.159	-32.6%	0.089	-43.8%
<10km Employment Share	0.416	0.332	-20.3%	0.374	-10.3%	0.324	-13.4%
Central: Population Share	0.46	0.356	-22.6%	0.374	-18.9%	0.298	-20.1%
Central: Employment Share	0.576	0.487	-15.6%	0.54	-6.2%	0.49	-9.3%
Log Pop. Density Gradient	0.075	0.053	-29.4%	0.058	-22.7%	0.038	-33.5%
Log Emp. Density Gradient	0.094	0.08	-14.8%	0.088	-5.8%	0.081	-8.9%
Central $\frac{\text{Employment}}{\text{Population}}$	1.252	1.366	9.1%	1.447	15.5%	1.643	13.6%
Closed Economy: Welfare							
							Pct. Pt. Chg.
Welfare (excl. construction)	100%	79.4%	-20.6%	125.1%	25.1%	110.6%	-14.4%
Welfare (2% discount rate)		16.4%	–		–	10.3%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (2% discount rate)		4.917	–		–	3.449	–
Welfare (4% discount rate)		14.5%	–		–	8.4%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (4% discount rate)		3.381	–		–	2.372	–
Welfare (6% discount rate)		9%	–		–	2.9%	–
$\frac{\text{Benefit}}{\text{Cost}}$ (6% discount rate)		1.78	–		–	1.248	–
With Migration							
Total Population	1	0.711	-28.9%	1.33	33.0%	1.143	-18.7%
Total Land Value	1	0.663	-33.7%	1.289	28.9%	1.067	-22.2%

Table 13: Impact of Removing The Train System when Automobile Speeds Match the U.S.

Notes: "Baseline" reports the baseline removal of the entire train system for the model with shared floorspace and spillovers, as in Section 6. In the "US Car Speeds", "Train" column, we report the equilibrium where the speed of cars in Tokyo match those of the typical U.S. city. The "US Car Speeds", "No Train" then removes the train in this environment and reports the percent changes in centralization statistics relative to the equilibrium with the train system but with U.S. car speeds. Welfare figures report consumption equivalents at baseline prices.
"Central" refer to regions within 20 km of the Imperial Palace. "Log Pop. Density Gradient" and "Log Emp. Density Gradient" reports the slope of the log population and employment density with respect to distance (in kilometers) from the center. "Welfare (excl. construction)" reports the consumption equivalent change in welfare for the household for a closed city if the construction, maintenance, and operating costs are not considered. Underneath, welfare rows report closed-city consumption equivalents (accounting for construction, operating, and maintenance costs) and associated benefit–cost ratios for discount rates of 2%, 4%, and 6% when annualizing construction and land costs. Migration rows allow Tokyo's total population and land value to adjust, holding constant Japan's population and assuming construction costs are paid by the national government.

References

Gabriel M. Ahlfeldt, Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6):2127–2189, 2015. ISSN 1468-0262. doi: 10.3982/ECTA10876. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10876. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10876.

Treb Allen and Costas Arkolakis. The Welfare Effects of Transportation Infrastructure Improvements. *The Review of Economic Studies*, 89(6):2911–2957, November 2022. ISSN 0034-6527. doi: 10.1093/restud/rdac001. URL https://doi.org/10.1093/restud/rdac001.

Treb Allen and Costas Arkolakis. Economic Activity across Space: A Supply and Demand Approach. *Journal of Economic Perspectives*, 37(2):3–28, May 2023. ISSN 0895-3309. doi: 10.1257/jep.37.2.3. URL https://pubs.aeaweb.org/doi/10.1257/jep.37.2.3.

Treb Allen and Costas Arkolakis. Quantitative Regional Economics, January 2025. URL https://papers.ssrn.com/abstract=5122155.

Treb Allen, Costas Arkolakis, and Yuta Takahashi. Universal Gravity. *Journal of Political Economy*, 128(2):393–433, February 2020. ISSN 0022-3808. doi: 10.1086/704385. URL https://www.journals.uchicago.edu/doi/full/10.1086/704385. Publisher: The University of Chicago Press.

Treb Allen, Costas Arkolakis, and Xiangliang Li. On the Equilibrium Properties of Spatial Models. *American Economic Review: Insights*, 6(4):472–489, December 2024. ISSN 2640-205X, 2640-2068. doi: 10.1257/aeri.20230495. URL https://pubs.aeaweb.org/doi/10.1257/aeri.20230495.

Costas Arkolakis, Arnaud Costinot, and Andrés Rodríguez-Clare. New Trade Models, Same Old Gains? *American Economic Review*, 102(1):94–130, February 2012. ISSN 0002-8282. doi: 10.1257/aer.102.1.94. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.1.94.

Kenzo Asahi, Andrea Herrera, and Hugo E Silva. The effects of transport infrastructure on housing supply and the role of land-use regulation. 2023.

Mark Bamba. Comparative-Static Heuristics in Spatial Models. 2025.

Abhijit Banerjee, Esther Duflo, and Nancy Qian. On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145:102442, June 2020. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2020.102442. URL https://www.sciencedirect.com/science/article/pii/S0304387820300171.

Nathaniel Baum-Snow. Did Highways Cause Suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805, 2007. ISSN 0033-5533. URL https://www.jstor.org/stable/25098858. Publisher: Oxford University Press.

Nathaniel Baum-Snow, Matthew E. Kahn, and Richard Voith. Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000 [with Comment]. *Brookings-Wharton Papers on Urban Affairs*, pages 147–206, 2005. ISSN 1528-7084. URL https://www.jstor.org/stable/25067419. Publisher: Brookings Institution Press.

Nathaniel Baum-Snow, Loren Brandt, J. Vernon Henderson, Matthew Turner, and Qinghua Zhang. Roads, Railroads, and Decentralization of Chinese Cities. *The Review of Economics and Statistics*, 99(3):435–448, 2017. URL https://econpapers.repec.org/article/tprrestat/v_3a99_3ay_3a2017_3ai_3a3_3ap_3a435-448.htm. Publisher: MIT Press.

Thor Berger and Kerstin Enflo. Locomotives of local growth: The short- and long-term impact of railroads in Sweden. *Journal of Urban Economics*, 98:124–138, March 2017. ISSN 0094-1190. doi: 10.1016/j.jue.2015.09.001. URL https://www.sciencedirect.com/science/article/pii/S0094119015000595.

Andrew B. Bernard, Andreas Moxnes, and Yukiko U. Saito. Production Networks, Geography, and Firm Performance. *Journal of Political Economy*, 127(2):639–688, April 2019. ISSN 0022-3808, 1537-534X. doi: 10.1086/700764. URL https://www.journals.uchicago.edu/doi/10.1086/700764.

Olivia Bordeu. Commuting Infrastructure in Fragmented Cities. 2024.

Leah Brooks and Zachary D. Liscow. Infrastructure Costs, December 2021. URL https://papers.ssrn.com/abstract=3428675.

Lucas Conwell, Fabian Eckert, and Ahmed Mushfiq Mobarak. Transportation Infrastructure and City-Center Accessibility in the US and Europe. 2023.

Sergio Correia. REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. *Statistical Software Components*, August 2023. URL https://ideas.repec.org//c/boc/bocode/s457874.html. Publisher: Boston College Department of Economics.

Sergio Correia, Paulo Guimarães, and Thomas Zylkin. ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects. *The Stata Journal: Promoting communications on statistics and Stata*, 20(1):95–115, March 2020. ISSN 1536-867X, 1536-8734. doi: 10.1177/1536867X20909691. URL http://arxiv.org/abs/1903.01690. arXiv:1903.01690 [econ].

Morris A. Davis and François Ortalo-Magné. Household expenditures, wages, rents. *Review of Economic Dynamics*, 14(2):248–261, April 2011. ISSN 1094-2025. doi: 10.1016/j.red.2009.12.003. URL https://www.sciencedirect.com/science/article/pii/S1094202509000830.

Robert Dekle, Jonathan Eaton, and Samuel Kortum. Global Rebalancing with Gravity: Measuring the Burden of Adjustment. *IMF Staff Papers*, 55(3):511–540, 2008. ISSN 1020-7635. URL https://www.jstor.org/stable/40377786. Publisher: Palgrave Macmillan Journals.

Eric Donald, Masao Fukui, and Yuhei Miyauchi. Unpacking Aggregate Welfare in a Spatial Economy. 2025.

Dave Donaldson. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934, April 2018. ISSN 0002-8282. doi: 10.1257/aer.20101199. URL https://www.aeaweb.org/articles?id=10.1257/aer.20101199.

Dave Donaldson. Transport Infrastructure and Policy Evaluation. 2025.

Dave Donaldson and Richard Hornbeck. Railroads and American Economic Growth: A "Market Access" Approach *. *The Quarterly Journal of Economics*, 131(2):799–858, May 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw002. URL https://doi.org/10.1093/qje/qjw002.

Gilles Duranton, Geetika Nagpal, and Matthew Turner. Transportation Infrastructure in the US, May 2020. URL https://www.nber.org/papers/w27254.

Benjamin Faber. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*, 81(3):1046–1070, July 2014. ISSN 0034-6527. doi: 10.1093/restud/rdu010. URL https://doi.org/10.1093/restud/rdu010.

Pablo D. Fajgelbaum and Edouard Schaal. Optimal Transport Networks in Spatial Equilibrium. *Econometrica*, 88(4):1411–1452, 2020. ISSN 1468-0262. doi: 10.3982/ECTA15213. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15213. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA15213.

Pablo D. Fajgelbaum, Cecile Gaubert, Nicole Gorton, Eduardo Morales, and Edouard Schaal. Political Preferences and the Spatial Distribution of Infrastructure: Evidence from California's High-Speed Rail, July 2023. URL https://www.nber.org/papers/w31438.

Robert William Fogel. *Railroads and American economic growth: essays in econometric history*. Johns Hopkins Univ. Pr, Baltimore, 2. print edition, 1970. ISBN 978-0-8018-1148-7 978-0-8018-0201-0.

Edward Glaeser, Matthew E. Kahn, and Jordan Rappaport. Why Do The Poor Live In Cities? The Role of Public Transportation. *Journal of Urban Economics*, 2008. ISSN 0094-1190. doi: 10.1016/j.jue.2006.12.004. URL https://dash.harvard.edu/handle/1/2958224. Accepted: 2009-05-14T18:30:49Z Publisher: Elsevier.

Edward L. Glaeser and Matthew E. Kahn. Decentralized Employment and the Transformation of the American City, February 2001. URL https://www.nber.org/papers/w8117.

Edward L. Glaeser and Matthew E. Kahn. Chapter 56 - Sprawl and Urban Growth. In J. Vernon Henderson and Jacques-François Thisse, editors, *Handbook of Regional and Urban Economics*, volume 4 of *Cities and Geography*, pages 2481–2527. Elsevier, January 2004. doi:

10.1016/S1574-0080(04)80013-0. URL https://www.sciencedirect.com/science/article/pii/S1574008004800130.

Marco Gonzalez-Navarro and Matthew A. Turner. Subways and urban growth: Evidence from earth. *Journal of Urban Economics*, 108:85–106, November 2018. ISSN 0094-1190. doi: 10.1016/j.jue.2018.09.002. URL https://www.sciencedirect.com/science/article/pii/S009411901830072X.

Caitlin Gorback. Ridesharing and the Redistribution of Economic Activity, 2024. URL https://www.ssrn.com/abstract=4956559.

Arpit Gupta, Stijn Van Nieuwerburgh, and Constantine Kontokosta. Take the Q Train: Value Capture of Public Infrastructure Projects. Technical Report w26789, National Bureau of Economic Research, March 2020. URL https://www.nber.org/papers/w26789.

Keith Head and Thierry Mayer. The United States of Europe: A Gravity Model Evaluation of the Four Freedoms. *Journal of Economic Perspectives*, 35(2):23–48, May 2021. ISSN 0895-3309. doi: 10.1257/jep.35.2.23. URL https://pubs.aeaweb.org/doi/10.1257/jep.35.2.23.

Stephan Heblich, Stephen J Redding, and Daniel M Sturm. The Making of the Modern Metropolis: Evidence from London*. *The Quarterly Journal of Economics*, 135(4):2059–2133, November 2020. ISSN 0033-5533. doi: 10.1093/qje/qjaa014. URL https://doi.org/10.1093/qje/qjaa014.

HAYAKAWA Kazunobu, Hans R A Koster, TABUCHI Takatoshi, and Jacques-François Thisse. High-speed Rail and the Spatial Distribution of Economic Activity: Evidence from Japan's Shinkansen. 2021.

Benny Kleinman, Ernest Liu, and Stephen J. Redding. International Friends and Enemies, July 2020. URL https://www.nber.org/papers/w27587.

Benny Kleinman, Ernest Liu, and Stephen J. Redding. International Friends and Enemies. *American Economic Journal: Macroeconomics*, 16(4):350–385, October 2024. ISSN 1945-7707, 1945-7715. doi: 10.1257/mac.20220223. URL https://pubs.aeaweb.org/doi/10.1257/mac.20220223. Publisher: American Economic Association.

Cameron LaPoint. You Only Lend Twice: Corporate Borrowing and Land Values in Real Estate Cycles. *SSRN Electronic Journal*, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3633606. URL https://www.ssrn.com/abstract=3633606.

John R. Meyer, Merton J. Peck, John Stenason, and Charles Zwick. *Economics of Competition in the Transportation Industries*. Harvard University Press, Cambridge Massachusetts London, England, January 1959. ISBN 978-0-674-23251-8.

John Robert Meyer. *The Urban Transportation Problem*. Harvard Univ Pr, Cambridge, Mass, pencil margin notes edition edition, January 1965. ISBN 978-0-674-93120-6.

Yuhei Miyauchi, Kentaro Nakajima, and Stephen Redding. The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data. Technical Report w28497, National Bureau of Economic Research, Cambridge, MA, February 2021. URL http://www.nber.org/papers/w28497.pdf.

Yuhei Miyauchi, Kentaro Nakajima, and Stephen J Redding. The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data*. *The Quarterly Journal of Economics*, 140(4): 2507–2570, November 2025. ISSN 0033-5533. doi: 10.1093/qje/qjaf038. URL https://doi.org/10.1093/qje/qjaf038.

Ferdinando Monte, Stephen J. Redding, and Esteban Rossi-Hansberg. Commuting, Migration, and Local Employment Elasticities. *American Economic Review*, 108(12):3855–3890, December 2018. ISSN 0002-8282. doi: 10.1257/aer.20151507. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151507.

Kentaro Nakajima, Iichiro Uesugi, Kaoru Hosono, and Takeshi Mizuta. Estimated Housing Supply Price Elasticities by Urban Employment Area (Beta Version), 2018. URL https://www.ier.hit-u.ac.jp/hit-refined/English/database/elas.html. Published: Web page.

Stephen Redding and Anthony J. Venables. Economic geography and international inequality. *Journal of International Economics*, 62(1):53–82, January 2004. ISSN 0022-1996. doi: 10.1016/j.jinteco.2003.07.001. URL http://www.sciencedirect.com/science/article/pii/S0022199603000965.

Stephen J. Redding. Evaluating Transport Improvements in Spatial Equilibrium. Technical Report w33532, National Bureau of Economic Research, March 2025. URL https://www.nber.org/papers/w33532.

Stephen J. Redding and Esteban Rossi-Hansberg. Quantitative Spatial Economics. *Annual Review of Economics*, 9(1):21–58, 2017. doi: 10.1146/annurev-economics-063016-103713. URL https://doi.org/10.1146/annurev-economics-063016-103713. _eprint: https://doi.org/10.1146/annurev-economics-063016-103713.

Stephen J. Redding and Daniel M. Sturm. The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review*, 98(5):1766–1797, December 2008. ISSN 0002-8282. doi: 10.1257/aer.98.5.1766. URL https://www.aeaweb.org/articles?id=10.1257/aer.98.5.1766.

Stephen J. Redding and Matthew A. Turner. Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity. In Gilles Duranton, J. Vernon Henderson, and

William C. Strange, editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 1339–1398. Elsevier, January 2015. doi: 10.1016/B978-0-444-59531-7.00020-X. URL http://www.sciencedirect.com/science/article/pii/B978044459531700020X.

Christopher Severen. Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification. *The Review of Economics and Statistics*, 105(5):1073–1091, September 2023. ISSN 0034-6535. doi: 10.1162/rest_a_01100. URL https://doi.org/10.1162/rest_a_01100.

Nick Tsivanidis. Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio. 2023.

Akos Valentinyi and Berthold Herrendorf. Measuring Factor Income Shares at the Sector Level. *Review of Economic Dynamics*, 11(4):820–835, 2008. URL https://ideas.repec.org//a/red/issued/07-50.html. Publisher: Elsevier for the Society for Economic Dynamics.

Román D. Zárate. *Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City*. World Bank, Washington, DC, March 2022. doi: 10.1596/1813-9450-9990. URL https://hdl.handle.net/10986/37274.

A Model Details

A.1 Model with Shared Floorspace Ahlfeldt et al. (2015)

Commuting costs are parametrized as

$$\tau_{ijm} = \exp(\kappa^O_{im} + \kappa^D_{jm} + \kappa_m t_{ijm}) \tag{A.1}$$

where t_{ijm} is the travel time for the commute from i to j on commuting mode m; κ_m governs the elasticity of commuting costs with respect to time for mode m and is expressed in units of 100 percent (or log) of wages per unit of time; κ^O_{im} and κ^D_{jm} are residence-mode-specific and workplace-mode-specific constants, respectively, and represent additional costs associated with a particular mode in a particular location that is not captured in travel times, such as the quality of highways or train stations in a particular location or the cost of car parking in a particular location.

The idiosyncratic preference shocks ϵ_{ijm} are drawn from an i.i.d. Fréchet distribution with shape parameter θ and unit scale:

$$F(\epsilon_{ijm}) = e^{-\epsilon_{ijm}^{-\theta}}, \quad \theta > 1,\ \epsilon_{ijm} > 0, \tag{A.2}$$

which implies that the number of commuters for each residence-workplace-mode combination is given by:

$$L_{ijm} = \bar{L}\left(\frac{\gamma}{\bar{V}}\right)^{\theta}\left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ijm}}\right)^{\theta} \tag{A.3}$$

where $\bar{L}$, the mass of workers, will be set to equal the total number of workers in the Tokyo metro, and $\bar{V}$ is the average indirect utility:

$$\bar{V} = \gamma(1+k)\left[\sum_j \sum_i \sum_m \left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ijm}}\right)^{\theta}\right]^{1/\theta} \tag{A.4}$$

and $\gamma \equiv \Gamma\left(\frac{\theta-1}{\theta}\right)$ is a constant that is implied by the Fréchet distribution. $\Gamma(\cdot)$ is the Gamma function. The above expression for average utility is implied by integrating the individual indirect utilities over the Fréchet distribution of idiosyncratic preferences. Equation (A.3) also implies the following expression for the probability that a worker resides in location i and works in location j and commutes using mode m:

$$\lambda_{ijm} \equiv \frac{L_{ijm}}{\sum_{i,j,m} L_{ijm}} = \frac{\left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ijm}}\right)^{\theta}}{\sum_{i,j,m}\left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ijm}}\right)^{\theta}}. \tag{A.5}$$

According to equation (A.3), θ is therefore the elasticity of workers with respect to commuting

costs as well as the elasticity with respect to wages and amenities. The reason for this result is that because the Fréchet shape parameter θ determines the heterogeneity of the distribution of idiosyncratic preferences (a higher θ leads to a lower variance in idiosyncratic preference shocks), the parameter dictates the responsiveness of the population to changes in location characteristics, and hence the elasticity of the population with respect to location characteristics, including wages.

Defining the overall commuting cost between locations i and j as:

$$\tau_{ij} \equiv \left[\sum_m \tau_{ijm}^{-\theta} \right]^{-1/\theta} \tag{A.6}$$

yields similar expressions for commuter flows and commuting probabilities between location i and j:

$$L_{ij} = \bar{L} \left(\frac{\gamma}{\bar{V}} \right)^{\theta} \left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ij}} \right)^{\theta} \tag{A.7}$$

$$\lambda_{ij} \equiv \sum_m \lambda_{ijm} = \frac{\left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ij}} \right)^{\theta}}{\sum_{i,j} \left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ij}} \right)^{\theta}}. \tag{A.8}$$

Summing equation (A.7) across locations yields the supply of residents and the supply of workers to each location:

$$L_{Ri} = \bar{L} \left(\frac{\gamma}{\bar{V}} \right)^{\theta} \left(U_i r_{Ri}^{\beta-1} \right)^{\theta} \Phi_{Ri} \tag{A.9}$$

$$L_{Fj} = \bar{L} \left(\frac{\gamma}{\bar{V}} \right)^{\theta} w_j^{\theta} \Phi_{Fj} \tag{A.10}$$

where we refer to $\Phi_{Ri} \equiv \sum_j (w_j/\tau_{ij})^{\theta}$ as residential commuter market access (RCA) and $\Phi_{Fj} \equiv \sum_i (U_i r_{Ri}^{\beta-1}/\tau_{ij})^{\theta}$ as firm commuter market access (FCA), following Tsivanidis (2023). RCA summarizes a location i's geographical access to firms, while FCA summarizes a location j's access to residents. In the gravity literature, these terms are the multilateral resistance terms such that conditional on transportation costs d_ij between two locations i and j, commuter flows between those locations increases if they are more remote, as measured by RCA and FCA. Thus, the supply of residents to a location is increasing in the location's amenities and the commuting access to firms and is decreasing in the price of floor space. Meanwhile, the supply of workers to a location is increasing in the wages of the firms in the location and the commuting access to residents.

Commuting market clearing implies that the number of workers in location j is equal to the number of residents commuting to location j from across the city:

$$L_{Fj} = \sum_i L_{ij} = \sum_i L_{Ri} \lambda_{ij|i} \tag{A.11}$$

where $\lambda_{ij|i}$ is the probability that a worker residing in location i works in location j (i.e. the

probability of working in location j, conditional on residing in location i):

$$\lambda_{ij|i} \equiv \frac{L_{ij}}{L_{Ri}} = \frac{(w_j/\tau_{ij})^\theta}{\sum_{j'}(w_{j'}/\tau_{ij'})^\theta} = \frac{(w_j/\tau_{ij})^\theta}{\Phi_{Ri}} \tag{A.12}$$

Labor market clearing implies that the population of workers across locations must sum to the total population, and likewise for the population of residents: $\sum_i L_{Ri} = \bar{L}$ and $\sum_j L_{Fj} = \bar{L}$.

Residents spend a $1-\beta$ share of their income on housing, and the demand for housing is therefore given by:

$$H_{Ri} = L_{Ri}\bar{y}_i \frac{1-\beta}{r_{Ri}} \tag{A.13}$$

where $\bar{y}_i \equiv E[w|i]$ is the average income of residents who reside in location i, and is equal to the average of wages across workplaces, weighted by the share of the residents in location i that work in each workplace:

$$\bar{y}_i = \sum_j \lambda_{ij|i} w_j. \tag{A.14}$$

Profit maximization and normalizing the price of the good ($p = 1$) implies the unit cost function:

$$1 = \frac{1}{A_j}\left(\frac{w_j}{\alpha}\right)^\alpha \left(\frac{r_{Fj}}{1-\alpha}\right)^{1-\alpha} \tag{A.15}$$

where r_{Fj} is the price of commercial floor space. Rearranging implies the following expression for the price of floor space:

$$r_{Fj} = (1-\alpha)\left(\frac{\alpha}{w_j}\right)^{\alpha/(1-\alpha)} A_j^{1/(1-\alpha)}, \tag{A.16}$$

and the firm's first-order conditions implies the following expressions for labor demand and demand for floorspace:

$$L_{Fj} = \left(\frac{\alpha A_j}{w_j}\right)^{1/(1-\alpha)} H_{Fj} \tag{A.17}$$

$$H_{Fj} = \left(\frac{(1-\alpha)A_j}{r_{Fj}}\right)^{1/\alpha} L_{Fj} \tag{A.18}$$

The above equations are satisfied whenever there is production in a region(corresponding to when $L_{Fj} > 0$ in our data).

Floor space can be used commercially or residentially, and the endogenous fraction of floor space allocated to commercial and residential use are denoted υ_i and $1-\upsilon_i$ respectively.

The no-arbitrage condition across alternative land uses implies that floor space is used entirely commercially, used entirely residentially, or used in a mix of both if the commercial price of floor space equals the residential price of floor space. The share of floor space used commercially is given

by:

$$\begin{aligned} &\upsilon_i = 1 \text{ if } r_{Fi} > r_{Ri} \\ &\upsilon_i \in [0,1] \text{ if } r_{Fi} = r_{Ri} \\ &\upsilon_i = 0 \text{ if } r_{Fi} < r_{Ri} \end{aligned} \tag{A.19}$$

The observed price of floor space, R_i is assumed to be the maximum of the commercial and residential price of floor space $R_i = \max(r_{Fi}, r_{Ri})$, such that the observed price of floor space is given by:

$$\begin{aligned} &R_i = r_{Fi}, \quad \upsilon_i = 1 \text{ if } r_{Fi} > r_{Ri} \\ &R_i = r_{Ri}, \quad \upsilon_i \in [0,1] \text{ if } r_{Fi} = r_{Ri} \\ &R_i = r_{Fi}, \quad \upsilon_i = 0 \text{ if } r_{Fi} < r_{Ri} \end{aligned} \tag{A.20}$$

Floorspace is produced by a competitive construction sector with a Cobb-Douglas production function that uses immobile land, D, and the final good K as inputs: $H_i = K_i^{\mu} D_i^{1-\mu}$, implying the following supply function of floorspace:

$$H_i = D_i \mu^{\frac{\mu}{1-\mu}} r_i^{\frac{\mu}{1-\mu}} \tag{A.21}$$

$$R_i = \chi R_{Li}^{1-\mu}, \tag{A.22}$$

where R_{Li} is the price of land, r_k is the price of the final good, which is normalized to one, r_i is the price of floorspace, and χ is a constant, and $\varphi_i = K_i^{\mu}$ is referred to as the density of development because it determines the relationship between floor space and land area. Residential land market clearing implies that the supply of floor space allocated for residential use $(1 - \upsilon_i)H_i$ is equal to the demand for floor space (H_{Ri}):

$$(1 - \upsilon_i)H_i = \frac{(1-\beta)\bar{y}_i L_{Ri}}{r_{Ri}}, \tag{A.23}$$

where $\bar{y} = E[w|i]$ is as in equation (A.14). Commercial land market clearing implies that the supply of floor space allocated for commercial use $\upsilon_i H_i$ is equal to the demand for commercial floor space. The Cobb-douglas production function then implies:

$$\upsilon_j H_j = L_{Fj} \left(\frac{A_i(1-\alpha)}{r_{Fi}} \right)^{1/\alpha} \tag{A.24}$$

A.2 Equilibrium

Defining the vectors: $\mathbf{D} \equiv [D_i]$, and $\mathbf{t} \equiv [t_{ijm}]$, $\kappa^{\mathbf{O}} \equiv [\kappa_{im}^{O}]$, $\kappa^{\mathbf{D}} \equiv [\kappa_{jm}^{D}]$, $\kappa \equiv [\kappa_m]$, $\mathbf{A} \equiv [A_i]$, and $\mathbf{U} \equiv [U_i]$, we then have:

Given the model's parameters, $\{\alpha, \beta, \mu, \theta, \kappa^{\mathbf{O}}, \kappa^{\mathbf{D}}, \kappa, \mu_U, \mu_A\}$, exogenous location characteristics

$\{\mathbf{A}, \mathbf{U}, \mathbf{D}\}$, travel times $\{\mathbf{t}\}$ and total population $\bar{L}$, a closed-economy equilibrium satisfies equations (A.4) (average utility), (A.9) (supply of residents), (A.10) (labor supply), (A.23) (residential floor-space demand), (A.24) (commercial floor-space demand), (A.21) (floor-space supply), (A.17) (labor demand), (A.14) (worker income), (A.19) (floor-space no-arbitrage condition), (13) (amenity spillovers), (12) (production spillovers), and associated expressions, (A.6) (commuting-cost aggregation), conditional probability, $\lambda_{ij|i}$, and RCA and FCA. These equations are reproduced below.

$$\bar{V} = \gamma(1+k)\left[\sum_j \sum_i \sum_m \left(\frac{U_i w_j r_{Ri}^{\beta-1}}{\tau_{ijm}}\right)^{\theta}\right]^{1/\theta}$$

$$L_{Ri} = \bar{L}\left(\frac{\gamma}{\bar{V}}\right)^{\theta}\left(U_i r_{Ri}^{\beta-1}\right)^{\theta}\Phi_{Ri}$$

$$L_{Fj} = \bar{L}\left(\frac{\gamma}{\bar{V}}\right)^{\theta} w_j^{\theta}\Phi_{Fj}$$

$$(1-\upsilon_i)H_i = \frac{(1-\beta)\bar{y}_i L_{Ri}}{r_{Ri}}$$

$$\upsilon_j H_j = L_{Fj}\left(\frac{A_i(1-\alpha)}{r_{Fi}}\right)^{1/\alpha}$$

$$H_i = D_i \mu^{\frac{\mu}{1-\mu}} r_i^{\frac{\mu}{1-\mu}}$$

$$\upsilon_i = 1 \text{ if } r_{Fi} > r_{Ri}$$

$$\upsilon_i \in [0,1] \text{ if } r_{Fi} = r_{Ri}$$

$$\upsilon_i = 0 \text{ if } r_{Fi} < r_{Ri}$$

$$L_{Fj} = \left(\frac{\alpha A_j}{w_j}\right)^{1/(1-\alpha)} H_{Fj}$$

$$\bar{y}_i = \sum_j \lambda_{ij|i} w_j$$

$$\lambda_{ij|i} = \frac{(w_j/\tau_{ij})^{\theta}}{\sum_{j'}(w_{j'}/\tau_{ij'})^{\theta}}$$

$$U_i = \bar{U}_i L_{Ri}^{\mu_U}$$

$$A_j = \bar{A}_j L_{Fj}^{\mu_A}$$

$$\Phi_{Ri} \equiv \sum_j (w_j/\tau_{ij})^{\theta}$$

$$\Phi_{Fj} \equiv \sum_i (U_i r_{Ri}^{\beta-1}/\tau_{ij})^{\theta}$$

$$\tau_{ij} = \left[\sum_m \exp(-\theta(\kappa_{im}^O + \kappa_{jm}^D + \kappa_m t_{ijm}))\right]^{-1/\theta}$$

For the case of exogenous productivities and amenities (no spillovers), Ahlfeldt et al. (2015) provide proofs guaranteeing existence of a unique general equilibrium of this model. They also show that there is a unique mapping from the observed variables to unobserved location characteristics including production and residential fundamentals even when we allow for spillovers, as is described in the following proposition, which we state without proof.

Proposition 2. Define $\mathbf{R} \equiv [R_{Li}]$, $\mathbf{L_R} \equiv [L_{Ri}]$, $\mathbf{L_F} \equiv [L_{Fj}]$, Given known values for the parameters $\{\alpha, \beta, \mu, \theta, \kappa^{\mathbf{O}}, \kappa^{\mathbf{D}}, \kappa, \mu_U, \mu_A\}$ and the observed data $\{\mathbf{R}, \mathbf{L_R}, \mathbf{L_F}, \mathbf{t}\}$, then there exists unique vectors of unobserved location characteristics $\{\mathbf{A}, \mathbf{U}, \mathbf{D}\}$ that are consistent with the data being an equilibrium of the model.

This proposition guarantees that with our data, we are able to recover location fundamentals, which we can then use to compute counterfactuals and perform welfare analysis. Interested readers are referred to the supplementary appendix of Ahlfeldt et al. (2015) where the authors prove the proposition in steps using Lemmas S.6-S.11. The differences between our proposition and theirs is as follows. First, the present paper defines multiple modes (thus, commute costs are $\tau_{ij} = \left[\sum_m \exp(-\theta(\kappa^O_{im} + \kappa^D_{jm} + \kappa_m t_{ijm}))\right]^{-1/\theta}$) whereas Ahlfeldt et al. (2015) use a single mode (with commute costs: $\tau_{ij} = \exp(\kappa t_{ij})$). Second, the present paper uses a different choice of normalization.[27] Last, the present paper recovers quality-adjusted land rather than density of development φ_i as in Ahlfeldt et al. (2015).

B Model Extensions

B.1 Model with Road Congestion

We model road congestion using the routing-choice framework from Allen and Arkolakis (2022). In addition to the choice of residence, i, workplace j, and transportation mode m, workers also choose a route r. Indirect utility is given by:

$$V_{ijmr}(\omega) = \frac{U_i w_j r_{Ri}^{\beta-1} \epsilon_{ijmr}(\omega)}{\tau_{ijmr}} \tag{B.1}$$

where we have replaced commuting costs τ_{ijm} with τ_{ijmr} and idiosyncratic preference shock ϵ_{ijm} with ϵ_{ijmr}. Agents choose their workplace, residence, and transportation mode first, and then choose their route in a nested structure. In other words, preference shocks are drawn from a nested Fréchet distribution with shape parameter θ in the upper nest and parameter ρ in the lower nest. The joint cumulative distribution function is from the generalized extreme-value family and is given

[27] Specifically, Ahlfeldt et al. (2015) define a set of adjusted variables in Section S.3 that the authors prove are mapped uniquely from the data. However, those adjusted variables are precisely our variables $\{\mathbf{A}, \mathbf{U}, \varphi\}$ given the assumptions we have made in our model. More specifically, if we use the notation from Ahlfeldt et al. (2015), our setting sets $E_i = T_i = \xi_i = \zeta_{Ri} = 1$, and then the adjusted variables are equal to the unadjusted variables (which we use), and therefore their results translate directly to our setting.

by:

$$F(\{\epsilon_{ijmr}\}) = \exp\left(-\sum_{i,j,m}\left(\sum_{r\in\mathscr{R}_{ijm}}\epsilon_{ijmr}^{-\rho}\right)^{\frac{-\theta}{\rho}}\right) \quad \text{where } \theta > 1,\ \theta < \rho,\ \epsilon_{ijmr} > 0. \tag{B.2}$$

where $\mathscr{R}_{ijm}$ is the set of routes between residence i and workplace j using mode m.

Expected indirect uility from choosing residence i, workplace j, mode m is then proportional to:

$$V_{ijm}(\omega) \equiv E[V_{ijmr}|i,j,m] \propto \frac{U_i w_j r_{Ri}^{\beta-1}\epsilon_{ijm}(\omega)}{\tau_{Cijm}} \tag{B.3}$$

where $E[V_{ijmr}|i,j,m]$ is the expected indirect utility from choosing residence i, workplace j, mode m where the expectation is taken across routes $r \in \mathscr{R}_{ijm}$. τ_{Cijm} is the commuting cost for mode m between residence i and workplace j, analagous to the commuting cost, τ_{ijm}, used in the previous section. We make the distinction between the commuting costs in this section and the previous section using the subscript C; note that C is not an index, whereas i,j,m are indices. ϵ_{ijm} is an i.i.d. Fréchet shock with shape parameter θ and unit scale, as in the previous section.

We focus our attention on automobile routing, as other modes (bicycle, walking, bus) have negligible congestion or usage relative to cars. Thus, we assume that there is only one route for modes other than cars and therefore commuting costs for these modes are parameterized as: $\tau_{ijmr} = \tau_{Cijm} = \exp(\kappa_{0m} + \kappa_{1m} t_{ijm})$ for $m \in \{\text{train, walk, cycle, bus}\}$, where κ_{1m} mediates the value of time for the mode and κ_{0m} is a mode-specific preference shifter. For cars, commuting costs between locations i and j are implied by the nested Fréchet structure:

$$\tau_{Cijm} = \left[\sum_{r\in\mathscr{R}_{ijm}}\tau_{ijmr}^{-\rho}\right]^{\frac{-1}{\rho}}. \tag{B.4}$$

Each location (in our case, a census tract) is a node on the network and is connected to other nodes either directly (if there is a direct link between the two nodes) or indirectly (via a series of other nodes). Associated with each link is a travel time, and a route is defined by a sequence of links such that the total travel time is the sum of the travel times of each link in the route.

We denote the commuting costs for cars between two nodes on a direct link as d_{kl}^L where k is the origin node and l is the destination node.[28] The commuting costs for cars between two nodes taking a route r is then given by:

$$\tau_r = \prod_{(k,l)\in r} d_{kl}^L \tag{B.5}$$

where the product is taken over all links in the route r. Analogous to the previous section, we

[28] We suppress the i,j,m subscript for notational clarity. m = car and the link-level cost is independent of the the trip origin i or trip destination j.

parametrize link-level costs as $d_{kl}^L = \exp(\kappa_{1,\text{car}} t_{kl})$, such that route-level costs are given by:

$$\tau_r = \exp\left(\kappa_{1,\text{car}} \sum_{(k,l)\in r} t_{kl}\right). \tag{B.6}$$

It proves convenient to describe the network using the matrix:

$$\mathbf{A} \equiv \left[\left(d_{kl}^L\right)^{-\rho}\right] \tag{B.7}$$

where

$$d_{kl}^L = \exp(\kappa_{1,\text{car}} t_{kl}) \tag{B.8}$$

whenever the nodes k and l are directly connected, and $d_{kl}^L = \infty$ otherwise. "Self-loops," links that start and end in the same node, are forbidden: $d_{kk}^L = \infty$, such that a commuter must leave the node once and then return in order to commute to the same workplace as their residence. Allen and Arkolakis (2023) show that the Fréchet distribution implies that commuting costs for cars (equation (B.4)) can be expressed in terms of the Leontief inverse of the matrix $\mathbf{A}$. Denoting $\mathbf{B} \equiv \left[\tau_{Cijm}^{-\rho}\right]$, they show:

$$\mathbf{B} = (\mathbf{I} - \mathbf{A})^{-1} \tag{B.9}$$

where $\mathbf{I}$ is the identity matrix. The commuting costs for cars between two nodes i and j is then given by $\tau_{Cijm} = B_{ij}^{-\frac{1}{\rho}}$ where B_{ij} are the (i,j)-th element of matrix $\mathbf{B}$, where once again, we have suppressed the index m as it is assumed that we are only considering car commutes.

In addition, the authors show that *link intensity*, the expected number of times that a link k,l is used for commutes between an origin i and destination j, can be expressed as

$$\pi_{ij}^{kl} = \left(\frac{\tau_{Cijm}}{\tau_{Cikm} d_{kl}^L \tau_{Cljm}}\right)^{\rho} \tag{B.10}$$

Traffic flows through link k,l are then the sum across all pairs of nodes i,j of the link intensity, weighted by the number of commuters between i and j:

$$Q_{kl} = \sum_{i,j} \pi_{ij}^{kl} L_{ij\text{car}}. \tag{B.11}$$

Travel times between link kl are parameterized to be a sum of an exogenous component of travel time, $\bar{t}_{kl}$, and an endogenous component that is a function of traffic flows, $t' = G_{kl} Q_{kl}^{\sigma}$, such that the total time passing through the link is:

$$t_{kl} = \bar{t}_{kl} + G_{kl} Q_{kl}^{\sigma}. \tag{B.12}$$

The exogenous component, $\bar{t}_{kl}$, represents the travel time on the link if there was no traffic congestion, the parameter σ is the elasticity of the endogenous component of travel times with respect to

traffic flows, and G_{kl} is a link-specific constant that captures the width of roads and other link-level attributes that mediate the impact of traffic on travel times. Allen and Arkolakis (2023) provide sufficient conditions for the existence of equilibria and sufficient conditions for a unique equilibrium for a similar model with fixed commercial and residential land.

B.2 Road Congestion Equilibrium

Relative to the baseline model, the road congestion model introduces two new parameters, σ and ρ, as well as primitives $\{\bar{t}_{kl}, G_{kl}\}$ that are the exogenous travel times and travel-time shifter for each link kl. The model also introduces a new set of endogenous variables, $\{Q_{kl}\}$, which are the traffic flows through each link k, l in the network.

Given the model's parameters, $\{\alpha, \beta, \mu, \theta, \kappa_{\mathbf{0}}, \kappa_{\mathbf{1}}, \mu_U, \mu_A, \sigma, \rho\}$, exogenous location characteristics $\{\mathbf{A}, \mathbf{U}, \mathbf{D}\}$, travel times $\{\mathbf{t}\}$ for non-car trips, and link-level characteristics $\{\bar{\mathbf{t}}, \mathbf{G}\}$ and total population $\bar{L}$, a closed-economy equilibrium satisfies equations (A.4) (average utility), (A.9) (supply of residents), (A.10) (labor supply), (A.23) (residential floor-space demand), (A.24) (commercial floor-space demand), (A.21) (floor-space supply), (A.17) (labor demand), (A.14) (worker income), (A.19) (floor-space no-arbitrage condition), (13) (amenity spillovers), (12) (production spillovers), and associated expressions, (A.6) (commuting-cost aggregation), (B.7) (network matrix), B.8 (link cost), B.9 (Leontief inverse), B.10 (link intensity), B.11 (traffic flows), B.12 (link-level travel time), and equations for conditional probability, $\lambda_{ij|i}$, and RCA and FCMA.

B.3 First-Stage Migration Choice

In the previous model, the equilibrium assumed that the population of Tokyo is fixed (i.e. that the city is closed). An alternative assumption is to allow agents to make a decision to move to Tokyo or the rest of Japan prior to making their other decisions.[29] Living outside of Tokyo awards them with utility $U_{rest}\epsilon_{rest}$ while living in Tokyo awards them with utility $V\epsilon_{Tokyo}$ where V is the average utility as discussed in the main text, and ϵ_{rest} and ϵ_{Tokyo} are i.i.d. Fréchet preference shocks with shape parameter η. Averaging across Frechet shocks implies average utility of the worker in Japan as:

$$U_{Japan} = \Gamma \left[U_{rest}^{\eta} + V^{\eta}\right]^{1/\eta} \tag{B.13}$$

and the population of Tokyo is given by:

$$\bar{L} = \frac{\bar{V}^{\eta}}{\bar{V}^{\eta} + U_{Rest}^{\eta}} \bar{L}_{\text{total}} \tag{B.14}$$

where Γ is a constant implied by the Fréchet distribution. The population of Japan is fixed at $\bar{L}_{JP}$. In this setup, the equilibrium equations change such that the exogenous population of Tokyo, $\bar{L}$, is determined endogenous by equation (B.14).

[29] We restrict workers to living in Japan.

C Details on Data and Assumptions

In this section, we describe additional details about the data and what assumptions are made to transform the data from the raw data to the processed data that is used in the analysis.

C.1 Travel Time Calculation

Travel times reported in the survey are those reported by commuters as their commuting times between their home and workplace. We believe these times are a more accurate measure of workers' commute times than those that we compute using GIS tools (discussed next) and subject to less systematic bias. Our general strategy for determining travel times in estimation and counterfactuals is therefore to use the survey travel times wherever possible to calibrate our (GIS) tools such that our calculated such that our calculated times are consistent with what is reported in the travel survey.

In estimating our gravity equation using the commuter flow data, we require GIS tools because the commuter-survey data contains travel times for pairs of locations that had positive commuter flows. However, our Poisson Pseudo-Maximum Likelihood (PPML) model requires travel times for all pairs of locations including those pairs with zero trips. We must therefore fill these missing travel times using GIS tools. In calibrating the rest of our model, we also need to calculate travel times at the census-tract level, since the commuter-flow survey unit of analysis is several times larger.

To calculate these travel times using GIS tools, we take the following steps. First, we calculate travel times for cars, cycling, and walking between the centroids of origin-destination pairs using OpenRoute Services, an open-source tool that utilizes the street network provided by OpenStreetMap to calculate the shortest routes between locations and their associated travel times.

Next, we utilize data on the train network provided by MLIT. The data includes GIS shapefiles for the train lines and train stations, in addition to the beginning and ending dates of operation of segments of train lines from the 1800s to the present. We exclude high-speed rail lines (shinkansen) from our analysis, as they account for a negligible share of commuting.[30]

We use the GIS shapefiles to construct a network on NetworkX, a network analysis library in Python and construct minimum travel times between the centroids of origin-destination pairs for trains. The average train speed (inclusive of station stops) for 24 of the major train lines were found online while the remaining train lines were assumed to travel at 40 km/hr, which was the approximate median travel speed of the train lines that we found. Passengers can transfer between train lines, but there is a 4-minute penalty for each transfer in addition to the time it takes to walk, which is assumed to occur at 4 km/hr. We calculate the travel times from the centroid to the train station using ORS by assuming that households walk to the station.

We then adjust the travel times from ORS and our train network to match the travel times

[30] See Bernard et al. (2019) for a study on the impact of high-speed rail on firms' buyer-seller linkages and Kazunobu et al. (2021) for a study on the impact of high-speed rail on the distribution of economic activity across Japan.

reported in the survey. To fill the missing travel times in the gravity estimation, we average across census-tract-level travel times to the commuter-flow-district level and regress the survey-reported times on the calculated travel times. We then extrapolate for the travel times that are missing.

For the remaining model, we multiply each of the travel times by a constant factor specific to each commuting-zone pair such that the average travel times computed across origin and destination census-tract pairs between commuter-zone pairs are equated to the average travel times reported in the survey.[31] Whenever there are no reported travel times between two zones (because there are no commuters between the zones), we once again extrapolate using the predicted values from the regression of reported travel times (from the commuter-flow data) on computed travel times.

Additional details on these calculations can be found below.[32]

C.2 Computing Travel Times for Cars, Walking, Cycling, and Buses

We use Openroute Service to compute travel times between the centroids of every pair of regions in Tokyo (for determining the automobile travel time) and also between each centroid and nearby train stations (for the last-mile mode of transport). Openroute Service is an opensource tool that utilizes the street network provided by OpenStreetMap to calculate routes between locations.[33] The tool typically assumes that vehicles travel at the speed limit of a road segment.[34] It is thus generally more accurate for travel times in low-traffic settings when the speed limit is a good proxy for the actual speed of travel. Whenever the centroid of a location does not land on a road segment, the tool will use the closest road segment within a 350-meter radius of the centroid.[35]

C.3 Computing Train Travel Times

We use longitudinal data on the train network provided by the MLIT to compute travel times. The data includes geographical information on each train line and station of each train line. We use this geographical data to construct a graph on Python's NetworkX library. Each station of a train line is a node and each segment of the train line between two neighboring stations forms an edge. We assume that passengers can transfer between any two stations within 1 km of each other by walking. If they do so, they must wait 4 minutes on the next train platform in addition to the time it takes to walk, assuming a walking speed of 4 km/hr using a straight-line distance.

If a station is a hub for multiple train lines, each train line receives its own node, and these nodes are connected by edges that represent the time it takes to transfer between train lines, following the same rules as stated above.

[31] This approach follows Miyauchi et al. (2021).

[32] Services that calculate travel times between bilateral pairs of locations, such as Google Routes API, may be useful for calculating the travel times between locations in the current equilibrium. However, the Google Maps API for public transit are not readily available for public-transport routes in Tokyo and therefore is not used.

[33] OpenStreetmap is a free and open geographic database of maps (including street maps) from across the world

[34] The documentation of Openroute Service writes "The travel time is calculated for each segment by using speed-limits for different waytypes and adjusting them for different surfaces or surface qualities of the road."

[35] There were two tracts however where the centroids were not within 350 meters of a road segment. For these tracts, we manually selected a point on a nearby road segment instead and used this point instead.

Each edge in the graph is weighted by the time it takes to travel between the two nodes. This time is calculated by dividing the distance (provided by the GIS map) between the two nodes by the effective train speed, which were looked up for 24 of the major train lines and were assumed to be 40 km/hr for other train lines.[36]

Households are located at the centroid of each district and are able to travel to their nearest stations by either walking or taking the bus. If they walk, they once again travel at 4 km/hr using the straight-line distance, whereas if they travel by bus, they will travel through the street network at half the speed of a car in addition to paying a fixed time cost of 6 minutes, which represents the additional time from walking to the bus stop and waiting for the bus.

For commutes that are short (e.g. commuting within the same census-tract or to a neighboring tract), commutes are typically not done by train (e.g. it would be much faster to walk or cycle). We must therefore reflect these relatively larger commute costs in our travel time calculations for trains. To do this, we assume that a household that uses a train mode must at least go to a train station on their way to reaching their destination (i.e. they cannot directly walk to another census tract if they use the "train" mode), although they do not have to actually get on a train. If census tracts are adjacent, this assumption may mean that the household will walk to the nearest train station and walk from this train station to their destination. With this assumption, commuting costs on trains for short trips are also impacted whenever the train network changes.

For commutes that are within one's own census tract, we assume that the travel time is half the time to its closest census tract. This assumption means that when the train network changes, the commute times to one's own census tract also changes.

C.4 Incorporating Reported Travel Times from Survey

As discussed in the main text, we believe times reported in the travel survey (*reported time* henceforth) are more accurate than the computed times (from Openroute Service or NetworkX) because the reported times do not rely on ad-hoc assumptions about travel speeds used in the computed time, and they should also take into account congestion, last-mile travel times, and other nuances of travel that may not be captured by the computed time. The two drawbacks of the reported time are that travel times are reported only for pairs of locations where trips have been reported and the reported time is only reported at the geographic unit of the commuter-flow data (about 15 thousands residents, referred to as the "commuting district" in the main body and "small-zone" in data manual) as opposed to the census-tract level (about 2 thousand residents) which is the level of the model analysis. [37]

In order to retain the accuracy of reported travel time while having a full matrix of travel times between census tracts for each mode, we take the following steps to incorporate both the reported and computed times. First, we calculate the average computed travel time between small zones

[36] 40 km/hr was chosen because this was the typical speed that we observed for the major train lines

[37] Another reason that the computed time is typically needed is that it is used to compute travel times for counterfactual transportation networks, though in the present analysis, we remove the entire train network, and so an alternative train network did not need to be computed.

by taking the simple average of computed travel times across pairs of census tracts between the any two small zones. When the data is available (i.e. there are reported trips at the small-zone level), we calculate the ratio between the computed travel time and the reported travel time. We then use this factor to scale the computed time at the small-zone level. Since the reported time is missing for many pairs of small zones, we then aggregate the computed and reported times to three levels of geographical aggregation. At the highest level of aggregation, each region contains on average 250 thousand residents (less than one percent of the population). We then compute the ratio between the computed travel time and the reported travel time for each level of aggregation, drop the top and bottom one percent of ratios across small-zone pairs for each mode, and use this factor to scale the computed time at the census-tract level, using the lowest level of aggregation that is available(e.g. the ratio at the small-zone level is used whenever it is available because it is smallest level of aggregation for the survey data).[38] From an initial dataset with 1.6 percent of pairs filled in (these are the diade-mode combinations that have at least one trip), the above procedure fills in an additional 42 percent of location-pair and mode combinations. The remaining observations are then filled in by running a regression of reported time on computed time for each mode and using the predictions from this regression model. These regressions and corresponding scatter plots are shown in Table C.1 and Figure C.1. For the predictive regressions, we use a simple linear model with a constant, because aside from an intercept (reflecting a fixed cost of travel), our prior was that reported and computed times should scale linearly. Scatter plots from Figure C.1 support this hypothesis.

C.5 Additional Information on the Official Land Price Announcements

The primary source of land price data is the Official Land Price Announcements, which is published annually by the Japanese Ministry of Land, Infrastructure, Transport and Tourism (MLIT). Land prices across Japan are collected by real estate appraisers hired by the national and local government to provide policymakers with reliable land values that are important for calibrating national and local property tax levies. The dataset provides a panel of 11288 land plots across the Greater Tokyo Area in 2020.

The national and prefectural governments select land plots for appraisal that are representative of a land-use category (residential, commercial, or industrial) at the municipality level (which has employment of approximately 100 thousand on average), and are typically tracked over time, unless the government deems the plot to no longer be representative of its category in a municipality.[39] Official municipality-level land-price indices then take a simple average of the land plots across a municipality, a methodology that we adopt to calculate the average municipality-level land price (used in our estimates of the model with congestion). For our census-tract-level analysis, we spatially interpolate the land prices to infer prices at each census tract.[40]

[38] Thus, these computed times are scaled such that the weighted-average travel time at the higher-level of aggregation matches that of the reported time, whenever the time has been reported.

[39] See LaPoint (2020) for additional details on the land appraisal method.

[40] Variation in log land prices across municipalities account for 84 percent of the total variance in log land prices

C.6 Merging Employment and Population Censuses

Both the employment and population census data appear at the block levels, but they do not utilize the same block identification codes. Because both datasets contain the address of the block, we merge the two datasets by geocoding the addresses using geocode software, and then merging the blocks using a spatial merge with GIS tools.[41] We geocode the addresses using the *CSV Address Matching Service*, a tool from Tokyo University's Center for Spatial Information Science.

The same challenge arises with the 1995 population data and 1996 employment data because the definitions of blocks and census tracts are not consistent across time. We therefore rely once again on the *CSV Address Matching Service* by looking up the coordinates from the addresses that are listed on the website[42]

Specifically, we use the following procedure. After looking up the coordinates of the 1995 population tracts, 1996 employment tracts, and 2021 employment tracts, we proceed by merging the 2020 population tracts with the 2021 employment coordinates, then merge the resulting districts with the 1996 employment coordinates, and then merge the resulting districts with the 1995 population coordinates. When we merge the 2020 population tracts (which are polygons in GIS jargon) with the 2021 employment coordinates (points in GIS jargon), we first perform a spatial inner join between the polygons and points. Thus, all coordinates that are contained in a 2020 population tract are joined. Then polygons that did not match with any points are matched with the closest employment point. If the nearest employment point has already been matched to another polygon, then the two polygons are merged (i.e. aggregated) together. If there are employment points that do not join with residential tracts, they are dropped. We then repeat this procedure to merge these new polygons with the 1996 employment coordinates, and then again with the 1995 population coordinates.

We continue to use the term "tract" for these new district definitions, because the difference the definitions are negligible.

C.7 Cleaning Data for Gravity Estimation

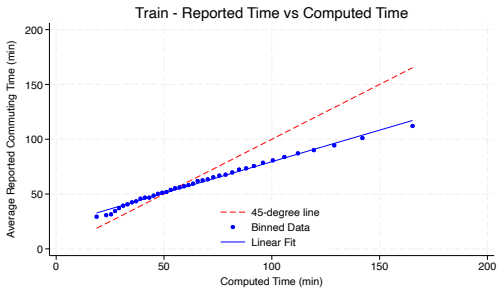
Prior to estimating gravity, we take the following steps to clean the data:

- We found 65 origin-destination observations in the bilateral flows data where the survey reported that commuters walked when the reported commuted times are unrealistically short. We believe these trips were misreported as walking trips when they were in fact trips taken

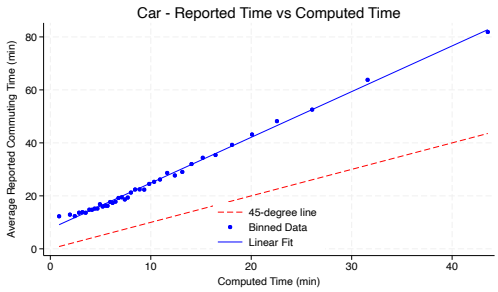
across plots. At the *oaza* level, the figure is 93 percent.

[41] Directly merging the datasets using the addresses is avoided because the addresses are not standardized across the two datasets.

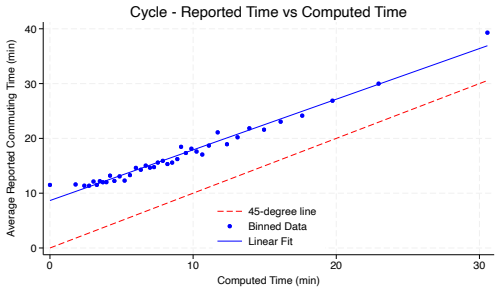
[42] These addresses are also not consistent over time. However, the *CSV Address Matching Service* is able to geolocate addresses that used earlier address names. For each address lookup the service provides an estimate of the accuracy: whether the coordinate is correct up to the prefecture, municipality ("shichooson" + "tokubetsuku"), or census tract ("ooaza"). We found that 0.2 percent of the 2021 employment addresses were matched up to the municipality level, 0.2 percent of the 1996 employment addresses were matched to this level, and 5.8 percent of the 1995 population addresses were matched to this level. All other addresses were matched to the census-tract level. Our census



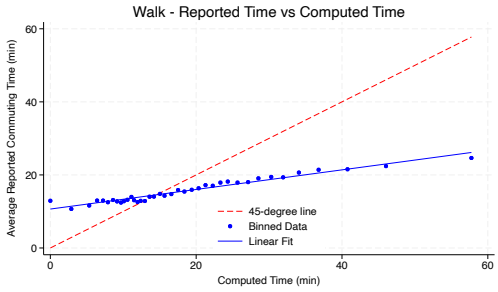
(a) $R^2 = 0.52$



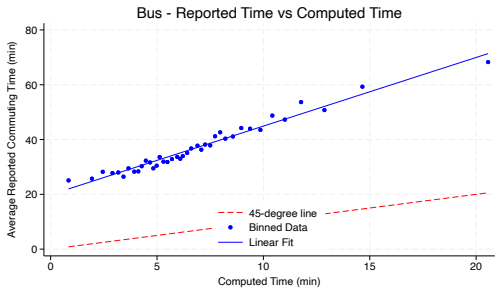
(b) $R^2 = 0.57$



(c) $R^2 = 0.3$



(d) $R^2 = 0.24$



(e) $R^2 = 0.25$

Figure C.1: Binscatter and OLS Fit of Reported Time on Computed Time

Binscatter plots and OLS fit from a regression of reported time on computed time for each mode. The R^2 of the regression is shown in the caption. Data on commuter trips and reported times are obtained by Tokyo's Person-Trip Survey. The unit of observation is the "small" zone, defined by the survey to contain on average 15 thousand residents. Computed travel times by car, bicycle, and walking are calculated using OpenRoute Services while travel times by train are calculated using the train network and OpenRoute Service. Computed bus travel time is the same as the car travel time. The regression is weighted by the number of trips. We restrict the regression and binscatter to the pairs of locations that are within the 99th percentile of the distribution of computed travel times by mode. See Data section and Appendix for details.

	Train	Weighted	Weighted (capped)	Car	Weighted	Weighted (capped)
Computed Time	0.547	0.568	0.575	1.587	1.634	1.719
	(0.002)	(0.002)	(0.002)	(0.013)	(0.015)	(0.011)
Intercept	25.444	22.408	21.991	12.085	8.453	7.666
	(0.167)	(0.150)	(0.153)	(0.187)	(0.160)	(0.132)
N	83040	83040	82058	35270	35270	34633
R-squared	0.50	0.56	0.53	0.55	0.60	0.57

	Bus	Weighted	Weighted (capped)	Cycle	Weighted	Weighted (capped)
Computed Time	1.394	1.412	2.445	0.691	0.627	0.922
	(0.080)	(0.106)	(0.073)	(0.043)	(0.053)	(0.015)
Intercept	29.157	27.128	20.321	13.097	11.257	8.691
	(0.632)	(0.739)	(0.498)	(0.602)	(0.467)	(0.142)
N	5996	5996	5921	12738	12738	12379
R-squared	0.22	0.22	0.26	0.22	0.25	0.29

	Walk	Weighted	Weighted (capped)
Computed Time	0.128	0.194	0.254
	(0.019)	(0.011)	(0.008)
Intercept	18.595	12.133	11.039
	(0.618)	(0.222)	(0.156)
N	8286	8286	7854
R-squared	0.08	0.20	0.24

Table C.1: Regressions of Reported Time on Computed Time

OLS regressions of reported time on computed time for each mode. The unit of observation is the "small" zone, defined by the survey to contain on average 15 thousand residents. Weighted regression use frequency weights based on the number of trips of that pair. "Weighted (capped)" restricts the sample to pairs of locations that are within the 99th percentile of the distribution of computed travel times by mode. Data on commuter trips and reported times are obtained by Tokyo's Person-Trip Survey. Computed travel times by car, bicycle, and walking are calculated using OpenRoute Services while travel times by train are calculated using the train network and OpenRoute Service. Computed bus travel time is the same as the car travel time. See Data section and Appendix for details. Robust standard errors in parentheses.

by train.[43] For these observations, we recoded the mode as train trips.

C.8 Costs of the Train System

The total land that utilizes the train system is valued at 14t yen (USD 140B using a PPP rate of 94 yen per dollar), or an average of 2.7b yen (USD 27m) per kilometer of track.[44]

Table C.2: Recent Urban Rail Investments in Greater Tokyo

Phase	Completion year	Length (km)	Tunnel share	Cost/km (2023 USD m)
Tama Monorail first phase	2000	16.0	0%	181.6
Monorail to Tokyo Station	2025	3.0	0%	429.7
Fukutoshin Line	2008	8.9	100%	335.3
Minatomirai Line	2004	4.1	100%	738.2
Oedo Line ring	2000	27.8	100%	416.3
Shin-Yokohama JR Line	2019	2.7	78%	515.1
Shin-Yokohama Tokyu Line	2022	10.0	80%	350.7
Tsukuba Express	2005	58.3	28%	215.7

The implied total land is 172 square kilometers. Dividing this number by the total track length (3900km) implies a crude measure of 44meters for the average width of the track, which seems plausible.

C.8.1 Subway and Radial Line Cost Calculations

In calculating the costs for the removal of subway and radial lines, we make the following assumptions about construction, land, and operating and maintenance costs. For construction costs, we use the two different per-kilometer cost estimates for recent rail projects that did or did not require significant tunneling. Projects that required tunneling cost approximately USD 470mm per km (2023 US dollars) while projects that required little to no tunneling cost 35% less at 306mm per km. The opportunity cost of land is calculated by first calculating the length of track that falls within each census tract. This length is multiplied by the average width of track that was calculated from satellite data (as discussed in the previous section)–44 meters–and then multiplied with the land price of that tract. Subways in Tokyo typically run underneath the main roads (which were typically built first). Hence, the cost of land is not as straightforward (since it is shared with roads) to calculate. For the current draft of the paper, we assume a price of land equal to the average land cost of the entire system, and multiply by the average width of track, as with the other radial lines. Operating and maintenance costs extrapolate using the cost per kilometer and cost per station that was discussed in section 3. As in Section 6, we express these costs as a percentage of Tokyo's GDP.

[43] It is expected that many respondents misreported train trips as walking trips because reporting a train trip takes significantly more time to report. A respondent reporting a train trip is required to report every transfer station and train line used along the route, which can be time consuming, relative to simply reporting that one walked.

[44] `https://www.ceicdata.com/en/japan/exchange-rate-forecast-oecd-member-quarterly/jp-purchasing-power-parity-national-currency-per-usd`

C.8.2 Note on Figure 2

We describe the sources and calculation methodology used in Figure 2. Data on historical track lengths and construction are from the MLIT. We use the date of construction for any segment of a train line as the date of construction. The reported track length may therefore differ from the actual track length present at any given point in time, because of train-line extensions that occur throughout history. In Panel (a), all train-line extensions are reported to have occurred on the date that the first segment of the train line was completed.

Population data for Panel (a) is from the prefecture-level population census data from 1920 onwards for the prefectures Ibaraki, Tochigi, Gunma, Saitama, Chiba, Tokyo, and Kanagawa, which differs slightly from the definition of Greater Tokyo used in the remaining paper. Hence, the population figures differ slightly.

Spatially-linked population data for 1930 for Panel (b) is obtained from Yuji Murayama's Historical Regional Statistical Database (http://giswin.geo.tsukuba.ac.jp/teacher/murayama/data.html). Other years are obtained from e-Stat.

D Details on Calibration and Estimation

D.1 Gravity Estimation

Given travel times and observed variables, our model predicts a perfect fit of the commuting flow data according to equation (A.3). We follow the gravity literature and assume that there is measurement error in the commuting shares, ε_{ijm}, yielding the regression model:

$$\ln \pi_{ijm} = \delta_m + \gamma_i + \vartheta_j - \theta\kappa t_{ijm} + \varepsilon_{ijm} \tag{D.1}$$

where we assume that the error ε_{ijm} is uncorrelated with travel times t_{ijm}.

The gravity model is estimated using trips reported on the Tokyo Person Trip Survey data. For the righthand-side variable (travel time), we use the methodology discussed in Section C.1, particularly Section C.4, where we fill in missing times using a prediction using the GIS-computed times. Observations whose travel times were calculated using the prediction algorithm are weighted proportionally to the R^2 of the prediction regression (while other observations have a weighting of 1).

In addition, our model predicts commuting flows at the census tract level, whereas the commuter-flows data is at the commuting-district level, whose population is about 7-8 times larger than that of a census tract on average. We therefore require additional assumptions to take our model to the data. We follow the strategy in Ahlfeldt et al. (2015) and first take means across pairs of blocks to obtain

$$\ln \pi_{IJm} = \gamma_{Im} + \vartheta_{Jm} - \theta\kappa t_{IJm} + \varepsilon_{IJm} \tag{D.2}$$

where I and J are commuting districts, t_{IJm} is the average travel time between the two commuting

districts across the census-tract pairs in those districts, and π_{IJm} is the approximated average of log commuting probabilities across all pairs of tracts in commuting districts I and J (i.e. the average of $\ln \pi_{ijm}$). We approximate the average of $\ln \pi_{ijm}$ using the observed log district commuting probabilities, and the average of the travel times across census tracts t_{IJm} with the average travel time across commuters, which is given in the survey data.[45]

Given estimates of $\{\theta\kappa_m\}_{m\in\mathcal{M}}$, $\left\{\theta\kappa^O_{im}, \theta\kappa^D_{im}\right\}_{i\in\mathcal{N},m\in\mathcal{M}}$, and travel times $\{t_{ijm}\}_{i,j\in\mathcal{N},m\in\mathcal{M}}$, we compute τ^θ_{ij} and use the labor market clearing condition (A.11) to obtain:

$$L_{Fj} = \sum_i L_{Ri}\lambda_{ij|i} = \sum_i L_{Ri}\frac{(w_j/\tau_{ij})^\theta}{\sum_{j'}(w_{j'}/\tau_{ij'})^\theta} \tag{D.3}$$

Defining transformed variables $\tilde{w}_i = w_i^\theta$ and $\tilde{\tau}_{ij} = \tau^\theta_{ij}$, we rewrite the above equation as:

$$L_{Fj} = \sum_i L_{Ri}\frac{\tilde{w}_j/\tilde{\tau}_{ij}}{\sum_{j'}\tilde{w}_{j'}/\tilde{d}_{ij'}} \tag{D.4}$$

Given estimates of $\{\theta\kappa_m\}_{m\in\mathcal{M}}$, $\left\{\theta\kappa^O_{im}, \theta\kappa^D_{im}\right\}_{i\in\mathcal{N},m\in\mathcal{M}}$, and travel times $\{t_{ijm}\}_{i,j\in\mathcal{N},m\in\mathcal{M}}$, data on travel times, employment, and population data are then used to recover transformed wages: w_i^θ for the N locations using these N equations (up to a choice of normalization). We then calibrate the heterogeneity parameter θ to match the heterogeneity in log earnings by residence location across the municipalities of Tokyo.

Because the Fréchet parameter applies a monotonic transformation to wages which adjusts the inequality in wages across locations, we use data on the dispersion of log earnings across districts, which are a weighted average of workplace wages, to calibrate θ. In particular, we use municipality-level data on average taxable earnings and set θ to equate the variance of log average earnings in the model (the population-weighted average of $\bar{y}_i \equiv E[w|i]$ from equation (A.14) within a municipality) with that of the taxable earnings data.[46] Once the Fréchet parameter is estimated, then we recover the semi-elasticity of commute costs with respect to travel time, κ_m, and mode-destination and mode-origin specific fixed costs, $\kappa^O_{im}, \kappa^D_m$.[47]

[45] The approximation for the average of $\ln \pi_{ijm}$ is based on taking the log of the averages (the commuter-flow data), rather than the log of the averages (the true commuting probability), and therefore the discrepancy depends on the extent to which Jensen's inequality leads to a discrepancy between the average of the logs and the log of the average. In practice, Ahlfeldt et al. (2015) find that the approximation is close in the sense that it leads to a negligible difference in parameter estimates.

[46] Ahlfeldt et al. (2015) instead match the variance of log workplace wages between the model and the Berlin data and obtain an estimate of between 6.5 to 6.8. We are not aware of reliable workplace wage data.

[47] Ahlfeldt et al. (2015) obtain a value of $\kappa = 0.01$. Tsivanidis (2023) obtain a similar value of 0.012 by estimating a discrete choice model of commuting mode and destination choice using a mobility survey of Bogota, where κ is identified from the sensitivity of individuals' mode choices to differences in times across modes within particular commutes.

D.2 Recovering Productivity, Amenities, and Quality-adjusted land

Given our estimate of θ, wages for each location can be recovered from the transformed wages that were obtained from solving .

We continue to use the gravity structure of the model to recover productivities, amenities, and quality-adjusted land. The unit cost function (equation (A.15)) implies:

$$\ln A_j = (1-\alpha)\ln r_{Fj} + \alpha \ln w_j - \alpha \ln \alpha - (1-\alpha)\ln(1-\alpha) \tag{D.5}$$

Given values for the price of floor space and wages, we recover productivities up to a normalization. For calibrating the model, we set the geometric mean of floor-prices to 1, and the geometric mean of wages to 1, which then implies values for productivities. To recover amenities, we use the supply of residents in each location (equation (A.9)) and divide by its geometric mean to obtain the following expression for amenities:

$$\ln \frac{U_i}{\bar{U}} = \frac{1}{\theta}\ln \frac{L_{Ri}}{\bar{L}_R} + (1-\beta)\ln \frac{r_{Ri}}{\bar{r}_R} - \frac{1}{\theta}\ln \frac{\Phi_{Ri}}{\bar{\Phi}_R} \tag{D.6}$$

where $\Phi_{Ri} = \sum_j (w_j/\tau_{ij})^\theta$ as defined previously and an overbar $(\bar{\cdot})$ refers to the geometric mean across census tracts. Note that the reservation utility has been eliminated from the above equation because it is common across locations. We then set the left hand side as our amenities respectively (i.e. $\ln U_i = \frac{1}{\theta}\ln \frac{L_{Ri}}{\bar{L}_R} + (1-\beta)\ln \frac{r_{Ri}}{\bar{r}_R} - \frac{1}{\theta}\ln \frac{\Phi_{Ri}}{\bar{\Phi}_R}$). By dividing the supply-of-residents equation by the geometric mean, we are setting a normalization for amenities such that the geometric mean of amenities is equal to 1.

Combining commercial and residential land market clearing conditions (equations (A.24) and (A.23)) yields the following expression for floorspace market clearing:

$$H_i = D_i \mu^{\frac{\mu}{1-\mu}} r_i^{\frac{\mu}{1-\mu}} = (1-\upsilon_i)H_i + \upsilon_i H_i = \frac{(1-\beta)\bar{y}_i L_{Ri}}{r_{Ri}} + L_{Fj}\left(\frac{A_i(1-\alpha)}{r_{Fi}}\right)^{1/\alpha} \tag{D.7}$$

where, as defined previously, $\bar{y}_i = \sum_j \lambda_{ij|i} w_j$, and we have that $r_i = r_{Ri} = r_{Fi}$ wherever both commercial and residential floor space are used in location i. Thus, D_i can be recovered given floorspace demand (right-hand-side), floor-prices r_i, and μ.

D.3 Calibrating the Road Congestion Model

We use the reported and computed car-travel times together with the previously computed commuting costs to calibrate the congestion model. We use GIS software to identify each location's adjacent locations and travel times from the survey. We believe the surveyed time is typically an overestimate of the link-level travel time, because there are likely fixed costs at the beginning and the end of a trip to enter and exit the network. Thus, we assume that the reported times should

be scaled by a fixed factor such that the link-level costs are given by:

$$d_{kl}^{L} = \exp(\kappa_{1,\text{car}}(t_{kl}^{rep} - \tilde{t})) \tag{D.8}$$

where t_{kl}^{rep} is the reported travel time between locations k and l, and $\tilde{t}$ is the fixed travel time to get on and off the road network. For hypothetical $\tilde{t}$, the commuting-cost matrix can be constructed using equation (B.9). We then choose $\tilde{t}$ to minimize the sum of squared differences between the conditional commuting-cost shares in the congestion model and the model without congestion:

$$\min_{\tilde{t}} \sum_{i,j} \left(\frac{d_{Cijcar}^{-\theta}(\tilde{t})}{\sum_j d_{Cijcar}^{-\theta}(\tilde{t})} - \frac{d_{ijcar}^{-\theta}}{\sum_j d_{ijcar}^{-\theta}} \right)^2 \tag{D.9}$$

where $d_{Cijcar}(\tilde{t})$ denotes the commuting cost for cars between locations i and j in the model with congestion as a function of the fixed time cost $\tilde{t}$, while d_{ijcar} denotes the commuting cost between locations i and j in the previous model that holds congestion fixed. Given an estimate of $\tilde{t}$ and implied link-level costs d_{kl}^{L}, link intensity is calculated using equation (B.10), and traffic flows using equation (B.11). To match the overall commuting share of cars ($\sim$3%), we then introduce a preference shifter for cars, ξ_m, such that the model-implied automobile commuting share matches the commuting share in the data:

$$\frac{\sum_{i,j} \left(\frac{U_i w_j r_{Ri}^{\beta-1} \xi_m}{d_{Cijcar}} \right)^{\theta}}{\sum_{i,j,m} \left(\frac{U_i w_j r_{Ri}^{\beta-1} \xi_m}{d_{Cijm}} \right)^{\theta}} = 0.23 \tag{D.10}$$

where ξ_m is equal to one for modes other than cars (as the commuting shares for these modes are already calibrated to the data).

We view the OpenRouteService travel time, t_{kl}^{OR}, as the time that excludes traffic congestion, because the service advertises their travel times in this way. However, this travel time also contains the fixed cost of getting on the network. We therefore scale the OpenRouteService travel times by the same time-factor that the reported travel time is scaled, so that

$$\bar{t}_{kl} = t_{kl}^{OR} \frac{t_{kl}}{t_{kl}^{rep}} \tag{D.11}$$

where $\bar{t}_{kl}$ is the exogenous component of travel time, as discussed in equation (B.12), and t_{kl} is the link-level travel time that was recovered from the minimum distance estimation. Given these variables we can rearrange equation (B.12) to recover the link-specific constant G_{kl}, and our model is calibrated.

D.4 Parameters

The parameters used in our counterfactual with congestion are shown in Table D.1, including both the estimated $\tilde{t}$, and ρ and σ. We set the elasticity of the endogenous component of travel times with respect to traffic flows, σ, to 0.35, which is about halfway between the two IV estimates from Allen and Arkolakis (2023) between 0.1 and 0.5. Our estimate of the fixed travel time to enter the street network is about 6 min. These were estimated using the municipality-level data. While we are not aware of comparable studies, we believe this figure is plausible given the time it takes to drive to the main road or high way from a residence.

Parameter	Value	Source
Elasticity of Travel Time w.r.t Traffic Flows (σ)	0.35	Allen and Arkolakis (2022)
Route Heterogeneity Parameter (ρ)	50	Chosen
Fixed Travel Time ($\tilde{t}$)	6min	Estimated

Table D.1: Additional Parameters Used in Counterfactual with Road Congestion

E Proofs

E.1 Proof of Proposition 1

We restate the objects from Section 7. Let Π^R denote the matrix of residence-based commuter shares, with element $\Pi^R_{ij} \equiv L_{ij}/L_{Ri}$, and let Π^F denote the matrix of workplace-based commuter shares, $\Pi^F_{ij} \equiv L_{ij}/L_{Fj}$. Define the direct-effect changes in residential and firm commuter access as

$$\Delta \mathbf{\Phi}_R^{DE} \equiv \begin{bmatrix} \Delta \ln \Phi_{R1}^{DE} \\ \vdots \\ \Delta \ln \Phi_{RN}^{DE} \end{bmatrix}, \qquad \Delta \mathbf{\Phi}_F^{DE} \equiv \begin{bmatrix} \Delta \ln \Phi_{F1}^{DE} \\ \vdots \\ \Delta \ln \Phi_{FN}^{DE} \end{bmatrix},$$

where $\Delta \ln \Phi_{Ri}^{DE} = \theta \sum_j \Pi^R_{ij} \Delta \ln \tau_{ij}$ and $\Delta \ln \Phi_{Fj}^{DE} = \theta \sum_i \Pi^F_{ij} \Delta \ln \tau_{ij}$ follow from equation (CA). Stack these vectors as $\Delta \mathbf{\Phi}^{DE} \equiv [(\Delta \mathbf{\Phi}_R^{DE})^\top \ (\Delta \mathbf{\Phi}_F^{DE})^\top]^\top$.

Let $\Delta \boldsymbol{\ell} \equiv [(\Delta \ln \tilde{\boldsymbol{L}}_R)^\top \ (\Delta \ln \tilde{\boldsymbol{L}}_F)^\top]^\top$ be the stacked vector of first-order changes in mobility-adjusted population and employment. Linearizing equations (9) and (10) around the baseline equilibrium yields

$$\Delta \ln \tilde{\boldsymbol{L}}_R = \frac{1}{1-\tilde{\phi}_U} \left(\Delta \mathbf{\Phi}_R^{DE} + \tilde{\phi}_A \Pi^R \Delta \ln \tilde{\boldsymbol{L}}_F \right), \tag{E.1}$$

$$\Delta \ln \tilde{\boldsymbol{L}}_F = \frac{1}{1-\tilde{\phi}_A} \left(\Delta \mathbf{\Phi}_F^{DE} + \tilde{\phi}_U (\Pi^F)^\top \Delta \ln \tilde{\boldsymbol{L}}_R \right), \tag{E.2}$$

where $\tilde{\phi}_U \equiv \theta \phi_U$ and $\tilde{\phi}_A \equiv \theta \phi_A$ govern the strength of amenity and productivity spillovers. Equa-

tions (E.1)–(E.2) can be written compactly as

$$\Delta\boldsymbol{\ell} = \tilde{\boldsymbol{\Gamma}}^{-1}\left(\Delta\boldsymbol{\Phi}^{DE} + \boldsymbol{\alpha}\Delta\boldsymbol{\ell}\right), \qquad \tilde{\boldsymbol{\Gamma}}^{-1} \equiv \begin{bmatrix} \frac{1}{1-\tilde{\phi}_U}\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \frac{1}{1-\tilde{\phi}_A}\mathbf{I}_N \end{bmatrix}, \quad \boldsymbol{\alpha} \equiv \begin{bmatrix} \mathbf{0} & \tilde{\phi}_A\Pi^R \\ \tilde{\phi}_U(\Pi^F)^\top & \mathbf{0} \end{bmatrix}. \tag{E.3}$$

Define $\boldsymbol{\alpha}_X \equiv \boldsymbol{\alpha}\,\tilde{\boldsymbol{\Gamma}}^{-1}$; explicitly,

$$\boldsymbol{\alpha}_X = \begin{bmatrix} \mathbf{0} & \frac{\tilde{\phi}_A}{1-\tilde{\phi}_A}\Pi^R \\ \frac{\tilde{\phi}_U}{1-\tilde{\phi}_U}(\Pi^F)^\top & \mathbf{0} \end{bmatrix}.$$

Because Π^R and Π^F are row-stochastic, $\|\Pi^R\|_\infty = \|\Pi^F\|_\infty = 1$ and therefore $\|\boldsymbol{\alpha}_X\|_\infty = \zeta^{DE} \equiv \max\left\{\left|\frac{\tilde{\phi}_A}{1-\tilde{\phi}_A}\right|, \left|\frac{\tilde{\phi}_U}{1-\tilde{\phi}_U}\right|\right\}$.

Starting from (E.3),

$$\Delta\boldsymbol{\ell} = \tilde{\boldsymbol{\Gamma}}^{-1}\Big(\Delta\boldsymbol{\Phi}^{DE} + \boldsymbol{\alpha}\,\Delta\boldsymbol{\ell}\Big),$$

which implies

$$\Delta\boldsymbol{\ell} = \tilde{\boldsymbol{\Gamma}}^{-1}\Delta\boldsymbol{\Phi}^{DE} + \tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\alpha}\,\Delta\boldsymbol{\ell}.$$

Substituting the right-hand side for $\Delta\boldsymbol{\ell}$ recursively yields the Neumann series

$$\Delta\boldsymbol{\ell} = \tilde{\boldsymbol{\Gamma}}^{-1}\sum_{k=0}^{\infty}\boldsymbol{\alpha}_X^k\Delta\boldsymbol{\Phi}^{DE},$$

which converges because $\|\boldsymbol{\alpha}_X\|_\infty = \zeta^{DE} < 1$.

Under the maintained assumption $|\frac{\tilde{\phi}_U\tilde{\phi}_A}{(1-\tilde{\phi}_U)(1-\tilde{\phi}_A)}| < 1$, we have $\zeta^{DE} < 1$ and the Neumann series exists:

$$\Delta\boldsymbol{\ell} = \tilde{\boldsymbol{\Gamma}}^{-1}\sum_{k=0}^{\infty}\boldsymbol{\alpha}_X^k\Delta\boldsymbol{\Phi}^{DE}. \tag{E.4}$$

Part (i): Direct-Effect proxy. Truncating (E.4) after the zeroth term yields

$$\Delta^{DE}\boldsymbol{\ell} \equiv \tilde{\boldsymbol{\Gamma}}^{-1}\Delta\boldsymbol{\Phi}^{DE},$$

and the remainder is

$$\Delta\boldsymbol{\ell} - \Delta^{DE}\boldsymbol{\ell} = \tilde{\boldsymbol{\Gamma}}^{-1}\sum_{k=1}^{\infty}\boldsymbol{\alpha}_X^k\Delta\boldsymbol{\Phi}^{DE}.$$

Taking norms and using $\|\tilde{\boldsymbol{\Gamma}}^{-1}\|_\infty = \max\left\{\left|\frac{1}{1-\tilde{\phi}_U}\right|, \left|\frac{1}{1-\tilde{\phi}_A}\right|\right\}$ together with $\left\|\sum_{k=1}^{\infty}\boldsymbol{\alpha}_X^k\right\|_\infty \le \frac{\zeta^{DE}}{1-\zeta^{DE}}$ delivers

$$\|\Delta\boldsymbol{\ell} - \Delta^{DE}\boldsymbol{\ell}\|_\infty \le \max\left\{\left|\frac{1}{1-\tilde{\phi}_U}\right|, \left|\frac{1}{1-\tilde{\phi}_A}\right|\right\} \cdot \max\{\Delta\ln\Phi_{Ri}^{DE}, \Delta\ln\Phi_{Fj}^{DE}\} \cdot \frac{\zeta^{DE}}{1-\zeta^{DE}}.$$

Because the ∞-norm bounds each component of the vector, this inequality implies the first pair of bounds claimed in Proposition 1.

Part (ii): Augmented Proxy. Define the augmented-access terms

$$\Delta \ln \Phi_{Ri}^{AP} \equiv \Delta \ln \Phi_{Ri}^{DE} + \frac{\tilde{\phi}_A}{1-\tilde{\phi}_A} \sum_j \Pi_{ij}^R \Delta \ln \Phi_{Fj}^{DE}, \tag{E.5}$$

$$\Delta \ln \Phi_{Fj}^{AP} \equiv \Delta \ln \Phi_{Fj}^{DE} + \frac{\tilde{\phi}_U}{1-\tilde{\phi}_U} \sum_i \Pi_{ij}^F \Delta \ln \Phi_{Ri}^{DE}. \tag{E.6}$$

Stacking these objects as $\Delta \mathbf{\Phi}^{AP}$, we have $\Delta^{AP} \boldsymbol{\ell} = \tilde{\mathbf{\Gamma}}^{-1} \Delta \mathbf{\Phi}^{AP}$. Recall from (9)–(10) that the log-linearized accessibility equations can be written as

$$\Delta \ln \mathbf{\Phi}_R = \Delta \ln \mathbf{\Phi}_R^{DE} + \frac{\tilde{\phi}_A}{1-\tilde{\phi}_A} \Pi^R \Delta \ln \tilde{\boldsymbol{L}}_F, \qquad \Delta \ln \mathbf{\Phi}_F = \Delta \ln \mathbf{\Phi}_F^{DE} + \frac{\tilde{\phi}_U}{1-\tilde{\phi}_U} (\Pi^F)^\top \Delta \ln \tilde{\boldsymbol{L}}_R.$$

Replacing $\Delta \ln \tilde{\boldsymbol{L}}_F$ and $\Delta \ln \tilde{\boldsymbol{L}}_R$ in these expressions using (E.1)–(E.2) yields

$$\Delta \ln \mathbf{\Phi}_R = \Delta \ln \mathbf{\Phi}_R^{AP} + c\, \Pi^R (\Pi^F)^\top \Delta \ln \mathbf{\Phi}_R, \tag{E.7}$$

$$\Delta \ln \mathbf{\Phi}_F = \Delta \ln \mathbf{\Phi}_F^{AP} + c\, (\Pi^F)^\top \Pi^R \Delta \ln \mathbf{\Phi}_F, \tag{E.8}$$

where $c \equiv \frac{\tilde{\phi}_U \tilde{\phi}_A}{(1-\tilde{\phi}_U)(1-\tilde{\phi}_A)}$ and $\Delta \ln \mathbf{\Phi}_R$ collects the true first-order changes in $\{\ln \Phi_{Ri}\}$. Equations (E.7)–(E.8) show that $\Delta \ln \mathbf{\Phi}_R$ (and analogously $\Delta \ln \mathbf{\Phi}_F$) are Neumann series in $c\Pi^R(\Pi^F)^\top$, whose ∞-norm equals $|c|$ because the share matrices are row-stochastic. Consequently,

$$\|\Delta \ln \mathbf{\Phi}_R - \Delta \ln \mathbf{\Phi}_R^{AP}\|_\infty \le \frac{|c|}{1-|c|} \|\Delta \ln \mathbf{\Phi}_R^{AP}\|_\infty, \qquad \|\Delta \ln \mathbf{\Phi}_F - \Delta \ln \mathbf{\Phi}_F^{AP}\|_\infty \le \frac{|c|}{1-|c|} \|\Delta \ln \mathbf{\Phi}_F^{AP}\|_\infty,$$

with $|c| = \zeta^{AP}$. Since $\Delta \ln \tilde{L}_{Ri} = \frac{1}{1-\tilde{\phi}_U} \Delta \ln \Phi_{Ri}$ and $\Delta^{AP} \ln \tilde{L}_{Ri} = \frac{1}{1-\tilde{\phi}_U} \Delta \ln \Phi_{Ri}^{AP}$ (and analogously for firms), multiplying both sides of the inequalities above by $\left|\frac{1}{1-\tilde{\phi}_U}\right|$ or $\left|\frac{1}{1-\tilde{\phi}_A}\right|$ and taking maxima over locations gives

$$|\Delta \ln \tilde{L}_{Rl} - \Delta^{AP} \ln \tilde{L}_{Rl}| \le \max_i \left\{ \Delta \ln \Phi_{Ri}^{AP} \right\} \frac{\zeta^{AP}}{1-\zeta^{AP}},$$

$$|\Delta \ln \tilde{L}_{Fl} - \Delta^{AP} \ln \tilde{L}_{Fl}| \le \max_j \left\{ \Delta \ln \Phi_{Fj}^{AP} \right\} \frac{\zeta^{AP}}{1-\zeta^{AP}},$$

which completes the proof. □

E.2 Proof of Uniqueness for Model Without Floorspace

Proof. We prove uniqueness for the more general case where:

1. $\tau_{ij} \in \mathbb{R}_{++} \cup \{\infty\}$,
2. For all $i \in S$, $\tau_{ij} < \infty$ for at least one j,
3. For all $j \in S$, $\tau_{ij} < \infty$ for at least one i,

4. $\{\tilde{A}_i\} \in \mathbb{R}^N_{++}$, $\{\tilde{U}_i\} \in \mathbb{R}^N_{++}$

5. $\theta \in \mathbb{R}_+$

6. $\frac{|\phi_A \theta \phi_U \theta|}{|(1-\phi_U \theta)(1-\phi_A \theta)|} < 1$

, then there exists a unique-to-scale solution with $\{L_{Ri}\} \in \mathbb{R}^N_{++}$, $\{L_{Fj}\} \in \mathbb{R}^N_{++}$, to the system.

Equations (9) and (10) together with the average utility expression (8) lead to the following system of $2N + 1$ equations:

$$\zeta = \frac{\bar{L}}{\left[\sum_i \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j L_{Fj}^{\tilde{\phi_A}} L_{Ri}^{\tilde{\phi_U}}\right]} \tag{E.9}$$

$$L_{Ri} = \zeta \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j L_{Fj}^{\tilde{\phi_A}} L_{Ri}^{\tilde{\phi_U}} \tag{E.10}$$

$$L_{Fj} = \zeta \sum_i \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j L_{Fj}^{\tilde{\phi_A}} L_{Ri}^{\tilde{\phi_U}} \tag{E.11}$$

Allen et al. (2024) characterize the equilibrium existence and uniqueness properties of systems of the above type. In this case, we define the change of variables:

$$y_{Ri} \equiv L_{Ri}^{1-\tilde{\phi_U}} \zeta^{\tilde{D}_R} \tag{E.12}$$

$$y_{Fj} \equiv L_{Fj}^{1-\tilde{\phi_A}} \zeta^{\tilde{D}_F} \tag{E.13}$$

where

$$\begin{bmatrix} \tilde{D}_R \\ \\ \tilde{D}_F \end{bmatrix} = \begin{bmatrix} 1 & \frac{-\tilde{\phi_A}}{1-\tilde{\phi_A}} \\ \\ \frac{-\tilde{\phi_U}}{1-\tilde{\phi_U}} & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ \\ -1 \end{bmatrix}$$

which is well defined if $\tilde{\phi_U} \neq 1, \tilde{\phi_A} \neq 1$ and if the determinate of the matrix $\begin{bmatrix} 1 & \frac{-\tilde{\phi_A}}{1-\tilde{\phi_A}} \\ \\ \frac{-\tilde{\phi_U}}{1-\tilde{\phi_U}} & 1 \end{bmatrix}$ is non-zero: $1 - \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \frac{\tilde{\phi_U}}{1-\tilde{\phi_U}} \neq 0$, or equivalently, that $\tilde{\phi_A} + \tilde{\phi_U} \neq 1$ when $\tilde{\phi_A} \neq 1, \quad \tilde{\phi_U} \neq 1$, which our hypothesis satisifes.

Substitution from the change-of-variables yields the following system of $2 \times N$ equations of $2 \times N$ unknowns, where ζ is no longer present in the equations:

$$y_{Ri} = \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Fj}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}} \tag{E.14}$$

$$y_{Fj} = \sum_i \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Ri}^{\frac{\tilde{\phi_U}}{1-\tilde{\phi_U}}} \tag{E.15}$$

The hypothesis of Theorem 1 (i) in Allen et al. (2024) with Remark 1, concerns the bounds to the following expressions when $y_{Ri} \in \mathbb{R}_{++}, y_{Fi} \in \mathbb{R}_{++}$: $\sum_k \left| \frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Fj}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}}}{\partial \ln y_{Rk}} \right|$, $\sum_k \left| \frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Fj}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}}}{\partial \ln y_{Fk}} \right|$, $\sum_k \left| \frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Ri}^{\frac{\tilde{\phi_U}}{1-\tilde{\phi_U}}}}{\partial \ln y_{Rk}} \right|$, $\sum_k \left| \frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Ri}^{\frac{\tilde{\phi_U}}{1-\tilde{\phi_U}}}}{\partial \ln y_{Fk}} \right|$.

We have: $\frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Fj}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}}}{\partial \ln y_{Rk}} = 0$, $\frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Fj}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}}}{\partial \ln y_{Fk}} = \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \frac{d_{ik}^{-\phi} \tilde{U}_i \tilde{A}_k y_{Fk}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}}}{\sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Fj}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}}}$, whose denominator is well defined if for all i, $\tau_{ij} < \infty$ for at least one j, given that the variables $\tilde{U}_i, \tilde{A}_j, y_{Fj}$ are strictly positive. Thus, $\sum_k \left| \frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Fj}^{\frac{\tilde{\phi_A}}{1-\tilde{\phi_A}}}}{\partial \ln y_{Fk}} \right| = \left| \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \right|$. Similarly, $\sum_k \left| \frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Ri}^{\frac{\tilde{\phi_U}}{1-\tilde{\phi_U}}}}{\partial \ln y_{Rk}} \right| = \left| \frac{\tilde{\phi_U}}{1-\tilde{\phi_U}} \right|$ if for all j, $\tau_{ij} < \infty$ for at least one i, and $\sum_k \left| \frac{\partial \ln \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j y_{Ri}^{\frac{\tilde{\phi_U}}{1-\tilde{\phi_U}}}}{\partial \ln y_{Fk}} \right| = 0$

The hypothesis of Theorem 1 (i) in Allen et al. (2024) with Remark 1 requires that the spectral radius of the matrix of bounds: $\begin{bmatrix} 0 & \left| \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \right| \\ \left| \frac{\tilde{\phi_U}}{1-\tilde{\phi_U}} \right| & 0 \end{bmatrix}$, is less than one. The largest eigenvalue of the matrix can be shown to be: $\rho\left(\begin{bmatrix} 0 & \left| \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \right| \\ \left| \frac{\tilde{\phi_U}}{1-\tilde{\phi_U}} \right| & 0 \end{bmatrix} \right) = \sqrt{\left| \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \right| \left| \frac{\tilde{\phi_U}}{1-\tilde{\phi_U}} \right|}$.

Our hypothesis states that $\left| \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \frac{\tilde{\phi_U}}{1-\tilde{\phi_U}} \right| < 1$, which implies that $\sqrt{\left| \frac{\tilde{\phi_A}}{1-\tilde{\phi_A}} \right| \left| \frac{\tilde{\phi_U}}{1-\tilde{\phi_U}} \right|} < 1$.

Thus by Theorem 1 (i) in Allen et al. (2024) with Remark 1, there exists a unique solution to the system (E.14) and (E.15), implying that there exists a unique-to-scale solution to the system (E.10) and (E.11). Then the exogenous total population $\bar{L}$ pins down the scale of the solution.

Substituting back the original variables, and now defining $\tilde{L}_{Ri} \equiv \kappa_R L_{Ri}$ and $\tilde{L}_{Fj} \equiv \kappa_F L_{Fj}$ where $\kappa_R \equiv \zeta^{\frac{\tilde{D}_R}{1-\tilde{\phi_U}}}$ and $\kappa_F \equiv \zeta^{\frac{\tilde{D}_F}{1-\tilde{\phi_A}}}$, we can rewrite the system as:

$$\tilde{L}_{Ri} = \sum_j \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j \tilde{L}_{Fj}^{\tilde{\phi_A}} \tilde{L}_{Ri}^{\tilde{\phi_U}}$$

$$\tilde{L}_{Fj} = \sum_i \tau_{ij}^{-\theta} \tilde{U}_i \tilde{A}_j \tilde{L}_{Ri}^{\tilde{\phi_U}} \tilde{L}_{Fj}^{\tilde{\phi_A}}$$

as desired.

□

F Additional Figures and Tables

Train Costs

	Annual costs (MM JPY)	Track length (km)	Stations
Panel A. Lines			
Tokyo Metro	284,490	195.0	179
Tsukuba	37,151	58.0	20
Asakusa	24,111	18.3	20
Mita	26,954	26.5	27
Shinjuku	31,521	23.5	21
Oedo	56,498	40.7	38
Yokohama	55,060	53.0	42
Hokuso	19,876	32.0	15
Panel B. OLS regression (no intercept)			
Dependent variable: Annual costs (MM JPY)			
Track length (km)	98.9	(218.6)	
Stations	1,446.3	(248.7)	
Observations	8		
R^2	0.995		

Table F.1: Annual costs and network characteristics; OLS without intercept

Notes: Costs are annual costs in millions of JPY (MM JPY). Track length is in kilometers. OLS without an intercept; standard errors in parentheses.

Model with Congestion

	With Train	Initial Train Removal		With Congestion	
	Level	Level	Pct. Chg	Level	Pct. Chg
Centralization					
<10km Population Share	0.236	0.206	-12.7%	0.201	-15.1%
<10km Employment Share	0.416	0.349	-16.1%	0.356	-14.5%
Central: Population Share	0.46	0.427	-7.3%	0.425	-7.6%
Central: Employment Share	0.576	0.504	-12.5%	0.51	-11.6%
Log Pop. Density Gradient	0.075	0.069	-8.1%	0.068	-8.8%
Log Emp. Density Gradient	0.094	0.082	-12.1%	0.083	-11.3%
Central $\frac{\text{Employment}}{\text{Population}}$	1.252	1.182	-5.6%	1.198	-4.3%
Closed Economy: Welfare					
Welfare (excl. construction)	1	0.676	-32.4%	0.655	-34.5%
Welfare (2% discount rate)	–	28.2%		30.3%	
$\frac{\text{Benefit}}{\text{Cost}}$ (2% discount rate)		7.741		8.231	
Welfare (4% discount rate)	–	26.3%		28.4%	
$\frac{\text{Benefit}}{\text{Cost}}$ (4% discount rate)		5.323		5.66	
Welfare (6% discount rate)	–	20.9%		22.9%	
$\frac{\text{Benefit}}{\text{Cost}}$ (6% discount rate)		2.802		2.979	
With Migration					
Total Population				0.526	
Total Land Value				0.51	

Table F.2: Train Removal with Road Congestion

Notes: "With Train" reports the 2020 equilibrium, "Initial Train Removal" reports the immediate effects of removing the train network from 2020 using the shared floorspace model with car travel times calibrated using method described in Section B.1 and D.3, and "With Congestion" reports the final equilibrium after road congestion effects have adjusted. Percentage changes are relative to the 2020 equilibrium. "Central" refers to regions within 20 km of the Imperial Palace. "Log Pop. Density Gradient" and "Log Emp. Density Gradient" report the slopes of log density with respect to distance (km) from the center. "Welfare (excl. construction)" reports the closed-city consumption-equivalent change when construction, maintenance, and operating costs are ignored. Welfare rows list closed-city consumption equivalents (net of construction/operating costs) plus benefit–cost ratios at 2%, 4%, and 6% discount rates. Migration rows allow Tokyo's total population and land value to adjust while holding Japan's population fixed and assuming construction costs are paid nationally.

Supplemental Subway and Radial Line Figures

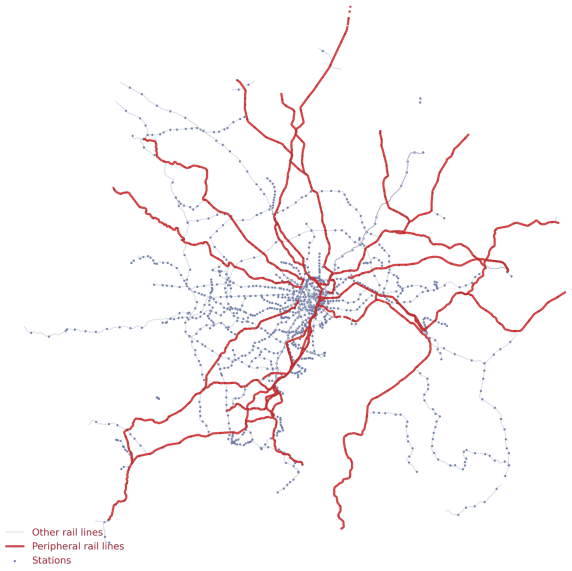


Figure F.1: Map of Radial Train Lines Removed in Section 8

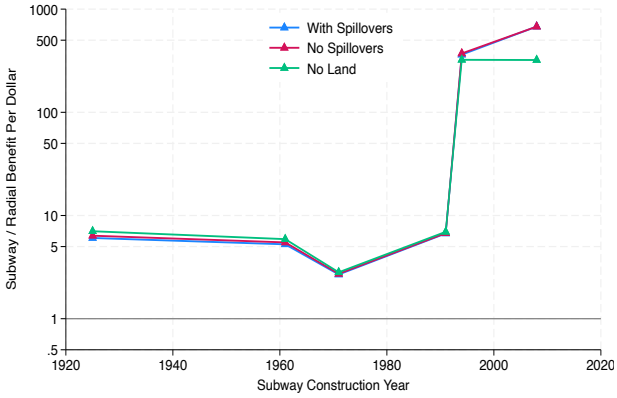


Figure F.2: Comparison of Welfare Impact of Subway Across Model Assumptions

Instrumental Variables Estimates

	(1) OLS	(2) Reduced Form	(3) First Stage	(4) IV
Log Change in Population	0.437***			0.352***
	(311.14)			(13.71)
First Station within 1km		0.170***	0.458***	
		(3.65)	(4.24)	
First Station within 3km		0.0752	0.304***	
		(1.53)	(2.80)	
First Station within 5km		0.116*	0.225	
		(1.67)	(1.63)	
Constant	-0.0546***	-0.0229***	0.0688***	
	(-437.39)	(-12.98)	(18.08)	
R-squared	0.983	0.221	0.121	0.941
Obs.	8594	8594	8594	8594
Municipality FE	✓	✓	✓	✓

Table F.3: Residential Spillover Estimates (First Station)

Cluster-robust Standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Observations are at the census-tract level. Tracts that do not have employment data in both periods are omitted. The dependent variable is the model-implied change in log amenity value, and the endogenous explanatory variable is log change in population. The instruments are 3 dummies for whether or not a tract received its first train station (during the sample period) within 1km, 3km, or 5km of the tract, respectively.

	(1) OLS	(2) Reduced Form	(3) First Stage	(4) IV
Log Change in Employment	0.349***			0.304***
	(376.11)			(21.24)
First Station within 1km		0.142***	0.451***	
		(5.08)	(5.50)	
First Station within 3km		0.127***	0.428***	
		(3.33)	(3.83)	
First Station within 5km		-0.0662	-0.126	
		(-1.57)	(-0.92)	
Constant	-0.0357***	-0.0142***	0.0585***	
	(-499.51)	(-14.34)	(19.53)	
R-squared	0.984	0.184	0.086	0.964
Obs.	8719	8719	8719	8719
Municipality FE	✓	✓	✓	✓

Table F.4: Production Spillover Estimates (First Station)

Cluster-robust Standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Observations are at the census-tract level. The dependent variable is the model-implied change in log productivity, and the endogenous explanatory variable is log change in employment. The instruments are 3 dummies' for whether or not a tract received its first train station (during the sample period) within 1km, 3km, or 5km of the tract, respectively. Tracts that do not have employment data in both periods are omitted.

	(1) OLS	(2) Reduced Form	(3) First Stage	(4) IV
Log Change in Population	0.437***			0.367***
	(311.14)			(16.83)
New Station within 1km		0.100***	0.272***	
		(3.33)	(3.77)	
New Station within 3km		0.0340**	0.0969***	
		(2.17)	(2.61)	
New Station within 5km		0.0242	0.0620	
		(1.03)	(1.30)	
Constant	-0.0546***	-0.0464***	0.00585	
	(-437.39)	(-3.54)	(0.21)	
R-squared	0.983	0.220	0.118	0.953
Obs.	8594	8594	8594	8594
Municipality FE	✓	✓	✓	✓

Table F.5: Residential Spillover Estimates (New Station)

Cluster-robust Standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Observations are at the census-tract level. The dependent variable is the model-implied change in log amenity value, and the endogenous explanatory variable is log change in population. The instruments are 3 dummies' for whether or not a tract there was an increase in the number of train stations (during the sample period) within 1km, 3km, or 5km of the tract, respectively. Tracts that do not have employment data in both periods are omitted.

	(1) OLS	(2) Reduced Form	(3) First Stage	(4) IV
Log Change in Employment	0.349***			0.277***
	(376.11)			(16.24)
New Station within 1km		0.0686***	0.236***	
		(4.43)	(5.24)	
New Station within 3km		0.00710	0.0321	
		(0.64)	(0.98)	
New Station within 5km		0.00304	0.0354	
		(0.21)	(0.76)	
Constant	-0.0357***	-0.0190***	0.0285	
	(-499.51)	(-2.60)	(1.15)	
R-squared	0.984	0.178	0.079	0.939
Obs.	8719	8719	8719	8719
Municipality FE	✓	✓	✓	✓

Table F.6: Production Spillover Estimates (New Station)

Cluster-robust Standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Observations are at the census-tract level. The dependent variable is the model-implied change in log productivity, and the endogenous explanatory variable is log change in employment. The instruments are 3 dummies for whether or not the tract experienced an increase in the number of train stations (during the sample period) within 1km, 3km, or 5km of the tract, respectively. Tracts that do not have employment data in both periods are omitted.

	New Station		First Station	
	Prod.	Amen.	Prod.	Amen.
Log Change in Employment	0.287		0.337	
	(0.012)		(0.008)	
Log Change in Population		0.362		0.388
		(0.016)		(0.022)
N	2673	2606	158	154
K–P rk F-Stat	41.1	29.5	10.5	7.1
Municipality FE	Yes	Yes	Yes	Yes
# Stations <1km	835	835	229	229
# Stations <3km	2683	2683	174	174

Table F.7: Summary of Agglomeration IV Results (New Station within 3km)

This table summarizes the IV results from the robustness exercise where we restrict the sample to census tracts that received a new train station (left two columns) or its first train station (right two columns) within 3 km of the census tract. The first two columns correspond to the results using the "New Station" instruments (i.e. whether or not a tract experienced an increase in the number of train stations within 1km), while the last two columns correspond to the results using the "First Station" instruments (i.e. whether or not a tract experienced its first train station within 1km). The last two rows report the number of census tracts that experienced an increase from zero to at least ones station within X km (for the First Station instrument) or the number of census tracts that experienced an increase in the number of stations within X km (for the New Station instrument). All specifications include municipality fixed effects. Standard errors are clustered at the municipality level. K–P rk F-Stat refers to the Kleibergen–Paap rk F statistic for weak instruments.