Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio*

Nick Tsivanidis†
University of California, Berkeley

August, 2023

Abstract

This paper estimates the effects of improving public transit infrastructure on city structure and welfare. It begins by developing a quantitative urban model with multiple groups of workers and transit modes. A special case of this model admits a sufficient statistics approach that measures aggregate welfare gains from improved transit in a broader class of models. The paper then estimates the reduced-form elasticities needed to implement the approach using data spanning the construction of the world's largest Bus Rapid Transit system in Bogotá, Colombia. This class of models performs well in explaining the adjustment of economic activity to the system. The standard approach for measuring the welfare gains from new infrastructure based on the value of travel time saved only accounts for 54% of the total welfare gain. Using the more general model to assess the distributional consequences, there is little impact on inequality after accounting for reallocation and general equilibrium effects.

*Thanks to Chang-Tai Hsieh, Marianne Bertrand, Erik Hurst and Robert E. Lucas Jr. for their guidance and support. I also thank Rodrigo Adao, Treb Allen, David Atkin, Jonathan Dingel, Cecile Gaubert, Rick Hornbeck, Daniel Sturm, Felix Tintelnot, Owen Zidar, and numerous seminar participants for their helpful comments. Additionally, I am thankful to Andres Villaveces, Marco Antonio Llinas Vargas, Helmut Menjura and those at DANE for help with data access. I gratefully acknowledge financial support from the Chicago Booth Social Enterprise Initiative, the University of Chicago Urban Network, the Energy Policy Institute at Chicago and the North American Regional Science Council. A previous version of this paper circulated under the title "The Aggregate and Distributional Effect of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio". The results in this paper use confidential microdata from DANE and have been reviewed to ensure no confidential information has been disclosed. The opinions and conclusions expressed herein are my own and do not necessarily represent the views of DANE. All errors are mine.

†Haas School of Business, University of California at Berkeley. Email: ntsivanidis@berkeley.edu. First version: November 2017.

1 Introduction

How large are the economic gains from improving public transit systems in cities? With 2.5 billion people predicted to move mostly into developing country cities by 2050, governments will spend vast sums on mass transit to reduce the congestion associated with rapid urban growth.[1] While existing approaches focus on the value of travel time saved (VTTS), measuring the benefits of these systems is challenging.[2] Individuals may change where they live and work, firms may expand or enter in newly accessible locations, and wages and house prices may adjust to this reallocation. Such effects are missed by time savings, and indirect effects may be felt throughout the city even on those who do not use the system. The lack of detailed intra-city data in less developed countries coinciding with the opening of large transit systems makes the task of evaluating these impacts even more daunting.

This study investigates the impact of new transit infrastructure on the structure of cities and the welfare of their inhabitants. It does so in the context of the construction of the world's most-used Bus Rapid Transit (BRT) system–TransMilenio–in Bogotá, Colombia. Opened in 2000, TransMilenio has a daily volume of over 2.2 million trips and operates similar to a subway. Buses run in dedicated lanes with express and local services, and passengers board buses at stations which they pay to enter using smart cards. BRT provides an attractive alternative to subways in rapidly growing developing country cities: they can deliver similar reductions in commuting times at a fraction of the cost, and are much faster to build.[3] This paper uses new sources of data covering 2,800 census tracts on residence, employment, commuting patterns, and land markets spanning the system's construction.

Prior to TransMilenio, poor and low-educated workers relied on a network of informal buses which, on average, were 30% slower than cars. This suggests new transit may affect the distribution of welfare across the rich and poor. To better understand the implications of improving transit infrastructure, this paper develops a quantitative urban model. Multiple worker skill groups have non-homothetic preferences over transit modes and residential locations, and make decisions over where to live and work and which travel mode to use to commute. Cars are faster than public transit, but are expensive. In equilibrium, rich, high-educated workers are more likely to buy cars while the poor rely on public transit. While this suggests the poor are likely to benefit most from improved public transit, worker types differ in their willingness to substitute between alternative residential and employment locations and are exposed to equilibrium effects on wages and house prices.

While this model is rich enough to speak to distributional impacts, a special case admits an even simpler sufficient statistics approach to measure the aggregate from new transit infrastructure and how it it reshapes economic activity across the city. This approach has appeal since these statistics are transparently estimated through linear regression, and because the approach is applicable in a

[1]McKinsey (2016) suggests that a need for $40 trillion of spending to close the transport infrastructure gap. Combining the average subway distance from Gonzalez-Navarro and Turner (2018) and cost estimates from Baum-Snow and Kahn (2005) indicates that the average subway system costs $27.81bn in 2017 dollars to build.

[2]E.g. Train and McFadden (1978), Small and Verhoef (2007), also used the World Bank (Mackie et. al. 2005)

[3]The per mile construction cost of the subway in Colombia's second largest city, Medellín, was 10 times that of TransMilenio, with similar system speeds. TransMilenio took less than 18 months to construct, compared to the 12 years taken by Metro Medellín. The average per mile construction cost of BRT is one-tenth of rail (Menckhoff 2005).

wide class of log-linear models that allow for endogenous firm location choice, endogenous housing supply, capital in the production function, and preferences over leisure (among others). These statistics are (i) a location's change in "commuter market access" (CMA), which summarizes worker and firm access to each other through the commuting network, and (ii) the elasticities of residential population, employment and floorspace prices to CMA, and the elasticity of commute flows to commute costs. Across alternative models, the structural parameters in the reduced-form elasticities of economic activity to CMA differ, but the reduced-form elasticities and the change in CMA are sufficient statistics to specify the impacts of changes in transit infrastructure on economic activity.[4]

The construction of TransMilenio provides variation in commute costs that can be used to estimate these elasticities, but concerns remain that these were endogenous to local unobserved economic fundamentals. Instead of leaning on a single approach, this paper exploits a variety of TransMilenio's institutional features to establish its causal impact on Bogotá's structure. First, I digitize four different plans from the 1980s and 1990s for a new transit network in Bogotá and include as regressors both the realized change in CMA due to TransMilenio and the hypothetical change had the network been built under these plans. This serves both as a falsification check (by showing the hypothetical changes had no impact on economic activity conditional on the realized CMA change) and controls for the omitted variable bias that can arise from locations' non-random exposure to infrastructure changes (Borusyak and Hull 2021). Second, I exploit TransMilenio's staggered rollout across three phases through event studies and falsification tests and demonstrate that there is no growth in outcomes prior to line openings. Third, I use variation in CMA induced by changes in the network more than 1.5km from a location, which is less likely to be correlated with local unobservables. Fourth, I condition on distance to the closest station to assess whether effects are driven by changes in accessibility rather than by other features of stations (e.g. changes in foot traffic or pollution). Fifth, I construct cost-shifting instruments to predict TransMilenio's routes based both on a historical tram network and engineering estimates of the cost to build BRT on different types of land.

After showing that the log-linear relationships between changes in outcomes and CMA predicted by this class of models are borne out in the data, I use the sufficient statistics approach to quantify the aggregate effects of the new infrastructure. A key theoretical result that arises through the application of the envelope theorem to the social planner's problem in an efficient economy is that the elasticity of welfare to a change in transit infrastructure is proportional to a weighted average of time savings. This is precisely the VTTS expression used in the literature: when the equilibrium is efficient and the change in infrastructure is infinitesimally small, only the direct effects of time saved matter. However, my results show that the VTTS only accounts for 54% of the total welfare gains under the equilibrium model. The size of the shock explains one-third of the gap and the externalities explain two-thirds. Welfare rose by 2.28% in the baseline case where the BRT does not cause migration into Bogotá from the rest of Colombia, and 0.6% with migration. GDP per capita rose by 2.5-5% in these cases respectively, net of construction and operating costs. Overall, TransMilenio can account for

[4] More precisely, as Proposition 1 establishes, the reduced form elasticities and changes in CMA are sufficient statistics across all such models to compute the relative change in economic activity in any location. To pin down the overall level of changes, one or two additional parameters are usually needed. These differ by model and are often readily calibrated.

between 2.96-14.36% of GDP growth in Bogotá from 2000 to 2016, and up to 29.24% of observed population growth. While these findings are specific to Bogotá, the framework can be applied to other cities in both developed and developing countries. While I do not find evidence TransMilenio impacted travel times on other modes, an extension of the model allowing for traffic congestion leaves these results qualitatively unchanged.

I next estimate the full model to understand how these welfare gains are shared between the rich and poor. Ultimately, welfare inequality rises by a mild 0.55% due to the BRT. On the one hand, low-skilled workers benefit from improved transit through higher use. On the other hand, the incidence also depends on how easily individuals substitute between different employment and residential locations and the extent to which each group faces lower wages through the increased supply of commuters traveling through the network. These forces favor high-skilled workers. This result is robust to allowing for employment in domestic services and alternative home ownership assumptions.

Two sets of counterfactuals draw additional policy insights. First, I evaluate a "land value capture" (LVC) scheme under which development rights to increase building densities near stations are sold by the government to developers. This increases housing supply and raises government revenue, and similar schemes have seen great success in Asian cities like Hong Kong and Tokyo. However, one of the main criticisms of TransMilenio was that the city experienced such a large change in transit without any adjustment to zoning laws to allow housing supply to respond. A well-targeted scheme would have increased the welfare gains from TransMilenio by around 44%, while government revenues would have covered 6-23% of the BRT's capital costs depending on the migration response from the rest of Colombia. This highlights the return to cities pursuing an integrated transit and land use policy. Second, by measuring the impacts of counterfactual networks I find that the system of feeder buses, which run on regular roads and connect dense, outlying residential neighborhoods with TransMilenio terminals, have greater welfare gains than either of the two key trunk lines (conditional on the rest of the network being built). This emphasizes the importance of cheap, last-mile services that increase access to mass rapid transit infrastructure.

A large body of work examines the impact of transportation infrastructure on economic activity. One strand examines the impact of new transit infrastructure and typically measures changes in population and property prices as a function of distance to the central business district (Baum-Snow 2007; Gonzalez-Navarro and Turner 2018; Baum-Snow et. al. 2017) or distance to stations (Gibbons and Machin 2005; Glaeser et. al. 2008; Billings 2011). This paper adds to this work by developing a theory-consistent sufficient statistics approach to measure the impacts of transit infrastructure. The CMA measures used in this approach embrace the spillovers across spatial units induced through a commuting network that can invalidate identification assumptions in distance-based regressions.[5]

A second strand of this literature explores the effect of infrastructure between regions on economic development through goods market access in models where agents live and work in the same location (Redding and Sturm 2008; Bartelme 2018; Donaldson and Hornbeck 2016; Donaldson 2018;

[5]In addition, since the change in accessibility from a station depends on the geography of the city and the transit network, average treatment effects based on distance to stations in one context might not be externally valid in another. The CMA approach predicts different treatment effects from different transit networks based on the specific network structure.

Alder 2019). This paper considers a different class of urban models where individuals can live and work in separate locations. This distinction leads to meaningful differences in the way the same transit network might affect firm access to workers and resident access to jobs in any location.[6] I use the context provided by a large, real world change in transit infrastructure to show these differential shocks to employment and residence capture the reallocation of economic activity in the city.

This paper also contributes to the growing body of work on quantitative spatial models (Ahlfeldt et. al. 2015; Allen et. al. 2015; Bird and Venables 2019; Fajgelbaum and Schaal 2020; Monte et. al. 2018; Owens et. al. 2020; Severen 2021; Bryan and Morten 2019; Heblich et. al. 2020; Adao et. al. 2019; Allen and Arkolakis 2021). Its main contribution lies in the development of a model in which multiple worker groups have non-homothetic preferences over transit modes and residential amenities. This allows the model to capture how new transit can affect the distribution of welfare across groups through their differential reliance on public transit, and through residential neighborhood choice and gentrification when house prices rise in response to better transit access.

Lastly, this paper relates to work in transportation economics measuring the benefits of improved transportation through the VTTS (Train and McFadden 1978; Small and Verhoef 2007). It connects with work measuring agglomeration externalities, providing intra-city estimates of productivity and amenity spillovers in a developing country city, identified using an expansion in the transit network that separately shifts the supply of labor and residents across the city.[7]

The paper proceeds as follows. Section 2 discusses the context of TransMilenio and the data. Section 3 develops the model. Section 4 presents and estimates the sufficient statistics approach to assess the BRT's aggregate effects. Section 5 estimates the full model to measure its distributional effects. Section 6 concludes.

2 Background and Data

2.1 TransMilenio: The World's Most-Used BRT System

Background Bogotá is the economic center of Colombia, accounting for 16% and 25% of population and GDP respectively. In 1995, the average work commute took 55 minutes, more than double that in US cities. The vast majority were taken by bus (73%), followed by car (17%) and walking (9%).[8] Despite its importance, public transit was highly inefficient. Bus companies operated routes allocated to them by the city, but a lack of entry controls led to a large over-supply of vehicles. Low enforcement meant that up to half of the city's bus fleet operated illegally (Cracknell 2003). Disregard of bus stops led to frequent boarding and alighting along curbs, further reducing traffic flows.

[6] In trade models, firm and consumer market access often equal each other under balanced trade (e.g. Donaldson and Hornbeck 2016). One can show in my setting that it is precisely the absence of balanced trade in commuters (which would imply residence equals employment in each location, which clearly fails in the data) that delivers the BRT's very distinct shocks to resident and firm CMA shown in Figure 1.

[7] Rosenthal and Strange (2004) provide a review. Other papers using potentially exogenous sources of variation in the density of (i) employment include Combes et. al. (2010), Greenstone et. al. (2010), Kline and Moretti (2014), Ahlfeldt et. al. (2015) and (ii) residence include Bayer et. al. (2007), Guerrieri et. al. (2013), Diamond (2016), Giannone (2021).

[8] Bicycles and motorbikes account for the remaining 1% of commutes.

At the start of his first term as mayor of Bogotá, Enrique Peñalosa wasted no time in transforming the city's transit infrastructure. TransMilenio was approved in March 1998, and its first phase opened a mere 21 months later, adding 42 km along Avenida Caracas and Calle 80, two arteries of the city.[9] Phases 2 and 3 added an additional 70km in 2006 and 2012, creating a network spanning the majority of the city. Today, the system is recognized as the "gold standard" of BRT and with more than 2.2mm riders a day using its 147 stations. It is the most heavily utilized system of its kind in the world (Cervero et. al. 2013).[10] Its average operational speed of 26.2kmh reported during phase one is on par with that of the New York subway (Cracknell 2003), and far surpassed the reported 10kmh speeds on the incumbent bus network (Wright and Hook 2007).

The system involves exclusive dual bus lanes running along the median of arterial roads in the city separated from other traffic. Buses stop only at stations which are entered using a smart card so that fares are paid before arriving at platforms. Dual lanes allow for both express and local services, and passing at stations. Accessibility for poorer citizens in the urban periphery is increased through a network of feeder buses that use existing roads to bring passengers to "portals" at the end of trunk lines at no additional cost. Free transfers and a fixed fare further enhance the subsidization of the poor (at the periphery) while the government sets fares close to those offered by existing buses.

BRT is a particularly attractive alternative to subways in developing country cities since it (i) delivers similar reductions in commute times at a fraction of the cost and (ii) is much faster to build. These features have led to systems being built in more than 200 cities, the vast majority constructed over the past 15 years in Latin America and Asia (BRT Data 2017).

Route Selection and System Rollout The corridors built during the first phase of the system were consistently mentioned in 30 years of transportation studies as first-priority for mass transit (Cracknell 2003). These studies chose routes based on current and future demand levels and expected capital costs. The result was a network that connected the city center with dense residential areas in the north, northwest and south of the city (Hidalgo and Graftieux 2005). The number of car lanes was left unchanged either because existing busways were converted or due to road widening.[11]

Three features make TransMilenio an attractive context for empirical analysis. First, since 1980 multiple administrations worked on proposals for a subway system. These can be used as placebo checks. Second, having identified neighborhoods in the city's periphery to be connected with the center, the final routes were chosen largely to minimize construction costs. Lines were placed along wide arterial roads, which were cheaper to convert and determined by the the city's historical evolution. Third, TransMilenio was was rolled out quickly, primarily so that a portion could be completed within Mayor Peñalosa's term that ran between 1998 and 2001. The unanticipated nature of the system's construction and the staggered opening of lines across three phases provide sources of time

[9]While the anticipation of a system may predate its inauguration, TransMilenio went from a "general idea" to implementation in only 35 months (Hidalgo and Graftieux 2008). A "pico y placa" driving restriction implemented two years prior to TransMilenio had little impact on overall car use (Lawell et. al. 2017).

[10]A map of each system component and their opening date is provided in Figure A.1, while Figure A.2 shows a station before and after TransMilenio was built. For comparison, the London tube carries 5 million passengers per day over a network of 402km, giving it a daily ridership per km of 12,000 compared to TransMilenio's 20,000.

[11]See Cracknell (2003) for discussion. This was confirmed through inspection of satellite images.

series variation used in the analysis.

One central criticism of TransMilenio was its singular focus on improving urban mobility without coordinated changes in land use regulation (Bocarejo et. al. 2013): Appendix G shows that housing supply did not respond to the system's construction. An integrated land use and transit policy that increases housing densities near stations allows more residents and firms to take advantage of improved commuting infrastructure, and sales of development rights can finance construction. In Section 5.3, I assess the impact of TransMilenio had Bogotá pursued a such an integrated policy.

Trip Characteristics Appendix G summarizes TransMilenio use. First, it is a quantitatively important mode of transit used more for longer trips than other modes. Second, TransMilenio provides an increase in door-to-door speeds of around 17% over existing buses, but remains 8.1% slower than cars. Third, when compared to other modes the BRT is used more for work commutes than leisure trips. TransMilenio's outsized role in commuting motivates the focus on access to jobs in this paper.

Yet this improvement in public transit may have differentially affected the rich and poor. Table A.19 shows that prior to TransMilenio, commutes by car were around 35% faster than bus but that low-educated Bogotanos were about 29% more likely to use buses than cars. Both facts are robust to controlling for origin-destination pair fixed effects to adjust for differences in trip composition.

2.2 Data

This section summarizes the data used in the analysis, with further details in Appendix F. The primary geographic unit used in the analysis is the census tract ("sección"). Bogotá is partitioned into 2,799 tracts, with an average size of 133,303 square meters and a mean population of 2,429 in 2005. These are contained within larger spatial units including 19 localities and 113 planning zones (UPZs).

The primary source of population data is the Department of Statistics' (DANE) General Censuses of 1993, 2005 and 2018. This provides the residential population of each block by education level. College-educated individuals are defined as those with some post-secondary education.

Employment data come from two sources. The first is a census covering the universe of establishments from DANE's 2005 General Census and 1990 Economic Census which report the location, industry and employment of each unit. The second is a database of establishments registered with the city's Chamber of Commerce (CCB) in 2000 and 2015. The data from 2015 contain the location, industry and employment of each establishment, but in 2000 employment is not provided. I therefore use establishment counts to proxy for employment, but show that establishment count and employment densities are highly correlated in years where both are available. An additional concern is that the spatial distribution of registered establishments may be different from that of total establishments. Figure A.7 shows that the employment and establishment densities in both years of the CCB data are highly correlated with the 2005 census. Coverage is even across rich and poor neighborhoods, suggesting both that the CCB data is fairly representative of overall employment. The main specifications examine changes from the CCB data, allowing employment over 10 years to respond to the first two phases of the system, but additional analyses use the economic census data to examine the impacts of phase 1 on employment growth in the 4 years following TransMilenio's opening.

Housing market data between 2000 and 2018 come from Bogotá's Cadastre. Its mission is to keep the city's geographical information up-to-date; all parcels, formal or informal, are included and the dataset covers 98.6% of the city's more than two million properties (Ruiz and Vallejo 2010).[12] It reports the use, floorspace and land area, and value per square meter of land and floorspace, as well as a number of property characteristics. Values in the cadastre are important for the government since they determine property taxes which comprise a substantial portion of city revenue. In developed countries, these valuations are typically determined using information on market transactions. However, Bogotá, like most developing cities, lacks comprehensive records of such data and those available may be subject to systematic under-reporting. The city addresses this through an innovative approach involving sending officials to pose as potential buyers in order to negotiate a sales price under the premise of a cash payment (Anselin and Lozano-Gracia 2012). Professional assessors are also sent to value at least one property in one of each of the city's more than 16,000 "homogenous zones" (Ruiz and Vallejo 2010). As a result, Figure A.8 shows the average price per square meter of floorspace in the cadastre is highly correlated with the average purchase price per room reported in a DANE worker survey. Importantly, the relationship is constant across rich and poor neighborhoods which would not be the case were the cadastre over- or under-valuing expensive properties.

Microdata on commuting behavior come from the city's Mobility Survey administered by the Department of Mobility and overseen by DANE in 2005, 2011 and 2015. For 1995, I obtained the Mobility Survey undertaken by the Japan International Cooperation Agency (JICA) to similar specifications as the DANE surveys in later years. These are representative household surveys in which each respomdent was asked to complete a travel diary for the previous day. The survey reports the demographic information of each traveler and household, including age, education, gender, industry of occupation, car ownership and in some years income. For each trip, the data report the departure time, arrival time, purpose of the trip, travel mode, and origin and destination UPZs.

Employment data by worker come from DANE's Continuing Household Survey (ECH) between 2000 and 2005, and its extension into the Integrated Household Survey (GEIH) for the 2008-2015. These are monthly, repeated cross-sectional labor market surveys covering approximately 10,000 households in Bogotá annually. Commute times between each pair of census tracts by mode are computed in ArcGIS using shapefiles of each mode's network from the city. Figure A.10 shows the computed times correlate well with observed door-to-door times from the Mobility Surveys.

3 A Quantitative Model of a City with Heterogeneous Skills

This section develops a quantitative model that captures the impact of transit infrastructure on the spatial organization of economic activity within a city. It departs from recent work (e.g. Ahlfeldt et. al. 2015) by incorporating multiple skill groups of workers, commute modes and industries.

[12] High coverage was confirmed by overlaying the shapefile for available properties over satellite images. Underlining the importance of property taxes, in 2008 they accounted for 19.8% of Bogotá's tax revenues (Uribe Sanchez 2010).

Locations differ in terms of commute times, housing floorspace, and (exogenous) amenities and productivities.[13] Workers decide where to live, whether to own a car, where to work, and which mode of transit to use to commute. Public transit is available to everyone, but only those with a car have the option to drive. Non-homothetic preferences for car ownership and residential amenities mean the rich are more likely to own cars and live in high amenities neighborhoods. Amenities and productivities also have components that depend on local economic activity. In equilibrium, floorspace use, floorspace prices and wages adjust to clear markets.

3.1 Workers

The city is populated by worker groups indexed by $g \in G = \{L, H\}$ with a fixed population $\bar{L}_g$. A worker ω in group g chooses a location i in which to live, a location j in which to work, whether or not to own a car $a \in \{0, 1\}$, and the mode of transport m to use to commute to work. Individuals derive utility from consumption of a freely traded numeraire good ($C_i(\omega)$); consumption of residential floorspace ($H_{Ri}(\omega)$); and an amenity reflecting the average preference of each group to live in i under car ownership a (u_{iag}). Owning a car provides an additional mode to use for commuting and an amenity benefit, but comes at a fixed cost p_{car}. Workers are heterogeneous in their match-productivity with firms in each location ($\epsilon_j(\omega)$), their preference for each residence-car ownership pair ($\nu_{ia}(\omega)$), and their disutility from commuting that reduces their productivity at work ($d_{ijm}(\omega) \geq 1$). Land is owned by residents and rents are redistributed lump sum through payment π.[14]

Individuals have Stone-Geary preferences in which they need a minimum amount of floorspace $\bar{h}$ in which to live. The indirect utility of a worker who has made choice (i, j, a, m) is then

$$U_{ijamg}(\omega) = u_{iag} \left(\frac{w_{jg}\epsilon_j(\omega)}{d_{ijm}(\omega)} - p_a a - r_{Ri}\bar{h} + \pi \right) r_{Ri}^{\beta-1} \nu_{ia}(\omega) \tag{1}$$

where w_{jg} is the wage per effective unit of labor, r_{Ri} is the price of residential floorspace in i, and $p_a = p_{car}$ for $a = 1$ and 0 otherwise.

The fixed expenditures on cars and housing allow me to match the Engel curves I document for car ownership and housing expenditure (Figure A.9) and drive sorting of workers over car ownership and residential neighborhoods by income. When cars are quicker than public transit, the rich are more willing to pay the fixed cost since their value of time is higher. The fixed expenditure on subsistence housing means that the poor spend a greater share of their income on housing and are attracted to low amenity neighborhoods where it is cheap.

Workers first choose where to live and whether or not to own a car, then where to work, and finally which transportation mode to use.[15] I solve their problem by backward induction.

Mode Choice Having chosen where to live and work and whether to own a car, individuals choose

[13] Since housing supply was unaffected by TransMilenio (Appendix G), total floorspace in a location is taken as given.

[14] Specifically $\pi = \bar{L}^{-1} \sum_i (r_{Ri}H_{Ri} + r_{Fi}H_{Fi})$. This ensures that all the gains are accounted for within the model while avoiding inefficiencies introduced by absentee landlords that would impact the application of Proposition 2.

[15] While allowing for joint decisions greatly complicates inversion and estimation of the model in the presence of fixed components of expenditure and income, Appendix A.4 solves such a model and finds qualitatively similar results.

which mode of transport to use to commute to work. Commuters have nested logit demand across modes. A nest of public modes $\mathcal{B}_{Pub} \equiv \{\text{Walk}, \text{Bus}, \text{TransMilenio}\}$ is available to everyone while a nest of private modes $\mathcal{B}_{Priv} \equiv \{\text{Car}\}$ is available only to car owners. Therefore, the set of modes available depends on car ownership with $\mathcal{M}_0 = \mathcal{B}_{Pub}$ and $\mathcal{M}_1 = \mathcal{B}_{Pub} \cup \mathcal{B}_{Priv}$. Individuals have idiosyncratic preferences across modes $\upsilon_{ijm}(\omega)$ such that the realized commute cost for individual ω is given by $d_{ijm}(\omega) = \exp(\kappa t_{ijm} - b_m + \upsilon_{ijm}(\omega))$, where t_{ijm} is the time it takes to travel from i to j using mode m, b_m is a preference shifter for mode m and κ controls the mapping between commute times and costs. The commuter's problem conditional on choice (i, j, a) is simply $\min_{m \in \mathcal{M}_a} \{d_{ijm}(\omega)\}$.

Following McFadden (1974), $\upsilon_{ijm}(\omega)$ are drawn from a generalized extreme value (GEV) distribution

$$F(\upsilon_{ij1}, \ldots, \upsilon_{ijN}) = 1 - \exp\left(-\sum_k \left(\sum_{m \in \mathcal{B}_k} \exp(\upsilon_{ijm}/\lambda_k)\right)^{\lambda_k}\right) \quad \text{for } k \in \{\text{Public}, \text{Private}\}.$$

This allows for correlation of preference shocks within nests, with $\lambda_k \to 0$ under perfect correlation.

Standard results for GEV distributions imply that this leads to nested logit demand for travel modes. Expected utility prior to drawing the shocks $\upsilon_{ijm}(\omega)$ is given by

$$U_{ijamg}(\omega) = u_{iag}\left(\frac{w_{jg}\epsilon_j(\omega)}{d_{ija}} - p_a a - r_{Ri}\bar{h} + \pi\right) r_{Ri}^{\beta-1} \nu_i(\omega)$$

where $d_{ija} = \exp(\kappa t_{ija})$ and

$$t_{ij0} = -\frac{\lambda}{\kappa} \ln \sum_{m \in \mathcal{B}_{\text{Public}}} \exp\left(b_m - \frac{\kappa}{\lambda} t_{ijm}\right) \tag{2}$$

$$t_{ij1} = -\frac{1}{\kappa} \ln\left(\exp(b_{car} - \kappa t_{ij\text{Car}}) + \exp(\kappa t_{ij0})\right). \tag{3}$$

Intuitively, the expected commute cost d_{ija} can be expressed as the inclusive value of commute times available to the individual with car ownership a.[16]

Employment Choice Having chosen where to live and whether to own a car, individuals draw a vector of match-productivities with firms across the city iid from a Frechet distribution $F(\epsilon_j) = \exp\left(-\tilde{T}_g \epsilon_j^{-\theta_g}\right)$. Here θ_g measures the dispersion of productivities while $\tilde{T}_g$ controls their level.

With these draws in hand, linearity of (1) means that workers choose to work in the location that offers the highest income net of commute costs $\max_j\{w_{jg}\epsilon_j(\omega)/d_{ija}\}$. Standard results imply that the probability a type-g worker who has made choice (i, a) decides to work in j is given by

$$\pi_{j|iag} = \frac{(w_{jg}/d_{ija})^{\theta_g}}{\sum_s (w_{sg}/d_{isa})^{\theta_g}} \equiv \frac{(w_{jg}/d_{ija})^{\theta_g}}{\Phi_{Riag}}. \tag{4}$$

[16]The set of shifters are normalized so that $\bar{t}_{iia} = 0\ \forall i, a$. This is equivalent to always taking a weighted average travel time of available modes, where the weights are the preference shifters b_m.

The term $\Phi_{Riag} \equiv \sum_s (w_{sg}/d_{isa})^{\theta_g}$ is defined as Residential Commuter Market Access (RCMA). It captures residents' access to well-paid jobs from location i. Individuals are more likely to commute to locations with a high wage net of commute costs (the numerator) relative to other locations (the denominator). The sensitivity of commute decisions to commute costs is governed by the dispersion of productivities, with a greater dispersion (lower θ_g) making choices less sensitive. Differences in θ_g across groups will be important in determining the incidence of improved infrastructure, since it controls the extent to which individuals are willing to bear high commute costs to work in a location.

Expected income prior to drawing the vector of match productivities is related to RCMA through

$$\bar{y}_{iag} = T_g \Phi_{Riag}^{1/\theta_g}, \tag{5}$$

where T_g is a transformation of the location parameter of the Frechet distribution.[17]

Residence and Car Ownership Choice In the first stage, individuals choose where to live and whether or not to own a car to maximize expected indirect utility. The idiosyncratic preferences $\nu_{ia}(\omega)$ are drawn from a Frechet distribution with shape parameter $\eta_g > 1$ and unit scale. The supply of type-g individuals to location i and car ownership a is then

$$L_{Riag} = \lambda_{U,g} \left(u_{iag} \tilde{y}_{iag} r_{Ri}^{\beta-1} \right)^{\eta_g} \tag{6}$$

where $\tilde{y}_{iag} \equiv \bar{y}_{iag} - p_a a - r_{Ri}\bar{h} + \pi$ is expected net income and $\lambda_{U,g}$ is an equilibrium constant.[18]

3.1.1 Aggregation

Firm Commuter Market Access and Labor Supply The supply of workers to any location is found by summing over the number of residents who commute there $L_{Fjg} = \sum_{i,a} \pi_{j|iag} L_{Riag}$. This implies

$$L_{Fjg} = w_{jg}^{\theta_g} \Phi_{Fjg} \tag{7}$$
$$\text{where } \Phi_{Fjg} = \sum_{i,a} d_{ija}^{-\theta_g} \frac{L_{Riag}}{\Phi_{Riag}}$$

Labor supply is log-linear and depends on two forces. First, more workers commute to destinations paying higher wages. Second, conditional on wages firms attract workers when they have better access to them through the commuting network. This is captured through Φ_{Fjg} which I refer to as Firm Commuter Market Access as it reflects firms' access to workers. This is because individuals care about wages net of commute costs. Total effective labor supply to a location is given by

[17]The constants in this section are $T_g \equiv \gamma_{\theta,g} \tilde{T}_g^{1/\theta_g}$, $\gamma_{\theta,g} = \Gamma\left(1 - \frac{1}{\theta_g}\right)$, $\lambda_{U,g} = \bar{L}_g (\gamma_{\eta,g}/\bar{U}_g)^{\eta_g}$ and $\gamma_{\eta,g} = \Gamma\left(1 - \frac{1}{\eta_g}\right)$ where $\Gamma(\cdot)$ is the gamma function. Expected utility prior to learning match productivities is $U_{iag}(\omega) = u_{iag}\left(\bar{y}_{iag} - p_a a - r_{Ri}\bar{h}\right) r_{Ri}^{\beta-1} \nu_{ia}(\omega)$.

[18]The model requires that $\pi > p_a a + r_{Ri}\bar{h}\ \forall i$ such that $\bar{u}_{iag} > 0$, since the Frechet distribution implies there will always be a positive mass of individuals with income arbitrarily close to zero. This is satisfied when the model is taken to the data.

$\tilde{L}_{Fjg} = \bar{\epsilon}_{jg} L_{Fjg}$, where $\bar{\epsilon}_{jg}$ is the average productivity of type-g workers who decide to work in j.[19]

Worker Welfare Properties of the Frechet distribution imply that average welfare in each location is equal to the expected utility prior to the first stage given by

$$\bar{U}_g = \gamma_{\eta,g} \left[\sum_{i,a} \left(u_{iag} \tilde{y}_{iag} r_{Ri}^{\beta-1} \right)^{\eta_g} \right]^{1/\eta_g} \tag{8}$$

3.1.2 Firms

Technology There are $s \in \{1, \ldots, S\}$ industries that produce varieties differentiated by location under perfect competition. Output is freely traded, and consumers aggregate each variety into the numeraire under CES with elasticity of substitution $\sigma_D > 1$. Firms produce using a Cobb-Douglas technology over labor and commercial floorspace

$$Y_{js} = A_{js} N_{js}^{\alpha_s} H_{Fjs}^{1-\alpha_s}$$
$$\text{where } N_{js} = \left(\sum_g \alpha_{sg} \tilde{L}_{Fjgs}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}.$$

The labor input is a CES aggregate over each skill group's effective labor with elasticity of substitution σ, $\alpha_s = \sum_g \alpha_{sg}$ is the total labor share, and A_{js} is the productivity of location j for firms in industry s which firms take as given.

Industries differ in terms of the intensity in which they use different types of workers α_{sg}. All else equal, industries like real estate and financial services demand more high-skilled workers while others, such as hotels and restaurants, rely on the low-skilled.

Factor Demand Perfect competition implies that the price of each variety is equal to its marginal cost $p_{js} = W_{js}^{\alpha_s} r_{Fj}^{1-\alpha_s} / A_{js}$, where r_{Fj} is the price of commercial floorspace in j and

$$W_{js} = \left(\sum_g \alpha_{sg}^{\sigma} w_{jg}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

is the cost of labor for firms of industry s in location j. Intuitively, labor costs differ by industry due to their differential skill requirements. Solving the firm's cost minimization problem and letting X_{js} denote firm sales, the demand for labor and commercial floorspace is[20]

$$\tilde{L}_{Fjgs} = \left(\frac{w_{jg}}{\alpha_{sg} W_{js}} \right)^{-\sigma} N_{js} \tag{9}$$

$$H_{Fjs} = (1 - \alpha_s) \frac{X_{js}}{r_{Fj}}. \tag{10}$$

[19] In particular, $\bar{\epsilon}_{jg} = T_g \sum_{i,a} \frac{\pi_{j|iag}^{-1/\theta_g}}{d_{ija}} \frac{\pi_{j|iag} L_{Riag}}{\sum_{r,o} \pi_{j|rog} L_{Rrog}}$.

[20] Here $X_{js} = p_{js}^{1-\sigma_D} X$ where $X = \sum_i \beta(E_i - \bar{h} r_{Ri} L_{Ri})$ is total spending on goods in the city and $E_i = \sum_{g,a} (\bar{y}_{iag} - p_a a + \pi) L_{Riag}$ is total spending on goods and housing from residents in i.

3.1.3 Floorspace

Market Clearing In each location there is a fixed amount of floorspace H_i, a fraction ϑ_i of which is allocated to residential use and $1-\vartheta_i$ to commercial use. Market clearing for residential floorspace requires its supply $H_{Ri} = \vartheta_i H_i$ equals demand:

$$r_{Ri} = (1-\beta)\frac{E_i}{H_{Ri} - \beta\bar{h}L_{Ri}} \tag{11}$$

where $L_{Ri} = \sum_{g,a} L_{Riag}$ is the total number of residents in i. Likewise, the supply of commercial floorspace $H_{Fj} = (1-\vartheta_i)H_j$ must equal that which is demanded by firms:

$$r_{Fj} = \frac{\sum_s (1-\alpha_s)\left(W_{js}^{\alpha_s} r_{Fj}^{1-\alpha_s}/A_{js}\right)^{1-\varsigma} X}{H_{Fj}}. \tag{12}$$

Floorspace Use Allocation Landowners allocate floorspace ϑ_i to maximize profits. They receive r_{Ri} per unit allocated to residential use, but land use regulations limit the return to each unit allocated to commercial use to $(1-\tau_i)r_{Fi}$. Since they maximize profits, we have

$$\begin{aligned} \vartheta_i = 1 &\text{ if } r_{Ri} > (1-\tau_i)r_{Fi} \\ (1-\tau_i)r_{Fi} = r_{Ri} &\ \forall\{i : \vartheta_i \in (0,1)\} \\ \vartheta_i = 0 &\text{ if } (1-\tau_i)r_{Fi} > r_{Ri} \end{aligned} \tag{13}$$

3.1.4 Externalities

Productivities A location's productivity depends on both an exogenous component, $\bar{A}_{js}$, which reflects features independent of economic activity (e.g. access to roads, slope of land), and the density of employment in that location

$$A_{js} = \bar{A}_{js}\left(\tilde{L}_{Fj}/T_j\right)^{\mu_A}, \tag{14}$$

where $\tilde{L}_{Fj} = \sum_s \tilde{L}_{Fjs}$ is the total effective labor supplied to that location and T_j is its land area. The strength of agglomeration externalities is governed by the parameter μ_A.

Amenities Amenities depend on an exogenous component, $\bar{u}_{iag}$, which varies by car ownership (e.g. leafy streets, proximity to getaways surrounding the city) and a residential externality that depends on the college share of residents

$$u_{iag} = \bar{u}_{iag}\Big(L_{RiH}/L_{Ri}\Big)^{\mu_{U,g}}. \tag{15}$$

In contrast to existing urban models (e.g. Ahlfeldt et. al. 2015), endogenous amenities depend on demographic composition across skill groups rather than the total density of residents. This seems especially applicable in developing country cities that lack strong public goods provision. In Bogotá, where crime is a significant problem, the rich often pay for private security around their buildings which increases the sense of safety in those areas. This externality provides an additional force toward residential segregation, since the high-skilled are more willing to pay to live in high-amenity neighborhoods and by doing so increase amenities further.[21]

3.1.5 Equilibrium

Definition. *Given vectors of exogenous location characteristics* $\{H_i, \bar{u}_{iag}, \bar{A}_{js}, t_{ija}, \tau_i\}$*, city group-wise populations* $\{\bar{L}_g\}$ *and model parameters* $\{\bar{h}, \beta, \alpha, p_a, \kappa, \theta_g, T_g, \eta_g, \alpha_{sg}, \sigma_D, \sigma, \mu_A, \mu_U\}$*, an equilibrium is defined as a vector of endogenous objects* $\{L_{Riag}, L_{Fjg}, w_{jg}, r_{Ri}, r_{Fi}, \vartheta_i, \bar{U}_g, \pi\}$ *such that*

1. **Labor Market Clearing** *The supply of labor by individuals (7) is consistent with demand for labor by firms (9);*
2. **Floorspace Market Clearing** *The market for residential floorspace clears (11) and its price is consistent with residential populations (6), the market for commercial floorspace clears (12) and floorspace shares are consistent with land owner optimality (13);*
3. **Closed City** *Populations add up to the city total, i.e.* $\bar{L}_g = \sum_{i,a} L_{Riag}\ \forall g$*, and rents are redistributed lump sum to residents.*

4 Empirical Analysis and Aggregate Effects

This section turns to a reduced-form analysis of how TransMilenio reshaped the organization of economic activity in Bogotá. To guide this analysis, I use the insight that a special case of the model delivers a log-linear reduced form between CMA and endogenous variables. The coefficients of these regressions are in fact sufficient statistics to analyze the impact of transit on the distribution of economic activity across the city and aggregate welfare. This approach has appeal in that it can speak to the BRT's aggregate effects, as its parameters can be transparently estimated via reduced form and because it holds for a broad class of models. I then turn to estimating the full model in Section 5 to measure its distributional impacts, which is of specific interest in this paper.

4.1 Reduced Form in a Special Case of the Model

Proposition 1 in Appendix C.2 shows that when there is one group of workers and firms and no fixed elements of expenditure or income, the equilibrium can be written as

$$\ln \hat{\mathbf{y}}_{Ri} = \boldsymbol{\beta}_R \ln \hat{\Phi}_{Ri} + e_{Ri} \tag{16}$$

$$\ln \hat{\mathbf{y}}_{Fi} = \boldsymbol{\beta}_F \ln \hat{\Phi}_{Fi} + e_{Fi}. \tag{17}$$

The outcome variables $\hat{\mathbf{y}}_{Ri} = [\hat{L}_{Ri}, \hat{r}_{Ri}]$ and $\hat{\mathbf{y}}_{Fi} = [\hat{L}_{Fi}, \hat{r}_{Fi}]$ are changes in residential and commercial

[21] TransMilenio may directly impact productivities $\bar{A}_{js}$ and amenities $\bar{u}_{iag}$ (e.g. through street improvements, crime, pollution). My empirical results which control for distance to station (which captures such effects) show little effect on CMA coefficients, motivating their exclusion from the model. Since Ahlfeldt et. al. (2015) find amenity and productivity spillovers decay rapidly across a few blocks, and the median tract contains 14 blocks, I omit cross-tract effects.

outcomes, consisting of residential population $\hat{L}_{Ri}$, residential floorspace prices $\hat{r}_{Ri}$, employment $\hat{L}_{Fi}$ and commercial floorspace prices $\hat{r}_{Fi}$. The right-hand side variables $\hat{\Phi}_{Ri}$ and $\hat{\Phi}_{Fi}$ are changes in CMA. The coefficients $\boldsymbol{\beta}_R = [\beta_{L_R}, \beta_{r_R}]$ and $\boldsymbol{\beta}_F = [\beta_{L_F}, \beta_{r_F}]$ reflect both the direct and indirect effects of improving CMA as it filters through land and labor markets. Finally, the residuals contain clusters of unobserved location characteristics that are exogenous to economic activity. For residential outcomes, e_{Ri} contains changes in amenities and residential floorspace supplies while for commercial outcomes, e_{Fi} contains changes in productivities and commercial floorspace supplies.[22]

This system shows that the transit network only matters for equilibrium outcomes through the two CMA variables. In fact, the change in the entire distribution of economic activity across the city only depends on the change in CMA and on a structural residual that reflects changing exogenous location fundamentals (productivities, amenities and floorspace supplies).[23] Proposition 1 also shows that this specification is shared by a broad class of urban models that include housing supply, firm mobility, capital as a productive input, and leisure in utility. Moreover, it shows that (i) the change in CMA terms and (ii) the reduced-form elasticities of outcomes to CMA are sufficient statistics for the *relative* change in economic activity across the city in response to changes in transit infrastructure. The overall *level* of the changes are pinned down by an assumption on population mobility into the city from the rest of the country, as well as by values for two parameters σ and β that cannot be estimated from the reduced form. These must be specified in some other way by the researcher, for example by calibrating them to external values or aggregate moments.

The CMA terms can be easily recovered (to scale) using data on residential populations, employment, commute costs d_{ij}, and the commuting elasticity θ from the following system of equations[24]

$$\Phi_{Ri} = \sum_j d_{ij}^{-\theta} \frac{L_{Fj}}{\Phi_{Fj}} \tag{18}$$

$$\Phi_{Fj} = \sum_i d_{ij}^{-\theta} \frac{L_{Ri}}{\Phi_{Ri}}. \tag{19}$$

RCMA reflects access to well-paid jobs. It is greater when a location is close (in terms of having low commute costs) to other locations with high employment, particularly so when these other locations lack access to workers (increasing the wages that firms there are willing to pay). FCMA reflects access to workers through the commuting network. It is greater when a location is close to other locations with high residential populations, particularly so when these other locations lack access to jobs (lowering the wages that individuals are willing to work for there).

[22]As discussed in Appendix C.2, there is one first order approximation $\tilde{\Phi}_{Fi} \approx \Phi_{Fi}^{\frac{\theta-1}{\theta}}$ involved to get this reduced form. The reduced form this approximation comes from involves $\Phi_{Ri}, \Phi_{Fi}, \tilde{\Phi}_{Fi}$ on the right hand side and holds exactly. Since $\Phi_{Fi}, \tilde{\Phi}_{Fi}$ are very highly correlated in the data (correlation coefficient of 0.98), this is not empirically feasible to implement and so the approximation is used to generate this simpler reduced form. The coefficients can be mapped between the two, however, so that the unapproximated version is used to conduct counterfactuals.

[23]The contents of the residual and reduced form parameters are outlined in Appendix C.7. All contents in the residual are exogenous to the model, except two equilibrium constants (average utility and total expenditure) that are absorbed into the regression constant.

[24]In this single-group model, there is a single commute elasticity θ.

I now turn to the specific context of Bogotá to visualize the change in CMA and how it differs from the distance-based measures of treatment effects commonly found in the literature. Figure 1 plots the distribution of changes in CMA induced by the construction of the first two phases of the system.[25] The system increased access to jobs much more in the outskirts of the city, which were far from the high-employment densities in the center. Firms' access to workers rose more in the center, where firms stood to benefit from the increased labor supply along all spokes of the network.

4.2 Measuring CMA

Computing changes in CMA induced by TransMilenio requires values for θ and $d_{ij} = \exp(\kappa t_{ij})$.

Identifying κ, λ, b_m. The mode choice parameters are estimated via maximum likelihood using standard expressions for choice shares in the nested logit model from Section 3.1 (see Appendix C.3). The data come from the 2015 Mobility Survey when all modes are available. κ is identified from the sensitivity of choices to differences in travel time across options, λ is identified from the differential sensitivity within public modes, and the preference shifters b_m are identified from differences in choice shares conditional on observed travel times. The results are in Panel A of Table 1. The estimate of $\kappa = 0.011$ is very close to the 0.01 reported in Ahlfeldt et. al. (2015). The value $\lambda = 0.157$ indicates a sizable correlation of draws within the public nest. Conditional on travel time, cars are most attractive, followed by buses and TransMilenio. That TransMilenio is least desirable likely reflects high crowds on the system, and the inconvenience of having to walk between stations and final origins and destinations. With these parameters, t_{ij0} and t_{ij1} can be obtained from (2) and (3). However, in the simple model considered in this section there is no car ownership. As described in Appendix C.4, average commute times t_{ij} are computed by assuming residents become car owners according to a Bernoulli distribution, with probability equal to the share of car owners in Bogotá.

Identifying θ. As shown in Appendix C.4, the special case of the model yields a simple gravity equation that identifies the parameter cluster $\theta\kappa$ from the sensitivity of the change in commute flows to the change in commute times between any two locations, controlling for origin and destination fixed effects. Estimating this relationship via PPML to account for zeros in the data yields a value of $\theta = 3.398$ in Panel B of Table 1, similar to existing estimates (Monte et. al. 2018; Heblich et. al. 2020). I use this as the baseline value, but use alternative values in robustness checks (see Appendix C.4).

4.3 Empirical Results

Identifying TransMilenio's Reduced Form Effect on Economic Activity. The regressions (16) and (17) may be biased if Bogotá's government chose routes in a way that targeted neighborhoods with differential trends in unobserved characteristics (such as if trying to stimulate lagging regions or to

[25] The figure plots the change in CMA holding population and employment fixed at their initial level in 1993 and 1990 respectively and changing only commute costs. This isolates the change due only to TransMilenio (discussed in Section 4.3). FCMA increases toward the center-North due to the high density of (low-skilled) workers in the South.

support thriving ones). Instead of leaning on a single approach, I exploit a variety of TransMilenio's institutional features to establish its causal impact on Bogotá's structure.

First, I include a rich set of controls, including locality fixed effects to (partially) control for changes in unobservables. Second, I use variation in CMA induced by changes in the network more than 1.5km from a location, which is less likely to be correlated with changes in surrounding unobservables. Third, I condition on distance to closest TransMilenio station to assess whether the effects are driven by changes in accessibility rather than by other station features (e.g. changes in foot traffic, pollution or complementary infrastructure). Fourth, I digitize four different historical plans for Bogotá's transit network and run specifications including both the realized change in CMA and the change induced by these (hypothetical) planned networks. The coefficients on the planned CMA variables can be interpreted as a placebo check that the planned-but-unbuilt locations do not grow differentially in the absence of new transit. The stability of the coefficients on the realized CMA variables addresses any omitted variable bias (not captured by the controls) that can arise from a location's non-random exposure to transport infrastructure (Borusyak and Hull 2021). Fifth, I exploit TransMilenio's staggered rollout across three phases by using event studies and falsification tests which assess whether there is growth in outcomes prior to line openings. Sixth, I construct cost-shifting instruments to predict TransMilenio's routes based on a historical tram network and on engineering estimates of the cost to build BRT on different types of land.

An additional challenge is that changes in CMA contain population and employment in both periods. Since productivity and amenity shocks that determine residential population and employment are contained in the error terms, they will be mechanically correlated with changes in CMA. I thus construct versions of the change in CMA by solving (18) and (19) while holding population and employment fixed at their initial levels, allowing only commute costs to change, and use these throughout the empirical analysis. This isolates the variation in CMA due only to changing commute costs through TransMilenio's construction. After solving for the CMA terms, I construct the change in CMA for a given location by excluding the location itself in the summation. This addresses the possibility that changes in unobservables may be correlated with a location's initial level of economic activity. The main specifications use these CMA measures as regressors, but later in this section I use these to instrument for the "realized" change in CMA.

Main Specification. Table 2 presents the baseline results. Each entry corresponds to the coefficient from a regression of the change in each outcome on the change in RCMA or FCMA in each census tract. Since the data do not all line up, each specification relies on changes over different periods. However, the changes in CMA are always measured using changes in commute times due to TransMilenio routes constructed between the two periods over which the outcome is measured.[26] Es-

[26] In population regressions, the outcome is the log change in residential population between 1993 and 2018. The change in CMA is that induced by all three phases of TransMilenio, holding residential population and employment fixed at their levels in 1993 and 1990 respectively. In land market regressions, outcomes are log changes between 2018 and 2000 and the change in CMA is that induced by all three phases holding residential population and employment fixed at their levels in 2000 (population in 2000 is a linear interpolation from the 1993 and 2005 census; employment is from the 2000 CCB data). Establishment regressions regress changes between 2000 and 2015 from the CCB data against the same CMA measures

tablishment regressions are weighted by the share of establishments in 2000 in each tract to increase precision.[27] Since some establishment results are noisy, I include the share of floorspace used for commercial purposes as an outcome to provide supplemental evidence for TransMilenio's impact on the reallocation of employment.[28]

Column (1) includes controls for locality fixed effects, basic tract characteristics, and log distance to CBD interacted with region dummies.[29] Changes in CMA due to TransMilenio have strong, positive impacts on all outcomes. These relationships remain mostly stable as more controls are added in columns (2) and (3), sometimes becoming sharper. The exception is log establishments in the final row, whose coefficient falls by a third with the full set of controls. I consider column (3) to be the baseline specification continued in later tables, as it includes the full set of controls.

Column (4) excludes tracts that are closer than 500m from an endpoint of a TransMilenio route (a "portal") or the CBD. The intent of the government was to connect outlying neighborhoods with the CBD, so the location of these portals may have been endogenous to underlying trends in local economic activity. The coefficients remain largely stable in this subsample of tracts, suggesting that endogeneity in the locations directly targeted by TransMilenio is not driving the results.

Column (5) uses the change in CMA to locations farther than 1.5km away from a tract. Network additions at this distance from a tract are less likely to be linked to local trends in unobservables. The results remain robust and, for the most part, stable. Column (6) assumes users take the quickest mode of public transit available, and shows the results are robust to alternative forms of aggregation.

Lastly, column (7) conditions on distance to stations to establish that the effects are primarily driven by changes in accessibility rather than by station features (e.g. changes in foot traffic, pollution or complementary infrastructure). This finding supports the model's emphasis on accessibility.

Visualizing the Relationship. Figure 2 plots the non-parametric relationship between residual growth in outcomes and CMA. The relationship appears approximately log-linear for each outcome, as predicted by the model. The simple model seems to capture the heterogeneous effects observed in the data: tracts with large improvements in accessibility experience large growth in outcomes.

Hypothetical Changes in CMA from Historical Network Plans. The location of the TransMilenio network was not random. The government may have located the network to support or spur existing local trends in economic activity. To provide additional evidence of TransMilenio's causal impact, I leverage four distinct historical plans for a transit network digitized from planning documents.

as the land market regressions. This is preferred to the census employment data since it gives employment 10 years to respond to TransMilenio. Table 4 uses employment data from the census to examine the impact of TransMilenio lines built during phase 1 (by 2003) on employment growth between 1990 and 2005.

[27] Unweighted regressions are presented in Table A.5.

[28] In the simple model this should not change. An extension in Appendix E.3 allows for floorspace use shares to respond to TransMilenio (with total floorspace supply held constant, as observed in the data), which delivers heterogeneous elasticities of economic activity to CMA. The model with endogenous housing supply in Appendix C.6 also allows for endogenous floorspace use shares (via changes in relative supplies for residential and commercial floorspace) and admits the same reduced form as the baseline model.

[29] The North is richer and more educated than the West and South, so this allows for differential growth further away from the city center within each region.

Since 1980, multiple administrations had worked on proposals for building a subway or metro system in Bogotá. Four distinct plans for the network were prepared before Mayor Peñalosa agreed with the proposal by JICA to build a BRT, given that the cost of a subway would have been "ten times higher than the alternative of articulated buses". I obtained and digitized the maps for these four planned networks, shown in Figure A.3.[30] I then solve for the predicted change in CMA had TransMilenio been been built along each of the planned networks, and compute the average change in log RCMA and FCMA across all four plans.

The baseline specification (column 3 of Table 2) is then extended to include these expected changes in CMA under the plans as additional regressors. One interpretation of the results is as a placebo check. If the observed impacts are due to TransMilenio itself rather than to the selection of routes based on trends in unobservables in adjacent neighborhoods, there should be no impact of these planned-but-unbuilt networks. A second interpretation is that this controls for the omitted variable bias that can arise from a location's non-random exposure to transport infrastructure, as highlighted by Borusyak and Hull (2021). The idea is that some locations may receive systemically different changes in accessibility under any network realization. For example, central neighborhoods will tend to have greater increases in FCMA since they are close to where workers live by virtue of their central location. Identification requires that these "on average" more exposed locations do not differ in their trends in unobservables. While any such trends may already be controlled for by the rich fixed effects and controls used in this paper, controlling for the average change in CMA under these counterfactual networks conducts the exact "recentering" shown by the authors to remove the omitted variable bias. If the controls already capture any differential trend in unobservables in more "on average" exposed locations, then the coefficient on the realized CMA terms should be invariant to the inclusion of the expected change in CMA and the coefficient on the latter should be zero.

Table 3 presents the results. Column (1) repeats the baseline specification, while column (2) adds the control for the expected change in CMA across the four plans. In each case, the coefficient on the realized change in CMA due to TransMilenio is invariant to the inclusion of this additional control. The p-value testing for equality of coefficients on the realized CMA variable across both columns ranges from 0.24 to 0.96. The coefficient on the planned CMA variable is statistically insignificant from zero in all specifications. These results suggest two things: first, that the observed impacts of TransMilenio are unlikely due to pre-existing trends in neighborhoods selected by city planners, and second, that the existing set of controls does a sufficient job in controlling for any omitted variable bias that could arise from non-random exposure to the network.

Staggered Station Openings. TransMilenio was opened in three distinct phases during the 2000s and 2010s.[31] This section runs a set of falsification checks to test for changes in outcomes prior to the

[30] See ("Historia de TransMilenio") for the quote, and Alcaldía Mayor de Bogotá D.C. (2009) for the network maps. As discussed in Appendix F.3, I add predicted feeder routes under these networks by placing a 2km radius disk around each end point of the planned lines connecting the two with 8 "spokes", and create stops every 250m.

[31] Phase 1 consisted of 3 lines in 2000, and 1 line each in 2001, 2002 and 2003. Each year consisted of 47%, 26%, 6% and 21% of the stations opened in phase 1, respectively. Phase 2 consisted of 2 lines in 2005 and 1 line in 2006, with each year accounting for roughly half the stations opened in this phase. Lastly, phase 3 consisted of 2 lines in 2012 and 1 in 2013 with

opening of stations in later phases. The specification is

$$\Delta_{t,t-\ell} \ln y_i = \beta^{Current} \Delta_{t,t-\ell} \ln \Phi_i + \beta^{Future} \Delta_{t+k,t} \ln \Phi_i + \gamma' X_i + \varepsilon_i$$

The outcome is the growth in a variable, y_i, between two periods, t and $t-\ell$ (e.g. 2006 and 2000). This is regressed on (i) the change in CMA between t and $t-\ell$, (ii) the change in future CMA between $t+k$ and t (e.g. 2015 and 2006), as well as the same set of controls as the baseline specification. If there is no growth in outcomes prior to TransMilenio being built, the coefficient β^{Future} should be zero.

The time periods are chosen to best line up with the available data and the opening of TransMilenio lines. Since the openings of phases 1 and 2 are spread out between 2000 and 2006 (with every year except 2004 experiencing station openings), I focus the analysis on phase 3, which opened in 2012 and 2013. For land market outcomes, the change in outcomes is measured between 2008 and 2000. The right-hand side variables include CMA growth due to (i) phases 1 and 2 of the system open by 2006 (to identify $\beta^{Current}$) and (ii) phase 3 of the system open by 2013 (to identify β^{Future}). While prices may experience some anticipation effects, plans for phase 3 were mired by uncertainty and delays, with construction only beginning in late 2009. For residential population, the change is measured between the 1993 and 2005 censuses. The right-hand side variables include CMA growth due to phase 1 (open by 2003, with most stations opening by 2001), as well as the change in CMA due to phases 2 and 3. Lastly, for employment, I turn to the measures from the economic census rather than the CCB data. While the latter is available only in 2000 and 2015 (bookending the entire network construction), the economic census is available in 1990 and 2005. This permits me to separately examine the impacts of changes in CMA due to phase 1 versus phases 2 and 3, similar to residential population.[32]

In Table 4, Panel A presents the results for residential population and residential floorspace prices. Odd columns repeat the baseline specification but with outcomes measured over this different period (e.g. 1993 to 2005 for residential population, compared to 1993 to 2018 in the baseline results). The positive relationships remain significant, although the point estimates are somewhat attenuated. This might be expected given that there is less time for outcomes to respond to the change in CMA than in the baseline specification. Even columns then run the specification above. They maintain a significant relationship between outcome growth and CMA growth due to lines constructed over the period, but an insignificant impact due to accessibility from future lines. While insignificant, these estimates of β^{Future} can be noisy. Panel B finds similar patterns for commercial land market outcomes.

Panel C presents the impact on total and formal employment from the economic census.[33] In the odd columns, which regress on realized changes in CMA, I document positive but noisy coefficients (p-values of 0.123 and 0.118). These estimates are larger than the baseline estimates using the CCB data, but the difference is statistically insignificant given the imprecision of the estimates. The even

2012 representing 84% of the openings

[32]The downside of the economic census data is that there is less time for employment to adjust: on average across lines opening during phase 1, there are just under 4 years between the opening year and the 2005 economic census. This compares to 10.5 years between the average opening year and the 2015 CCB.

[33]Formal employment is defined as employment in establishments with 5 or more workers.

columns add in future CMA growth, which is statistically insignificant in both cases.

Floorspace Price Event Study. I leverage the annual cadastral data to examine more granular house price dynamics prior to the opening of TransMilenio's third phase. I run regressions of the form

$$\ln r_{Rit} = \alpha_i + \gamma_{\ell(i)t} + \sum_{\tau=-8}^{\tau=6} \beta^{\tau} \Delta_{12,06} \ln \Phi_{Ri} + \delta_t' X_i + \varepsilon_{it},$$

where α_i are tract fixed effects, $\gamma_{\ell(i)t}$ are locality-year fixed effects, and $\delta_t' X_i$ is a set of controls with time-varying coefficients. The controls include those from the baseline specification, but add the change in CMA due to the first two phases of the system to capture the impact of changes that these earlier lines had on house prices that is correlated with the change due to phase 3. The regression is weighted by initial floorspace price in 2000 to improve precision. The β^{τ} coefficients capture the response of residential floorspace prices in a tract τ years from the third phase lines opening to the change in CMA due to the lines that open during this phase.

Figure 3 plots the event study coefficients. Reassuringly, the change in CMA induced by the network expansion in phase 3 has no impact on floorspace price growth to the line openings. It is not clear ex ante that this would be the case. Prices could rise due to anticipatory effects as expectations around whether and where the line would open firm up. Alternatively they could fall due to the disamenities surrounding the construction from late 2009 through 2012. In fact, consistent with this possibility, there is a mild decrease in house prices in tracts that experienced a larger growth in accessibility due to phase 3 in the two or three years prior to opening. The year before the lines open, the responsiveness of prices to CMA jumps approximately 0.4 log points. This is potentially due to anticipation effects as the opening of the third phase became certain.[34] This effect is stable until two years after opening, after which the elasticity rises 1 log point until six years after opening.

While the difference between the short- and medium-run effects in this event study may reflect the multiplier effect due to the reallocation of population and employment to treated areas, it is important to note that (part of) this could also be due to the way the data is constructed. As described in Appendix F, part of the annual change in prices in the cadastral database is based on inflating prior years' values. Primary data are collected by the cadastral office to fully reassess properties—based on collecting information on properties for sale, making offers to elicit true sales values, and having in person visits by professional assessors—but this happens fairly infrequently (around three times over the period in question). This motivates the focus on long-run impacts in the rest of the paper.

Instrumental Variables to Predict TransMilenio's Placement. Lastly, I construct two cost-shifting instruments for TransMilenio routes. These in turn imply two instruments for the change in CMA.[35]

[34] Corruption cases surrounding the construction of the third phase had added to construction delays, which may have brought more uncertainty than usual to whether and when the lines would actually open.

[35] Additional details can be found in Appendix F.3. To compute the instruments, I first calculate the commute times had the system been built along each instrument. Plugging these into (18) and (19) and continuing to hold population and employment fixed at their initial level, I obtain the predicted CMA had TransMilenio been built along these routes. My instrument for the change in CMA is then the difference between this predicted CMA under TransMilenio and its value

The first takes as given the government's overall strategy of connecting portals at the edge of the city with the CBD, excludes those areas from the analysis, and constructs the routes that would have been built if the sole aim had been to minimize costs. This is done by using engineering estimates to compute the cost to build BRT in each parcel of land in Bogotá based on its land use in 1980. This is a valid instrument when these least-cost routes predict TransMilenio's placement but are uncorrelated with trends in unobserved amenities and productivities (conditional on controls).

The second instrument exploits the location of a tram system that opened in 1884, which was last extended in 1921 and stopped operating in 1951. I extend the 1921 lines to the present edge of the city to improve predictive fit, given the city's substantial expansion over the period. The tram was built along wide arterial roads, which are cheaper to convert to BRT than narrow ones. The tram may have had persistent direct effects on trends in unobservables that lasted well after its construction, which I capture by including historical controls. Conditional on these historical variables, the tram routes should be uncorrelated with changes in productivities and amenities between 2000 and 2012 to the extent that these were unanticipated by city planners in 1921.

The identification assumption is that the instruments have only an indirect effect on outcome growth through the predicted change in CMA. One concern is that features that make a location cheaper to build BRT, such as proximity to a main road, can have direct effects on outcomes. A key advantage of my approach is that I can control for distance to these features (distance to the tram, distance to main roads) and use only residual variation in predicted CMA growth for identification.

Table 5 presents the results. Column 1 reproduces the baseline results for reference. Column 2 shows the results are very similar when instrumenting the realized change in CMA (allowing residence and employment to change across periods and summing over all locations) with the measure from the baseline specification. Columns 3 and 4 instrument the realized change in CMA using the average change across the tram and least-cost path instruments, using either all tracts except the tract itself (column 3) or only tracts 1.5km away (column 4). The coefficients are mostly stable across these specifications, with the exception of residential floorspace prices which roughly double when moving to the final column. The population and commercial floorspace price coefficients are imprecise, but sharpen in the last column. Taken together, these specifications support the impression from the analyses above—that changes in CMA due to TransMilenio seem unrelated to trends in unobservables conditional on the rich set of controls. Given the broad stability of the estimates across specifications, I use the coefficients from the baseline specification in the next section but explore the robustness of the results to using the elasticities from the other columns in this table (see Table A.2).

Robustness Checks and Additional Results. Appendix G.6 presents robustness of these results to alternative i) methods of aggregating times; ii) commute elasticities; iii) clustering of standard errors; iv) additional controls and sample selection criterion and v) weighting procedures. Appendix G.4 provides evidence that TransMilenio increased wages but also led to a sorting response where the high-skilled moved into neighborhoods with improved market access. This is consistent with the

in the initial period without the system. Historical and least cost instruments are often used in the literature (Baum-Snow 2007; Duranton and Turner 2012; Faber 2014, Alder 2019).

model's Stone-Geary preferences, since the rich are more likely to move into appreciating neighborhoods, given that they spend a smaller fraction of their income on housing.

4.4 Aggregate Effects from Reduced Form Sufficient Statistics

Table 6 measures TransMilenio's aggregate effects by using the estimated reduced-form elasticities to implement the sufficient statistics approach outlined in Proposition 1.

First Order vs General Equilibrium Welfare Impacts. The standard approach to evaluate the gains from transit infrastructure is based on the "value of travel time savings" (e.g. Small and Verhoef 2007). Despite the rich channels captured in the general equilibrium model, Proposition 2 in Appendix C.5 shows that when the equilibrium is efficient, an application of the envelope theorem implies that this is precisely the first order welfare impact from a change in infrastructure.

Panel A of Table 6 simulates what Bogotá would have looked like in 2018 without TransMilenio, and then adds it back in under the different approaches.[36] The first column reports TransMilenio's gains under the first order approximation or VTTS approach from Proposition 2. This delivers a welfare increase of 1.26%, accruing solely through time savings. The second column shows the welfare gains using the full model from Proposition 1 and the estimated elasticities. These deliver a much larger gain of 2.34%. The VTTS thus accounts for only 54% of the total welfare gains, yielding one of the paper's central results—that equilibrium effects matter for valuing the gains from new transit infrastructure in cities. Confidence intervals for these main welfare effect are reported from a bootstrap procedure that accounts for the uncertainty in the model's parameter estimates (see Appendix C.9 for details). While there is meaningful uncertainty surrounding these estimates, I can reject the null that the fraction of welfare gains accounted for by the VTTS is not less than one (p-value of 0.04). The difference between the equilibrium and the first order welfare effects could be due either to the size of the shock (since the approximation may perform poorly for large shocks) or deviation from efficiency (due to amenity and productivity externalities). The final column shows that when externalities are turned off, the VTTS explains a larger portion of the equilibrium effects. The size of the shock explains about one third of the 46% gap between the VTTS and the general equilibrium welfare effect, with the externalities accounting for the remaining two thirds.

Aggregate Effects. Panel B presents TransMilenio's impact on aggregate outcomes using the results from Proposition 1.[37] Doing so requires values for α, β and σ in addition to the CMA elasticities. I estimate the $1 - \alpha = 0.206$ by computing the share of floorspace in total costs across establishments

[36] I refer to the "2018 equilibrium" as the post-TransMilenio equilibrium. Population data and land market data come from 2018, employment data from 2015, land market data come from 2018, and the TransMilenio network includes all phases. Since there may be multiple equilibrium in the presence of externalities, the selection rule used is to start the algorithm from the observed equilibrium when solving for counterfactual equilibria. This can be rationalized through path dependence in a dynamic model of a city.

[37] The percentage change in each variable under the counterfactual without the TransMilenio network are reported, i.e. $100 \times (X_{2015}^{NoTM}/X_{2015} - 1)$ for any variable X_{2015}. Numbers may therefore differ from Panel A which inverts the ordering by using the equilibrium without TransMilenio as the base. Table A.2 presents robustness of the welfare results in row 1 to alternative parameter values.

in each one digit non-agricultural industry, and averaging these by the sectoral employment shares in Bogotá.[38] I estimate $1-\beta = 0.274$ from the average expenditure share of housing in Bogotá. Lastly I set $\sigma = 6$ close to median estimates from Feenstra et. al. (2018), but vary this in robustness checks.

The first column presents the closed city results from the model developed above. The second column presents results from an extension outlined in Appendix E.1, which allows for an upward-sloping supply of migrants into the city from the rest of the country. There are large aggregate impacts on welfare and city output under either mobility assumption. Without migration into Bogotá, GDP and welfare rise by 2.98% and 2.21% respetively, with a slight fall in the level of floorspace prices.[39] With migration, the welfare gain falls to 0.6% since the increase in population of 9.51% bid up floorspace values by 5.28%. GDP rises by 12.71%, but this is mostly due to population growth: GDP per capita rises by 5.67%. The final two rows show that TransMilenio can account for between 2.96% and 13.36% of Bogotá's GDP growth from 2000 to 2016, and up to 34.9% of observed population growth. TransMilenio's effects are quantitatively important, but not implausibly large. The third row shows that it was also a profitable investment for the city, leading to an increase of at least 2.5% in the steady state level of GDP net of construction and operating costs (see Appendix F.4 for details).

Incorporating Congestion. While speeds for cars and other types of buses did not change on routes adjacent to TransMilenio (see Appendix G), the BRT could have had aggregate effects on road speeds that do not appear in a difference-in-difference specification. Appendix E.2 extends the baseline model to gauge the impact of the BRT in the presence of congestion. The extension blends elements from Allen and Arkolakis (2021) and Gaduh et. al. (2022). The "economic module" of the model is unchanged: the same system of equations governs the response of economic activity to a change in commute times. However, a new "traffic module" is added that allows the change in commute times to depend on both new physical infrastructure and any changes in commuting patterns via congestion. The result is one combined system of equations where the change in economic activity and commute times are jointly determined in response to new infrastructure.

The results are shown in Panel C of Table 6, using the congestion elasticity of 0.06 estimated for Bogotá by Akbar and Duranton (2017). The first two rows show TransMilenio's welfare effect in the model with and without congestion.[40] Allowing for congestion leads to a larger welfare gain: as some commuters substitute away from cars onto the BRT, roads become less congested and driving times fall. This effect is small, however, with a welfare increase that is only 0.55% larger than without congestion. The last row assesses the welfare impact had the TransMilenio lanes been used to add new car instead of BRT lanes. The welfare effects are tiny in comparison: the welfare change would have been only 0.6% of the gains caused by TransMilenio. Overall, these results suggest that the baseline welfare effects provide a lower bound of the BRT system's impact in the presence of congestion.

[38]The data on cost shares comes from the Encuesta Anual Manufacturera, Encuesta Anual de Servicios and the Encuesta Anual de Comercio in 2010. The sectoral employment shares are averages from 2000-2015 from the GEIH and ECH.

[39]It is typical to have little change in the overall level of house prices in closed city models with fixed housing supply since the supply of housing and overall population is fixed.

[40]The welfare effect in this model with the congestion elasticity set to zero differs from the baseline model, due to the different formulation of constructing commute costs.

5 Distributional Effects

While the sufficient statistics approach from the previous section can speak to the BRT's aggregate effects, it is silent on the distribution of these impacts across worker groups. This section therefore estimates the full model from Section 3 to answer this question.[41]

5.1 Parameter Estimation

5.1.1 Parameters Estimated without Solving the Model

Externally Calibrated Parameters $\{\sigma, \sigma_D\}$ I set the elasticity of substitution between labor skill groups to $\sigma = 1/0.7$ based on the review in Card (2009), and $\sigma_D = 6$ as described in Section 4.4.[42]

Share Parameters $\{\alpha_s, \beta, \alpha_{sg}\}$ I estimate $1 - \alpha_s = 0.206$ as described in Section 4.4 using data on the share of floorspace in total production costs, and set this to be equal across industries. I estimate $1 - \beta = 0.24$ to match Bogotá's long-run housing expenditure share.[43] As described in Appendix D.2, the labor shares α_{sg} are calibrated to match the relative wage bill for college-educated workers in each industry.

Commute Costs and Elasticity The estimates κ, λ, b_m were provided in Table 1, delivering commute times t_{ija} under each car ownership status from (2) and (3). The commute elasticities θ_g can be estimated by taking logs and first differences of the expression for commute flows (4) to yield

$$\Delta \ln \pi_{j|iag} = \gamma_{iag} + \delta_{jg} - \theta_g \kappa \Delta t_{ija} + \varepsilon_{ijag},$$

where γ_{iag} and δ_{jg} are fixed effects and ε_{ijag} is an unobserved component of commute costs. Given a value of κ, θ_g is identified off the sensitivity of changes in commute flows to changes in commute times induced by TransMilenio for each group.[44]

Table 7 presents the results. Across all specifications, high-skilled workers are found to have a lower θ_g (i.e. a higher dispersion of productivity shocks across locations) making their commute choices less sensitive to commute times. The overall magnitude and fact that more educated workers are estimated to have a greater dispersion of match-productivities line up with existing estimates (e.g. Lee 2020; Hsieh et. al. 2019; Galle et. al. 2022). Since the moment conditions in the following section use the instruments for TransMilenio's placement, I use the IV estimates in column 2 as preferred estimates, which yield values of $\theta_H = 2.655$ and $\theta_L = 3.98$. I explore the robustness of the results to using the OLS or PPML estimates in Table A.4.

[41] A list of all parameters and sources of variation used to identify them is provided in Appendix A.3.

[42] Estimating σ within the model would require a shock to relative labor supply and wage data by location of employment across space to measure the response in relative wages. Since the latter are not available, I calibrate σ instead and conduct robustness to both σ and σ_D in Table A.4.

[43] See the Engel curves presented in Appendix G.

[44] Another option here would have been to allow κ to vary by group. While this wouldn't matter for this particular moment of the sensitivity of commute flows to commute times, allowing for this possibility in estimating the mode choice model led to $\kappa_H = 0.0126$ (0.0062) and $\kappa_L = 0.0113$ (0.0054). Since these are statistically indistinguishable from each other (p-value for a test of equality of 0.22), I use the assumption of common κ across groups.

5.1.2 Parameters Estimated Solving the Full Model

It remains to estimate the parameters $\{\bar{h}, p_a, T_g, \eta_g, \mu_A, \mu_{U,g}\}$. Appendix D.3 shows how, given prior parameter estimates, there is a vector $\{\bar{h}, p_a, T_g\}$ that matches the average expenditure share on housing, the average expenditure on cars, and the college wage premium, respectively.

The residential supply elasticity η_g and the spillover parameters $\mu_A, \mu_{U,g}$ are estimated via GMM. The intuition for identification is very similar to that of the sufficient statistics approach of Section 4. TransMilenio provides a shock to the attractiveness of each residential neighborhood through increased RCMA. The response of residential inflows to this shock identifies the residential supply elasticity. The response of model-inferred amenities to the resultant change in neighborhood composition identifies the amenity spillover. TransMilenio also provides a shock to the supply of workers commuting to each employment location through increased FCMA. The response of model-inferred productivity to this change in employment identifies the productivity spillover.

Amenities Moments Taking logs of the expression for resident supply (6) delivers

$$\Delta \ln L_{Riag} = \eta_g \Delta \ln V_{iag} + \eta_g \mu_{U,g} \Delta \ln \frac{L_{RiH}}{L_{Ri}} + \gamma_{\ell,g} + \gamma'_{R,g} \text{Controls}_i + \Delta \ln \epsilon_{Riag} \tag{20}$$

where $\Delta \ln V_{iag} \equiv \Delta \ln \tilde{y}_{iag} - (1-\beta)\Delta \ln r_{Ri}$, $\gamma_{\ell g}$ are locality-group fixed effects, and Controls_i denote tract characteristics (that can have separate effects by group) used to partially control for changes in unobserved amenities. $\Delta \ln \epsilon_{Riag}$ reflects residual variation in unobserved amenity growth.

The residential supply elasticity η_g is identified off the responsiveness of residential populations to exogenous variation in the common utility from living in a location $\Delta \ln V_{iag}$. This comes from my instruments for RCMA, which I use to construct predicted change in net income using the instrument for TransMilenio to generate expected income in the post-period.[45] Let $\Delta \ln \tilde{\Phi}^{IV}_{Riag}$ denote the expected growth in net income growth averaged across the LCP and Tram instruments (as in Table 5).

Identification of the spillovers $\mu_{U,g}$ requires exogenous variation in a neighborhood's college share. I use two instruments to this end. First, tracts that experience a greater growth in CMA to high-skilled jobs relative to low-skilled jobs should experience a larger increase in the share of college residents. This is captured by $Z_{Diff,i} = \Delta \ln \bar{\tilde{\Phi}}^{IV}_{RiH} - \Delta \ln \bar{\tilde{\Phi}}^{IV}_{RiL}$ where $\bar{X}_i = \sum_a X_{ia}$. Second, tracts with expensive housing where CMA improves should see a greater rise in the college share. This comes directly from log-linearizing the expression for residential populations (6). Intuitively, poor low-skilled residents are less willing to pay for increased access to jobs in expensive neighborhoods due to their greater expenditure on housing.[46] I capture this by interacting the change for

[45] Letting $t-1$ and t reference the pre- and post-TM periods, adjusted RCMA is $\tilde{\Phi}_{Riag,t-1} \equiv T_{g,t-1}\Phi_{Riag,t-1}{}^{1/\theta_g} - p_{a,t-1}a + \pi_{t-1}$ and $\tilde{\Phi}^{IV}_{Riag,t} \equiv T_{g,t}(\Phi^{IV}_{Riag,t})^{1/\theta_g} - p_{a,t}a + \pi_t$. The change is simply $\Delta \ln \tilde{\Phi}^{IV,k}_{Riag} = \ln \tilde{\Phi}^{IV,k}_{Riag,t} - \ln \tilde{\Phi}_{Riag,t-1}$ for $k \in \{LCP, Tram\}$. Then $\Delta \ln \tilde{\Phi}^{IV}_{Riag} = E_k\left[\Delta \ln \tilde{\Phi}^{IV,k}_{Riag}\right]$.

[46] Log-linearizing the expression for residential populations (6) yields

$$\Delta \ln L_{Riag} \approx \mu^L_{iag} \frac{\eta_g}{\theta_g} \Delta \ln \Phi_{Riag} - \eta_g(1-\beta+\mu^R_{iag})\Delta \ln r_{Ri} + \epsilon_{iag}$$

where $\epsilon_{iag} \equiv a\mu^a_{iag}\Delta \ln p_a + \mu^\pi_{iag}\Delta \ln \pi + \eta_g \Delta \ln u_{iag}$. Here $\mu^L_{iag} \equiv T_g \Phi_{Ri}^{1/\theta_g}/\tilde{y}_{iag}$ and $\mu^R_{iag} = r_{Ri}\bar{h}/\tilde{y}_{iag}$ are the share of labor high-skilled residents with the house price in the initial period $Z_{Rents,i} = \Delta \ln \bar{\tilde{\Phi}}^{IV}_{RiH} \times \ln r^{2000}_{Ri}$.[47]

The moment conditions used to identify η_g and $\mu_{U,g}$ are therefore[48]

$$E\left[\Delta \ln \epsilon_{Riag} Z_{Riag}\right] = 0, \qquad Z_{Riag} \in \left\{\Delta \ln \tilde{\Phi}^{IV}_{Riag} \quad Z_{Diff,i} \quad Z_{Rents,i}\right\}.$$

Productivity Moments Composite productivity $A_{js} \propto W^{\alpha_s}_{js} r^{1-\alpha_s}_{Fj} X^{1/(\sigma_D - 1)}_{js}$ is the residual that ensures the model definition for sales holds. As shown in Proposition 3 in Appendix D.1, this can be recovered (to scale) using data on employment, residence, floorspace prices and commute costs. The model infers high productivity in locations where employment is high (reflected through high sales) relative to the observed price of commercial floorspace and the accessibility to workers through the commuting network (which determines wages). Taking logs of (14) and including a set of control variables to (partially) capture changing fundamentals yields

$$\Delta \ln A_{js} = \mu_A \Delta \ln \tilde{L}_{Fj} + \gamma_\ell + \gamma'_F \text{Controls}_j + \Delta \ln \epsilon_{Fjs}$$

where $\Delta \ln \epsilon_{Fjs}$ reflects residual variation in unobserved productivity growth.

The agglomeration elasticity is identified from the extent to which model-implied composite productivity depends on employment. Since employment will be correlated with unobserved components that make locations more productive, I use the instruments for FCMA growth as a labor supply shock. The moment conditions used to identify μ_A are therefore

$$E\left[\Delta \ln \epsilon_{Fis} Z_{Fig}\right] = 0, \qquad Z_{Fig} \in \left\{\Delta \ln \bar{\Phi}^{IV}_{FiL} \quad \Delta \ln \bar{\Phi}^{IV}_{FiH}\right\}.$$

Both sets of moments are stacked into a system of moment conditions which is estimated jointly in a single GMM estimation. I estimate standard errors via a block-bootstrap procedure, resampling at the tract-level to allow for arbitrary within-tract correlation in unobservables.[49]

GMM Results Table 8 presents the main results. The productivity externality of 0.242 is slightly larger than existing estimates, although it is slightly noisy (p-value of 0.058), and thus contains smaller values within its confidence intervals. This is also one of the first estimates outside of a developed country setting. The residential population elasticity is slightly larger for the high-skilled than the low-skilled. The spillovers for residential amenities are 0.730 and 1.002 for low- and high-skilled workers. While both groups value living near high-types, the college educated value it most.

income and fixed housing expenditure of total net income. Note that μ^R_{iag} is greater for poor individuals since they spend a greater fraction of income on housing. Thus, poor low-skilled workers are more sensitive to house price appreciation and are less willing to pay for improved CMA than the high-skilled.

[47] Controls for initial house prices are included to allow this characteristic to have its own impact on population growth; controls from the reduced form results are included and reported in Table 8.

[48] Orthogonality conditions with each control variable are also included. The time periods used for pre- and post-periods for each variable are the same as in the previous section, and use the full 2013 TransMilenio network.

[49] Bootstrapping is needed since units of observation vary across moment conditions, rendering the standard asymptotic variance formulas inapplicable. See Appendix D.4 for a benchmark of the amenity spillover estimates to Diamond (2016).

Model Validation The model's fit of two non-targeted moments provides additional confidence in its results. First, Figure A.4 plots the observed change in the share of floorspace used for residential purposes against that predicted by the model. While the two correlate well, the correct test of the model is not that the correlation or R2 is high but rather that the regression of the observed on the model-predicted changes has a slope of one.[50] The slope of this regression is 1.598 (0.822), which is statistically indistinguishable from one (p-value 0.47). Second, the model predicts that changes in income are related to RCMA through $d\ln \bar{y}_i = \frac{1}{\theta} d\ln \Phi_{Ri}$ with elasticity $1/\theta$. This regression is reported in column (3) of Table A.16. The coefficient of 0.522 (0.224) is statistically indistinguishable from $1/\theta$ after plugging in the estimate of $\theta = 3.398$ (p-value 0.31).

5.2 Results

Panel A of Table 9 presents the main result: welfare inequality increases by 0.55% as a result of TransMilenio. It should be noted that the confidence intervals convey uncertainty in this estimate, and the test of whether the high-skill gain more than the low-skill only has a p-value of 0.15. With this caveat in mind, I turn understanding the source of this result.

Why would the high-skilled benefit the most? Panel B decomposes the welfare gains, starting with a simplified case of the model and slowly adding its ingredients to isolate each one's impact.

The first row assumes workers share the same (average) value for η and θ and are perfect substitutes in production. This model is similar to the simple model used in Section 4.4 since it abstracts from heterogeneity across workers. Reassuringly, the average welfare effect of 2.19% is very close to the 2.335% reported from the sufficient statistics approach in Table 6. Low-skilled workers benefit the most, with inequality falling by 0.37%.

The second row allows workers to differ in their commuting elasticities. Recall that the high-skilled have a lower commute elasticity. This shifts the gains towards the high-skilled, with inequality now falling by 0.16%. A lower commuting elasticity tends to increase the incidence of high commute costs, since workers have very sticky preferences for workplace locations and are less able to substitute away to less costly options when transit is slow. The third row allows the residential choice parameters to equal their estimated values, with a modest reduction in the fall in inequality.

The last thing that changes as one moves to the result from the full model in Panel A is that workers are imperfect substitutes in production. Intuitively, the average welfare effect falls. For example, a large inflow of low-skilled workers will increase the supply of the labor bundle less than when both types are perfect substitutes. However, this also causes the sign of the impact on inequality to switch, with welfare inequality rising by 0.55%. This occurs for two reasons. First, high-skilled workers are now partially shielded from the reduction in wages due to the large labor supply shift of low-skilled workers who use public transit since they are no longer perfect substitutes. Second, it now matters whether each skill group is connected to locations where demand for their skill is

[50] Other shocks orthogonal to the model may cause the correlation or R2 between observed and model-predicted changes to fall. The model is trying to capture the counterfactual of how activity would have changed if the only shock had been the change in infrastructure. For discussion see Adao, Costinot and Donaldson (2023).

highest. For the geography of Bogotá and TransMilenio, this tends to benefit the high-skilled who are concentrated in the city's north which TransMilenio connected with the high skill-intensive industries in the center and center-north. Residence and employment for the low-skilled is more dispersed, so TransMilenio connected a smaller fraction of these workers with high-wage locations.

Overall, these results suggest that the incidence of improving public transit depends not only on how much each group uses it, but also on how willing each group is to bear high commute costs to work at a particular location, whether the system connects worker groups with their high-wage locations and the general equilibrium response of wages and house prices. In the context of Bogotá's TransMilenio, these reallocation and equilibrium effects are large enough to reverse the effects on inequality, which ultimately rose 0.55% as a result of the BRT.

Domestic Services and Alternative Home Ownership Assumptions. From 2000-2014, 7.3% of non-college educated Bogotanos worked as domestic helpers while almost no college-educated workers did. On the one hand, the model may underestimate the gains to the low-skilled by ignoring the fact that TransMilenio improved access to domestic services jobs in the homes of the college educated in the North. On the other hand, the high-skilled also benefitted from this increased labor supply, which lowered the cost of hiring domestic workers. Appendix E.5 extends the model to incorporate employment in domestic services, and Panel C of Table 9 presents the results. Overall, these two effects tend to balance out—the increase in inequality is very similar to the main model in Panel A. The last two rows of Panel C incorporate different assumptions over home ownership as outlined in Appendix E.6, with the results fairly invariant across the alternatives.[51]

5.3 Policy Counterfactuals

Impact of Alternative Networks. The first panel of Table 10 reports the impact on welfare, inequality, and output had the network been built without lines A and H, which connect the city's north and south with the CBD. The line to the south has a bigger effect on welfare (which would have been 0.3% lower without it), which is logical given the higher population density of poor and middle income workers there. For the same reason, the line to the north has the greater effect on inequality (which would have risen by 0.2% less without it). Intuitively, each group benefits relatively more from lines that improve accessibility from where they live.

The welfare gains from these trunk lines are exceeded by the benefits from the feeder bus network (as welfare would have been 0.94% lower without it). These buses connect outlying areas with portals and run on existing roadways. By providing complementary services that reach residents in outlying but dense residential areas, they can solve the last-mile problem of traveling between stations and final destinations. Given the low cost of feeder systems compared to the capital-intense BRT, these results suggest a high return to policy makers considering cheap, complementary services to increase access to mass rapid transit infrastructure.

[51] Table A.4 reports robustness of the main results to (i) allowing migration into the city, (ii) a larger elasticity of substitution between labor types, (iii) alternative employment data, (iv) alternative elasticities of demand, (v) alternative commuting elasticities and (iv) a decision over where to live and work.

A key trade-off policymakers face is whether to build fast rail, medium-speed BRT, or slower bus networks. The speed of such networks could affect the distributional consequences, for example, if the high-skilled were especially willing to pay to live near faster networks and priced out poorer residents. The last row runs a counterfactual that increases TransMilenio's speed to 35 km/h, close to the average operating speed of London's Underground. The increase in both welfare and inequality would have been much higher, confirming the intuition that faster systems benefit the rich relatively more. However, Figure A.5 compares the change in the college share near stations under this counterfactual. While it does increase in tracts closer than 500m from a station, the increase is very mild. This suggests the channels mentioned above, rather than gentrification, are responsible for why the rich benefit more from faster transit.

Land Value Capture One main criticism of TransMilenio was that its construction was not accompanied by an adjustment of zoning laws to allow housing supply to respond where it was needed. Appendix G shows that housing supply did not respond to the system's construction, consistent with other evidence on the restrictive role played by land use regulation (Cervero et. al. 2013). Many cities, such as Hong Kong and Tokyo, have had success in implementing LVC schemes which increase permitted densities around new stations but charge developers for the right to build there (Hong et. al. 2015). These policies increase housing supply and raise revenue to finance the infrastructure's construction.

I now evaluate the impact of TransMilenio had housing supply responded to the opening of the system. As a benchmark, I allow housing supply to adjust to the increase in floorspace values following a log-linear supply curve. Given that I do not observe a housing supply response in Bogotá that would permit me to measure a city-specific housing elasticity, I instead make a conservative choice and assume the housing supply in Bogotá is the same as that in Oakland, CA, the 6th most inelastic city in the US according to Saiz (2010). I then simulate the effect of two potential LVC schemes. First, I assume the government sells the rights to developers to increase floorspace by a maximum of 30% in tracts within 500m of stations, mimicking the "development rights sales" undertaken in certain Asian, European, and American cities.[52] Second, I assume the government sells permits that allow for the same change in total floorspace, but instead allocates the permitted floorspace changes according to a location's predicted change in CMA. Details on this model extension are provided in Appendix E.4. I compare the two equilibria by first removing TransMilenio (without housing adjustment) and then by adding it back under each housing supply model.

The last two panels of Table 10 present the results. Panel B shows the impacts on welfare. Under free adjustment, welfare would have been 44.04% higher than it is today. Under the LVC schemes, welfare would have been 24.47% or 43.82% higher than it is today under the distance- or CMA-based policies respectively (with similar relative effects on city output). These welfare improvements come from increasing housing supply where it is demanded the most as a result of new infrastructure,

[52]See Hong et. al. (2015) and Salon (2014) for further details. The parameters of this counterfactual are motivated by the example of Nanchang, China, where floor area ratios were increased by a uniform amount within 500m of stations. Revenues from the scheme covered 20.5% of costs, similar to my results.

tempering down floorspace price appreciation. The high return to the CMA-based instrument highlights how well-targeted zoning adjustments that allocate permits towards where they are needed most deliver bigger benefits. Panel C shows the fiscal benefits of the LVC schemes. Depending on how much the city population grows in response to the BRT, the distance-based instrument recoups 4-11% of construction costs, while the CMA-based scheme covers 6-21% of such costs.

These results suggest the potential for large welfare gains to governments pursuing a unified transit and land use policy. These policies can also be used to finance the construction of public transit, and targeting zoning adjustments based on where demand for housing will increase the most delivers the largest benefits.

6 Conclusion

This paper makes three contributions to our understanding of the aggregate and distributional effects of urban transit systems. First, it develops a sufficient statistics approach to evaluate the aggregate effects of new transit infrastructure in cities. Second, it shows that these statistics can be measured from readily available data and estimates them using the variation in accessibility induced by TransMilenio's construction. Third, it quantifies the welfare gains from the BRT under the equilibrium model and compares these gains with the VTTS to isolate the importance of reallocation and general equilibrium effects. Fourth, it estimates a richer model to that nests the sufficient statistics approach to quantify who the gains are shared between the rich and poor.

The study finds that the quantitative urban model performs well in explaining the adjustment of economic activity to transit infrastructure, with the log-linear relationships predicted by the model borne out in the data. The VTTS only account for around 57.5% of the total welfare gain from the new transit infrastructure. Thus, accounting for equilibrium effects matters for valuing the gains from new transit infrastructure in cities, and the framework developed in this paper provides a blueprint to do so. It also finds that the accounting for reallocation and general equilibrium effects acts against the benefits to poor workers who tend to use transit the most, which in the case of TransMilenio meant that welfare inequality rose by a mild 0.55%.

The paper also provides two key insights that can inform transit infrastructure policy. The first is that low-cost "feeder" bus systems that complement mass rapid transit by providing "last-mile" service for passengers' easy access to a system's terminals yield high returns. The second is that the welfare gains would have been around 40% higher had the the government implemented a more accommodative zoning policy, and government revenues from an LVC scheme could have raised a significant portion of construction costs. This underscores the benefits to cities from pursuing a unified transit and land use policy.

References

Adao, R., Arkolakis, C., and F. Esposito (2019), "Spatial Linkages, Global Shocks, and Labor Markets: Theory and Evidence," Working Paper.

Adao, R., Costinot, A., and D. Donaldson (2023), "Putting Quantitative Models to the Test: An Application to Trump's Trade War," Working Paper.

Ahlfeldt, G.M., Redding, S.J., Sturm, D.M. and N. Wolf (2015), "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83(6): 2127-2189.

Akbar, P., and G. Duranton (2017), "Measuring the Cost of Congestion in a Highly Congested City: Bogotá," Working Paper.

Akbar, P., G. Duranton, Couture, V., Ghani, E., and A. Storeygard (2021), "Accessibility and Mobility in Urban India," Working Paper.

Alcaldía Mayor de Bogotá D.C. (2009), "Evaluación de la Alternativas de la Red Metro Del SITP," https://www.metrodebogota.gov.co/sites/default/files/documentos/Producto%2015.%20Tomo%201.%20F

Alder, S., (2019), "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," Working Paper.

Allen, T., Arkolakis, C. and X. Li (2014), "On the Existence and Uniqueness of Trade Equilibria," Working Paper.

Allen, T., Arkolakis, C. and X. Li (2015), "Optimal City Structure," Working Paper.

Allen, T., and C. Arkolakis (2019), "The Welfare Effects of Transportation Infrastructure Improvements," NBER Working Paper 25487.

Allen, T., and C. Arkolakis (2021), "The Welfare Effects of Transportation Infrastructure Improvements," Working Paper.

Anselin, L., and N. Lozano-Gracia (2012), "Is the price right?: Assessing estimates of cadastral values for Bogotá, Colombia," *Regional Sciency Policy & Practice*, 4(4): 495-508.

Baltes, M., J. Barrios, A. Cain, G. Darido, P. Rodriguez (2006), "Applicability of Bogotá's Trans-Milenio BRT System to the United States," Federal Transit Administration Report Number FL-26-7104-01.

Barr, J., J. Beveridge, C. Clayton, A. Danahar, J. Gonsalves, B. Koziol, S. Rathwell (2010), "Designing Bus Rapid Transit Running Ways", American Public Transportation Association, Recommended Practice Report APTA-BTS-BRT-RP-003-10.

Bartelme, D. (2018), "Trade Costs and Economic Geography: Evidence from the U.S.," Working Paper.

Baum-Snow, N. (2007), "Did Highways Cause Suburbanization?"*Quarterly Journal of Economics*, 122: 775–805.

Baum-Snow, N. , and M. Kahn (2005),"Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000," *Brookings-Wharton Papers on Urban Affairs*, 147-206.

Baum-Snow, N., Brandt, L., Henderson, J., Turner, M., and Q. Zhang (2017), "Roads, Railroads and Decentralization of Chinese Cities," *Review of Economics and Statistics*, 99(3): 435-448.

Bayer, P., Ferreira, F., and R. McMillan (2007), "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 115(4): 588— 638.

Billings, S. (2011), "Estimating the value of a new transit option," *Regional Science and Urban Economics*, 41(6):525–536.

Bird, J and A. Venables. (2019), "The efficiency of land-use in a developing city: traditional vs modern tenure systems in Kampala, Uganda," *CEPR Discussion Papers 13563, C.E.P.R. Discussion Papers.*

Bocarejo, J-P., Perez, M.A., and I. Portilla (2013), "Impact of Transmilenio on Density, Land Use, and Land Value in Bogotá," *Research in Transportation Economics*, 40:78-86.

BORUSYAK, K., AND P. HULL (2021), "Non-Random Exposure to Exogeneous Shocks," Working Paper.

BRT DATA (2017), *Global BRT Statistics*, Retrieved from http://brtdata.org.

BRYAN, G., AND M. MORTEN (2019), "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," Journal of Political Economy, 127(5):2229-2268.

CARD, D. (2009),"Immigration and Inequality," *American Economic Review*, 99(2): 1–21.

CERVERO, R., IUCHI, K., AND H. SUZUKI (2013), *Transforming Cities with Transit: Transit and Land-Use Integration for Sustainable Urban Development*., Washington, DC: World Bank.

COMBES, P.-P., G. DURANTON, L. GOBILLON, AND S. ROUX (2010), "Estimating Agglomeration Economies with History, Geology, and Worker Effects," in Agglomeration Economies, ed. by E. L. Glaeser, Chicago University Press, Chicago.

CONLEY, T. (1999), "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics,* 92(1), 1–45.

CRACKNELL, J. (2003), "Transmilenio Busway-Based Mass Transit, Bogotá, Colombia," World Bank Working Paper.

DAVIS,M., AND F. ORTALO-MAGNÉ (2011), "Household expenditures, wages, rents," *Review of Economic Dynamics*,14(2):248-261.

DIAMOND, R. (2016), "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 106(3):479–524.

DONALDSON, D. (2015), "The Gains from Market Integration," *Annual Review of Economics*, 7:619-647.

DONALDSON, D., AND R. HORNBECK (2016), "Railroads and American Economic Growth: A Market Access Approach," *Quarterly Journal of Economics*, 131(2): 799-858.

DONALDSON, D. (2018), "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108(4-5):899-934.

EATON, J., AND S. KORTUM (2002), "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

FAJGELBAUM, P. D., AND E. SCHAAL (2020), "Optimal Transport Networks in Spatial Equilibrium," Econometrica, 88(4):1411-1452.

FEENSTRA, R., P. LUCK, M. OBSTFELD AND K. RUSS (2018), "In Search of the Armington Elasticity," The Review of Economics and Statistics, 100(1): 135-150.

FERNANDEZ-VAL, I. AND M. WEIDNER (2016), "Individual and Time Effects in Nonlinear Panel Data Models with Large N, T," *Journal of Econometrics*, 192(1): 291-312

FUJIMOTO, T., AND U. KRAUSE (1985), "Strong ergodicity for strictly increasing nonlinear operators," LINEAR ALGEBRA AND ITS APPLICATIONS, 71: 101-112.

GADUH, A., GRACNER, T., AND A ROTHENBERG (2022), "Life in the Slow Lane: Unintended Consequences of Public Transit in Jakarta," *Journal of Urban Economics*, 128.

GALLE, S., RODRIGUEZ-CLARE, A., AND M. YI (2022), "Slicing the Pie: Quantifying the Aggregate and Distributional Consequences of Trade," *The Review of Economic Studies*.

GANONG P., AND D. SHOAG (2017), "Why has regional income convergence in the U.S. declined?" *Journal of Urban Economics*, 102:76-90.

GIANNONE, E. (2021), "Skill-Biased Technical Change and Regional Convergence," Working Paper.

GIBBONS, S., AND S. MACHIN (2005), "Valuing rail access using transport innovations," *Journal of Urban Economics*, 57(1):148–169.

GLAESER, E., KAHN, M., AND J. RAPPAPORT (2008), "Why do the poor live in cities? The role of public transportation," *Journal of Urban Economics,* 63(1):1–24.

Gonzalez-Navarro, M., and M. Turner (2018), “Subways and urban growth: Evidence from Earth,” *Journal of Urban Economics*, 108:85-106.

Guerrieri, G., Hartley, D., and E. Hurst (2013), “Endogenous Gentrification and Housing Price Dynamics,” *Journal of Public Economics*, 100:45-60.

Greenstone, M., Hornbeck, R., and E. Moretti (2010), “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy*, 118(3), 536–598.

Heblich, S., Redding, S., and D. Sturm (2020), “The Making of the Modern Metropolis: Evidence from London,” *The Quarterly Journal of Economics*, 135(4):2059-2133.

Hidalgo, D., and P. Graftieaux (2008), “Bus Rapid Transit Systems in Latin America and Asia: Results and Difficulties in 11 Cities,” *Journal of the Transportation Research Board*, 2072(1):77-88.

"Historia de TransMilenio" TransMileno.gov.co. Feb 15, 2022. https://www.transmilenio.gov.co/pu de-transmilenio/.

Hong, Y-H., Murakami, J., Suzuki, H., and B. Tamayose (2015), *Financing Transit-Oriented Development with Land Values*, Washinton, DC: The World Bank.

Hsieh, C-T., Hurst, E., Jones, C., and P. Klenow (2019), “The Allocation of Talent and US Economic Growth,” *Econometrica*,87(5):1439-1474.

Kline, P., E. Moretti (2014), “Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority,” *Quarterly Journal of Economics*, 129(1), 275–331.

Lawell, C., Umanskaya, V., and Zhang, W., (2017), “The Effects of License Plate-Based Driving Restrictions on Air Quality: Theory and Empirical Evidence,” *Journal of Environmental Economics and Management*,82:181-220.

Lee, E. (2020), “Trade, Inequality, and the Endogenous Sorting of Heterogeneous Workers,” *Journal of International Economics*, 125.

Mackie, P., J. Nellthorp, and J. Laird (2005), “Notes on the Economic Evaluation of Transport Projects,” World Bank Transport Note No. TRN-5.

McFadden, D. (1974), “The measurement of urban travel demand”, *Journal of Public Economics,* 3(4): 303-328.

McKinsey (2016), *Bridging Global Infrastructure Gaps*, McKinsey Global Institute

Menckhoff, G. (2005), “Latin American Experience with Bus Rapid Transit,” presented at Institute of Transportation Engineers Annual Meeting, Melbourne, Australia.

Monte, F., Redding, S., and E. Rossi-Hansberg (2018), “Commuting, Migration and Local Employment Elasticities,” *American Economic Review*,108(12):3855-3890.

Morrison, A., (2007), “The Tramways of Bogotá, Colombia,” *http://www.tramz.com/co/bg/t/te.html*

Müller, J.M. (2014), Bus Rapid Transit: The Answer to Transport Problems in Megacities? The Example of TransMilenio (Bogotá, Colombia), in Kraas, F., Aggarwal, S., Coy, M., Mertins, G. (eds) Megacities, International Year of Planet Earth., Springer, Dordrecht.

Owens, R., E. Rossi-Hansberg, and P-D. Sarte (2020), “Rethinking Detroit,” American Economic Journal: Economic Policy, 12(2):258-305.

Redding, S., and D. Sturm (2008), “The Costs of Remoteness: Evidence from German Division and Reunification,” *American Economic Review*, 98(5), 1766–1797.

Rosenthal, S., and W. Strange (2004), “Evidence on the Nature and Sources of Agglomeration Economics,” in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. Thisse, vol. 4. Elsevier North Holland, Amsterdam.

Rossi-Hansberg, E., P. Sarte, and R. Owens III (2010), "Housing Externalities," Journal of Political Economy, 118(3): 485-535.

Ruiz, F., and G. Vallejo (2010), "Using Land Registration as a Tool to Generate Municipal Revenue: Lessons from Bogota," Cadastre of Bogota Working Paper.

Saiz, A. (2010), "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 125(3):1253-1296.

Salon, D. (2014), "Location Value Capture Opportunities for Urban Public Transport Finance," Transit Leadership Summit, London, White Paper.

Severen, C. (2021), "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification," *The Review of Economics and Statistics*.

Steiner, R., and H. Vallejo (2010), *Economy, in Colombia: a country study*, Library of Congress. Federal Research Division.

Small, K., and E. Verhoef (2007), *The Economics of Urban Transportation*, New York: Routledge.

Train, K., and D. McFadden (1978), "The goods/leisure tradeoff and disaggregate work trip mode choice models," *Transportation Research*, 12(5): 349-353.

Transportation Review Board (2003), "Bogotá, Colombia. Brief: TransMilenio BRT", Transit Cooperative Research Program, Report 90.

Transportation Review Board (2007), "Bus Rapid Transit Practitioner's Guide", Transit Cooperative Research Program, Report 118.

Uribe Sanchez, M.C. (2010), "Land Information Updating, a De Facto Tax Reform: Bringing up to Date the Cadastral Database of Bogota," Cadastre of Bogota Working Paper.

World Bank (2009), 2009 Annual Report: Energy Sector Management Assistance Program (ESMAP),Washington, DC: World Bank.

Wright , L., and W. Hook (2007), *Bus Rapid Transit Planning Guide*, New York: Institute for Transportation and Development Policy.

Tables

Table 1: Mode Choice and Commuting Parameter Estimates

Panel A: Mode Choice	
Parameter	**Estimate**
κ	0.012**
	(0.006)
b_{Bus}	-0.085*
	(0.051)
b_{Car}	0.853***
	(0.291)
b_{TM}	-0.212*
	(0.108)
λ	0.138**
	(0.067)
N	19,510
Panel B: Commute Semi-Elasticity (Aggregate)	
Parameter	**Estimate**
$\theta\kappa$	0.039**
	(0.016)
N	710

Note: Panel A shows estimation results from nested logit regression on mode choices from trip-level data from the 2015 Mobility Survey. Controls for hour of trip departure dummies and dummies for gender and quintiles of the age distribution are included for each mode, which is equivalent to allowing preferences for each mode to vary by these characteristics. Heteroscedasticity robust standard errors are reported in parentheses. Panel B shows gravity equation estimation results, estimated via PPML. The outcome is the log number of commuters between each origin and destination locality pair in 1995 or 2015. Fixed effects for each origin-destination pair, origin-year and destination-year pair are included. Reported coefficient is that on travel time. In both panels, only trips to work during rush hour (5-8am) by heads of households included. Controls of origin-destination pair characteristics interacted with year dummies include (i) the average number of crimes per year from 2007-2014, (ii) the average log house price in 2012 and (iii) the share of the trip that takes place along a primary road along the least-cost routes between origin and destination. Standard errors are clustered at the origin-destination locality. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2: Baseline Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Residents							
ln(Res Floorspace Price)	0.370**	0.382**	0.384**	0.342*	0.379***	0.228**	0.365*
	(0.179)	(0.172)	(0.169)	(0.174)	(0.147)	(0.090)	(0.202)
N	2,201	2,201	2,201	2,166	2,199	2,201	2,201
R^2	0.41	0.43	0.43	0.43	0.43	0.43	0.43
ln(Res Population)	0.722**	0.648*	0.746**	0.630*	0.683**	0.318*	1.086***
	(0.337)	(0.337)	(0.331)	(0.334)	(0.293)	(0.174)	(0.388)
N	2,256	2,256	2,256	2,219	2,255	2,254	2,256
R^2	0.34	0.35	0.37	0.37	0.37	0.37	0.37
Panel B: Firms							
ln(Comm Floorspace Price)	0.526**	0.540**	0.621**	0.580**	0.514**	0.362***	0.718**
	(0.253)	(0.251)	(0.248)	(0.249)	(0.214)	(0.135)	(0.301)
N	2,080	2,080	2,080	2,047	2,089	2,083	2,080
R^2	0.09	0.10	0.11	0.11	0.11	0.11	0.11
Comm Floorspace Share	0.290***	0.297***	0.291***	0.284***	0.197***	0.151***	0.286***
	(0.087)	(0.088)	(0.088)	(0.088)	(0.072)	(0.047)	(0.101)
N	2,230	2,230	2,230	2,195	2,239	2,233	2,230
R^2	0.14	0.15	0.15	0.15	0.15	0.15	0.15
ln(Establishments)	2.101***	1.787**	1.329*	1.283*	1.414**	0.924**	1.414
	(0.735)	(0.761)	(0.751)	(0.757)	(0.642)	(0.395)	(0.888)
N	2,028	2,028	2,028	1,995	2,028	2,028	2,028
R^2	0.65	0.67	0.68	0.68	0.68	0.68	0.68
Locality FE	X	X	X	X	X	X	X
Log Dist CBD X Region FE	X	X	X	X	X	X	X
Basic Tract Controls	X	X	X	X	X	X	X
Historical Controls		X	X	X	X	X	X
Land Market Controls			X	X	X	X	X
Exclude Portals+CBD				X			
Exclude Band					1.5km		
Alt Time Aggregation						X	
Distance to TM Controls							X

Note: Observation is a census tract. Each entry reports the coefficient from a regression of the change in the variable in each row on the change in firm or residential commuter market access (RCMA for residential outcomes, FCMA for commercial outcomes). CMA is always computed holding employment and population fixed at their initial levels and excluding the location itself from the summation. Each column corresponds to a specification. In land market regressions of row 1, 3 and 4, outcomes are log changes between 2018 and 2000 and the change in CMA is that induced by all three phases holding residential population and employment fixed at their levels in 2000 (population in 2000 is a linear interpolation from the 1993 and 2005 census; employment is from the 2000 CCB data). In population regressions of row 2, the outcome is the log change in residential population between 1993 and 2018. The change in CMA is that induced by all three phases of TransMilenio, holding residential population and employment fixed at their levels in 1993 and 1990 respectively (measured from the population and economic censuses). In establisment regressions of row 5, the outcome is the log change in the number of establishments between 2000 and 2015 from the CCB data against the same CMA measures as the land market regressions. Establishment specifications are weighted by the share of establishments in a tract in the initial period. CBD X Region controls are log distance to the CBD, interacted with dummies for whether the locality is in the North, West or South of the city. Basic tract controls include (i) log area, (ii) log distance to the main road, (iii) log distance to a main road interacted with log distance to the CBD, (iv) dummies for each quartile of 1993 population density, 1990 employment share (employment divided by employment plus population), and 1993 college share. Historical controls include dummies for each quartile of population density in 1918, and a dummy for whether the tract was closer than 500m to a main road in 1933. Land market controls include the share of land developed, floor area ratio, share of floorspace used for commercial purpose, and log average floorspace value in 2000. Any control that represents the initial value of an outcome variable is dropped from that specification. Columns (1) to (3) incrementally add controls. Column (4) restricts the sample to tracts more than 500m from a portal or the CBD. Column (5) computes the change in market access to tracts further than 1.5km from the tract itself. Column (6) assumes users take the quickest form of public transit (i.e. the minimum rather than the weighted average within the public nest). Column (7) includes a dummy for whether tract is closer than 500m from any TransMilenio station. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 3: Planned Networks

	(1)	(2)
Panel A: Residents		
ln(Res Floorspace Price)		
Δ ln(CMA)	0.384**	0.368**
	(0.169)	(0.175)
E[Δ ln(CMA Plan)]		0.084
		(0.225)
N	4,402	4,402
R^2	0.43	0.43
p-val		0.71
ln(Res Population)		
Δ ln(CMA)	0.746**	0.817**
	(0.331)	(0.337)
E[Δ ln(CMA Plan)]		-0.388
		(0.371)
N	4,512	4,512
R^2	0.37	0.37
p-val		0.29
Panel B: Firms		
ln(Comm Floorspace Price)		
Δ ln(CMA)	0.621**	0.687***
	(0.248)	(0.259)
E[Δ ln(CMA Plan)]		-0.360
		(0.426)
N	4,160	4,160
R^2	0.11	0.11
p-val		0.39
Comm Floorspace Share		
Δ ln(CMA)	0.291***	0.290***
	(0.088)	(0.089)
E[Δ ln(CMA Plan)]		0.005
		(0.099)
N	4,460	4,460
R^2	0.15	0.15
p-val		0.96
ln(Establishments)		
Δ ln(CMA)	1.329*	1.170
	(0.751)	(0.778)
E[Δ ln(CMA Plan)]		0.829
		(0.692)
N	4,056	4,056
R^2	0.68	0.68
p-val		0.24

Note: Column (1) repeats the baseline specification i.e. column (3) from Table 2. That is, each entry reports the coefficient from a regression of the change in the variable in each row on the change in firm or residential commuter market access (RCMA for residential outcomes, FCMA for commercial outcomes). Column (2) adds as an additional explanatory variable the average change in RCMA or FCMA (depending on the outcome, RCMA for residential and FCMA for commercial) each tract would have received had TransMilenio been built across the 4 historical plans. The p-value corresponds to a χ^2 test of equality of coefficients on Δ ln(CMA) in columns 1 and 2. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4: Staggered Station Openings

	(1)	(2)	(3)	(4)
Panel A: Residents				
	ResPr	ResPr	ResPop	ResPop
ln(RCMA)	0.183*	0.231**	0.396*	0.521**
	(0.101)	(0.109)	(0.237)	(0.253)
ln(RCMA) Later Phase		0.336		0.345
		(0.239)		(0.348)
N	2,144	2,144	2,207	2,207
R^2	0.45	0.45	0.29	0.29
Panel B: Firms (Land Markets)				
	CommPr	CommPr	CommSh	CommSh
ln(FCMA)	0.483**	0.478**	0.210***	0.206***
	(0.199)	(0.201)	(0.057)	(0.056)
ln(FCMA) Later Phase		0.281		0.220
		(0.666)		(0.172)
N	2,055	2,055	2,182	2,182
R^2	0.06	0.06	0.08	0.08
Panel C: Firms (Census Employment)				
	Emp	Emp	Form Emp	Form Emp
ln(FCMA)	1.640	1.791*	2.097	2.143
	(1.064)	(1.071)	(1.339)	(1.348)
ln(FCMA) Later Phase		1.497		0.438
		(1.288)		(1.745)
N	1,927	1,927	1,629	1,629
R^2	0.23	0.23	0.17	0.17

Note: Table repeats the baseline specification i.e. column (3) from Table 2. Outcomes are (growth in) residential floorspace prices (Res Pr), residential population (Res Pop), commercial floorspace prices (Comm Pr), commercial floorspace share (Comm Sh), employment from the census (Emp), employment in establishments with more than 10 workers (Form Emp). For land market outcomes, the change in outcomes are measured between 2008 and 2000. The right hand side variables include CMA growth due to (i) phases 1 and 2 of the system open by 2006 ($\ln(CMA)$) and (ii) phase 3 of the system open by 2013 ($\ln(CMA)$ Later Phase). For residential population, the change in outcome in measured between the 2005 and 1993 census. The right hand side variables include CMA growth due to phase 1 (open by 2003, with most opening by 2001), and the change in CMA due to phases 2 and 3 (opened in 2006 and 2013). For employment, the change in employment is measured from between the 2005 and 1990 economic censuses. The CMA variables are the same as for residential population. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 5: IV Estimates

	Baseline	IV	IV-LCP&Tram Exc Own	IV-LCP&Tram Exc 1.5km
Panel A: Residents				
ln(Res Floorspace Price)	0.384**	0.276***	1.117***	1.111***
	(0.169)	(0.102)	(0.247)	(0.229)
N	2,201	2,201	2,202	2,202
F-Stat		2,475.45	166.46	201.21
ln(Res Population)	0.746**	0.553**	0.554	0.653
	(0.331)	(0.225)	(0.483)	(0.438)
N	2,256	2,256	2,239	2,239
F-Stat		2,404.19	248.56	295.97
Panel B: Firms				
ln(Comm Floorspace Price)	0.621**	0.552***	0.397	0.607**
	(0.248)	(0.204)	(0.291)	(0.304)
N	2,080	2,080	2,085	2,085
F-Stat		3,165.71	664.43	746.85
Comm Floorspace Share	0.283***	0.257***	0.227**	0.204**
	(0.093)	(0.071)	(0.105)	(0.103)
N	2,231	2,230	2,235	2,235
F-Stat		3,112.12	670.54	733.74
ln(Establishments)	1.329*	1.229**	2.207**	1.954**
	(0.751)	(0.562)	(0.880)	(0.827)
N	2,028	2,028	1,995	1,995
F-Stat		2,878.76	402.94	494.76

Note: Observation is a census tract. Specification corresponds to column (3) of Table 2. Column 1 reproduces the baseline results. Column 2 instruments the true change in CMA (i.e. including the location itself in the summation and measure employment and population in both periods instead of holding them constant at their initial values) with the baseline change in CMA measure from column 1. Column 3 instruments for the change in CMA using the average change in CMA across the IV and tram instruments constructing excluding the tract itself in the summation, while column 4 excludes all tracts closer than 1.km from the tract. In this specification, only census tracts further than 500m from a portal and a dummy for whether a census tract is further than 1km from the historical tram system is included (to capture direct effects from the tram instrument). Column 1 reports HAC standard errors as in the baseline specification. Columns 2-4 report heteroscedasticity robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6: Aggregate Results Using Sufficient Statistics Approach

Panel A: VTTS Comparison

	VTTS	GE	GE (No Ext.)
Welfare Gain (%)	1.260	2.335	1.527
90% CI	(0.740,3.106)	(0.740,5.861)	
95% CI	[0.309,3.692]	[0.475,6.936]	
As Fraction of VTTS		53.95	82.49
90% CI		(44.20,71.74)	
95% CI		[39.22,85.96]	

Panel B: Aggregate Effects

	No Migration	Migration
Welfare	2.282	0.597
GDP	3.121	15.131
GDP Net of Costs	2.504	14.514
Population	0.000	9.514
Rents	-0.672	5.283
% of Obs GDP Growth	2.963	14.362
% of Obs Population Growth	0.000	34.886

Panel C: Incorporating Congestion

	% Change in Welfare	% of No Congestion Welfare Change
No Congestion	3.921	100.00
Congestion	3.943	100.55
Convert TM to Car Lanes	0.028	0.641

Notes: Table shows the welfare effects from TransMilenio using the sufficient statistics approach from Proposition 1. Panel A compares the GE welfare effects with those from the first order approximation (VTTS) in proposition 2. The % change in welfare is computed as adding TransMilenio back to the counterfactual equilibrium without it. Each entry is computed by first simulating the effect of removing TransMilenio (the initial equilibrium) and then adding it back in under the different approaches. In column 1, the change in travel times accounts for the discrete choice over modes used to aggregate mode-specific travel times. In column 2, the GE effects are reported using the main reduced form elasticities (column 3 in Table 2). 90% and 95% confidence intervals are provided by bootstrapping the quantitative exercise 200 times as described in Appendix C.9. The second row reports the fraction of GE gains are explained by VTTS, with confidence intervals also reported. The non-parametric test on the bootstrap sample of whether the fraction of gains explained by VTTS is greater than 1 rejects this null (p-value 0.04). Column 3 reports GE results from a model without externalities. It computes the reduced form elasticities using the expressions derived in Appendix C.8.1, using estimates for $\theta, \alpha, \beta, \sigma$ and setting $\mu_A = \mu_U = 0$. Confidence intervals are not reported since this removes sampling variation from the 4 estimated reduced form elasticities. Panel B shows the (negative of the) value of the percentage change in each variable from removing the TransMilenio network (phases 1 through 3) from the 2016 equilibrium, under both assumptions on population mobility. The scenario with migration assumes a migration elasticity of $\rho = 3$ (see Appendix E.1 for details). The last two rows show the fraction of observed growth of population and GDP between 2000 and 2016 that can be accounted for by TransMilenio under each scenario. Bogotá's GDP increased by 105.35% (average annual growth rate of 4.6%) while population grew by 27% over the period. GDP net of costs shows the NPV increase in GDP accounting for capital costs and the NPV of operating costs as described in Appendix F.4. Note the average welfare value in Panel B differs from that in Panel A, which uses the counterfactual equilibrium without TransMilenio as the initial equilibrium for ease of comparison with the VTTS. Lastly, Panel C reports welfare results from model allowing for congestion (see Appendix E.2 for details). A congestion elasticity of 0.06 is used, the average congestion elasticity estimated for Bogotá by Duranton and Akbar (2017). The first row shows the welfare effect (the absolute value of $\bar{U}^{NoTM}/\bar{U}^{TM} - 1$) in the closed city model in this model extension, when the congestion elasticity is set to zero. This differs slightly from the baseline number since the congestion elasticity is used when calibrating the unobserved traffic matrix for the observed equilibrium, and the construction of commute times is slightly different due to the routing model of commutes. The second row shows the welfare impact of TransMilenio with congestion, and the second column shows the welfare gains as a fraction of the baseline case without congestion in row 1. The third row shows the welfare impact had TransMilenio routes been made into car lanes instead of BRT (the absolute value of $\bar{U}^{NoTM}/\bar{U}^{ReplaceTMWithRoads} - 1$).

Table 7: Commuting Elasticities

	OLS	IV-LCP&Tram	PPML	PPML	PPML
HighSkill X ln Commute Cost	-0.0250**	-0.0295**	-0.0154***	-0.0054	-0.0253***
	(0.0116)	(0.0120)	(0.0028)	(0.0104)	(0.0089)
LowSkill X ln Commute Cost	-0.0278**	-0.0460***	-0.0292***	-0.0534***	-0.0663***
	(0.0121)	(0.0150)	(0.0027)	(0.0109)	(0.0096)
N	1,738	1,738	1,444	2,608	4,032
Years	1995,2015	1995,2015	2015	1995,2015	1995,2011,2015
Origin-Destination-Skill-Car Ownership FE	X	X		X	X
Destination-Skill-Year FE	X	X	X	X	X
Origin-Skill-Car Ownership-Year FE	X	X	X	X	X

Note: Outcome is the conditional commuting shares. Observation is an origin-destination-skill-car ownership-year cell. Skill corresponds to college or non-college educated workers. Only trips to work during rush hour (5-8am) by individuals aged 18-55. Columns 1 and 2 estimate OLS and IV models between 1995 and 2015. Columns 3-5 run PPML models on alternative sets of years: 2015, 1995 and 2015, and 1995, 2011 and 2015 respectively. Since the coefficient for high-skill workers is imprecise in the main specification using two years in column 4, the final column 5 pools data from 3 years. Travel times are measured according to the network in each year e.g. travel times for TransMilenio in 2011 come from the 2006 network, while those in 2015 come from the 2012 network. Standard errors are clustered at the origin-destination locality. * p < 0.1; ** p < 0.05; *** p < 0.01

Table 8: GMM Results

Parameter	Estimate
Panel A: Firms	
μ_A	0.242*
	(0.128)
Panel B: Workers	
η_L	2.070**
	(0.805)
η_H	2.250***
	(0.734)
μ_U^L	0.730**
	(0.372)
μ_U^H	1.002***
	(0.335)

Note: Estimates are from joint GMM procedure as described in text. In Panel A, controls are locality fixed effects, log distance to CBD interacted with region fixed effects, basic tract controls (log area, log distance to the main road, log distance to a main road interacted with log distance to the CBD, dummies for each quartile of 1993 population density, 1990 employment share (employment divided by employment plus population), and 1993 college share), land market controls (share of land developed, floor area ratio, share of floorspace used for commercial purpose, and log average floorspace value in 2000) and historical controls (dummies for each quartile of population density in 1918). In Panel B, controls are locality fixed effects, basic tract controls (log area, log distance to the main road, log distance to a main road interacted with log distance to the CBD, dummies for each quartile of 1990 employment share), land market controls (share of land developed, floor area ratio, share of floorspace used for commercial purpose, and dummies for each quartile of floorspace value in 2000) and historical controls (dummies for each quartile of population density in 1918). All controls and fixed effects are interacted with group-specific dummies. Tracts closer than 500m to a TransMilenio portal are excluded. Instruments exclude the tract itself in summations, and are averages across the LCP and Tram measures as in the reduced form results. Standard errors clustered by tract obtained from 200 block-bootstrapped replications resampled at the tract-level.* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 9: Main Quantitative Results & Distributional Effects

	Average Welfare	**Inequality**
Panel A: Main Results		
Diff θ, η, Imperf Sub	1.007	0.546
90% CI	(0.254, 4.022)	(-0.159, 0.875)
95% CI	[0.017, 4.407]	[-0.441, 1.138]
P-value $\widehat{\widehat{U}}_H > \widehat{\widehat{U}}_L$	0.15	
Panel B: Decomposing the Role of Elasticities		
Same η, θ, Perf Sub	2.191	-0.371
Diff θ, same η, Perf Sub	2.261	-0.163
Diff θ, η, Perf Sub	2.510	-0.139
Panel C: Model Extensions		
Domestic Services	0.832	0.585
Local Home Ownership	0.854	0.596
All Renters	0.867	0.619

Note: Table shows the percentage welfare and inequality change (defined as $\widehat{\widehat{U}}_H/\widehat{\widehat{U}}_L$) from TransMilenio under models. Each entry is computed by first simulating the effect of removing TransMilenio, and reports the absolute value of the percentage welfare change from moving from the TM to no TM equilibrium. Panel A reports results from the full model, where θ_g, η_g are set to their estimated values and $\sigma = 1/0.7$ as described in the text. Confidence intervals from 200 bootstrap replications are reported (using the same procedure as described in Appendix C.9), as well as the p-value from a non-parametric test of whether the high-skill gain more than the low-skilled across these bootstraps. Panel B reports results decomposing the role of these elasticities. The first row assumes θ, η are equal across groups (set to their average value) and labor types are perfect substitutes in production. The second and third rows allow θ and η to differ across groups (set to their estimated values). Panel C shows results from model extensions to allow for employment of the low-skilled in domestic services, as well as alternative assumptions over home ownership. See Appendix E.5 and E.6 for further details.

Table 10: Policy Counterfactuals

Panel A: Alternative Networks

	% Δ **Welfare**	% Δ**Inequality**	% Δ**Output**
Remove Line South	-0.298	-0.060	-0.318
Remove Line North	-0.084	-0.204	-0.699
Remove Feeders	-0.942	-0.196	-1.014
Faster TM	1.355	0.698	2.790

Panel B: Land Value Capture Welfare Effects

	% **Increase Relative to Baseline**	
	Welfare	**Output**
Free Adjustment	44.04	15.78
LVC, Bands	24.47	9.17
LVC, CMA	43.82	11.95

Panel C: Land Value Capture Revenue Effects

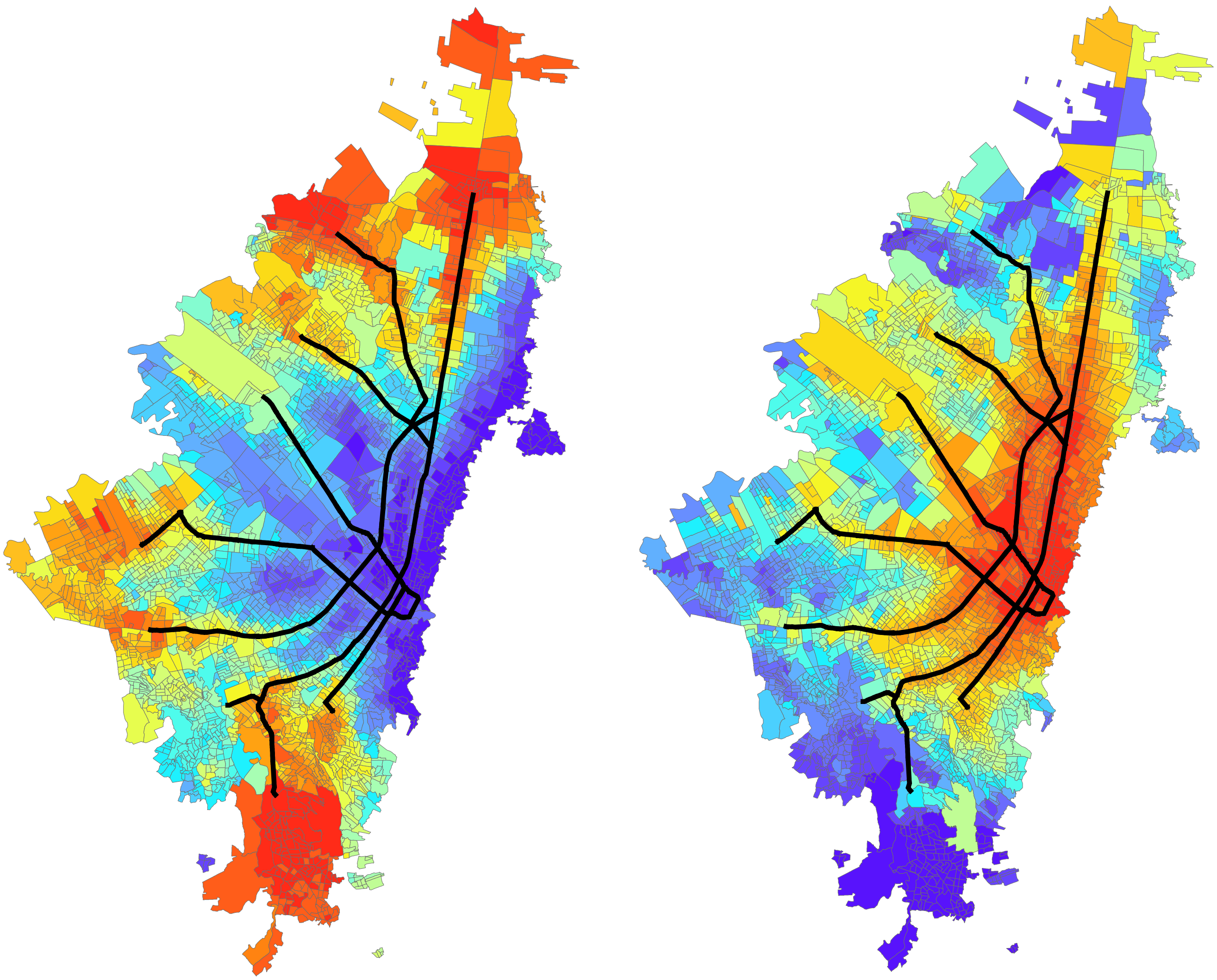
	Closed City	**Open City**
LVC Band Revenue (mm)	58.62	152.77
As share of capital costs	4.04	10.54
LVC CMA Revenue (mm)	88.31	297.57
As share of capital costs	6.09	20.53

Note: Panel A shows the impact of particular network components relative to the full network using the full model. The numbers report the percentage change in each variable from moving from the full TransMilenio network to the counterfactual one. The last row reports results from making the TransMilenio faster, with an operational speed of 35km/h. Panel B shows the impacts of alternative housing supply models, using the model extension from Section E.4. I first solve for the counterfactual equilibrium without TransMilenio. I then compute the equilibrium returning to the TransMilenio network under each housing supply model, and report the percentage change in each variable as a fraction of returning to the observed network under the fixed housing supply assumption (minus one, since the change in each variable in each counterfactual scenario exceeds the value under fixed housing supply). The first row is the case with freely adjusting housing. The second row is the distance-band based land value capture (LVC) scheme, where the government sells rights to construct up to 30% new floorspace in tracts closer than 500m from stations. The third row is the CMA-based scheme where the same number of permits are issued by distributed instead by a tract's relative change in CMA as described in the text. These figures are all from the closed city model, relative comparisons are similar in the open city model. Panel C shows the government revenue earned under the land value capture policies, in levels and as a fraction of TransMilenio's construction costs. These are reported for the closed and open city model separately since the results vary by assumption. Numbers in millions of 2016 USD.

Figures

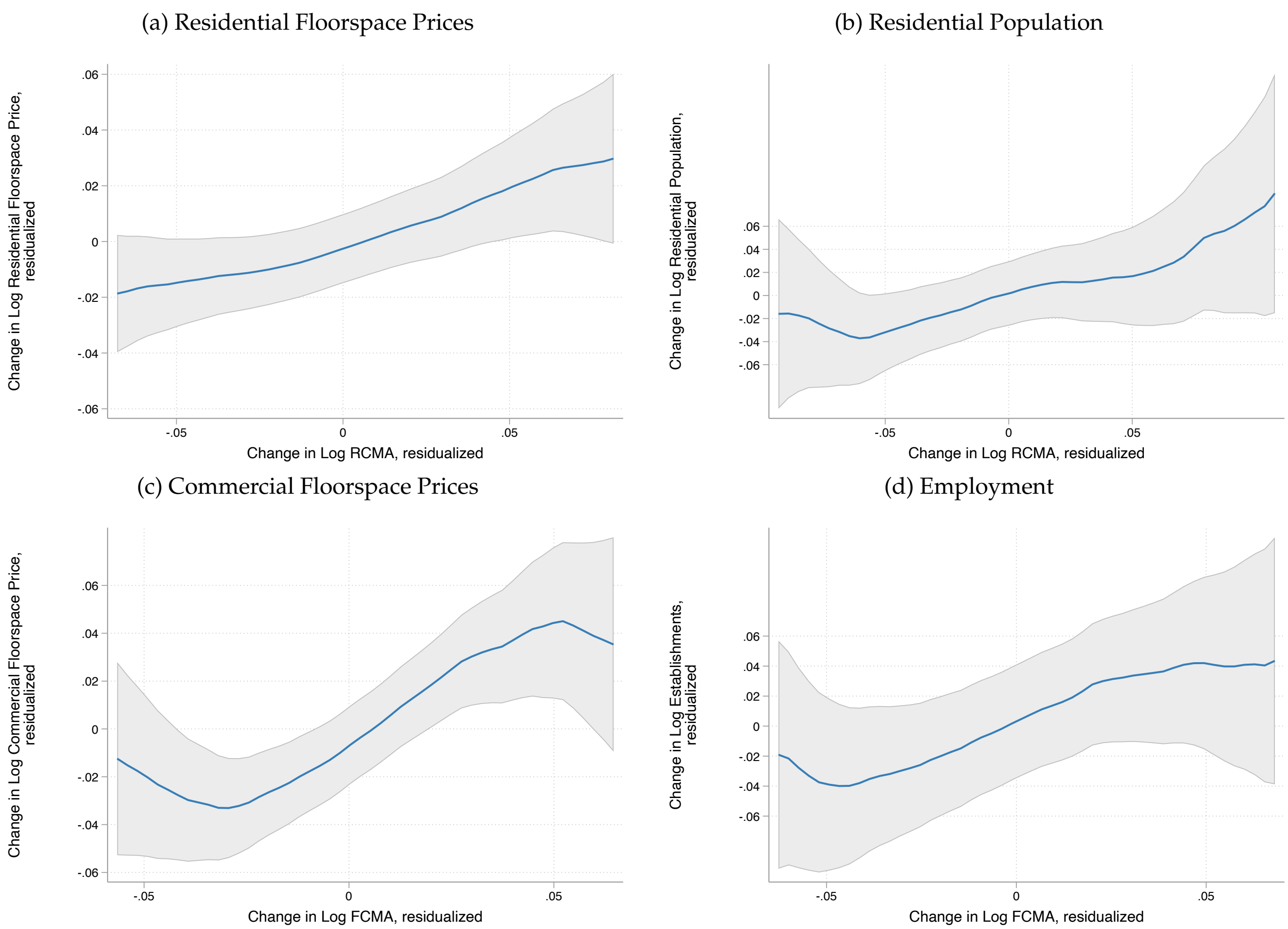
Figure 1: Change in Commuter Market Access from TransMilenio

(a) Resident CMA (b) Firm CMA



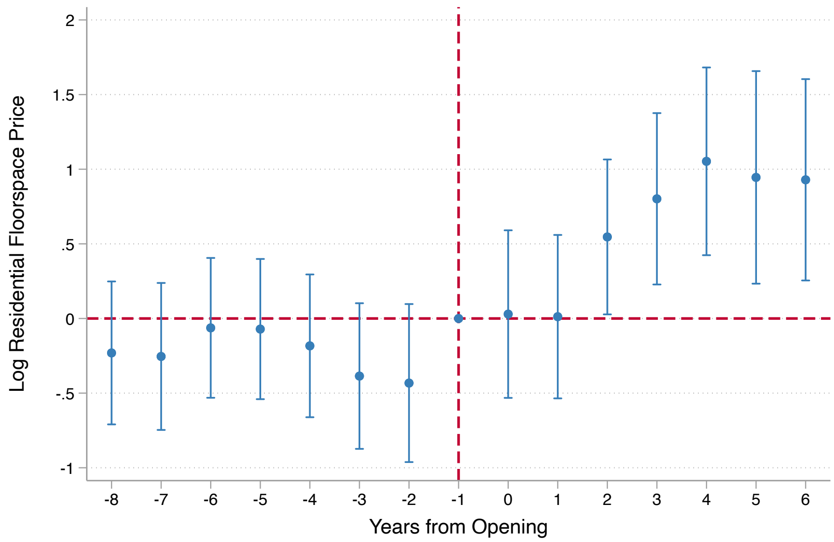
Note: Plot shows the change in CMA from the baseline specification. Population and employment arefixed at their initial level and changing only commute costs (to the full TransMilenio network as of phase 3). Tracts are grouped into vigintiles based on the the change in CMA, with warmer colors indicating a larger increase in CMA. Black line shows the TransMilenio routes as of 2013. The changes in CMA are normalized to have mean zero. For the change in RCMA, the min is -.198, the max is .375, the standard deviation is 0.097 and the average range of each vigintile is .028. For the change in FCMA, the min is -.147, the max is .246, the standard deviation is 0.068 and the average range of each vigintile is .020.

Figure 2: Non-Parametric Relationship Between Outcomes and Commuter Market Access



Note: Plot shows the non-parametric relationship between outcomes and CMA. Specifications correspond to the reduced form from column (3) of Table 2. Top and bottom 2% of the change in CMA are trimmed to reduce noise at the tails and zoom in on main relationship.

Figure 3: Residential Floorspace Price Event Study



Note: See discussion in Section 4.3 for details. The year before opening is the omitted category. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported.

For Online Publication: Appendix to "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio"

Contents

A Additional Tables

Table A.1: Robustness

	(1) Baseline	(2) Simple Avg Time	(3) Slow	(4) Fast	(5) High θ	(6) Low θ	(7) <3km TM	(8) Sect-Clus SE	(9) Exc. 1km Nodes	(10) TM X CBD
Panel A: Residents										
ln(Res Floorspace Price)	0.384***	0.368***	0.401***	0.467***	0.194*	0.749***	0.492***	0.384*	0.320**	0.399**
	(0.141)	(0.134)	(0.141)	(0.134)	(0.113)	(0.205)	(0.151)	(0.224)	(0.189)	(0.170)
N	2,201	2,200	2,198	2,201	2,202	2,205	2,051	2,201	2,008	2,201
R^2	0.43	0.43	0.43	0.43	0.43	0.43	0.40	0.43	0.44	0.43
ln(Res Population)	0.746**	0.709**	0.460	0.624**	0.424*	0.817*	0.869***	0.746*	0.578*	1.119***
	(0.304)	(0.289)	(0.301)	(0.293)	(0.243)	(0.467)	(0.309)	(0.399)	(0.332)	(0.348)
N	2,256	2,255	2,276	2,276	2,276	2,276	2,102	2,256	2,061	2,256
R^2	0.37	0.37	0.37	0.37	0.37	0.37	0.36	0.37	0.37	0.37
Panel B: Firms										
ln(Comm Floorspace Price)	0.621**	0.665***	0.710***	0.646***	0.479**	1.090***	0.567**	0.621**	0.552**	0.736**
	(0.248)	(0.255)	(0.274)	(0.233)	(0.219)	(0.337)	(0.253)	(0.268)	(0.253)	(0.300)
N	2,080	2,083	2,073	2,077	2,078	2,083	1,964	2,080	1,902	2,080
R^2	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.12	0.11
Comm Floorspace Share	0.291***	0.257***	0.302***	0.231***	0.270***	0.172	0.323***	0.291***	0.282***	0.285***
	(0.088)	(0.089)	(0.103)	(0.084)	(0.080)	(0.113)	(0.088)	(0.105)	(0.089)	(0.101)
N	2,230	2,233	2,225	2,229	2,228	2,232	2,091	2,230	2,042	2,230
R^2	0.15	0.15	0.15	0.15	0.15	0.14	0.16	0.15	0.17	0.15
ln(Establishments)	1.329*	1.769**	1.449*	1.579**	0.867	3.203***	1.190	1.329*	1.267	1.467
	(0.751)	(0.768)	(0.845)	(0.690)	(0.694)	(0.968)	(0.760)	(0.713)	(0.773)	(0.892)
N	2,028	2,028	2,028	2,028	2,028	2,028	1,913	2,028	1,845	2,028
R^2	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.27	0.67	0.68

Note: Specification includes same controls as baseline in column 3 of Table 2, repeated in column (1). Column (2) uses a simple arithmetic average for times across car owners and non-car owners as described in Appendix C.3. Column (3) uses slower commute times that best match average speeds in the post-period. Column (4) uses faster ones that do the same for the pre-period. Recall the baseline specification uses an average of the two. Column (5) uses a larger value of θ equal to 1.5 times its baseline estimate, while column (6) scales it down by the same factor. Column (7) considers only census tracts closer than 3km from a TransMilenio station. Column (8) clusters standard errors by sector (560 administrative units above census tract). Column (9) excludes tracts within 1km of a portal (compared with 500m in the main table). Column (10) controls for log distance to CBD interacted with a dummy for whether a tract is closer than 500m to a TM station as of 2013 to examine whether all the CMA effect is simply due to heterogeneity of the distance effect at different distances from the CBD. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported in parentheses in columns other than (8).

Table A.2: Aggregate Welfare Effects: Robustness

	No Migration	Migration
Panel A: Alt. Estimated Params		
Baseline	2.28	0.60
IV	2.25	0.39
IV-Loc	2.45	0.67
Alternative Times	2.07	0.18
θ OLS	2.57	0.69
θ IV	1.16	0.29
Panel B: Alt. Calibrated Params		
$\sigma = 4$	2.53	1.19
$\sigma = 8$	2.17	0.48
$\beta = 0.8$	2.51	0.70
$\beta = 0.7$	2.20	0.56
$\rho = 6$	2.28	0.57

Note: Table shows the percentage change in average welfare (as defined in Table 6) under alternative parameter values using the sufficient statistics approach. Panel A examines sensitivity to alternative values of estimated parameters. The first row recreates the baseline results. The second row uses the CMA elasticities from the second column of Table 5 which instrument for the realized change in CMA (i.e. the term that does not hold residential population and employment fixed at their initial value in the post-period) using the baseline measure. The third row uses the CMA elasticities from the third column of Table 5 when instrumenting for the realized change in CMA using the LCP and Tram instrument. The fourth row uses the coefficients from column 6 of Table 2, using an alternative method to aggregate mode-specific commute times. The fifth row uses an alternative value for $\theta = 3.97$ estimated via OLS in column 3 of Table A.20. The six row uses a value for $\theta = 6.15$ estimated via IV using the LCP and Tram instrument in column 4 of Table A.20. Panel B varies the value of calibrated parameters.

Table A.3: List of Parameters and Sources of Identifying Variation

Panel A: Externally Calibrated Parameters

Parameter	Description	Identification Source
σ	Elasticity of substitution between labor types	Card (2009)
σ_D	Elasticity of demand	Feenstra et. al. (2018)

Panel B: Internally Calibrated Parameters

Parameter	Description	Identification Source
α_s	Cost Share of Commercial Floorspace	Same as description
β	Long-run housing expenditure share	Expenditure share on housing at high income levels
α_{sg}	Skill-specific labor demand shifters	Share of industry wage bill paid to high-skill workers
$\bar{h}$	Subsistence housing requirement	Average expenditure on housing
p_a	Cost of cars	Average expenditure on cars
T_g	Location parameter of worker productivity distribution	College wage premium

Panel C: Estimated Parameters

Parameter	Description	Identification Source
b_m	Travel mode preference shifter	Mode choice shares conditional on travel times
κ	Dependence of commute costs on travel times	Sensitivity of mode choices (within commutes) to travel times
λ	Correlation of preference shocks in public mode nest	Differential sensitivity of mode choices to travel times amongst public modes
θ_g	Commuting elasticity	Sensitivity of commute choices to travel times (in changes)
η_g	Resident supply elasticity	Sensitivity of residential populations to instruments for RCMA
$\mu_{U,g}$	Amenity externality	Sensitivity of residential populations to shifts in the share of high-skilled residents induced by instruments*
μ_A	Productivity externality	Sensitivity of model productivity residual to shifts in labor supply induced by instruments for FCMA

*Note: Instruments are the differential growth inf instrumented RCMA for high-type vs low-type, and the growth of instrumented RCMA for high-skill interacted with initial house prices (controls allowing for a separate effect of initial house prices on population growth also included).

Table A.4: Multigroup Model: Robustness

Panel A: Main Robustness

	Avg Welfare	**Inequality**	**Output**	**Rents**
Baseline	1.007	0.546	2.091	2.143
Migration	0.146	0.044	4.496	5.032
$\sigma_L = 2.5$	1.294	0.444	2.045	2.188
Census Employment	1.009	0.545	2.092	2.148
$\sigma = 4$	0.916	0.595	2.137	2.049
$\sigma = 9$	1.077	0.512	2.060	2.201
θ PPML	2.657	0.493	2.998	3.194
θ PPML Diff	0.888	0.851	2.027	1.916
θ OLS	1.960	0.237	2.687	2.945
Joint Pref. Shock	0.831	0.917	0.440	0.585

Panel B: Net Benefit Under Multigroup Model

	No Migration	**Migration**
% Net Increase GDP	1.47	3.88

Notes: Panel A shows main results (constructed in the same way as Table 9. Row 1 reproduces the main results. Row 2 uses the open city model with migration elasticity of $\rho = 3$ for both groups. Row 3 uses a larger value of the elasticity of substitution between skill groups in production, using the value of 2.5 from Card (2009) estimated at the MSA-level in the US. Row 4 uses census employment measured in 2005 instead of the CCB employment measured in 2015 as the measure of employment in the baseline equilibrium. Rows 5 and 6 use alternative values for the elasticity of demand. Rows 7, 8 and 9 use alternative values of θ_g estimated in columns 1, 3 and 5 of Table 7 respectively. Row 10 has a joint decision over residence and workplace location (with workers having an idiosyncratic preference for each pair). Panel B recreates the net increase in GDP from Panel B of Table 6 for the multigroup model.

Table A.5: Unweighted Establishment Regressions

	(1)	(2)	(3)
Weighted	2.101***	1.787***	1.168*
	(0.611)	(0.619)	(0.604)
N	2,028	2,028	2,028
R^2	0.21	0.23	0.27
Unweighted	1.050*	1.175**	0.697
	(0.550)	(0.555)	(0.547)
N	2,028	2,028	2,028
R^2	0.24	0.24	0.27
Locality FE	X	X	X
Log Dist CBD X Region FE	X	X	X
Basic Tract Controls	X	X	X
Historical Controls		X	X
Land Market Controls			X

Note: First row reports the establishment regressions from the first three columns of the main table (Table 2), where observations are weighted by a tracts share of total establishments in the initial period. Second row reports the same specifications without weights. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.6: Gravity Equation: Single Group, Full Estimates

Panel A: Main Regression

	PPML	PPML	OLS	IV
ln Commute Cost	-0.036**	-0.039**	-0.035*	-0.071***
	(0.017)	(0.016)	(0.020)	(0.024)
N	710	710	576	576
Controls X Year FE		X	X	X

Panel B: Alternative Clustering

	PPML	PPML	PPML
ln Commute Cost	-0.039**	-0.039**	-0.039*
	(0.016)	(0.018)	(0.022)
N	710	710	710
Clusters	355	38	19
Clustering	O-D	O-t & D-t	O & D

Note: Panel A Outcome is the commute shares in levels (PPML) or logs (OLS). Observation is an origin-destination-year cell. Only trips to work during rush hour (hour of departure 4-8am) by individuals 18-55 are included. Data is from 1995 and 2015 mobility surveys. Columns 1-2 estimate PPML models, 3 and 4 OLS and IV models respectively. The last column instruments for travel times in the post-period using the the average change in times across the LCP and tram instruments. Route-level controls are (i) the average number of crimes per year from 2007-2014, (ii) the average log house price in 2012 and (iii) the share of the trip that takes place along a primary road along the least-cost routes between origin and destination. Robust standard errors are reported in parentheses. Panel B repeats the baseline specification (column 2 of Panel A) with alternative levels of clustering (origin-destination pair; origin-year and destination-year; origin and destination). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.7: Costs and Benefits

	No Migration	**Migration**
NPV Increase GDP (mm)	43619.74	211452.29
Capital Costs (mm)	1449.75	1449.75
NPV Operating Costs (mm)	7180.53	7180.53
NPV Total Costs (mm)	8630.28	8630.28
NPV Net Increase GDP (mm)	34989.46	202822.00
% Net Increase GDP	2.50	14.51

Table A.8: Note: Table shows the costs and net benefits, computing net present values (NPV) over a 50 year time horizon with a 5% interest rate. All numbers are in millions of 2016 USD. The NPV of the increase in GDP is simply the NPV of the change in Bogotás GDP in dollar values. Capital costs are the one-time infrastructure costs of building the network. Total costs are the one-time capital costs associated with building the network combined with the NPV of operating costs. The NPV net increase in GDP nets this out from the gross gains in the first row, while the final row converts this back into a fraction of 2016 GDP.

Table A.9: α_{Hs} Across Industries

Industry	α_{Hs}	Relative HS Wage Bill
Domestic Services	0.160	0.055
Hotels & Restaurants	0.420	0.376
Social & Health Services	0.508	0.623
Transport & Storage	0.515	0.647
Construction	0.552	0.802
Wholesale, Retail, Repair	0.583	0.959
Manufacturing	0.599	1.056
Real Estate	0.601	1.066
Agriculture	0.628	1.254
Arts, Entertainment & Recreation	0.639	1.342
Other Services	0.701	2.016
Water Treatment and Distribution	0.729	2.441
Public Administration	0.769	3.322
Foreign Orgs	0.773	3.430
Elec, Gas	0.800	4.303
Social & Health Services	0.801	4.343
Information & Communication	0.804	4.458
Professional, Scientific and Technical Activities	0.837	6.154
Mining	0.846	6.761
Education	0.854	7.436
Financial Brokerage	0.865	8.455

Note: See Section D.2 for details.

Table A.10: Employment Data Summary Statistics

Year	N Est.	Mean Emp.	p10	p50	p90
Panel A: Census					
1990	219,812	5.41	1	2	7
2005	625,852	4.93	1	2	6
Panel B: Chamber of Commerce					
2000	34,322				
2015	126,867	2.37	1	1	4

Note: The first column provides the number of establishments in each dataset, column (2) provides the average employment while columns (3)-(5) report percentiles of the firm size distribution. Employment is not reported in the raw 2000 Chamber of Commerce establishment data.

Table A.11: Trip Characteristics in 2015

	Bus	Car	Walk	TM
Share of all trips	0.343	0.137	0.360	0.161
Mean Distance (km)	6.683	6.178	1.526	10.487
Share of (trip purpose)				
To work	0.478	0.150	0.158	0.214
Business trips	0.289	0.333	0.184	0.193
To school	0.292	0.042	0.502	0.164
Private matters	0.267	0.163	0.450	0.120
Shopping	0.149	0.121	0.678	0.052

Note: Table created using data from the 2015 Mobility Survey.

Table A.12: Commute Characteristics over Time

Mode	Bus	Car	Walk	TM
Panel A: Commute Shares				
1995	0.74	0.17	0.09	
2005	0.66	0.17	0.07	0.11
2011	0.46	0.16	0.19	0.19
2015	0.48	0.15	0.16	0.21
Panel B: Commute Speeds (kmh)				
1995	16.31	25.37	8.20	
2005	12.88	15.65	6.53	16.88
2011	10.49	14.02	7.95	13.08
2015	10.37	12.95	6.36	13.04
Panel C: Ownership shares				
1995		0.29		
2005		0.28		
2011		0.25		
2015		0.25		

Note: Only trips to work included in trip-level data (car ownership is at the household level).

Table A.13: Effect of TransMilenio on other Mode Speeds

Outcome: $\ln(\text{Speed})$	(1)	(2)	(3)	(4)
Panel A: Car Trips				
TM Route X Post	-0.107	-0.060	0.014	0.052
	(0.086)	(0.089)	(0.064)	(0.065)
R^2	0.80	0.80	0.80	0.80
N	9,916	9,916	9,916	9,916
Panel B: Bus Trips				
TM Route X Post	-0.164***	-0.074	-0.064	-0.020
	(0.046)	(0.047)	(0.041)	(0.040)
R^2	0.72	0.72	0.72	0.72
N	38,616	38,616	38,616	38,616
Route Measure	Share TM	Share TM	TM>75%	TM>75%
Baseline Controls	X	X	X	X
Locality Origin X Post FE	X	X	X	X
Locality Destination X Post FE	X	X	X	X
Log Distance X Post FE		X		X

Note: Observation is a UPZ Origin-UPZ Destination-Year. Outcome is log reported speed from the 1995 and 2015 Mobility Surveys. Share TM is the share of a car trip's least cost route that lies along a TM line. TM>75% is a dummy equal to one if the share is greater than 75%. Baseline controls are a gender dummy, hour of departure dummies and age quantile dummies, each interacted with year dummies. Only trips to work included during rush hours included. Panel A includes only trips by car, while panel B includes only those by bus. Standard errors clustered at the origin-destination pair-level. $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.14: Relationship between Predicted and Observed Times Over Time

	(1)	(2)	(3)	(4)
ln(Predicted Time)	0.705***	0.511***	0.655***	0.697***
	(0.034)	(0.020)	(0.032)	(0.023)
Post	0.317*	-0.662***	0.151	
	(0.190)	(0.126)	(0.216)	
ln(Predicted Time) X Post	0.018	0.187***	0.046	
	(0.051)	(0.030)	(0.052)	
Car				-0.037
				(0.167)
TM				0.020
				(0.193)
ln(Predicted Time) X Car				0.026
				(0.044)
ln(Predicted Time) X TM				0.003
				(0.047)
R^2	0.42	0.34	0.39	0.42
N	2,219	6,671	2,419	5,005
Mode	Car	Bus	TM	All
Post Only				X

Note: Observation is a UPZ Origin-UPZ Destination-Year. Outcome is log reported time from Mobility Survey. Post is a dummy equal to one in 2015 and zero in 1995 (2005 for TM). Trips include journeys to and from work during rush hour (hour of departure between 5 and 8 am, hour of return between 4 and 6pm). Individual observations averaged to the trip-year level, and regressions weighted by number of individual observations in each trip-year-mode. Columns (1)-(3) include observations for pre- and post years and consider only one mode; column (4) includes only observations from the post period and includes all modes. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.15: Effect of TransMilenio on Growth in Floorspace

Outcome: Floorspace Growth	(1)	(2)	(3)	(4)
Δ ln RCMA	-0.084			
	(0.211)			
Δ ln FCMA		-0.106		
		(0.286)		
ln Distance F1			0.014	0.013
			(0.014)	(0.018)
ln Distance F2			0.016	0.009
			(0.015)	(0.020)
ln Distance F3			-0.014	-0.019
			(0.026)	(0.027)
ln Distance F1 X Far CBD				0.005
				(0.025)
ln Distance F2 X Far CBD				0.014
				(0.026)
ln Distance F3 X Far CBD				0.002
				(0.039)
N	2,235	2,233	2,205	2,205
R^2	0.34	0.34	0.33	0.33

Note: Specification is baseline specification from main table with full controls (column (3)), but outcome is growth in floorspace between 2018 and 2000 using the Davis-Haltiwanger measure. In column 3 the coefficients report the log distance from the closest station in each phase of TransMilenio. Column 4 interacts this with a dummy for whether the tract is above the median distance from the CBD (Far CBD). The full interaction is included. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported in parentheses. *p<0.1; ** $p < 0.05$; *** $p < 0.01$

Table A.16: TransMilenio's Effect on Income

	(1)	(2)	(3)
Panel A: Clustered by UPZ X Post			
ln(RCMA)	0.982***	0.510**	0.522**
	(0.349)	(0.224)	(0.224)
N	87,674	87,673	87,673
R^2	0.48	0.56	0.1563
P-val Coef = $1/\theta$			0.31
Panel B: Clustered by UPZ			
ln(RCMA)	0.982**	0.510*	0.522*
	(0.451)	(0.288)	(0.287)
N	87,674	87,673	87,673
R^2	0.48	0.56	0.1563
P-val Coef = $1/\theta$			0.43
UPZ FE	X	X	X
Region X Year FE	X	X	X
Log Dist CBD X Region X Year FE	X	X	X
Basic Tract Controls X Year FE	X	X	X
Historical Controls X Year FE	X	X	X
Land Market Controls X Year FE	X	X	X
Basic Worker Demographics X Year FE	X	X	X
Education X Year FE		X	X
Hours Worked X Year FE			X

Note: Outcome variable is the log average weekly labor income for full-time, working age (18-65) individuals reporting more than 40 hours worked per week. Data covers 2000-2005 in the pre-period and 2015-2019 in the post period and comes from the ECH and GEIH. Post is a dummy for the post period. RCMA is measured at the UPZ-level using the pre-TM network in the pre-period, and using the 2013 network in the post-period, and at the UPZ-level. Region are dummies for the North, West and South of the city. Controls present are the same as in the main specification (interacted with year dummies), other than basic worker demographics which contain dummies for age (ine 10 year bins) and gender. Columns 2 and 3 contain dummies for each category of highest education level attained. Column 3 contains dummies for hours worked per week in 10 hour bins. Standard errors are clustered by UPZ and period. The p-value tests the null that the coefficient on log RCMA equals $1/\theta$ as predicted by the model, with $\theta = 3.39$. Standard errors are clustered by UPZ and Post in Panel A, and by UPZ in Panel B. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.17: TransMilenio's Effect on the College Share of Residents

	(1)	(2)
Δ lnRCMA	0.053*	0.061**
	(0.031)	(0.031)
N	2,106	2,106
R^2	0.15	0.18
Init. Coll Share		X

Note: Outcome is the change in the share of college educated residents in a tract between 1993 and 2018. Specification includes all controls from baseline specification, excluding the initial college share in column 1 but including it in column 2. HAC standard errors are reported with a 500m bandwidth. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.18: Impacts of Both Types of CMA

	(1)	(2)	(3)
Panel A: Residents			
	Δ ln(Res Price)	Δ ln(Res Pop)	
ΔlnRCMA	0.187	1.048***	
	(0.176)	(0.388)	
ΔlnFCMA	0.307	-1.107	
	(0.242)	(0.677)	
N	2,161	2,228	
R^2	0.43	0.37	
Panel B: Firms, Floorspace			
	Δ ln(Comm Price)	Δ Comm Share	
ΔlnFCMA	0.441	0.553***	
	(0.321)	(0.101)	
ΔlnRCMA	0.160	-0.352***	
	(0.279)	(0.070)	
N	2,048	2,194	
R^2	0.11	0.16	
Panel C: Firms, Employment			
	Δ ln(Est, CCB)	Δ ln(Emp, Census)	Δ ln(Emp Formal, Census)
ΔlnFCMA	-1.127	1.384	2.562
	(0.832)	(1.179)	(1.577)
ΔlnRCMA	3.419***	0.036	-0.869
	(0.769)	(0.519)	(0.801)
N	2,028	1,943	1,653
R^2	0.28	0.23	0.16

Note: Table repeats the baseline specification i.e. column (3) from Table 2 and columns (1) and (3) from Table 4 for census employment, including both the change in RCMA and FCMA in the same regression. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors (Conley (1999)) with a 0.5km bandwidth reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.19: Commuting in 1995

	lnSpeed	lnSpeed	Bus	Bus
Bus	-0.363***	-0.309***		
	(0.020)	(0.016)		
Low-Skill			0.287***	0.163***
			(0.010)	(0.011)
R^2	0.06	0.76	0.18	0.47
N	14,945	12,975	18,843	16,461
UPZ O-D FE		X		X
Time of day Controls	X	X	X	X
Demographic Controls	X	X	X	X

Note: Data is from 1995 Mobility Survey. Low-Skill is a dummy for having no post-secondary education. Bus is a dummy for whether bus is used during a commute, relative to the omitted category of car. Time of day controls are dummies for hour of departure, and demographics are log age and a gender dummy. UPZ O-D FE are fixed effects for each upz origin-destination. Only trips to work during rush hour (hour of departure between 5-8am) included. Standard errors clustered at upz origin-destination pair. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.20: Aggregate Gravity Equation

	PPML	PPML	OLS	IV
ln Commute Cost	-0.036**	-0.039**	-0.035*	-0.071***
	(0.017)	(0.016)	(0.020)	(0.024)
N	710	710	576	576
Controls X Year FE		X	X	X

Note: Outcome is the log number of commuters between each origin and destination locality pair in 1995 or 2015. Only trips to work during rush hour (5-8am) by heads of households included. Fixed effects for each origin locality-year, destination locality-year, and origin-destination pair included in each specification. Controls include (i) the average number of crimes per year from 2007-2014, (ii) the average log house price in 2012 and (iii) the share of the trip that takes place along a primary road along the least-cost routes between origin and destination. Columns 1 and 2 run PPML specifications (with column 2 corresponding to the main value from the text), column 3 runs OLS and column 4 runs an IV using the same instrument as column 3 of Table 5. Standard errors clustered at the origin-destination pair-level are reported.* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

B Additional Figures

Figure A.1: TransMilenio Network and Bogotá

(a) TransMilenio Routes

(b) Population Density in 1993

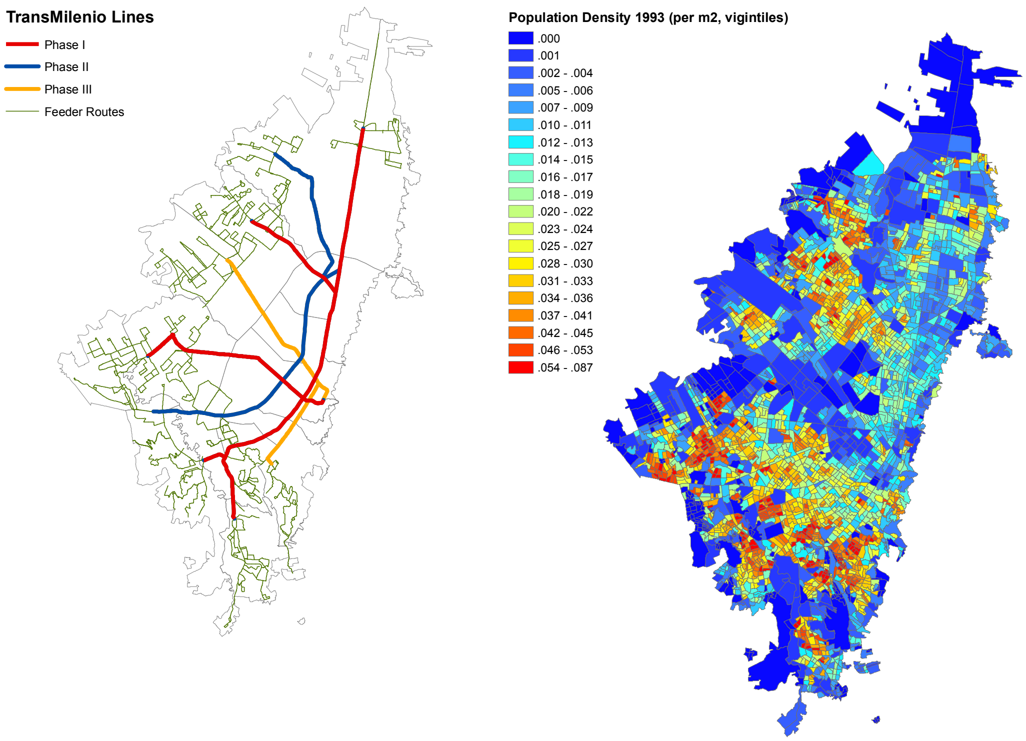


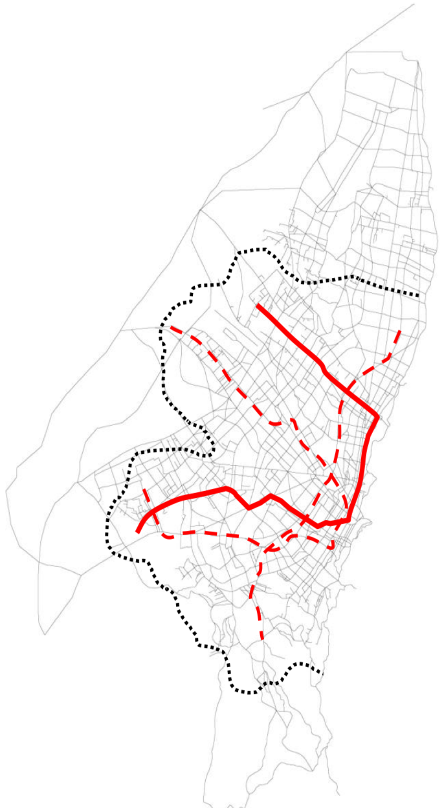
Figure A.2: TransMilenio Routes: Before and After

(a) Previous bus lanes, Avenida Caracas (Sur)

(b) TransMilenio Station, Avenida Caracas (Norte)

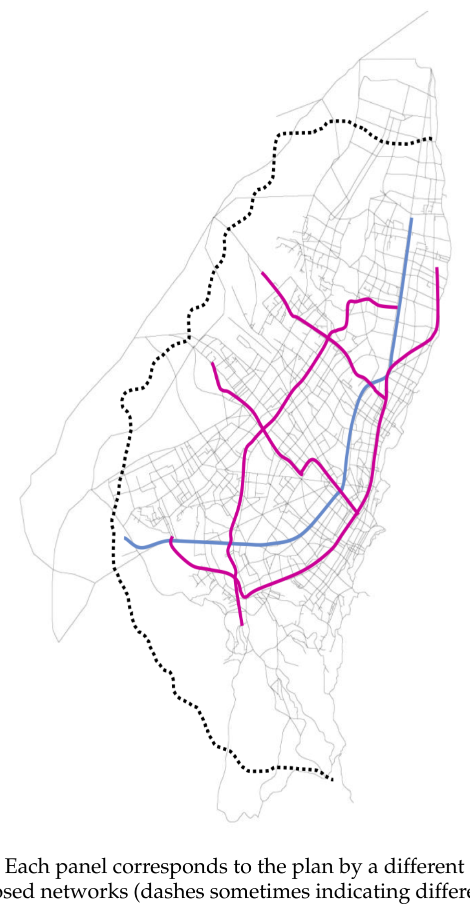
Figure A.3: Planned Networks From Previous Studies

(a) Ineco-Sofretug, 1981

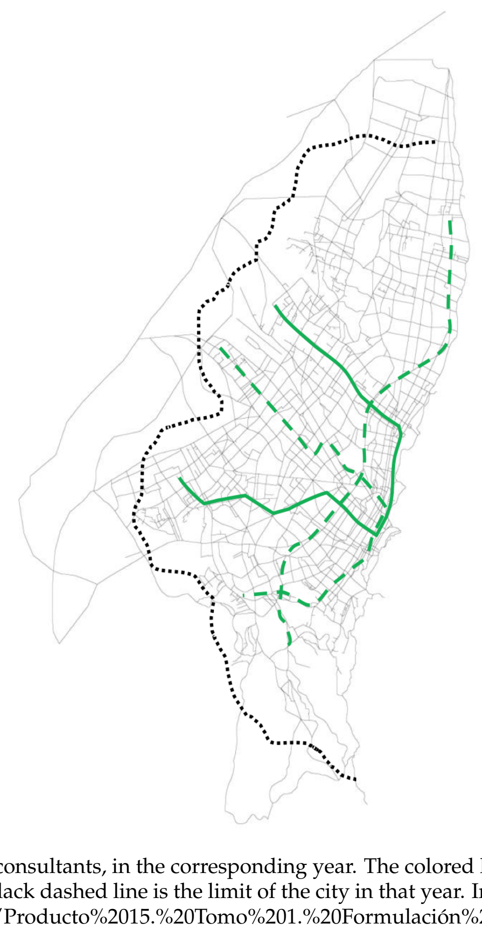


(b) Intermetro-SPA, 1987

(c) JICA, 1996

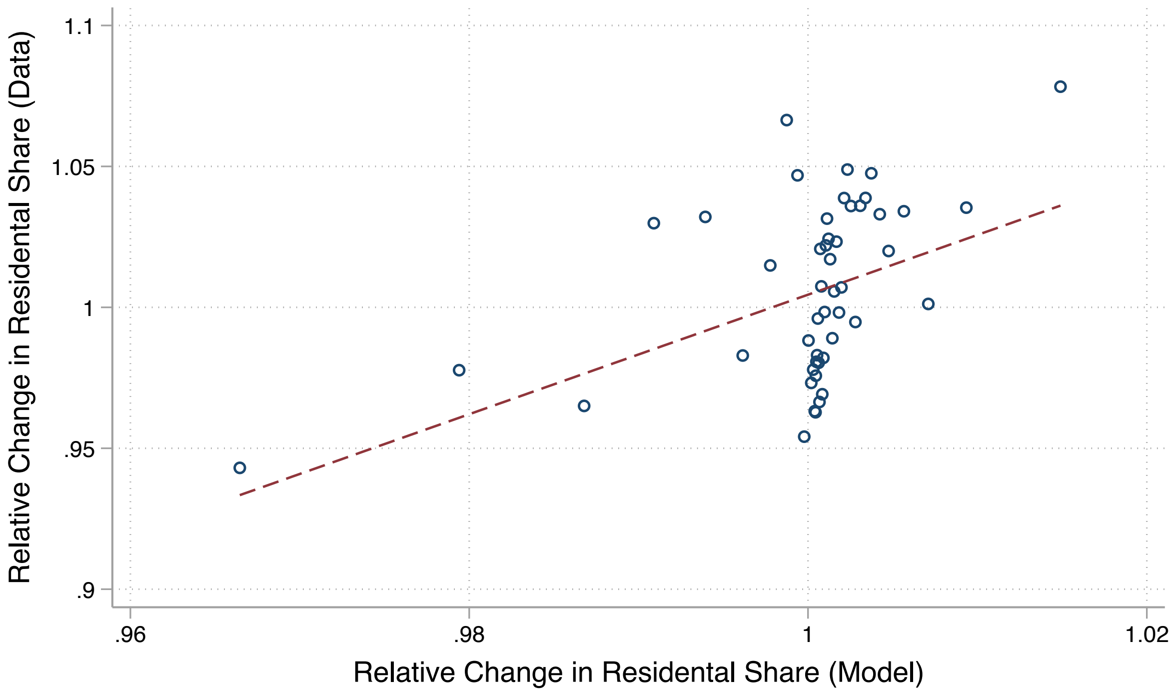


(d) Bechtel-Systra-Ingetec, 1997



Note: Each panel corresponds to the plan by a different consortium of consultants, in the corresponding year. The colored lines are the proposed networks (dashes sometimes indicating different lines), the black dashed line is the limit of the city in that year. Images obtained from https://www.metrodebogota.gov.co/sites/default/files/documentos/Producto%2015.%20Tomo%201.%20Formulación%20y%20caracterización%20de%20las

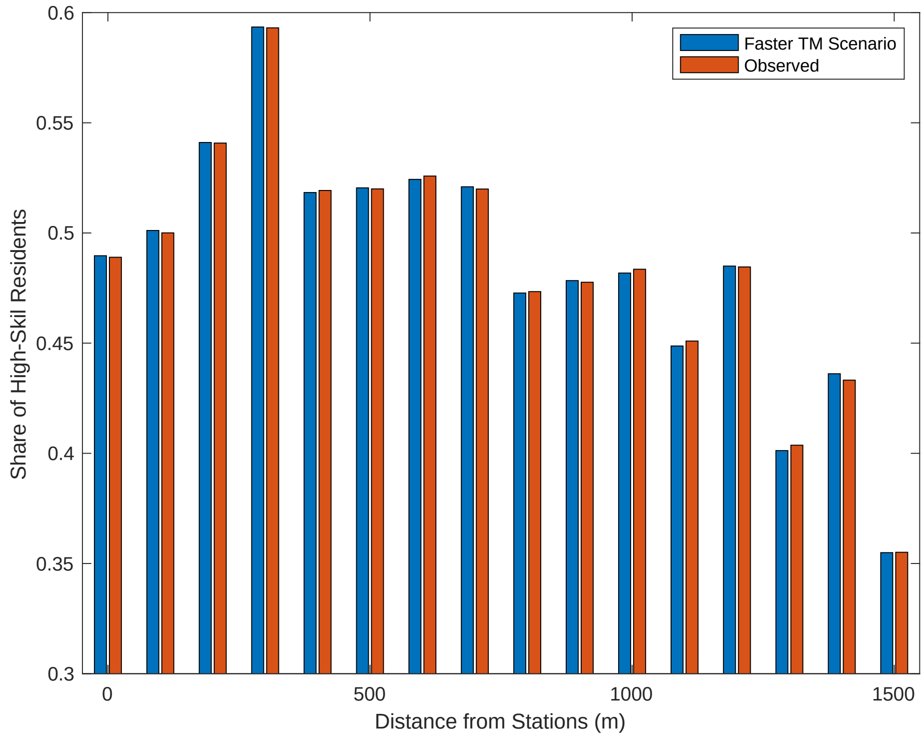
Figure A.4: Planned Networks From Previous Studies



Regression of the observed change in the data on the model-predicted change has coefficient 1.598 (0.822),and a p-value for the F-test that the slope is 1 of 0.467.

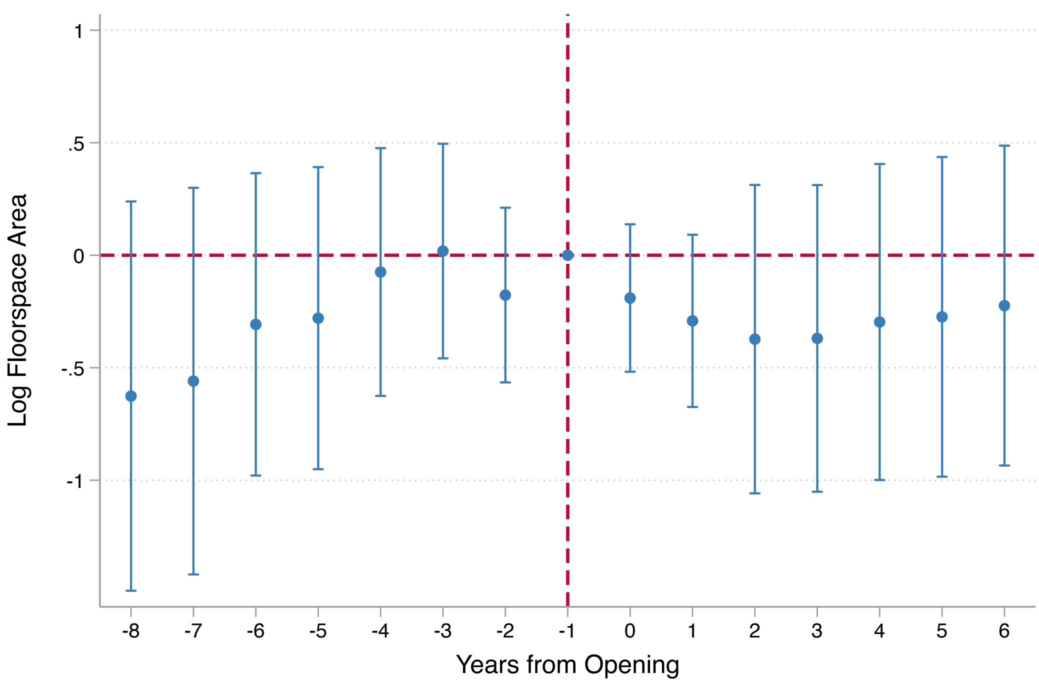
Note: Graph plots a binscatter (50 bins) the observed change in the share of floorspace used for residential purposes in the data (y-axis) vs the model (x-axis). Both are normalized to have unit mean on the plot. Graph caption also reports results from regression of the change in the share of floorspace used for residential purposes in the data on that from the model.

Figure A.5: College Share in Observed vs Counterfactual Equilibrium with Faster TM



Note: Graph shows share of high-skill residents in tracts in 100m cells from their nearest TransMilenio station. 1500m cell includes all tracts 1500m or more from their nearest station. Red bars show the observed shares in the post-period, blue bars show those from a counterfactual where TransMilenio runs at 35 km/h.

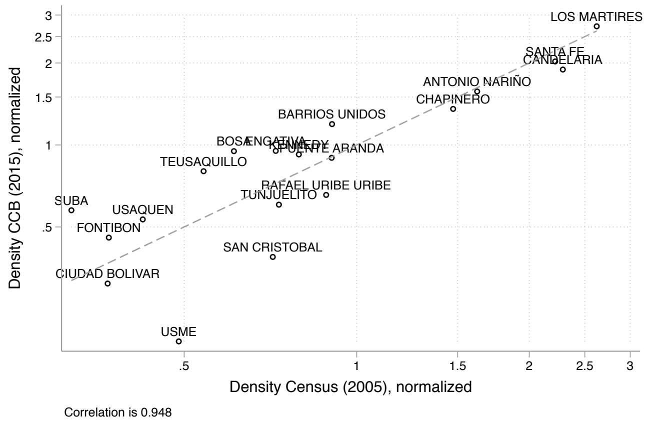
Figure A.6: Event Study on Floorspace Area



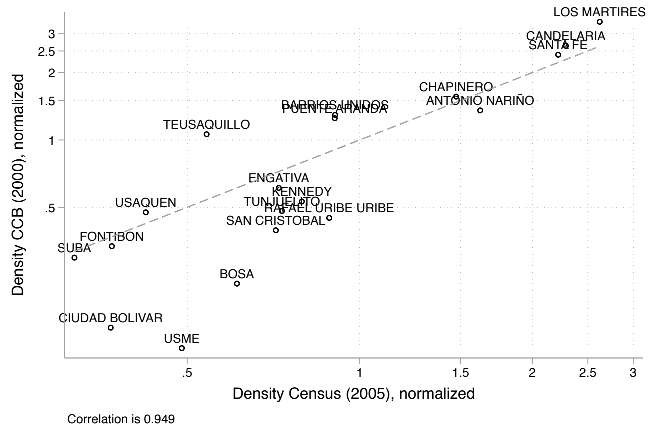
Note: Figure plots event study similar to Figure 3 but using log floorspace area as the outcome.

Figure A.7: Employment in Chamber of Commerce vs Census

(a) 2015 Establishment Comparison by Locality



(b) 2000 Establishment Comparison by Locality



(c) Establishment Comparison by Sector

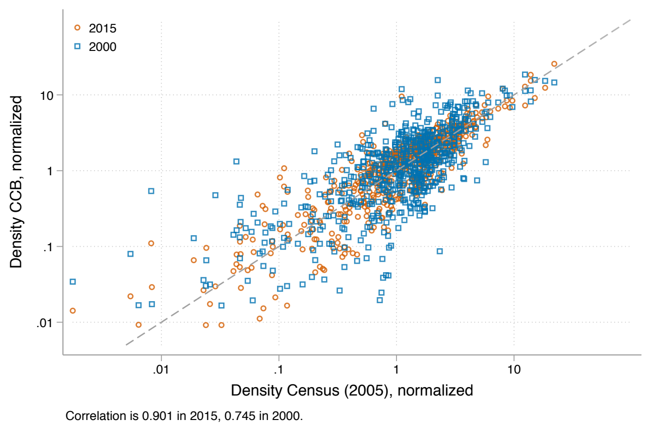
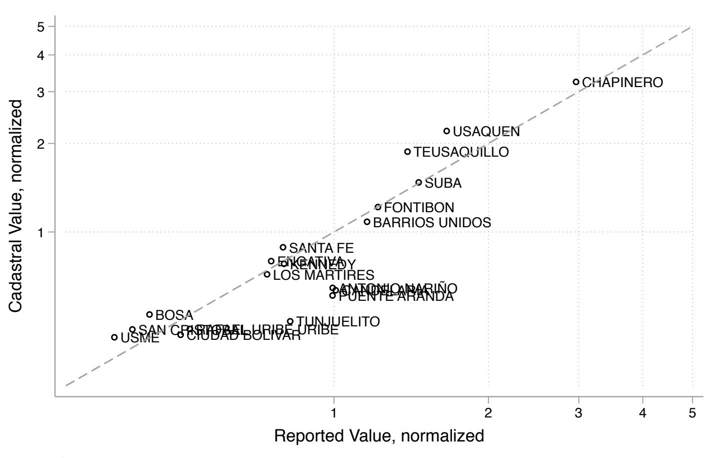


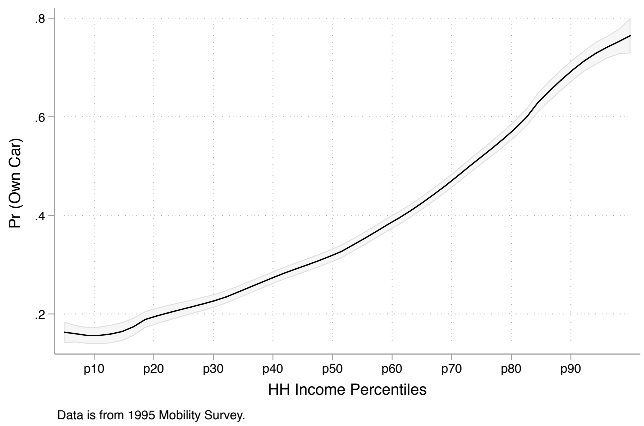
Figure A.8: Cadastral vs Reported Property Values



Note: Reported value is the reported purchase price per room as observed in the Multipurpose survey in 2014, for properties bought after 2005 (both the purchase price and year are reported). The cadastral value is the average residential property value per m2 in the locality in that year. Prices are averaged over the period, and normalized so that each variable has mean one.

Figure A.9: Engel Curves for Car Ownership and Housing

(a) Car Ownership



(b) Housing Expenditure

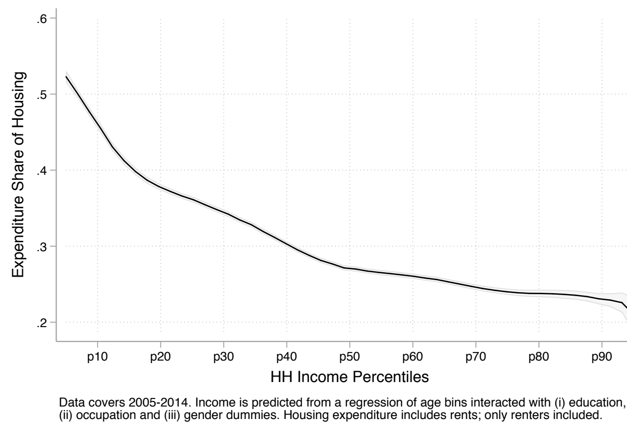
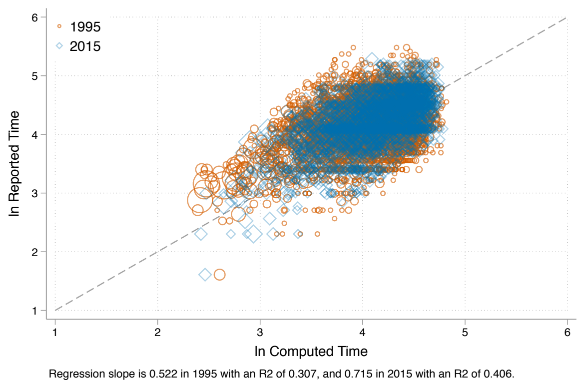
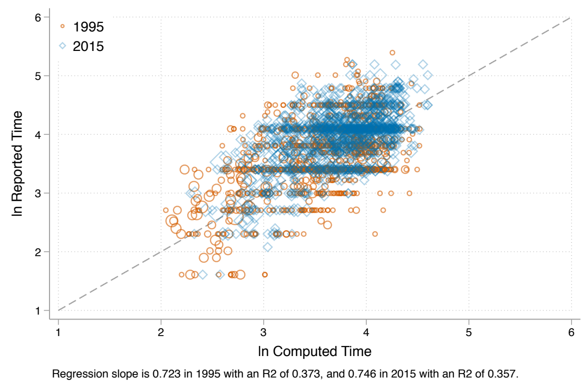


Figure A.10: Computed vs Observed Commute Times

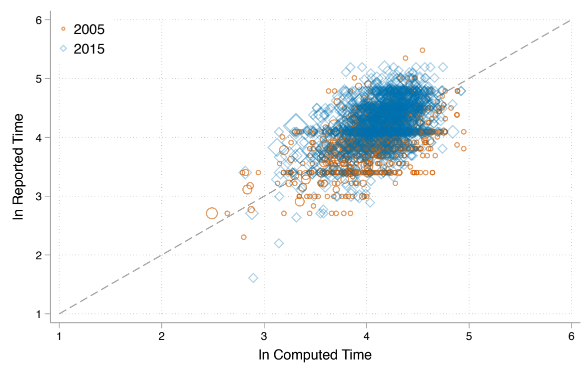
(a) Buses



(b) Cars



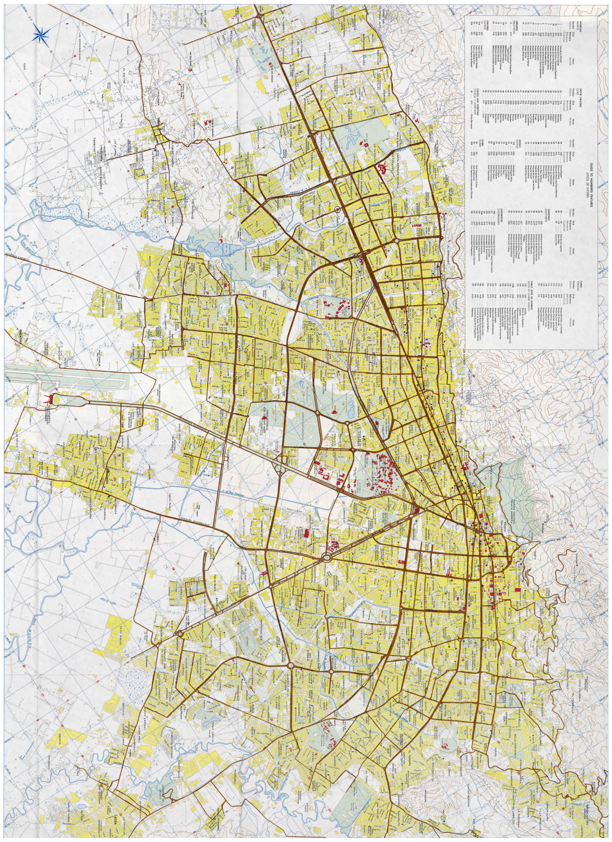
(c) TransMilenio



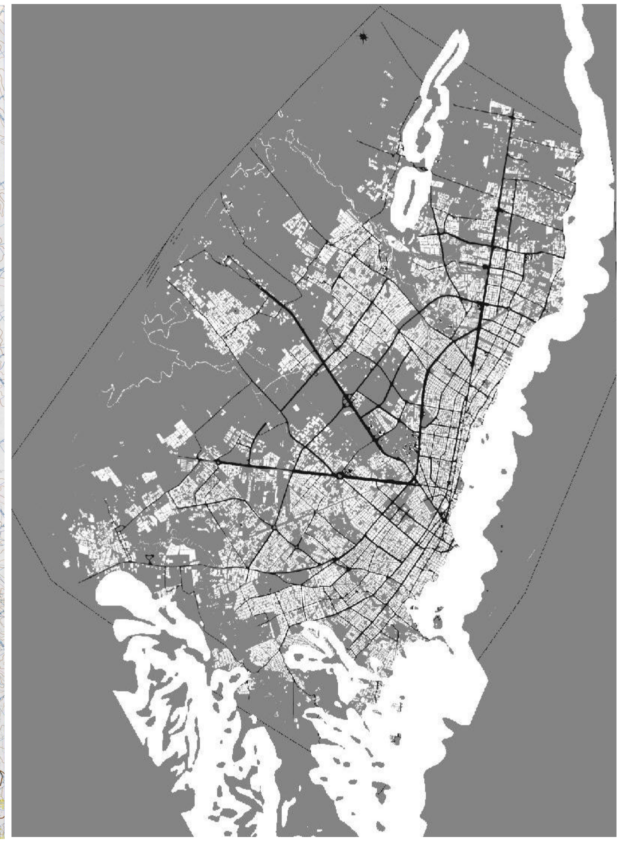
Note: Figures plot the average reported trip time between pairs of UPZs in the Mobility Survey versus the times computed in ArcMap using the pre speeds for 1995 and post speeds for 2015. Only trips to and from work during rush hour included. Marker size is proportional to the number of commuters in each pairwise combination (reported coefficients from regressions weighted by this number).

Figure A.11: Instruments

(a) Raw Land Use Map 1980

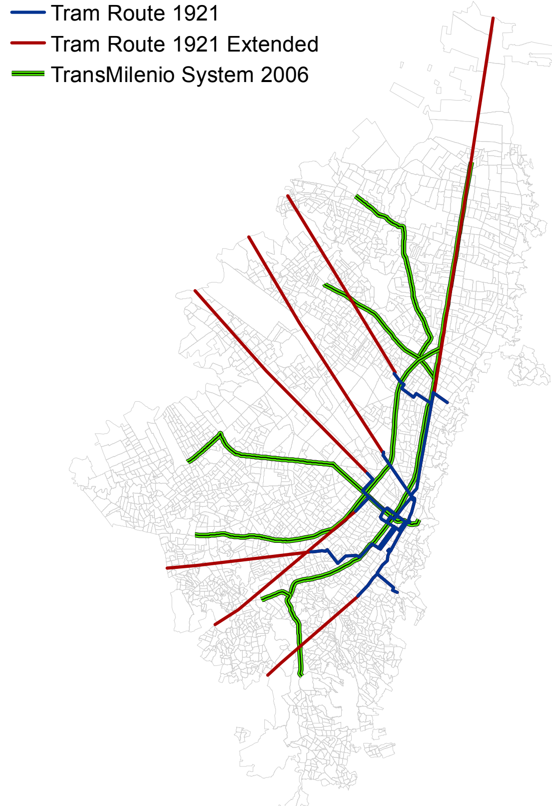


(b) Cost Raster



(c) LCP Instrument

(d) Tram Instrument



C Using A Special Case of the Model to Derive Sufficient Statistics for the Impact of Transit Infrastructure on Economic Activity

This section considers a special case of the model where there is one type of worker and firm, no fixed element of expenditure or income and a fixed allocation of floorspace to residential and commercial use. For simplicity, I assume workers make a joint decision over home and workplace but this is later relaxed to have separate decisions as in the main model. This special case is shared by a wider class of quantitative urban models. Section C.1 sets up and characterizes this simple model from scratch, and shows it admits a reduced form representation where changes in endogenous variables can be written as log-linear functions of changes in CMA. Section C.2 shows that (i) the change in CMA and elasticities of economic activity to CMA turn out to be sufficient statistics that speak to the impact of transit infrastructure on aggregate outcomes (such as house prices, output and welfare) as well as the reorganization of activity across space. Section C.5 derives a relationship between first order welfare effects in this class of general equilibrium models and the value of time savings approach typically used to evaluate gains from transit infrastructure. Section C.8 provides proofs for the results in this section.

C.1 A Simple Quantitative Urban Model

Setup. I consider a simple quantitative model of a city in the spirit of Ahlfeldt et. al. (2015) and Allen et. al. (2015). There are $i \in I$ locations that differ in their exogenous amenities $\bar{u}_i$, productivities $\bar{A}_i$, residential and commercial floorspace supplies H_{Ri}, H_{Fi} and the time t_{ij} it takes to commute to any other location.[53] A continuum of workers with mass $\bar{L}$ choose where to live and work and have Cobb-Douglas preferences over a freely-traded numeraire good and housing. Commuting reduces effective labor supply at workplace so that an individual living in i and working in j receives income w_j/d_{ij}, where $d_{ij} = \exp(\kappa t_{ij})$ converts commute times into commute costs. In each location, a representative firm produces a freely traded variety under perfect competition that are aggregated by consumers in CES fashion to form the final numeraire good.

Individuals. Indirect utility across pairs of residential and employment locations (i, j) is given by

$$U_{ij}(\omega) = \frac{u_i w_j r_{Ri}^{\beta-1}}{d_{ij}} \epsilon_{ij}(\omega), \tag{21}$$

where $\epsilon_{ij}(\omega)$ is an idiosyncratic productivity for worker ω on commute (i, j), $1-\beta$ is the expenditure share on housing, and u_i is the amenity enjoyed by residents who live in i. To allow for the possibility of local spillovers, amenities depend on both exogenous location characteristics $\bar{u}_i$ and the number of residents through $u_i = \bar{u}_i L_{Ri}^{\mu_U}$. Workers choose the commute pair that maximizes their utility. Assuming these are drawn iid from a Frechet distribution with shape parameter θ yields a simple expression for the number of commuters for each live-work pair

$$L_{ij} = \bar{L}\bar{U}^{-\theta} \left(\frac{u_i w_j r_{Ri}^{\beta-1}}{d_{ij}} \right)^{\theta}, \tag{22}$$

[53] Appendix G shown housing supply was unaffected by TransMilenio, so I consider these as fixed location characteristics. This assumption is relaxed in Section 5.3. Appendix G also shows that there were no significant relative changes in car and bus speeds along routes most affected by TransMilenio, so I assume travel times are fixed in the baseline model. This is relaxed in Panel C of Table 6.

where $\bar{U} = \gamma \left[\sum_{ij} (u_i w_j r_{Ri}^{\beta-1}/d_{ij})^\theta \right]^{1/\theta}$ is average utility, $\gamma = \Gamma\left(\frac{\theta-1}{\theta}\right)$ and $\Gamma(\cdot)$ is the Gamma function. The supply of residents and workers to each location can be computed by summing these flows over all destinations and origins respectively to get

$$L_{Ri} = \bar{L}\bar{U}^{-\theta} \left(u_i r_{Ri}^{\beta-1} \right)^\theta \Phi_{Ri} \tag{23}$$

$$L_{Fj} = \bar{L}\bar{U}^{-\theta} w_j^\theta \Phi_{Fj}. \tag{24}$$

The Φ_{Ri} and Φ_{Fi} terms are what I refer to as commuter market access terms. Residential commuter market access (RCMA) $\Phi_{Ri} = \sum_j (w_j/d_{ij})^\theta$ reflects residents' access to well-paid jobs from location i. Firm commuter market access (FCMA) $\Phi_{Fj} = \sum_i (u_i r_{Ri}^{\beta-1}/d_{ij})^\theta$ reflects firms' access to workers from location j (i.e. being close to locations with high amenities or low rents). The resident supply curve (23) therefore tells us that more residents will move to locations with high amenities, low house prices, and better access to well-paid jobs through the commuting network. The labor supply curve (24) tells us that firms will attract more workers to locations with high wages and better access to workers via the commuting network.

The supply of effective labor units to a location can be computed by leveraging that, under the Frechet distribution, the average productivity of workers who have chosen (i, j) is inversely related to the share of workers choosing that pair $\bar{\epsilon}_{ij} \propto {\pi_{ij}}^{-1/\theta}$ where $\pi_{ij} = L_{ij}/\bar{L}$. Total effective labor supply is simply $\tilde{L}_{Fj} = \bar{L} \sum_i \pi_{ij}^{\frac{\theta-1}{\theta}}/d_{ij}$, which simplifies to

$$\tilde{L}_{Fj} = \bar{L}\bar{U}^{-(\theta-1)} w_j^{\theta-1} \tilde{\Phi}_{Fj} \tag{25}$$

where $\tilde{\Phi}_{Fj} = \sum_i (u_i r_{Ri}^{\beta-1})^{\theta-1} d_{ij}^{-\theta}$ is adjusted FCMA capturing access to effective units of labor.

Consumers spend a constant fraction $1-\beta$ on housing, so that residential floorspace (inverse) demand is given by

$$r_{Ri} = \frac{1-\beta}{H_{Ri}} \bar{y}_i L_{Ri}, \tag{26}$$

where $\bar{y}_i \equiv \Phi_{Ri}^{1/\theta} L_{Ri}^{-1/\theta}$ is average income of residents in i.[54]

Firms. The production side of the model assumes an Armington structure with no trade costs. In each location, a representative firm produces a differentiated variety using the Cobb-Douglas technology $Y_i = A_i \tilde{L}_{Fi}^\alpha H_{Fi}^{1-\alpha}$. As for amenities, I allow for the possibility of productivity externalities of the form $A_i = \bar{A}_i \tilde{L}_{Fi}^{\mu_A}$.[55] Solving firms' profit maximization problem delivers labor demand

$$\tilde{L}_{Fi} = \frac{1}{\alpha} w_i^{\alpha(1-\sigma)-1} A_i^{\sigma-1} r_{Fi}^{(1-\sigma)(1-\alpha)} E \tag{27}$$

where $E = \sum_i \bar{y}_i L_{Ri}$ is aggregate expenditure and σ is the elasticity of demand across varieties. Firm (inverse)

[54] See Appendix C.8.4 for a derivation. The model with separate residential and employment location decisions covered in Appendix C.6 has the more familiar form $\bar{y}_i \equiv \Phi_{Ri}^{1/\theta}$.

[55] Given evidence on highly localized spatial spillovers (Rossi-Hansberg et. al. 2010; Ahlfeldt et. al. 2015), I do not allow for spillovers across locations given the size of census tracts. Previous versions of the paper show how the regression framework in that model still holds but outcomes depend both on a location's own CMA and those nearby.

demand for commercial floorspace is given by

$$r_{Fi} = \left(\frac{A_i^{\sigma-1} w_i^{-\alpha(\sigma-1)} P^{\sigma-1} E}{(1-\alpha) H_{Fi}} \right)^{\frac{1}{1+(\sigma-1)(1-\alpha)}} \tag{28}$$

Equilibrium. Given model parameters $\{\alpha, \beta, \sigma, \theta, \kappa, \mu_U, \mu_A\}$ and location characteristics $\{H_{Ri}, H_{Fi}, t_{ij}, \bar{u}_i, \bar{A}_i\}$, an equilibrium of the model is a vector $\{L_{Ri}, \tilde{L}_{Fj}, w_j, r_{Ri}, r_{Fj}, \bar{U}\}$ such that (i) the supply of residents and labor is consistent with worker optimality (23) and (25), (ii) the demand for labor is consistent with firm optimality (27), (iii) demand for floorspace is consistent with firm and worker optimal and equals supply (26) and (28) and (iv) the population of the city $\bar{L}$ is fixed, and welfare $\bar{U}$ is given by $\bar{U} = \gamma \left[\sum_{ij} (u_i w_j r_{Ri}^{\beta-1} / d_{ij})^{\theta} \right]^{1/\theta}$.[56]

C.2 Sufficient Statistics for Impacts of Transit Infrastructure

The following proposition shows how the model and related extensions admit a simple reduced form and sufficient statistics approach to quantify the impacts of changes in transit infrastructure.

Proposition 1. *Consider a change in commute costs from* $\mathbf{d}$ *to* $\mathbf{d}'$*, and let* $\hat{x} \equiv x'/x$ *denote relative changes in a variable between the pre- and post-period. Then*

Part 1: Reduced Form. *The model yields a reduced form where endogenous variables can be written as log-linear functions of CMA as*

$$\begin{aligned} \ln \hat{\mathbf{y}}_i &= \boldsymbol{\beta}_R \ln \hat{\Phi}_{Ri} + \tilde{\boldsymbol{\beta}}_{1,F} \ln \hat{\Phi}_{Fi} + \tilde{\boldsymbol{\beta}}_{2,F} \ln \hat{\tilde{\Phi}}_{Fi} + \mathbf{e}_i \\ &\approx \boldsymbol{\beta}_R \ln \hat{\Phi}_{Ri} + \boldsymbol{\beta}_F \ln \hat{\Phi}_{Fi} + \mathbf{e}_i \end{aligned}$$

where $\mathbf{y}_i = [L_{Ri}, r_{Ri}, r_{Fi}, L_{Fi}]$ *and* $\mathbf{e}_i$ *is a vector of structural residuals.* $\boldsymbol{\beta}_F$ *and* $\boldsymbol{\beta}_R$ *have zero elements in the first and last two entries respectively, so this is a system of 4 univariate regressions yielding 4 coefficients* $\beta_{L_R}, \beta_{r_R}, \beta_{r_F}, \beta_{L_F}$*. Unique (to-scale) values of the CMA terms* Φ_{Ri}, Φ_{Fi} *can be computed given data* $\{L_{Ri}, L_{Fi}, d_{ij}\}$ *and the commuting elasticity* θ*. While the first line holds exactly (given the values for* $\hat{\Phi}_{Ri}, \hat{\Phi}_{Fi}, \hat{\tilde{\Phi}}_{Fi}$ *which also depend on* $\hat{L}_{Ri}, \hat{L}_{Fi}$*), the second lines uses the first-order approximation* $\ln \hat{\tilde{\Phi}}_{Fi} \approx \frac{\theta-1}{\theta} \ln \Phi_{Fi}$ *around* $d_{ij}^{-\theta} = 0$*.*

Part 2: Relative Impacts of Transit Infrastructure*. Assuming that exogenous, location-specific characteristics are unchanged by the infrastructure, relative changes in endogenous variables* $\hat{\hat{\mathbf{y}}}_i \equiv \hat{\mathbf{y}}_i / \left(\prod_r \hat{\mathbf{y}}_i \right)^{1/I}$ *can be computed using (i) estimates of* $\beta_{L_R}, \beta_{r_R}, \beta_{r_F}, \beta_{L_F}, \theta$*, (ii) data on the initial distribution of economic activity* $\{L_{Ri}, L_{Fi}, d_{ij}\}$ *and (iii) data on the change in commute costs* $\{\hat{d}_{ij}\}$*.*

Part 3: Level Impacts of Transit Infrastructure*. Level changes in endogenous variables* $\hat{\mathbf{y}}_i$ *and endogenous constants* $\hat{\bar{L}}, \hat{\bar{U}}$ *can be computed from the relative changes obtained in part 2 with (i) an assumption on population mobility between the city and the rest of the country, and (ii) values for* σ, β*.*

Part 4: Isomorphisms*. Parts 1 and 2 apply to a more general class of models which feature (i) a gravity equation for commute flows and (ii) an equilibrium that can be written as a system of* K *equations in* K *endogenous variables* $\{y_{1i}, \ldots, y_{ki}\}_{i=1}^{I}$

[56] Existence of the equilibrium and conditions for uniqueness were established in a previous version of the paper. Alternative assumptions over population mobility between Bogotá and the rest of the country are covered in Proposition 1.

of the form

$$\prod_{k=1}^{K} y_{ki}^{\alpha_{kh}} = \lambda_h \Phi_{Ri}^{b_h^R} \Phi_{Fi}^{b_h^F} e_{ih} \quad \textit{for } h = 1, \ldots, K.$$

These models will yield the same counterfactual changes in outcomes (relative to city-wide averages) as those from the baseline model, given estimates of β_R, β_F, θ. This class includes models with endogenous firm location choice, Eaton and Kortum production, capital in the production function, endogenous housing supply, leisure, preference rather than productivity shocks, and alternative residential and employment supply elasticities and timing assumptions. However, the overall level of changes and changes in endogenous constants will depend on (a subset of) the particular structural parameters of the model $\{\{\alpha_{kh}\}_k, b_h^R, b_h^F\}_h$, and are not determined by the reduced form elasticities alone.

The implications of these results are now discussed in turn.

Reduced Form Representation. The first part of Proposition 1 shows that the transit network only matters for equilibrium outcomes through the two CMA variables. In fact, the change in the entire distribution of economic activity across the city between two periods depends only on the change in CMA as well as a structural residual that reflects changing location fundamentals (productivities, amenities and floorspace supplies).[57] This system reduces to a system of 4 univariate regressions, where residential outcomes depend on RCMA and commercial outcomes depend on FCMA.

These CMA terms can be easily recovered using data on residential populations, employment, commute costs and the commuting elasticity θ. This ensures estimation of the reduced form is straightforward, even if CMA is not directly observed in the data. The proof of Proposition 1 shows that the CMA terms are the unique to-scale solution to the system given in (18) and (19) in the paper. It also discusses the approximation used collapse the reduced form that contains three CMA terms $\Phi_{Ri}, \Phi_{Fi}, \tilde{\Phi}_{Fi}$ into one with just Φ_{Ri}, Φ_{Fi}. This choice is made both for parsimony and empirical feasibility (the correlation between Φ_{Fi} and $\tilde{\Phi}_{Fi}$ is 0.98 in the data). The unapproximated reduced form is used to conduct counterfactuals, with a simple adjustment made to the coefficients from the approximated reduced form to map them to the coefficients from the unapproximated system (see proof in Appendix C.8.1 for details).

Counterfactual Impacts of Transit Infrastructure. Part 2 of Proposition 1 shows that relative changes in endogenous variables across the city in response to a change in commute costs can be computed using data on the initial distribution L_{Ri}, L_{Fi}, d_{ij}, the change in commute costs $\hat{d}_{ij}$, the commuting elasticity θ, and the reduced form parameters $\beta_{L_R}, \beta_{r_R}, \beta_{r_F}, \beta_{L_F}$. In other words, these data and parameters are sufficient statistics for the change in economic activity across the city in response to changes in transit infrastructure. As shown in the proof, the elasticities and the change in CMA are the sufficient statistics; the data on initial economic activity and changes in commute costs are necessary to compute the change in CMA.

Part 3 shows that computing both the level change in endogenous variables as well as the change in equilibrium constants requires slightly more structure. These require an assumption on population mobility into the city from the rest of the country, and values for two parameters σ and β that cannot be estimated from the reduced form. These must be specified in some other way by the researcher, for example by calibrating to external values or aggregate moments.

[57] The contents of the residual and reduced form parameters are outlined in Appendix C.7. The residual contains changes in unobserved amenities and residential floorspace for residential outcomes, and changes in unobserved productivities and commercial floorspace for commercial outcomes.

Part 4 shows that some of these results apply more generally to a wider class of models which feature a gravity equation for commute flows and a log-linear equilibrium representation. Despite having different underlying structural parameters, these models yield the same log-linear reduced form. Since part 2 requires only values of these reduced form elasticities to compute relative changes in activity across the city in response to changes in the transit network, they yield the same (relative) counterfactual impacts as the baseline model. This result is particularly useful because the researcher does not need to take a stand on which particular modeling assumption is true; each will yield the same counterfactual impact on relative outcomes as the baseline model conditional on the reduced form estimates $\boldsymbol{\beta}_R, \boldsymbol{\beta}_F$. Where the modeling assumptions do come into play is in determining the overall level of changes and aggregate effects (such as welfare). As the example in part 3 shows, this depends on the underlying structural parameters of the model. However if the researcher is ready to take a stand on the value of those parameters in their model, then these aggregate impacts can be computed using the procedure shown in the proof of part 3 and the values of the particular structural parameters of that model.

C.3 Estimating Demand for Travel Modes

Standard results on GEV distributions imply that the choice probabilities are

$$\begin{aligned}\pi_{m|ija} &= \pi_{k|ija} \times \pi_{m|ijka} \\ &= \frac{\left(\sum_{n\in\mathcal{B}_k} \exp\left(b_n - \frac{\kappa}{\lambda_k} t_{ijn}\right)\right)^{\lambda_k}}{\sum_{k'}\left(\sum_{n\in\mathcal{B}_{k'}} \exp\left(b_n - \frac{\kappa}{\lambda_{k'}} t_{ijn}\right)\right)^{\lambda_{k'}}} \times \frac{\exp\left(b_m - \frac{\kappa}{\lambda_k} t_{ijm}\right)}{\sum_{n\in\mathcal{B}_k} \exp\left(b_n - \frac{\kappa}{\lambda_k} t_{ijn}\right)}\end{aligned}$$

where $b_m \equiv -\tilde{b}_m/\lambda_k$. That is, the probability a worker chooses mode m can be decomposed into the probability they choose the nest containing m and the probability they choose the mode from the options available in that nest. This is estimated via MLE as described in the main text.

C.4 Estimating the Commute Elasticity θ

Taking logs and first differences of the expression for commute flows (22) yields a gravity equation relating the change in commute flows to changes in commute times

$$\ln L_{ijt} = \alpha_{ij} + \gamma_{it} + \delta_{jt} - \theta\kappa t_{ijt} + \varepsilon_{ijt}, \tag{29}$$

where α_{ij}, γ_{it} and δ_{jt} are origin-destination, origin-year and destination-year fixed effects. While other estimation approaches typically leverage cross- sectional variation, this paper uses the change in commute times induced by TransMilenio to difference out time-invariant characteristics potentially correlated with commute times. Changes in origin- or destination- specific unobservables—such as amenities and productivities—are absorbed in the fixed effects.

Commute times t_{ijt} are formed using the same mode choice model as in the general model, but incorporating car ownership according to an exogenous probability rather than an endogenous decision. Workers become car owners according to a Bernoulli distribution with parameter ρ_{car}. Expected utility conditional on car ownership is

$$U_{ijm|a}(\omega) = \frac{u_i w_j r_{Ri}^{\beta-1} \epsilon_{ij}(\omega)}{\exp\left(\kappa t_{ijm} + \upsilon_{ijm}(\omega)\right)}.$$

Expected utility prior to drawing the mode-specific preference shocks shocks is given by

$$\begin{aligned} E_a\left[\max_m \left\{U_{ijm|a}(\omega)\right\}\right] &= u_i w_j r_{Ri}^{\beta-1}\epsilon_{ij}(\omega) \times \left[\rho_{car} E\left[\max_{m\in\mathcal{M}_1}\left\{1/d_{ijm}(\omega)\right\}\right] + (1-\rho_{car})E\left[\max_{m\in\mathcal{M}_0}\left\{1/d_{ijm}(\omega)\right\}\right]\right] \\ &= \frac{u_i w_j r_{Ri}^{\beta-1}\epsilon_{ij}(\omega)}{\exp\left(\kappa \bar{t}_{ij}\right)} \end{aligned}$$

where

$$\begin{aligned} t_{ij} &= -\frac{1}{\kappa}\ln\left[\rho_{car}\exp\left(-\kappa\bar{t}_{ij1}\right) + (1-\rho_{car})\exp\left(-\kappa\bar{t}_{ij0}\right)\right] \\ \text{where } \bar{t}_{ij0} &= -\frac{\lambda}{\kappa}\ln\sum_{m\in\mathcal{B}_{\text{Public}}}\exp\left(b_m - \frac{\kappa}{\lambda}t_{ijm}\right) \\ \bar{t}_{ij1} &= -\frac{1}{\kappa}\ln\left(\exp(b_{car} - \kappa t_{ij\text{Car}}) + \exp\left(\kappa\bar{t}_{ij0}\right)\right). \end{aligned}$$

The expressions $\bar{t}_{ij0}, \bar{t}_{ij1}$ are exactly the same car-ownership-specific commute time indices as in the baseline model. The only difference is that they are averaged using the parameter ρ_{car} which reflects the probability of owning a car. I then compute $\bar{t}_{ijt}$ for different years, where variation over time is induced by the changes in the TransMilenio network. I set $\rho_{car} = 0.181$ equal to the share of car owners in 2015.

The estimates for (29) are presented in Table A.6. Columns 1 and 2 run PPML regressions to account for the presence of zeros in the data. Controlling for route observables interacted with year fixed effects implies a value of $\theta = 3.398$ reported in Table 1. Column 3 runs the same regression via OLS which do not account for pairs with zero commute flows, finding similar but mildly smaller estimates. The last column instruments for the change in travel times using the instrument from Section 5 for travel times in the post-period, delivering a larger estimate.

C.5 First Order vs Equilibrium Effects

The standard approach to evaluate the gains from transit infrastructure is based on the Value of Travel Time Savings (e.g. Small and Verhoef 2007), in which its benefits are given by minutes saved times the value of time. The following proposition shows that under certain conditions, this is precisely the first order welfare impact from a change in infrastructure in the full general equilibrium model.

Proposition 2. *In a version of the baseline model with (i) no amenity or productivity spillovers, (ii) preference shocks over residential locations, (iii) workers owning an equal share of all floorspace and (iv) a labor income tax $1/(1+\theta)$ redistributed lump sum, the elasticity of welfare to a change in commute costs is*

$$d\ln\bar{U} = -\alpha\beta\kappa\sum_{ij}\frac{w_{ij}L_{ij}}{\sum_{rs}w_{rs}L_{rs}}dt_{ij}, \tag{30}$$

where w_{ij} is average labor income of commuters along pair (i,j).

The proof of the proposition first establishes that under these conditions the equilibrium is efficient. An application of the envelope theorem then shows that—to a first order—only the time savings from new infrastructure matter for welfare. This is simply proportional to a labor income-weighted average of the commute time reductions, scaled by κ and $\alpha\beta$. The former converts commute times to commute costs, while the latter reflects that a share of the gains go to floorspace owners rather than directly to workers.[58] Lastly, as explained in the proof of the proposition, technical reasons require the restrictions (ii)-(iv) to be imposed to derive this result. However, simulations of small shocks in the model from Section C.1 with only condition (i) imposed confirm this expression correctly captures the first order welfare effects in that model as well.

C.6 Examples of Isomorphic Models in Proposition 1

Sorting of Individual Entrepreneurs. Consider a production side where each variety is produced by a monopolist who can choose where to locate in the city. The entrepreneur has the same Cobb-Douglas production function over labor and commercial floorspace, so profits are a fraction $1/\sigma$ of sales. Entrepreneurs have idiosyncratic preferences for producing in each block so that the return from locating in j is given by

$$\begin{aligned} V_j(\omega) &= \pi_j \epsilon_j(\omega) \\ \text{where } \pi_j &= \bar{\sigma}\left(w_j^{\alpha} r_{Fj}^{1-\alpha}/A_j\right)^{1-\sigma} E \end{aligned}$$

where $\bar{\sigma} \equiv \sigma^{-\sigma}(\sigma-1)^{-(\sigma-1)}$ and $\epsilon_j(\omega)$ is the preference of entrepreneur ω in to produce in j. If these preferences are drawn from a Frechet distribution with shape $\theta_F > 1$, then (normalizing the mass of firms to 1) the number of firms producing in j is

$$N_j = \frac{\left(A_j/w_j^{\alpha} r_{Fj}^{1-\alpha}\right)^{\theta_F(\sigma-1)}}{\sum_s \left(A_s/w_s^{\alpha} r_{Fs}^{1-\alpha}\right)^{\theta_F(\sigma-1)}}$$

The wage bill is a fraction $\alpha\frac{\sigma-1}{\sigma}$ of sales so $w_j\ell_j = \alpha\left(\sigma/(\sigma-1)\right)^{-\sigma}\left(w_j^{\alpha} r_{Fj}^{1-\alpha}/A_j\right)^{1-\sigma} E$. Since total labor demand is simply $\tilde{L}_{Fj} = N_j\ell_j$, we find that

$$\tilde{L}_{Fj} = \alpha\left(\sigma/(\sigma-1)\right)^{-\sigma} \bar{U}_F^{-1/\theta_F(\sigma-1)} \times A_j^{(1+\theta_F)(\sigma-1)} w_j^{-(1+(1+\theta_F)\alpha(\sigma-1))} r_{Fi}^{-(\sigma-1)(1-\alpha)(1+\theta_F)} E$$

where $\bar{U}_F = \left[\sum_s \left(A_s/w_s^{\alpha} r_{Fs}^{1-\alpha}\right)^{\theta_F(\sigma-1)}\right]^{1/\theta_F(\sigma-1)}$. Using the same logic as for labor, demand for commercial floorspace is

$$H_{Fj} = (1-\alpha)\left(\sigma/(\sigma-1)\right)^{-\sigma} \bar{U}_F^{-1/\theta_F(\sigma-1)} \times A_j^{(1+\theta_F)(\sigma-1)} w_j^{-(1+\theta_F)\alpha(\sigma-1)} r_{Fi}^{-(\sigma-1)(1-\alpha)(1+\theta_F)-1} E.$$

Since floorspace is fixed, this is the commercial floorspace clearing condition.

Only the labor demand and commercial floorspace market clearing conditions have changed. Since they have the same log-linear parametric structure, the same reduced form representation as in the baseline model will hold. To see how, the equilibrium system becomes

$$\begin{aligned} L_{Ri} &= \bar{L}\bar{U}^{-\theta}\left(u_i r_{Ri}^{\beta-1}\right)^{\theta} \Phi_{Ri} \\ L_{Fj} &= \bar{L}\bar{U}^{-\theta} w_j^{\theta} \Phi_{Fj} \\ \tilde{L}_{Fj} &= \left(\bar{L}\bar{U}^{-\theta}\right)^{\frac{\theta-1}{\theta}} w_j^{\theta-1}\tilde{\Phi}_{Fj} \\ \tilde{L}_{Fj} &= \alpha\left(\sigma/(\sigma-1)\right)^{-\sigma} \bar{U}_F^{-1/\theta_F(\sigma-1)} w_j^{(1+\theta_F)\alpha(1-\sigma)-1} A_j^{(1+\theta_F)(\sigma-1)} r_{Fi}^{(1-\alpha)(1-\sigma)(1+\theta_F)} E \\ r_{Ri} &= \frac{1-\beta}{H_{Ri}} \Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}} \end{aligned}$$

[58] While these gains ultimately make their way back to workers who own the housing stock, these equilibrium price effects do not matter to a first order.

$$r_{Fi}=\left(\left(1-\alpha\right)\left(\sigma/(\sigma-1)\right)^{-\sigma}\bar{U}_F^{-1/\theta_F(\sigma-1)}\frac{A_j^{(1+\theta_F)(\sigma-1)}w_j^{-(1+\theta_F)\alpha(\sigma-1)}E}{H_{Fi}}\right)^{\frac{1}{1+(\sigma-1)(1-\alpha)(1+\theta_F)}}$$

where the CMA definitions are unchanged. Using the third line to substitute out for wages and ignoring the second line (which pins down L_{Fj} given the other variables of the model), we arrive at a system of 4 equations in $\{L_{Ri},\tilde{L}_{Fi},r_{Ri},r_{Fi}\}$ given $\{\Phi_{Ri},\tilde{\Phi}_{Fi}\}$

$$\begin{aligned}
L_{Ri}^{1-\theta\mu_U}r_{Ri}^{\theta(1-\beta)} &= \lambda_1\Phi_{Ri}\bar{u}_i^{\theta}\\
L_{Ri}^{-\frac{\theta-1}{\theta}}r_{ri} &= \lambda_2\Phi_{Ri}^{1/\theta}H_{Ri}^{-1}\\
r_{Fi}^{1+(\sigma-1)(1-\alpha)(1+\theta_F)}\tilde{L}_{Fi}^{(\sigma-1)\left(\frac{\alpha-\mu_A(\theta-1)}{\theta-1}\right)(1+\theta_F)} &= \lambda_3\frac{\bar{A}_i^{(\sigma-1)(1+\theta_F)}\tilde{\Phi}_{Fj}^{\frac{\alpha(\sigma-1)}{\theta-1}(1+\theta_F)}}{H_{Fi}}\\
r_{Fi}^{(\sigma-1)(1-\alpha)(1+\theta_F)}\tilde{L}_{Fj}^{\frac{\theta+(1+\theta_F)(\sigma-1)\left[\alpha-\mu_U(\theta-1)\right]}{\theta-1}} &= \lambda_4\bar{A}_j^{(1+\theta_F)(\sigma-1)}\tilde{\Phi}_{Fi}^{\frac{1+(1+\theta_F)\alpha(\sigma-1)}{\theta-1}}
\end{aligned}$$

where the (endogenous) constants are given by $\lambda_1\equiv\bar{L}\bar{U}^{-\theta}$, $\lambda_2=1-\beta$, $\lambda_3\equiv(1-\alpha)\left(\sigma/(\sigma-1)\right)^{-\sigma}\bar{U}_F^{-1/\theta_F(\sigma-1)}\left(\bar{L}\bar{U}^{-\theta}\right)^{-\frac{\alpha(\sigma-1)(1+\theta_F)}{\theta}}$ and $\lambda_4\equiv\alpha\left(\sigma/(\sigma-1)\right)^{-\sigma}\bar{U}_F^{-1/\theta_F(\sigma-1)}\left(\bar{L}\bar{U}^{-\theta}\right)^{-\frac{1+(1+\theta_F)\alpha(\sigma-1)}{\theta}}E$. This is of the same parametric form as the system (51), and thus admits the same reduced form as the baseline model. To see this explicitly for this example, write the system in changes and take logs to get

$$\begin{bmatrix}1-\theta\mu_U & \theta(1-\beta) & 0 & 0\\ -\frac{\theta-1}{\theta} & 1 & 0 & 0\\ 0 & 0 & 1+(\sigma-1)(1-\alpha)(1+\theta_F) & (\sigma-1)\left(\frac{\alpha-\mu_A(\theta-1)}{\theta-1}\right)(1+\theta_F)\\ 0 & 0 & (\sigma-1)(1-\alpha)(1+\theta_F) & \frac{\theta+(1+\theta_F)(\sigma-1)[\alpha-\mu_U(\theta-1)]}{\theta-1}\end{bmatrix}\begin{bmatrix}\ln\hat{L}_{Ri}\\ \ln\hat{r}_{Ri}\\ \ln\hat{r}_{Fi}\\ \ln\hat{\tilde{L}}_{Fi}\end{bmatrix}$$

$$=\begin{bmatrix}1\\ \frac{1}{\theta}\\ 0\\ 0\end{bmatrix}\ln\hat{\Phi}_{Ri}+\begin{bmatrix}0\\ 0\\ \frac{\alpha(\sigma-1)(1+\theta_F)}{\theta-1}\\ \frac{1+(1+\theta_F)\alpha(\sigma-1)}{\theta-1}\end{bmatrix}\ln\hat{\tilde{\Phi}}_{Fi}+\begin{bmatrix}\theta\ln\hat{\bar{u}}_i+\ln\hat{\bar{L}}\hat{\bar{U}}^{-\theta}\\ -\ln\hat{H}_{Ri}\\ (\sigma-1)(1+\theta_F)\ln\hat{\bar{A}}_i-\ln\hat{H}_{Fi}-\frac{\alpha(\sigma-1)(1+\theta_F)}{\theta}\ln\hat{\bar{L}}\hat{\bar{U}}^{-\theta}+\ln\hat{E}-\frac{1}{\theta_F(\sigma-1)}\ln\hat{\bar{U}}_F\\ (\sigma-1)(1+\theta_F)\ln\hat{\bar{A}}_i-\frac{1+(1+\theta_F)\alpha(\sigma-1)}{\theta}\ln\hat{\bar{L}}\hat{\bar{U}}^{-\theta}+\ln\hat{E}-\frac{1}{\theta_F(\sigma-1)}\ln\hat{\bar{U}}_F\end{bmatrix}$$

By the results of part (iv), the relative impacts of changes in the commuting network are the same in this model as the baseline model given estimates of θ and the reduced form elasticities. (Note the reduced form elasticities have the same parametric form in this model as the baseline, since $\boldsymbol{\beta}_F,\boldsymbol{\beta}_R$ have zero entries in the first and last two entries respectively.) The level effects would differ, however, since these depend on the structural parameters that appear in the A matrix and the error term $\mathbf{e}$.

Endogenous Housing Supply. Consider an extension of the model in which housing floorspace for each type of floorspace is produced using land T_i and capital K_i according to a Cobb-Douglas production function $H_i=T_i^{1-\eta}K_i^{\eta}$. Capital is freely traded across the city with price p_K. Each unit of land is owned by an atomistic developer who chooses $h_i=k_i^{\eta}$ units of housing to construct per unit of land, where k_i units of capital are used per unit of land. Profit maximization by developers yields a density of development $h_i=(\eta r_i/p_K)^{1/(1-\eta)}$. Total housing supply is therefore

$$H_{Ri}=T_i\left(\frac{\eta r_{Ri}}{p_K}\right)^{\frac{1}{1-\eta}}\quad\text{and}\quad H_{Fi}=T_i\left(\frac{\eta r_{Fi}}{p_K}\right)^{\frac{1}{1-\eta}}$$

All that changes in the model is that H_{Ri},H_{Fi} are now endogenous since they depend on floorspace prices.

Adding these equations into the system and rearranging yields

$$
\begin{aligned}
L_{Ri}^{1-\theta\mu_U} r_{Ri}^{\theta(1-\beta)} &= \bar{L}\bar{U}^{-\theta}\Phi_{Ri}\bar{u}_i^{\theta} \\
L_{Ri}^{-\frac{\theta-1}{\theta}} r_{Ri}^{1+\frac{1}{1-\eta}} &= (1-\beta)(p_K/\eta)^{\frac{1}{1-\eta}}\Phi_{Ri}^{1/\theta}T_i^{-1} \\
r_{Fi}^{1+(\sigma-1)(1-\alpha)+\frac{1}{1-\eta}}\tilde{L}_{Fi}^{\frac{(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1}} &= (1-\alpha)(p_K/\eta)^{\frac{1}{1-\eta}}\bar{A}_i^{\sigma-1}\left(\left(\bar{L}\bar{U}^{-(\theta-1)}\right)\tilde{\Phi}_{Fj}\right)^{\frac{\alpha(\sigma-1)}{\theta-1}}T_i^{-1}E \\
r_{Fi}^{(\sigma-1)(1-\alpha)}\tilde{L}_{Fi}^{\frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1}} &= \alpha\left(\left(\bar{L}\bar{U}^{-(\theta-1)}\right)\tilde{\Phi}_{Fj}\right)^{\frac{1+\alpha(\sigma-1)}{\theta-1}}\bar{A}_i^{\sigma-1}E
\end{aligned}
$$

This is of the same parametric form as the system (51). Writing the system in log changes yields

$$
\begin{bmatrix}
1-\theta\mu_U & \theta(1-\beta) & 0 & 0 \\
-\frac{\theta-1}{\theta} & 1+\frac{1}{1-\eta} & 0 & 0 \\
0 & 0 & 1+(\sigma-1)(1-\alpha)+\frac{1}{1-\eta} & \frac{(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1} \\
0 & 0 & (\sigma-1)(1-\alpha) & \frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1}
\end{bmatrix}
\begin{bmatrix}
\ln\hat{L}_{Ri} \\ \ln\hat{r}_{Ri} \\ \ln\hat{r}_{Fi} \\ \ln\hat{\tilde{L}}_{Fi}
\end{bmatrix}
$$
$$
=\begin{bmatrix} 1 \\ \frac{1}{\theta} \\ 0 \\ 0 \end{bmatrix}\ln\hat{\Phi}_{Ri}+
\begin{bmatrix} 0 \\ 0 \\ \frac{\alpha(\sigma-1)}{\theta-1} \\ \frac{1+\alpha(\sigma-1)}{\theta-1} \end{bmatrix}\ln\hat{\tilde{\Phi}}_{Fi}+
\begin{bmatrix}
\theta\ln\hat{\bar{u}}_i+\ln\hat{\bar{L}}-\theta\ln\hat{\bar{U}} \\
-\ln\hat{T}_i \\
(\sigma-1)\ln\hat{\bar{A}}_i-\ln\hat{H}_{Fi}+\frac{\alpha(\sigma-1)}{\theta-1}\left(\ln\hat{\bar{L}}-(\theta-1)\ln\hat{\bar{U}}\right)+\ln\hat{E} \\
(\sigma-1)\ln\hat{\bar{A}}_i+\frac{1+\alpha(\sigma-1)}{\theta-1}\left(\ln\hat{\bar{L}}-(\theta-1)\ln\hat{\bar{U}}\right)+\ln\hat{E}
\end{bmatrix}
$$

assuming the cost of capital p_K is unaffected by the system. This model admits exactly the same parametric form of regression equations as the baseline model, and so the results of part 4 apply. Note this model allows the share of floorspace used for commercial purposes in a census tract to respond to a change in commute costs. This would occur if the price of commercial floorspace changed relative to that of residential floorspace, since $\hat{\vartheta}_i = \frac{\hat{r}_{Fi}^{1/(1-\eta)}}{\vartheta_i\hat{r}_{Fi}^{1/(1-\eta)}+(1-\vartheta_i)\hat{r}_{Ri}^{1/(1-\eta)}}$ where $\vartheta_i \equiv H_{Fi}/(H_{Fi}+H_{Ri})$ is the share of floorspace allocated to commercial use in the initial equilibrium.

Eaton and Kortum. In the Eaton and Kortum (2002) setup, there is a continuum of goods $\omega \in [0,1]$. Each location has idiosyncratic draw for each good from a Frechet distribution with location parameter $A_j > 0$ and shape $\theta_F > 1$. There is perfect competition so that $p_j(\omega) = w_j/z_j(\omega)$. Goods market clearing implies that sales are given by

$$
X_j = \sum_i \frac{\left(w_j^{\alpha}r_{Fj}^{1-\alpha}/A_j\right)^{-\theta_F}}{\sum_s\left(w_s^{\alpha}r_{Fs}^{1-\alpha}/A_s\right)^{-\theta_F}}E_i = \left(w_j^{\alpha}r_{Fj}^{1-\alpha}\right)^{-\theta_F}A_j^{\theta_F}P^{\theta_F}E
$$

This yields the same system of equations as in the baseline model, with $\sigma-1$ replaced with θ_F, and thus the results of part 4 apply.

Capital. Consider an extension of the model in which firms can invest in capital to respond to changes in transit networks. Suppose firms use the production function $Y_i = A_i\tilde{L}_{Fi}^{\alpha_L}H_{Fi}^{\alpha_H}K_{Fi}^{\alpha_K}$. Capital is freely traded across the city and available at price p_K. Profit maximization implies firms spend constant fractions of sales on each factor, with factor demands given by

$$
\begin{aligned}
w_i \tilde{L}_{Fi} &= \frac{1}{\alpha_L} \left(\frac{w_i^{\alpha_L} r_{Ri}^{\alpha_H} p_K^{\alpha_K}}{A_i} \right)^{1-\sigma} E \\
r_{Fi} H_{Fi} &= \frac{1}{\alpha_H} \left(\frac{w_i^{\alpha_L} r_{Ri}^{\alpha_H} p_K^{\alpha_K}}{A_i} \right)^{1-\sigma} E \\
p_K K_{Fi} &= \frac{1}{\alpha_K} \left(\frac{w_i^{\alpha_L} r_{Ri}^{\alpha_H} p_K^{\alpha_K}}{A_i} \right)^{1-\sigma} E.
\end{aligned}
$$

We assume Colombia is a small open economy so that the price of capital is pinned down in international capital markets, i.e. p_K is a constant exogenous to the model. Only the condition for labor demand and commercial floorspace market clearing change. The equilibrium system is given by a system of $6 \times I$ equations in as many unknowns (given $\{\Phi_{Ri}, \Phi_{Fi}, \tilde{\Phi}_{Fi}\}$, themselves auxiliary variables of these same unknowns in the same system as the baseline model)

$$
\begin{aligned}
L_{Ri} &= \bar{L}\bar{U}^{-\theta} \left(u_i r_{Ri}^{\beta-1} \right)^{\theta} \Phi_{Ri} \\
L_{Fj} &= \bar{L}\bar{U}^{-\theta} w_j^{\theta} \Phi_{Fj} \\
\tilde{L}_{Fj} &= \left(\bar{L}\bar{U}^{-\theta} \right)^{\frac{\theta-1}{\theta}} w_j^{\theta-1} \tilde{\Phi}_{Fj} \\
\tilde{L}_{Fi} &= \frac{1}{\alpha_L} w_i^{\alpha_L(1-\sigma)-1} A_i^{\sigma-1} r_{Fi}^{\alpha_H(1-\sigma)} p_K^{\alpha_K(1-\sigma)} E \\
r_{Ri} &= \frac{1-\beta}{H_{Ri}} \Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}} \\
r_{Fi} &= \left(\frac{A_i^{\sigma-1} w_i^{-\alpha_L(\sigma-1)} p_K^{\alpha_K(1-\sigma)} E}{\alpha_H H_{Fi}} \right)^{\frac{1}{1+\alpha_H(\sigma-1)}}
\end{aligned}
$$

Note that with these solved for, capital demand can be recovered using the demand equation above. This is of the same parametric form as the system (51). Writing in relative changes and taking logs yields

$$
\begin{bmatrix}
1-\theta\mu_U & \theta(1-\beta) & 0 & 0 \\
-\frac{\theta-1}{\theta} & 1 & 0 & 0 \\
0 & 0 & 1+\alpha_H(\sigma-1) & \frac{(\sigma-1)(\alpha_L-\mu_A(\theta-1))}{\theta-1} \\
0 & 0 & \alpha_H(\sigma-1) & \frac{\theta+(\sigma-1)(\alpha_L-\mu_A(\theta-1))}{\theta-1}
\end{bmatrix}
\begin{bmatrix}
\ln \hat{L}_{Ri} \\ \ln \hat{r}_{Ri} \\ \ln \hat{r}_{Fi} \\ \ln \hat{\tilde{L}}_{Fi}
\end{bmatrix}
$$

$$
= \begin{bmatrix} 1 \\ \frac{1}{\theta} \\ 0 \\ 0 \end{bmatrix} \ln \hat{\Phi}_{Ri} + \begin{bmatrix} 0 \\ 0 \\ \frac{\alpha_L(\sigma-1)}{\theta-1} \\ \frac{1+\alpha_L(\sigma-1)}{\theta-1} \end{bmatrix} \ln \hat{\tilde{\Phi}}_{Fi} + \begin{bmatrix}
\theta \ln \hat{\bar{u}}_i + \ln \hat{\bar{L}} - \theta \ln \hat{\bar{U}} \\
-\ln \hat{H}_{Ri} \\
(\sigma-1)\ln \hat{\bar{A}}_i - \ln \hat{H}_{Fi} - \frac{\alpha(\sigma-1)}{\theta}\left(\ln \hat{\bar{L}} - \theta \ln \hat{\bar{U}} \right) - \alpha_K(\sigma-1)\ln \hat{p}_K + \ln \hat{E} \\
(\sigma-1)\ln \hat{\bar{A}}_i - \frac{\alpha(\sigma-1)+1}{\theta}\left(\ln \hat{\bar{L}} - \theta \ln \hat{\bar{U}} \right) - \alpha_K(\sigma-1)\ln \hat{p}_K + \ln \hat{E}
\end{bmatrix}
$$

The only changes from the baseline model are that the labor and housing elasticities have been relabeled, and the change in the price of capital has entered the residual. The results of part 4 apply.

Leisure. We consider an extension of the model where consumers derive utility over goods, housing and leisure. When preferences are Cobb-Douglas, the individual's problem is

$$
\max_{C,H,L} \quad u_i C^{\alpha} H^{\beta} L^{\gamma} \epsilon_{ij}(\omega) \text{ s.t. } C + r_{Ri}H + w_j L = w_j(1-t_{ij})
$$

Solving for commute flows yields

$$L_{ij} = \left(u_i w_j^{1-\gamma} r_{Ri}^{-\beta} / d_{ij}\right)^{\theta}$$

where $d_{ij} \equiv \frac{1}{1-t_{ij}}$. This has the same parametric form as the baseline model, but with alternative exponents on wages and house prices in the resident and labor supply terms and CMA definitions. The equilibrium can once again be written in the parametric form as the system (51), and the results of part 4 apply.

Preference Shocks. We consider an extension of the model in which consumers have preference rather than productivity shocks over each commute. Average income becomes $\bar{y}_i = \sum_j \pi_{j|i} w_j$ where $\pi_{j|i} = (w_j/d_{ij})^{\theta} / \sum_s (w_s/d_{is})^{\theta}$ is the probability of commuting to j conditional on living in i. Effective labor supply is simply L_{Fj}. The remaining equations of the model are unchanged. The equilibrium system becomes

$$\begin{aligned}
L_{Ri} &= \bar{L}\bar{U}^{-\theta}\left(u_i r_{Ri}^{\beta-1}\right)^{\theta} \Phi_{Ri} \\
L_{Fj} &= \bar{L}\bar{U}^{-\theta} w_j^{\theta} \Phi_{Fj} \\
L_{Fi} &= \frac{1}{\alpha} w_i^{\alpha(1-\sigma)-1} A_i^{\sigma-1} r_{Fi}^{(1-\sigma)(1-\alpha)} E \\
r_{Ri} &= \frac{1-\beta}{H_{Ri}} \bar{y}_i L_{Ri} \\
r_{Fi} &= \left(\frac{A_i^{\sigma-1} w_i^{-\alpha(\sigma-1)} E}{(1-\alpha) H_{Fi}}\right)^{\frac{1}{1+(\sigma-1)(1-\alpha)}}
\end{aligned}$$

Approximating $\bar{y}_i$ around the point $d_{ij}^{-\theta} = 0$ yields $\hat{\bar{y}}_i \approx \hat{\Phi}_{Ri}^{1/\theta}$, so the endogenous variables can again be expressed as log-linear functions of CMA and structural residuals. In particular, taking changes and logs yields a system exactly the same as the baseline model, but with the second entry in the first column of the A matrix changing from $-\frac{\theta-1}{\theta}$ to -1. The equilibrium can once again be written in the parametric form as the system (51), and the results of part 4 apply.

Alternative Labor and Residential Supply Elasticities and Timing Assumptions. We consider an extension of the model where commuters draw separate shocks over workplace and residence locations. Indirect utility across pairs of residential and employment locations (i, j) is given by

$$U_{ij}(\omega) = \frac{u_i w_j r_{Ri}^{\beta-1}}{d_{ij}} \epsilon_j(\omega) \nu_i(\omega),$$

where $\epsilon_j(\omega)$ is a productivity shock for employment in location j drawn from a Frechet distribution with shape θ and $\nu_i(\omega)$ is a preference shock for living in location i drawn from a Frechet distribution with shape η. Whichever choice is made first, the supply and residents and workers to locations is given by

$$\begin{aligned}
L_{Ri} &= \bar{L}\bar{U}^{-\theta}\left(u_i r_{Ri}^{\beta-1} \Phi_{Ri}^{1/\theta}\right)^{\eta} \\
L_{Fj} &= \bar{L}\bar{U}^{-\theta} w_j^{\theta} \Phi_{Fj}
\end{aligned}$$

where $\Phi_{Ri} = \sum_j (w_j/d_{ij})^{\theta}$ as before, but now $\Phi_{Fj} = \sum_i (u_i r_{Ri}^{\beta-1})^{\eta} d_{ij}^{-\theta} \Phi_{Ri}^{\frac{\eta}{\theta}-1}$. While these CMA terms look different from those in the original model, substituting the resident and labor supply curves back into them yield the same system of equations (18)-(19) defining CMA. The remaining model equations remain log-linear in endogenous variables and Φ_{Ri} and Φ_{Fi} (noting that now expected income is simply $\bar{y}_i = \gamma\Phi_{Ri}^{1/\theta}$). These results are independent of whether employment or residential locations are chosen first. The equilibrium can once again be written in the parametric form as the system (51), and the results of part 4 apply.

C.7 Reduced Form Coefficients and Residuals

This section makes explicit the structural content of the reduced form elasticities and residuals.

Residuals. As shown in the proof of Proposition 1, the residuals are given by $\mathbf{e}_i = A^{-1}\tilde{\mathbf{e}}_i$. Applying

$$A^{-1} = \begin{pmatrix} \frac{1}{\beta+\theta(1-\beta-\mu_U)} & -\frac{\theta(1-\beta)}{\beta+\theta(1-\beta-\mu_U)} & 0 & 0 \\ \frac{\theta-1}{\theta}\frac{1}{\beta+\theta(1-\beta-\mu_U)} & \frac{\theta(1-\mu_U)}{\beta+\theta(1-\beta-\mu_U)} & 0 & 0 \\ 0 & 0 & \frac{\theta+(\sigma-1)(\alpha-(\theta-1)\mu_A)}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha+\mu_A)} & -\frac{(\sigma-1)(\alpha-(\theta-1)\mu_A)}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha+\mu_A)} \\ 0 & 0 & -\frac{(1-\alpha)(\theta-1)(\sigma-1)}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha+\mu_A)} & \frac{(\theta-1)(\sigma(1-\alpha)+\alpha)}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha+\mu_A)} \end{pmatrix}$$

to the residual vector $\tilde{\mathbf{e}}_i$ yields

$$\mathbf{e}_i = \begin{bmatrix} \frac{1}{\beta+\theta(1-\beta-\mu_U)}\left[\theta\ln\hat{\bar{u}}_i + \ln\hat{\bar{L}} - \theta\ln\hat{\bar{U}}\right] + \frac{\theta(1-\beta)}{\beta+\theta(1-\beta-\mu_U)}\ln\hat{H}_{Ri} \\ \frac{\theta-1}{\theta}\frac{1}{\beta+\theta(1-\beta-\mu_U)}\left[\theta\ln\hat{\bar{u}}_i + \ln\hat{\bar{L}} - \theta\ln\hat{\bar{U}}\right] - \frac{\theta(1-\mu_U)}{\beta+\theta(1-\beta-\mu_U)}\ln\hat{H}_{Ri} \\ \frac{\theta}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha+\mu_A)}\left[(\sigma-1)\ln\hat{\bar{A}}_i - \frac{\theta+(\sigma-1)(\alpha-(\theta-1)\mu_A)}{\theta}\ln\hat{H}_{Fi} + \frac{(\sigma-1)(\alpha+\mu_A)}{\theta}\left(\ln\hat{L} - (\theta-1)\ln\hat{\bar{U}}\right) + \ln\hat{E}\right] \\ \frac{\theta-1}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha+\mu_A)}\left[(\sigma-1)\ln\hat{\bar{A}}_i + \frac{(1-\alpha)(\theta-1)(\sigma-1)}{\theta-1}\ln\hat{H}_{Fi} + \frac{\sigma}{\theta-1}\left(\ln\hat{L} - (\theta-1)\ln\hat{\bar{U}}\right) + \ln\hat{E}\right] \end{bmatrix}$$

where each entry corresponds to the residual for the specification with $L_{Ri}, r_{Ri}, r_{Fi}, \tilde{L}_{Fi}$ as the outcome, respectively. Residuals that vary across observations contain weighted sums of changes in (i) unobserved amenities and residential floorspace supplies for residential outcomes and (ii) unobserved productivities and commercial floorspace supplies for commercial outcomes.

CMA Elasticities. Computing $\boldsymbol{\beta}_R = A^{-1}b_R$ and $\boldsymbol{\beta}_F = A^{-1}b_F$ yields

$$\begin{pmatrix} \beta_{L_R} \\ \beta_{r_R} \\ \beta_{r_F} \\ \beta_{L_F} \end{pmatrix} = \begin{pmatrix} \frac{\beta}{\beta+\theta(1-\beta-\mu_U)} \\ \frac{1-\mu_U}{\beta+\theta(1-\beta-\mu_U)} \\ \frac{1}{1+\theta\left(\frac{\sigma}{\sigma-1}\frac{1}{\alpha+\mu_A}-1\right)} \\ \frac{\sigma}{(\sigma-1)(\alpha+\mu_A)}\frac{1}{1+\theta\left(\frac{\sigma}{\sigma-1}\frac{1}{\alpha+\mu_A}-1\right)} \end{pmatrix}$$

Rearranging these expressions yields $\mu_A = \frac{\sigma}{\sigma-1}\Big/\frac{\beta_{L_F}}{\beta_{r_F}} - \alpha$ and $\mu_U = 1 - \beta\Big/\frac{\beta_{L_R}}{\beta_{r_R}}$ as referenced in the text.

Given the reduced form estimates and the estimate of θ from the gravity equation, this is a system of 4 equations in 5 parameters $\beta, \mu_U, \sigma, \alpha, \mu_A$. However, even if one additional parameter is calibrated, these equations cannot be inverted for the remaining structural parameters. Consider first the system of equations determining β, μ_U in the first two lines. This can be rearranged into

$$\beta = \frac{\theta}{\theta - 1 + \frac{1}{\beta_{L_R}}}(1-\mu_U)$$

$$\beta = \frac{\theta - \frac{1}{\beta_{r_R}}}{\theta - 1}(1 - \mu_U).$$

These are two straight lines in the $(\beta, 1-\mu_U)$ space with the same intercept (at zero) but different slopes, other than the knife edge case where $\frac{\theta}{\theta-1+\frac{1}{\beta_{L_R}}} = \frac{\theta-\frac{1}{\beta_{r_R}}}{\theta-1}$ in which case there are an infinite number of solutions. For the second two equations, if α is calibrated to an external value then the system of equations is

$$\mu_A = \frac{\sigma}{\sigma-1}\Big/\left(1 + \frac{1-\beta_{r_F}}{\theta}\right) - \alpha$$
$$\mu_A = \frac{\sigma}{\sigma-1}\Big/\left(\frac{\beta_{L_F}(\theta-1)}{\beta_{L_F}\theta - 1}\right) - \alpha.$$

These are two straight lines in the $(\mu_A, \sigma/(\sigma-1))$ space with the same intercept but different slopes, other than the knife edge case where $1 + \frac{1-\beta_{r_F}}{\theta} = \frac{\beta_{L_F}(\theta-1)}{\beta_{L_F}\theta-1}$ in which case there are an infinite number of solutions.

Given that these equations cannot be inverted, one could try to calibrate one parameter (such as α) and jointly estimate the remaining 5 (including θ) to most closely match the CMA elasticities and the commuting semi-elasticity in the gravity equation. However, the match will not be exact given the results above. The sufficient statistics approach has the advantage that the researcher does not need to specify the value of all structural parameters and can conduct analysis using the commuting semi-elasticity, the CMA elasticities (and σ, β to obtain the overall level of changes). The researcher also does not need to take a stance on the particular model generating the data, i.e. what the specific cluster of structural parameters are that determine $\boldsymbol{\beta}_R, \boldsymbol{\beta}_F$.

C.8 Proofs & Additional Derivations

C.8.1 Proof of Proposition 1

Part 1: Reduced Form. Stacking the equilibrium conditions delivers

$$\begin{aligned}
L_{Ri} &= \bar{L}\bar{U}^{-\theta}\left(u_i r_{Ri}^{\beta-1}\right)^{\theta}\Phi_{Ri} \\
\tilde{L}_{Fj} &= \bar{L}\bar{U}^{-(\theta-1)} w_j^{\theta-1}\tilde{\Phi}_{Fj} \\
\tilde{L}_{Fi} &= \alpha w_i^{\alpha(1-\sigma)-1} A_i^{\sigma-1} r_{Fi}^{(1-\sigma)(1-\alpha)} E \\
r_{Ri} &= \frac{1-\beta}{H_{Ri}}\Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}} \\
r_{Fi} &= \left((1-\alpha)\frac{A_i^{\sigma-1} w_i^{-\alpha(\sigma-1)} E}{H_{Fi}}\right)^{\frac{1}{1+(\sigma-1)(1-\alpha)}} \\
\bar{U} &= \left[\sum_i \left(u_i \Phi_{Ri}^{1/\theta} r_{Ri}^{\beta-1}\right)^{\theta}\right]^{1/\theta}
\end{aligned}$$

where $E = \beta \sum_i \Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}}$ is total expenditure, and the CMA equations are

$$\Phi_{Ri} = \left(\bar{L}\bar{U}^{-\theta}\right)^{-1}\sum_j d_{ij}^{-\theta}\frac{L_{Fj}}{\Phi_{Fj}} \tag{31}$$

$$\Phi_{Fj} = \left(\bar{L}\bar{U}^{-\theta}\right)^{-1}\sum_i d_{ij}^{-\theta}\frac{L_{Ri}}{\Phi_{Ri}} \tag{32}$$

$$\tilde{\Phi}_{Fj} = \left(\bar{L}\bar{U}^{-\theta}\right)^{-\frac{\theta-1}{\theta}} \sum_i d_{ij}^{-\theta} \left(\frac{L_{Ri}}{\Phi_{Ri}}\right)^{(\theta-1)/\theta} \tag{33}$$

Using the second line to substitute out for wages we arrive at a system of 4 equations in $\{L_{Ri}, \tilde{L}_{Fi}, r_{Ri}, r_{Fi}\}$ given $\{\Phi_{Ri}, \tilde{\Phi}_{Fi}\}$

$$\begin{aligned}
L_{Ri}^{1-\theta\mu_U} r_{Ri}^{\theta(1-\beta)} &= \bar{L}\bar{U}^{-\theta}\Phi_{Ri}\bar{u}_i^{\theta} \\
L_{Ri}^{-\frac{\theta-1}{\theta}} r_{Ri} &= (1-\beta)\Phi_{Ri}^{1/\theta} H_{Ri}^{-1} \\
r_{Fi}^{1+(\sigma-1)(1-\alpha)} \tilde{L}_{Fi}^{\frac{(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1}} &= (1-\alpha)\bar{A}_i^{\sigma-1}\left(\left(\bar{L}\bar{U}^{-(\theta-1)}\right)\tilde{\Phi}_{Fj}\right)^{\frac{\alpha(\sigma-1)}{\theta-1}} H_{Fi}^{-1}E \\
r_{Fi}^{(\sigma-1)(1-\alpha)} \tilde{L}_{Fi}^{\frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1}} &= \alpha\left(\left(\bar{L}\bar{U}^{-(\theta-1)}\right)\tilde{\Phi}_{Fj}\right)^{\frac{1+\alpha(\sigma-1)}{\theta-1}} \bar{A}_i^{\sigma-1}E
\end{aligned}$$

Letting $\hat{x} = x'/x$ denote relative changes across two equilibria, we can take logs and rearrange to get

$$\underbrace{\begin{bmatrix} 1-\theta\mu_U & \theta(1-\beta) & 0 & 0 \\ -\frac{\theta-1}{\theta} & 1 & 0 & 0 \\ 0 & 0 & 1+(\sigma-1)(1-\alpha) & \frac{(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1} \\ 0 & 0 & (\sigma-1)(1-\alpha) & \frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \ln\hat{L}_{Ri} \\ \ln\hat{r}_{Ri} \\ \ln\hat{r}_{Fi} \\ \ln\hat{\tilde{L}}_{Fi} \end{bmatrix}}_{\ln\hat{\tilde{y}}_i}$$

$$= \underbrace{\begin{bmatrix} 1 \\ \frac{1}{\theta} \\ 0 \\ 0 \end{bmatrix}}_{b_R} \ln\hat{\Phi}_{Ri} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \frac{\alpha(\sigma-1)}{\theta-1} \\ \frac{1+\alpha(\sigma-1)}{\theta-1} \end{bmatrix}}_{b_F} \ln\hat{\tilde{\Phi}}_{Fi} + \underbrace{\begin{bmatrix} \theta\ln\hat{\bar{u}}_i + \ln\hat{\bar{L}} - \theta\ln\hat{\bar{U}} \\ -\ln\hat{H}_{Ri} \\ (\sigma-1)\ln\hat{\bar{A}}_i - \ln\hat{H}_{Fi} + \frac{\alpha(\sigma-1)}{\theta-1}\left(\ln\hat{\bar{L}} - (\theta-1)\ln\hat{\bar{U}}\right) + \ln\hat{E} \\ (\sigma-1)\ln\hat{\bar{A}}_i + \frac{1+\alpha(\sigma-1)}{\theta-1}\left(\ln\hat{\bar{L}} - (\theta-1)\ln\hat{\bar{U}}\right) + \ln\hat{E} \end{bmatrix}}_{\tilde{\mathbf{e}}_i} \tag{34}$$

Premultiplying by A^{-1} delivers the system

$$\ln\hat{\tilde{y}}_i = \boldsymbol{\beta}_R \ln\hat{\Phi}_{Ri} + \boldsymbol{\beta}_F \ln\hat{\tilde{\Phi}}_{Fi} + \mathbf{e}_i$$

where $\boldsymbol{\beta}_R = A^{-1}b_R$, $\boldsymbol{\beta}_F = A^{-1}b_F$ and $\mathbf{e}_i = A^{-1}\tilde{\mathbf{e}}_i$. Note that the last two elements of $\boldsymbol{\beta}_R$ are zero as are the first two elements of $\boldsymbol{\beta}_F$.[59] Since A^{-1} is block diagonal, the first two elements of $\mathbf{e}_i$ determining residential outcomes depend only on $\hat{\bar{u}}_i, \hat{H}_{Ri}, \hat{\bar{L}}, \hat{\bar{U}}$ while the second two elements determining commercial outcomes depend only on $\hat{\bar{A}}_i, \hat{H}_{Fi}, \hat{\bar{L}}, \hat{\bar{U}}, \hat{E}$. The exact reduced form (34) is the one which is used to conduct counterfactuals in parts 2 and 3.

However, in the data we observe L_{Fi} rather than $\tilde{L}_{Fi}$. Combining $L_{Fj} = \bar{L}\bar{U}^{-\theta}w_j^{\theta}\Phi_{Fj}$ and $\tilde{L}_{Fj} = \left(\bar{L}\bar{U}^{-(\theta-1)}\right)w_j^{\theta-1}\tilde{\Phi}_{Fj}$

[59]Note that solving these expressions yields

$$\begin{pmatrix} \beta_{L_R} \\ \beta_{r_R} \\ \beta_{r_F} \\ \beta_{L_F} \end{pmatrix} = \begin{pmatrix} \frac{\beta}{\beta+\theta(1-\beta-\mu_U)} \\ \frac{1-\mu_U}{\beta+\theta(1-\beta-\mu_U)} \\ \frac{1}{1+\theta\left(\frac{\sigma}{\sigma-1}\frac{1}{\alpha+\mu_A}-1\right)} \\ \frac{\sigma}{(\sigma-1)(\alpha+\mu_A)}\frac{1}{1+\theta\left(\frac{\sigma}{\sigma-1}\frac{1}{\alpha+\mu_A}-1\right)} \end{pmatrix}$$

Manipulating these expressions yields $\mu_A = \frac{\sigma}{\sigma-1}/\frac{\beta_{L_F}}{\beta_{r_F}} - \alpha$ and $\mu_U = 1 - \beta/\frac{\beta_{L_R}}{\beta_{r_R}}$ as referenced in the text.

yields the following relationship between the two

$$\hat{\tilde{L}}_{Fj} = \left(\frac{\hat{L}_{Fj}}{\hat{\Phi}_{Fj}}\right)^{\frac{\theta-1}{\theta}} \hat{\tilde{\Phi}}_{Fj}.$$

Substituting this in, we arrive at the following system

$$\underbrace{\begin{bmatrix} 1-\theta\mu_U & \theta(1-\beta) & 0 & 0 \\ -\frac{\theta-1}{\theta} & 1 & 0 & 0 \\ 0 & 0 & 1+(\sigma-1)(1-\alpha) & \frac{(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta} \\ 0 & 0 & (\sigma-1)(1-\alpha) & \frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \ln \hat{L}_{Ri} \\ \ln \hat{r}_{Ri} \\ \ln \hat{r}_{Fi} \\ \ln \hat{L}_{Fi} \end{bmatrix}}_{\ln \hat{\mathbf{y}}_i} = \underbrace{\begin{bmatrix} 1 \\ \frac{1}{\theta} \\ 0 \\ 0 \end{bmatrix}}_{b_R} \ln \hat{\Phi}_{Ri} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \frac{(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta} \\ \frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta} \end{bmatrix}}_{b_F} \ln \hat{\Phi}_{Fi}$$

$$+ \underbrace{\begin{bmatrix} 0 \\ 0 \\ \mu_A(\sigma-1) \\ \mu_A(\theta-1)(\sigma-1)-1 \end{bmatrix}}_{\tilde{b}_F} \ln \hat{\tilde{\Phi}}_{Fi} + \underbrace{\begin{bmatrix} \theta \ln \hat{\bar{u}}_i + \ln \hat{\bar{L}} - \theta \ln \hat{\bar{U}} \\ -\ln \hat{H}_{Ri} \\ (\sigma-1)\ln \hat{\bar{A}}_i - \ln \hat{H}_{Fi} + \frac{\alpha(\sigma-1)}{\theta-1}\left(\ln \hat{L} - (\theta-1)\ln \hat{\bar{U}}\right) + \ln \hat{E} \\ (\sigma-1)\ln \hat{\bar{A}}_i + \frac{1+\alpha(\sigma-1)}{\theta-1}\left(\ln \hat{L} - (\theta-1)\ln \hat{\bar{U}}\right) + \ln \hat{E} \end{bmatrix}}_{\tilde{\mathbf{e}}_i} \tag{35}$$

or, after premultiplying by A^{-1},

$$\ln \hat{\mathbf{y}}_i = \boldsymbol{\beta}_R \ln \hat{\Phi}_{Ri} + \boldsymbol{\beta}_F \ln \hat{\Phi}_{Fi} + \tilde{\boldsymbol{\beta}}_F \ln \hat{\tilde{\Phi}}_{Fi} + \mathbf{e}_i$$

This reduced form (35) along with the CMA definitions (31)-(33) hold globally to define a change in endogenous variables $\{\hat{L}_{Ri}, \hat{r}_{Ri}, \hat{r}_{Fi}, \hat{L}_{Fi}, \hat{\Phi}_{Ri}, \hat{\Phi}_{Fi}, \hat{\tilde{\Phi}}_{Fi}\}$ (and analogously the auxiliary variables $\hat{\bar{U}}, \hat{E}$ defined as a function of these variables above) given a change in exogenous (or "forcing") variables $\{\hat{\bar{u}}_i, \hat{\bar{A}}_i, \hat{H}_{Ri}, \hat{H}_{Fi}, \hat{\bar{L}}, \hat{d}_{ij}\}$. Note that in counterfactuals, all exogenous variables other than commute costs d_{ij} will be held constant.

However, the two FCMA terms, defined in (32) and (33), are very highly correlated in the data (correlation coefficient of 0.98). To make this regression simpler and to allow for enough residual variation to identify the coefficients on each term, I take a first order approximation of $\tilde{\Phi}_{Fi}$ around the point $d_{ij}^{-\theta} = 0$ which yields $\tilde{\Phi}_{Fj} \approx \Phi_{Fj}^{\frac{\theta-1}{\theta}}$. Substituting this in simplifies the system to

$$\underbrace{\begin{bmatrix} 1-\theta\mu_U & \theta(1-\beta) & 0 & 0 \\ -\frac{\theta-1}{\theta} & 1 & 0 & 0 \\ 0 & 0 & 1+(\sigma-1)(1-\alpha) & \frac{(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta} \\ 0 & 0 & (\sigma-1)(1-\alpha) & \frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta} \end{bmatrix}}_{A} \begin{bmatrix} \ln \hat{L}_{Ri} \\ \ln \hat{r}_{Ri} \\ \ln \hat{r}_{Fi} \\ \ln \hat{L}_{Fi} \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ \frac{1}{\theta} \\ 0 \\ 0 \end{bmatrix} \ln \hat{\Phi}_{Ri} + \begin{bmatrix} 0 \\ 0 \\ \frac{\alpha(\sigma-1)}{\theta} \\ \frac{1+\alpha(\sigma-1)}{\theta} \end{bmatrix} \ln \hat{\Phi}_{Fi} + \begin{bmatrix} \theta \ln \hat{\bar{u}}_i + \ln \hat{\bar{L}} - \theta \ln \hat{\bar{U}} \\ -\ln \hat{H}_{Ri} \\ (\sigma-1)\ln \hat{\bar{A}}_i - \ln \hat{H}_{Fi} + \frac{\alpha(\sigma-1)}{\theta-1}\left(\ln \hat{L} - (\theta-1)\ln \hat{\bar{U}}\right) + \ln \hat{E} \\ (\sigma-1)\ln \hat{\bar{A}}_i + \frac{1+\alpha(\sigma-1)}{\theta-1}\left(\ln \hat{L} - (\theta-1)\ln \hat{\bar{U}}\right) + \ln \hat{E} \end{bmatrix} \tag{36}$$

or, after premultiplying by A^{-1},

$$\ln \hat{\mathbf{y}}_i = \boldsymbol{\beta}_R \ln \hat{\Phi}_{Ri} + \boldsymbol{\beta}_F \ln \hat{\Phi}_{Fi} + \mathbf{e}_i$$

for $\mathbf{y}_i = [L_{Ri}, r_{Ri}, r_{Fi}, L_{Fi}]$. Compared with the unapproximated model, all that has happened is to approximate $\ln \hat{\tilde{\Phi}}_{Fj} \approx \frac{\theta-1}{\theta} \ln \hat{\Phi}_{Fj}$ to collapse the two FCMA terms into one.

Lastly, since we will use the system (34) to conduct counterfactuals with the estimated parameters, we need to relate the coefficients we will estimate in (36) to those in (34) . The only difference is that the last two elements of the 4th column of A and b_F have θ rather than $\theta - 1$ in the numerator. Computing $A^{-1}b_R$ after this adjustments yields the same coefficient as in the unapproximated model, but the commercial variable elasticities change to $A^{-1}b_F = \begin{pmatrix} 0 \\ 0 \\ \frac{\theta-1}{\theta} \frac{1}{1+\theta\left(\frac{\sigma}{\sigma-1}\frac{1}{\alpha+\mu_A}-1\right)} \\ \frac{\sigma}{(\sigma-1)(\alpha+\mu_A)} \frac{1}{1+\theta\left(\frac{\sigma}{\sigma-1}\frac{1}{\alpha+\mu_A}-1\right)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \frac{\theta-1}{\theta}\beta_{r_F} \\ \beta_{L_F} \end{pmatrix}$. So the only change needed is to replace β_{r_F} with $\frac{\theta}{\theta-1}\beta_{r_F}$ in the unapproximated model equations (where β_{r_F} is the elasticity estimated in the data).

Lastly, we show that unique (to-scale) values of CMA can be recovered given $d_{ij}, L_{Ri}, L_{Fi}, \theta$. Equations (31) and (32) can be written in the form

$$\Phi_{Ri} = \sum_j K^R_{ij} \Phi^{-1}_{Fj}$$
$$\Phi_{Fj} = \sum_i K^F_{ij} \Phi^{-1}_{Ri}$$

where $K^R_{ij} \equiv d^{-\theta}_{ij} L_{Fj}$ and $K^F_{ij} \equiv d^{-\theta}_{ij} L_{Ri}$. This satisfies the structure of the equations in theorem 1 in Allen et. al. (2014). In the notation of that theorem, $\Gamma = I$ and $B = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$. The spectral radius of the matrix $|B\Gamma^{-1}|$ (where $|\cdot|$ denotes the element-wise absolute value) is one. Parts (i) and (ii) of theorem 1 then imply that there exists unique (to-scale) solution Φ_{Ri}, Φ_{Fi}.[60]

Part 2: Relative Impacts of Transit Infrastructure. We now show we can use this system of equations to compute changes in economic activity relative to the citywide average in response to a transit shock using estimates of θ and the reduced form elasticities $\beta_{L_R}, \beta_{r_R}, \beta_{r_F}, \beta_{L_F}$, in addition to data on the initial equilibrium d_{ij}, L_{Ri}, L_{Fi} and the change in transit infrastructure $\hat{d}_{ij}$. Assuming unobservables are constant across equilibria, exponentiating the (unapproximated) system (34) and letting A^{-1}_{ij} denote the ij-th entry of A^{-1} yields

$$\hat{L}_{Ri} = \hat{\Phi}^{\beta_{L_R}}_{Ri} \left(\hat{\bar{L}}\hat{\bar{U}}^{-\theta}\right)^{A^{-1}_{11}} \tag{37}$$

$$\hat{r}_{Ri} = \hat{\Phi}^{\beta_{r_R}}_{Ri} \left(\hat{\bar{L}}\hat{\bar{U}}^{-\theta}\right)^{A^{-1}_{21}} \tag{38}$$

$$\hat{r}_{Fi} = \hat{\tilde{\Phi}}^{\beta_{r_F}}_{Fi} \left(\hat{\bar{L}}\hat{\bar{U}}^{-(\theta-1)}\right)^{\left(A^{-1}_{33}\frac{\alpha(\sigma-1)}{\theta} + A^{-1}_{34}\frac{\alpha(\sigma-1)+1}{\theta}\right)} \hat{E}^{A^{-1}_{33}+A^{-1}_{34}} \tag{39}$$

$$\hat{\tilde{L}}_{Fi} = \hat{\tilde{\Phi}}^{\beta_{L_F}}_{Fi} \left(\hat{\bar{L}}\hat{\bar{U}}^{-(\theta-1)}\right)^{\left(A^{-1}_{43}\frac{\alpha(\sigma-1)}{\theta} + A^{-1}_{44}\frac{\alpha(\sigma-1)+1}{\theta}\right)} \hat{E}^{A^{-1}_{43}+A^{-1}_{44}} \tag{40}$$

where $\hat{\Phi}_{Ri}, \hat{\Phi}_{Fi}, \hat{\tilde{\Phi}}_{Fi}, \hat{\tilde{L}}_{Fi}$ are given by

$$\hat{\Phi}_{Ri} = \left(\hat{\bar{L}}\hat{\bar{U}}^{-\theta}\right)^{-1} \sum_j \pi^R_{ij} \hat{d}^{-\theta}_{ij} \frac{\hat{L}_{Fj}}{\hat{\Phi}_{Fj}} \tag{41}$$

[60] Once these are recovered, a unique to-scale solution for $\tilde{\Phi}_{Fi}$ is simply recovered from (33).

$$\hat{\Phi}_{Fj} = \left(\hat{\bar{L}}\hat{\bar{U}}^{-\theta}\right)^{-1}\sum_i \pi^F_{ij}\hat{d}_{ij}^{-\theta}\frac{\hat{L}_{Ri}}{\hat{\Phi}_{Ri}} \tag{42}$$

$$\hat{\tilde{\Phi}}_{Fj} = \left(\hat{\bar{L}}\hat{\bar{U}}^{-\theta}\right)^{-\frac{\theta-1}{\theta}}\sum_i \tilde{\pi}^F_{ij}\hat{d}_{ij}^{-(\theta-1)}\left(\frac{\hat{L}_{Ri}}{\hat{\Phi}_{Ri}}\right)^{(\theta-1)/\theta} \tag{43}$$

$$\hat{L}_{Fj} = \left(\hat{\tilde{L}}_{Fj}/\hat{\tilde{\Phi}}_{Fj}\right)^{\frac{\theta}{\theta-1}}\hat{\Phi}_{Fj}. \tag{44}$$

Note that we are using that $\boldsymbol{\beta}_F$ and $\boldsymbol{\beta}_R$ have zeros in the first and last two entries respectively, otherwise both CMA terms would appear in each line. Here $\pi^R_{ij} = \frac{d_{ij}^{-\theta}\frac{L_{Fj}}{\Phi_{Fj}}}{\sum_j d_{ij}^{-\theta}\frac{L_{Fj}}{\Phi_{Fj}}}$, $\pi^F_{ij} = \frac{d_{ji}^{-\theta}\frac{L_{Rj}}{\Phi_{Rj}}}{\sum_j d_{ji}^{-\theta}\frac{L_{Rj}}{\Phi_{Rj}}}$ and $\tilde{\pi}^F_{ij} = \frac{d_{ji}^{-\theta}\left(\frac{L_{Rj}}{\Phi_{Rj}}\right)^{\frac{\theta-1}{\theta}}}{\sum_j d_{ji}^{-\theta}\left(\frac{L_{Rj}}{\Phi_{Rj}}\right)^{\frac{\theta-1}{\theta}}}$. Since these shares are homogenous of degree zero in Φ_{Ri}, Φ_{Fi}, their unique values are identified using values for $d_{ij}, L_{Ri}, L_{Fi}, \theta$ (since these determine unique to-scale solutions for the CMA terms). In my particular model, computing the terms in the A^{-1} matrix yields the system

$$\hat{L}_{Ri} = \hat{\Phi}_{Ri}^{\beta_{L_R}}\hat{\bar{L}}^{\beta_{L_R}}\hat{\bar{U}}^{-\frac{\beta_{L_R}\theta}{\beta}} \tag{45}$$

$$\hat{r}_{Ri} = \hat{\Phi}_{Ri}^{\beta_{r_R}}\hat{\bar{L}}^{\beta_{r_R}}\hat{\bar{U}}^{-\frac{\beta_{L_R}(\theta-1)}{\beta}} \tag{46}$$

$$\hat{r}_{Fi} = \hat{\tilde{\Phi}}_{Fi}^{\beta_{r_F}}\left(\hat{\bar{L}}\hat{\bar{U}}^{-(\theta-1)}\right)^{\beta_{r_F}}\hat{E}^{\frac{\beta_{L_F}\theta}{\sigma}} \tag{47}$$

$$\hat{\tilde{L}}_{Fi} = \hat{\tilde{\Phi}}_{Fi}^{\beta_{L_F}}\left(\hat{\bar{L}}\hat{\bar{U}}^{-(\theta-1)}\right)^{\beta_{L_F}}\hat{E}^{\beta_{L_F}\frac{\theta-1}{\sigma}} \tag{48}$$

The change in constants are given by

$$\hat{\bar{U}} = \hat{\bar{L}}^{(\beta_{L_R}-1)\frac{\beta}{\theta\beta_{L_R}}}\left[\sum_i \pi_{Ri}\hat{\Phi}_{Ri}^{\beta_{L_R}}\right]^{\frac{\beta}{\theta\beta_{L_R}}} \tag{49}$$

$$\hat{E} = \sum_i \pi^E_i\hat{\Phi}_{Ri}^{1/\theta}\hat{L}_{Ri}^{\frac{\theta-1}{\theta}} \tag{50}$$

where $\pi^{L_R}_i = \frac{L_{Ri}}{\sum_r L_{Rs}}$ and $\pi^E_i = \frac{\Phi_{Ri}^{1/\theta}L_{Ri}^{\frac{\theta-1}{\theta}}}{\sum_r \Phi_{Ri}^{1/\theta}L_{Ri}^{\frac{\theta-1}{\theta}}}$ are residential and expenditure shares from the initial equilibrium, where the expression for $\hat{\bar{U}}$ comes from summing up (45).

Now define $\hat{\hat{y}}_i = \hat{y}_i/\left(\prod_i \hat{y}_i\right)^{1/I}$ as the double-differenced change in y_i between two periods relative to the geometric average change across the whole city. Then this system becomes

$$\hat{\hat{L}}_{Ri} = \hat{\hat{\Phi}}_{Ri}^{\beta_{L_R}}$$

$$\hat{\hat{r}}_{Ri} = \hat{\hat{\Phi}}_{Ri}^{\beta_{r_R}}$$

$$\hat{\hat{r}}_{Fi} = \hat{\hat{\tilde{\Phi}}}_{Fi}^{\frac{\theta}{\theta-1}\beta_{r_F}}$$

$$\hat{\hat{\tilde{L}}}_{Fi} = \hat{\hat{\tilde{\Phi}}}_{Fi}^{\beta_{L_F}}$$

$$\hat{\hat{\Phi}}_{Ri} = \lambda_R\sum_j \pi^R_{ij}\hat{d}_{ij}^{-\theta}\frac{\hat{\hat{L}}_{Fj}}{\hat{\hat{\Phi}}_{Fj}}$$

$$\hat{\hat{\Phi}}_{Fi} = \lambda_F\sum_j \pi^F_{ij}\hat{d}_{ji}^{-\theta}\frac{\hat{\hat{L}}_{Rj}}{\hat{\hat{\Phi}}_{Rj}}$$

$$\hat{\hat{\tilde{\Phi}}}_{Fi} = \tilde{\lambda}_F \sum_j \tilde{\pi}^F_{ij} \hat{d}_{ji}^{-(\theta-1)} \left(\frac{\hat{\hat{L}}_{Rj}}{\hat{\hat{\Phi}}_{Rj}} \right)^{(\theta-1)/\theta}$$

$$\hat{\hat{L}}_{Fj} = \left(\hat{\hat{\tilde{L}}}_{Fj} / \hat{\hat{\tilde{\Phi}}}_{Fj} \right)^{\frac{\theta}{\theta-1}} \hat{\hat{\Phi}}_{Fj}$$

where $\lambda_R = \left[\prod_i \sum_j \pi^R_{ij} \hat{d}_{ij}^{-\theta} \frac{\hat{\hat{L}}_{Fj}}{\hat{\hat{\Phi}}_{Fj}} \right]^{-1/N}$, $\lambda_F = \left[\prod_i \sum_j \pi^F_{ij} \hat{d}_{ji}^{-\theta} \frac{\hat{\hat{L}}_{Rj}}{\hat{\hat{\Phi}}_{Rj}} \right]^{-1/N}$ and $\tilde{\lambda}_F = \left[\prod_i \sum_i \tilde{\pi}^F_{ij} \hat{d}_{ij}^{\theta-1} \left(\frac{\hat{\hat{L}}_{Ri}}{\hat{\hat{\Phi}}_{Ri}} \right)^{(\theta-1)/\theta} \right]^{-1/N}$. To solve this system, one can begin by solving for Φ_{Ri}, Φ_{Fi} using $\{\theta, d_{ij}, L_{Ri}, L_{Fi}\}$ following the procedure outlined above. With these in hand, $\pi^R_{ij}, \pi^F_{ij}, \tilde{\pi}^F_{ij}$ can be computed. Then, a change in the transit network $\hat{\mathbf{d}}$ can be fed into the system above which constitutes a system of $8N$ equations in as many unknowns $\{\hat{\hat{L}}_{Ri}, \hat{\hat{L}}_{Fi}, \hat{\hat{\tilde{L}}}_{Fi}, \hat{\hat{r}}_{Ri}, \hat{\hat{r}}_{Fi}, \hat{\hat{\Phi}}_{Ri}, \hat{\hat{\Phi}}_{Fi}, \hat{\hat{\tilde{\Phi}}}_{Fi}\}$ given data $\{L_{Fj}, L_{Ri}, d_{ij}\}$ and parameters $(\theta, \beta_{L_R}, \beta_{r_R}, \beta_{r_F}, \beta_{L_F})$. Any model with a gravity equation for commuting with commute costs d_{ij}, commuting elasticity θ, and the reduced form $\ln \hat{\mathbf{y}}_i = \boldsymbol{\beta}_R \ln \hat{\Phi}_{Ri} + \boldsymbol{\beta}_F \ln \hat{\Phi}_{Fi} + e_i$ will deliver the same distribution of relative changes to the shock across the city.

Part 3: Level Impact of Transit Infrastructure. Solving for the level effect of a counterfactual change in transit infrastructure requires solving for (i) the scale of each relative change variable from part 2 and (ii) the three endogenous scalars $\hat{\bar{U}}, \hat{\bar{L}}, \hat{E}$ until the system of equations (41)-(50) holds. This is a system of $8N+2$ equations in as many unknowns, if the value of either $\hat{\bar{U}}$or $\hat{\bar{L}}$ is known. This last condition is realized by alternative assumptions on population mobility. In the closed city case, city population is fixed so that $\hat{\bar{L}} = 1$. In the case with migration into the city, two equations in $\hat{\bar{U}}$or $\hat{\bar{L}}$ are provided in Appendix E.1. The additional data requirements to solve this system are the shares $\pi^{L_R}_i$ and π^E_i (which can be solved using $\{L_{Fj}, L_{Ri}, d_{ij}, \theta\}$. The additional parameters required are β, σ as can be seen from the exponents on the scalars in (45)-(48).

Part 4: Isomorphisms. Consider a model where the supply of commuters is determined by a gravity equation $L_{ij} = c\delta_j \gamma_i \kappa_{ij}$. Then the supply of residents and labor are given by $L_{Ri} = \gamma_i \Phi_{Ri}$ and $L_{Fi} = \delta_i \Phi_{Fi}$ where

$$\Phi_{Ri} = c \sum_j \frac{L_{Fj}}{\Phi_{Fj}} \kappa_{ij}$$

$$\Phi_{Fi} = c \sum_j \frac{L_{Ri}}{\Phi_{Ri}} \kappa_{ji}$$

Following the results in part 1, this solution has a unique to-scale solution.

Now suppose that in addition to these two equations pinning down CMA, the equilibrium can be written as a system of K equations in K endogenous variables $\{y_{1i}, \ldots, y_{ki}\}_{i=1}^I$ of the form

$$\prod_{k=1}^K y_{ki}^{\alpha_{kh}} = \lambda_h \Phi_{Ri}^{b^R_h} \Phi_{Fi}^{b^F_h} e_{ih} \quad \text{for } h = 1, \ldots, K \tag{51}$$

Then this system can be written of the form

$$\begin{bmatrix} \alpha_{11} & \cdots & \alpha_{K1} \\ \vdots & \ddots & \vdots \\ \alpha_{K1} & \cdots & \alpha_{KK} \end{bmatrix} \begin{bmatrix} \ln \hat{y}_{1i} \\ \vdots \\ \ln \hat{y}_{Ki} \end{bmatrix} = \begin{bmatrix} b^R_1 \\ \vdots \\ b^R_K \end{bmatrix} \ln \hat{\Phi}_{Ri} + \begin{bmatrix} b^F_1 \\ \vdots \\ b^F_K \end{bmatrix} \ln \hat{\Phi}_{Fi} + \begin{bmatrix} \ln \hat{e}_{1i} + \ln \hat{\lambda}_1 \\ \vdots \\ \ln \hat{e}_{Ki} + \ln \hat{\lambda}_K \end{bmatrix}$$

$$\Leftrightarrow A \ln \hat{\mathbf{y}}_i = b^R \ln \hat{\Phi}_{Ri} + b^F \ln \hat{\Phi}_{Fi} + \ln \hat{\hat{\mathbf{e}}}_i$$
$$\Leftrightarrow \ln \hat{\mathbf{y}}_i = \boldsymbol{\beta}_R \ln \hat{\Phi}_{Ri} + \boldsymbol{\beta}_F \ln \hat{\Phi}_{Fi} + A^{-1} \ln \hat{\hat{\mathbf{e}}}_i$$

Exponentiating the system and assuming unobservables are constant across equilibria yields[61]

$$\hat{y}_{ih} = \hat{\Phi}_{Ri}^{\beta_{R,h}} \hat{\Phi}_{Fi}^{\beta_{F,h}} \left[A^{-1} \hat{\boldsymbol{\lambda}} \right]_h \quad \text{for } h = 1, \ldots, K$$

Relative changes across the city are given by

$$\hat{\hat{y}}_{ih} = \hat{\hat{\Phi}}_{Ri}^{\beta_{R,h}} \hat{\hat{\Phi}}_{Fi}^{\beta_{F,h}}$$

where

$$\hat{\hat{\Phi}}_{Ri} = \rho_R \sum_j \pi_{ij}^R \frac{\hat{\hat{L}}_{Fj}}{\hat{\hat{\Phi}}_{Fj}} \hat{\kappa}_{ij}$$
$$\hat{\hat{\Phi}}_{Fi} = \rho_F \sum_j \pi_{ij}^F \frac{\hat{\hat{L}}_{Rj}}{\hat{\hat{\Phi}}_{Rj}} \hat{\kappa}_{ji}$$

where $\rho_R = \left[\prod_i \sum_j \pi_{ij}^R \hat{\kappa}_{ij} \frac{\hat{\hat{L}}_{Fj}}{\hat{\hat{\Phi}}_{Fj}} \right]^{-1/N}$, $\rho_F = \left[\prod_i \sum_j \pi_{ij}^F \hat{\kappa}_{ji} \frac{\hat{\hat{L}}_{Rj}}{\hat{\hat{\Phi}}_{Rj}} \right]^{-1/N}$, and $\pi_{ij}^R = \frac{\kappa_{ij} \frac{L_{Fj}}{\Phi_{Fj}}}{\sum_j \kappa_{ij} \frac{L_{Fj}}{\Phi_{Fj}}}$, $\pi_{ij}^F = \frac{\kappa_{ji} \frac{L_{Rj}}{\Phi_{Rj}}}{\sum_j \kappa_{ji} \frac{L_{Rj}}{\Phi_{Rj}}}$ can be solved using the to-scale versions of the CMA terms. Taken together, this is $K+2$ equations in the $K+2$ unknowns $\{\hat{\hat{y}}_{ih}\}_{h=1}^K, \hat{\hat{\Phi}}_{Ri}, \hat{\hat{\Phi}}_{Fi}$. Thus we have shown that parts (i) and (ii) apply to any model of this class. Appendix C.6 provides explicit examples of models that fall under it. ■

C.8.2 Proof of Proposition 2

This proof considers a slight modification of the baseline model, in which individuals (i) have separate productivity shocks over workplace locations and preference shocks over residential locations, (ii) own an equal share of the housing stock and (iii) face a labor income tax of $t_{ij} = 1/(1+\theta)$.

The reason for these changes is that efficiency requires lump sum redistribution to workers (i.e. part of income that does not depend on workplace location). In the decentralized equilibrium of the model in Section C.1, income always depends on workplace location. Even if total income is $y_j = w_j + e$ for some lump sum transfer e and productivity shocks are over pairs ij, then total income is $y_j/d_{ij} \times E[\epsilon_{ij}|\text{Choose } ij]$. Since this average productivity term depends on the choice of workplace location, there is no longer a location-independent portion of income.

Despite the slight difference between the model used in this proof and the baseline model, simulations that feed

[61] Note that

$$\hat{\Phi}_{Ri} = \hat{c} \sum_j \pi_{ij}^R \hat{\kappa}_{ij} \frac{\hat{L}_{Fj}}{\hat{\Phi}_{Fj}}$$
$$\hat{\Phi}_{Fj} = \hat{c} \sum_i \pi_{ij}^F \hat{\kappa}_{ji} \frac{\hat{L}_{Ri}}{\hat{\Phi}_{Ri}}$$

where $\pi_{ij}^R = \frac{\kappa_{ij} \frac{L_{Fj}}{\Phi_{Fj}}}{\sum_j \kappa_{ij} \frac{L_{Fj}}{\Phi_{Fj}}}$, $\pi_{ij}^F = \frac{\kappa_{ji} \frac{L_{Rj}}{\Phi_{Rj}}}{\sum_j \kappa_{ji} \frac{L_{Rj}}{\Phi_{Rj}}}$ can be solved using the to-scale versions of the CMA terms. So unique to-scale values for the changes in CMA terms are pinned down given values $\hat{L}_{Ri}, \hat{L}_{Fi}$, yielding the full system of equations that characterizes the equilibrium. Uniqueness (to-scale) of this solution in changes follows the same argument as for the solution in levels, given they have the same functional form.

in a very small shock (a constant $d\ln d_{ij} = 0.00001\ \forall ij$) into an efficient version of the baseline model (i.e. where $\mu_U = \mu_A = 0$) confirmed the expression for the welfare elasticity derived in the proof holds in the baseline model too.

Equilibrium Equations. The equilibrium equations in this model are

$$
\begin{aligned}
L_{Ri} &= \bar{L}\left(\frac{u_i \bar{y}_i r_{Ri}^{\beta-1}}{\bar{U} d_{ij}}\right)^{\theta} \\
L_{ij} &= L_{Ri}\frac{((1-t_{ij})w_j/d_{ij})^{\theta}}{\Phi_{Ri}} \\
\tilde{L}_{Fj} &= \sum_i L_{ij}\bar{\epsilon}_{ij} \\
w_i \tilde{L}_{Fi} &= \alpha p_i^{1-\sigma}\beta Y \\
p_i &= \tilde{\alpha}\frac{w_i^{\alpha} r_{Fi}^{1-\alpha}}{A_i} \\
r_{Ri} &= \frac{1-\beta}{H_{Ri}} L_{Ri}\bar{y}_i \\
r_{Fi}H_{Fi} &= (1-\alpha)p_i^{1-\sigma}\beta Y \\
\bar{U} &= \left[\sum_i \left(u_i \bar{y}_i r_{Ri}^{\beta-1}/d_{ij}\right)^{\theta}\right]^{1/\theta} \\
\bar{y}_i &= \Phi_{Ri}^{1/\theta} + e \\
\Phi_{Ri} &= \sum_j ((1-t_{ij})w_j/d_{ij})^{\theta} \\
e &= \frac{1}{\bar{L}}\left[(1-\alpha\beta)Y + \sum_{ij} t_{ij}L_{ij}\bar{w}_{ij}\right] \\
\bar{w}_{ij} &= \Phi_{Ri}^{1/\theta}
\end{aligned}
$$

where $Y = \sum_i Y_i$ is aggregate expenditure, $\tilde{\alpha} \equiv \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}$ is a constant, and $(1-\alpha\beta)Y = (1-\beta)Y + \beta(1-\alpha)Y$ is total expenditure on residential and commercial floorspace. Note that the preference draw $\nu_i(\omega)$ and productivity draw $\epsilon_j(\omega)$ are both drawn from a Frechet distribution with unit scale and shape $\theta > 1$, and workers choose their residential location before deciding where to work.

Planner Problem. The planner knows the distribution of individual heterogeneity, but not their specific draws. She announces a policy where workers receive some amount of the consumption and housing good per unit of effective labor supply based on where they work, as well as an amount based on where they live. In particular, the policy for someone who chooses to live in i and work in j with productivity ϵ is

$$
\begin{aligned}
c_{ij}(\epsilon) &= \tilde{c}_{ij}\frac{\epsilon}{d_{ij}} + \bar{c}_i \\
h_{ij}(\epsilon) &= \tilde{h}_{ij}\frac{\epsilon}{d_{ij}} + \bar{h}_i.
\end{aligned}
$$

Given these policies, individuals make free decisions about where to live and work. Utility from each choice is

$U_{ij}(\epsilon,\nu) = u_i \left(\frac{\tilde{c}_{ij}}{d_{ij}}\epsilon + \bar{c}_i\right)^{\beta} \left(\frac{\tilde{h}_{ij}}{d_{ij}}\epsilon + \bar{h}_i\right)^{1-\beta} \nu$. Since this is non-linear in ϵ, I constrain the planner to policies that make the two transfers proportional to one another (with a constant of proportionality that can vary by residential location), i.e. $\tilde{h}_{ij} = \iota_i \tilde{c}_{ij}$ and $\bar{c}_i = \iota_i \bar{h}_i$. Then $U_{ij}(\epsilon,\nu) = u_i \iota_i^{1-\beta} c_{ij}(\epsilon)\nu$.

The planner then chooses the consumption policies and supply of residents and workers to maximize utility subject to the following technological constraints

- Goods Feasibility: $\left(\sum_k c_{kij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} = L_{ij}(\tilde{c}_{ij}\epsilon_{ij} + \bar{c})$, where c_{kij} is the amount of variety k consumed by individuals choosing ij. (Each variety is a freely traded good from a particular location, i.e. Armington without trade costs).
- Residential Housing Feasibility: $H_{Ri} = \sum_j L_{ij}\iota_i(\tilde{c}_{ij}\epsilon_{ij} + \bar{c})$
- Commercial Floorspace Feasibility: $\tilde{H}_{Fi} = H_{Fi}$
- Production Technology: $A_i \tilde{L}_i^{\alpha} \tilde{H}_{Fi}^{1-\alpha} = \sum_{rs} c_{irs}$
- Effective Labor Feasibility: $\tilde{L}_j = \sum_i L_{ij}\epsilon_{ij}$
- Worker Mobility: $L_{Ri} = \bar{L}\left(\frac{u_i \iota_i^{1-\beta}\left(\Phi_{Ri}^{1/\theta} + \bar{c}\right)}{\bar{U}}\right)^{\theta}$.
- Commuting Mobility: $L_{ij} = \frac{(w_j/d_{ij})^{\theta}}{\Phi_{Ri}} L_{Ri}$
- Residential feasibility: $\bar{L} = \sum_i L_{Ri}$
- Effective Labor Technology: $\bar{\epsilon}_{ij} = \frac{\Phi_{Ri}^{1/\theta} d_{ij}}{w_j}$
- CMA Definition: $\Phi_{Ri} = \sum_j (\tilde{c}_{ij}/d_{ij})^{\theta}$

The Lagrangian is

$$
\begin{aligned}
\mathcal{L} &= \bar{U} \\
&+ \sum_{ij} \upsilon_{ij}\left(\left(\sum_k c_{kij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - L_{ij}(\tilde{c}_{ij}\epsilon_{ij} + \bar{c}_i)\right) + \sum_i \kappa_i\left(H_{Ri} - \sum_j L_{ij}\iota_i(\tilde{c}_{ij}\epsilon_{ij} + \bar{c}_i)\right) \\
&+ \sum_i \lambda_i\left(A_i \tilde{L}_i^{\alpha}\tilde{H}_{Fi}^{1-\alpha} - \sum_{rs} c_{irs}\right) + \sum_j \xi_j\left(\sum_i \bar{\epsilon}_{ij}L_{ij} - \tilde{L}_j\right) \\
&+ \sum_i \delta_i(H_{Fi} - \tilde{H}_{Fi}) \\
&+ \sum_i \rho_i\left(\left(\frac{L_{Ri}}{\bar{L}}\right)^{-1/\theta} u_i \iota_i^{1-\beta}\left(\Phi_{Ri}^{1/\theta} + \bar{c}_i\right) - \bar{U}\right) \\
&+ \sum_{ij} \psi_{ij}\left(\frac{(\tilde{c}_{ij}/d_{ij})^{\theta}}{\Phi_{Ri}} L_{Ri} - L_{ij}\right) \\
&+ \sum_i \tau_i\left(\sum_j (\tilde{c}_{ij}/d_{ij})^{\theta} - \Phi_{Ri}\right) \\
&+ \sum_{ij} \phi_{ij}\left(\frac{\Phi_{Ri}^{1/\theta} d_{ij}}{\tilde{c}_{ij}}\frac{1}{d_{ij}} - \bar{\epsilon}_{ij}\right) + \mu\left(\bar{L} - \sum_i L_{Ri}\right)
\end{aligned}
$$

The first order conditions with respect to the choice variables $\{\bar{U}, \tilde{c}_{ij}, \iota_i, \bar{c}_i, c_{kij}, L_{ij}, L_{Ri}, \tilde{L}_{Fi}, \tilde{H}_{Fi}, \bar{\epsilon}_{ij}, \Phi_{Ri}\}$ are

$$(v_{ij} + \kappa_i \iota_i)\tilde{c}_{ij}\bar{\epsilon}_{ij}L_{ij} = L_{ij}\left(\theta\psi_{ij} + \theta\tau_i\frac{\Phi_{Ri}}{L_{Ri}} - \frac{\phi_{ij}\bar{\epsilon}_{ij}}{L_{ij}}\right) \tag{$\tilde{c}_{ij}$}$$

$$\sum_j(\tilde{c}_{ij}\epsilon_{ij} + \bar{c}_i)\kappa_i L_{ij} = (1-\beta)\frac{\rho_i\bar{U}}{\iota_i} \tag{ι_i}$$

$$\sum_j L_{ij}\left(v_{ij} + \kappa_i\iota_i\right) = \rho_i\frac{\bar{U}}{\Phi_{Ri}^{1/\theta} + \bar{c}_i} \tag{$\bar{c}_i$}$$

$$\lambda_k = \nu_{ij}\left(\frac{c_{kij}}{C_{ij}}\right)^{-\frac{1}{\sigma}} \tag{c_{kij}}$$

$$\delta_i = (1-\alpha)\lambda_i A_i\left(\frac{\tilde{L}_i}{\tilde{H}_{Fi}}\right)^{\alpha} \tag{H_{Fi}}$$

$$\xi_j\bar{\epsilon}_{ij} = \left(v_{ij} + \kappa_i\iota_i\right)\left(\tilde{c}_{ij}\bar{\epsilon}_{ij} + \bar{c}\right) + \psi_{ij} \tag{L_{ij}}$$

$$\sum_j\psi_{ij}\frac{L_{ij}}{L_{Ri}} = \frac{1}{\theta}\frac{\rho_i\bar{U}}{L_{Ri}} + \mu \tag{L_{Ri}}$$

$$\phi_{ij} + L_{ij}\tilde{c}_{ij}\left(\nu_{ij} + \kappa_i\iota_i\right) = \xi_j L_{ij} \tag{$\bar{\epsilon}_{ij}$}$$

$$\tau_i = \frac{1}{\theta}\frac{\bar{U}}{\Phi_{Ri}}\frac{\Phi_{Ri}^{1/\theta}}{\Phi_{Ri}^{1/\theta} + \bar{c}} - \sum_j\psi_{ij}\frac{L_{ij}}{\Phi_{Ri}} + \sum_j\frac{1}{\theta}\phi_{ij}\frac{\epsilon_{ij}}{\Phi_{Ri}} \tag{Φ_{Ri}}$$

$$\xi_i = \alpha\lambda_i A_i\left(\frac{H_{Fi}}{\tilde{L}_i}\right)^{1-\alpha} \tag{$\tilde{L}_i$}$$

$$1 = \sum_i\rho_i \tag{$\bar{U}$}$$

and each of the constraint holds (to provide a condition for each multiplier).

Consumption and Housing. Define $\tilde{x}_{ij} = v\tilde{c}_{ij} + \kappa_i\tilde{h}_{ij} = \tilde{c}_{ij}\left(v + \kappa_i\iota_i\right)$ to be expenditure per unit of effective labor (as shown below, $v_{ij} = v\ \forall ij$). Likewise define $\bar{x}_i = \bar{c}_i\left(v + \kappa_i\iota_i\right)$ to be the expenditure on the fixed good so that $\bar{c}_i = \bar{x}_i/\left(v + \kappa_i\iota_i\right)$. Putting these into the mobility condition yield

$$\bar{U} = \left(\frac{L_{Ri}}{\bar{L}}\right)^{-1/\theta}u_i\frac{\iota_i^{1-\beta}}{(v + \kappa_i\iota_i)}\left(\tilde{\Phi}_{Ri}^{1/\theta} + \bar{x}_i\right),$$

where $\tilde{\Phi}_{Ri}^{1/\theta} \equiv \left[\sum_s(\tilde{x}_{is}/d_{ij})^{\theta}\right]^{1/\theta}$. To solve for ι_i, from its FOC we obtain

$$\frac{\kappa_i\iota_i}{v + \kappa_i\iota_i} = (1-\beta)\frac{\rho_i\bar{U}}{\sum_j(x_{ij}\bar{\epsilon}_{ij} + \bar{x}_i)L_{ij}}.$$

To solve this, we need a value for ρ_i. From the FOC for $\bar{c}_i$,

$$L_{Ri} = \rho_i\frac{\bar{U}}{\tilde{\Phi}_{Ri}^{1/\theta} + \bar{x}_i}.$$

The definition of $\bar{\epsilon}_{ij}$ yields $\tilde{x}_{ij}\bar{\epsilon}_{ij} = \tilde{\Phi}_{Ri}^{1/\theta}$ so that average income is constant across workplace locations within a residence location. Using this to simplify the denominator in the FOC for ι_i, combining these two conditions gives

$v + \kappa_i \iota_i = v/\beta$. Substituting this into $\tilde{c}_{ij} = \tilde{x}_{ij}/\left(v + \kappa_i \iota_i\right)$, $\bar{c}_i = \bar{x}_i/\left(v + \kappa_i \iota_i\right)$, $\tilde{h}_{ij} = \iota_i \tilde{c}_{ij}$ and $\bar{h}_i = \iota_i \bar{c}_i$ yields

$$\begin{aligned}
\tilde{c}_{ij} &= \beta \tilde{x}_{ij}/v \\
\tilde{h}_{ij} &= (1-\beta)\tilde{x}_{ij}/\kappa_i \\
\bar{c}_i &= \beta \bar{x}_i/v \\
\bar{h}_i &= (1-\beta)\bar{x}_i/\kappa_i.
\end{aligned}$$

Plugging this into the expression for utility gives residential supply

$$L_{Ri} \propto \bar{L} \left(\frac{u_i \bar{y}_i \kappa_i^{\beta-1}}{\bar{U}} \right)^{\theta}$$

where $\bar{y}_i = \tilde{x}_{ij}\bar{\epsilon}_{ij} + \bar{x}_i = \tilde{\Phi}_{Ri}^{1/\theta} + \bar{x}_i$ is average income of residents in i. Substituting this into the expression for residential feasibility provides an alternative expression for average welfare

$$\bar{U} \propto \left[\sum_i \left(u_i \bar{y}_i \kappa_i^{\beta-1} \right)^{\theta} \right]^{1/\theta}$$

Residential Floorspace. Using these results, the floorspace market clearing condition implies

$$H_{Ri} = (1-\beta)\frac{L_{Ri}\bar{y}_i}{\kappa_i}.$$

Production. The FOC for c_{kij} implies $c_{kij} = \left(\frac{\lambda_k}{v_{ij}}\right)^{-\sigma} C_{ij}$, which plugged into the definition of $C_{ij} = \left(\sum_k c_{kij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ yields $v_{ij} = v = \left(\sum_k \lambda_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ $\forall i, j$. The market clearing condition for goods implies

$$A_i \tilde{L}_i^{\alpha} H_{Fi}^{1-\alpha} = \lambda_i^{-\sigma} v^{\sigma-1} \beta Y$$

where $Y \equiv \sum_i L_{Ri}\bar{y}_i$ is aggregate expenditure. Combining the FOC for labor and commercial floorspace gives $H_{Fi} = \frac{\alpha}{1-\alpha}\frac{\xi_i}{\delta_i}\tilde{L}_i$, and substituting this back into the FOC yields both factor demands and an expression for λ_i

$$\begin{aligned}
\xi_i \tilde{L}_i &= \alpha \lambda_i^{1-\sigma} v^{\sigma-1} \beta Y \\
\delta_i H_{Fi} &= (1-\alpha)\lambda_i^{1-\sigma} v^{\sigma-1} \beta Y \\
\lambda_i &= \tilde{\alpha} \frac{\xi_i^{\alpha}\delta_i^{1-\alpha}}{A_i}.
\end{aligned}$$

Labor Supply. Finally we need to solve the spatial mobility condition, i.e. the FOC for L_{ij}. First, note the condition for L_{Ri} implies $\psi_{ij} = \psi_i$. The FOC for L_{Ri} and L_{ij} can then be combined to get

$$\xi_j \bar{\epsilon}_{ij} = \tilde{x}_{ij}\bar{\epsilon}_{ij} + \bar{x}_i + \frac{1}{\theta}\frac{\rho_i \bar{U}}{L_{Ri}} + \mu.$$

Substituting in the value for ρ_i from above gives

$$\tilde{x}_{ij}\bar{\epsilon}_{ij} + \bar{x}_i = \frac{\theta}{\theta+1}\xi_j \bar{\epsilon}_{ij} - \frac{\theta}{\theta+1}\mu.$$

Note this implies that expenditure per effective unit of labor depends only on workplace location ($\tilde{x}_{ij} = \frac{\theta}{\theta+1}\xi_j$), and expenditure per worker is constant across residential locations ($\bar{x}_i = -\frac{\theta}{\theta+1}\mu$). Substituting the expression for $\tilde{c}_{ij}$ into the commuting constraint gives

$$L_{ij} = L_{Ri}\frac{\left(\xi_j/d_{ij}\right)^\theta}{\sum_s \left(\xi_s/d_{is}\right)^\theta}.$$

Taking Stock. The solution to the planner's problem is the vector $(\bar{U}, \kappa_i, v, \xi_i, \delta_i, \lambda_i, L_{ij}, L_{Ri})$ that satisfies

$$\begin{aligned}
L_{Ri} &\propto \bar{L}\left(\frac{u_i \bar{y}_i \kappa_i^{\beta-1}}{\bar{U} d_{ij}}\right)^\theta \\
L_{ij} &= L_{Ri}\frac{\left(\xi_j/d_{ij}\right)^\theta}{\sum_s \left(\xi_s/d_{is}\right)^\theta} \\
\tilde{L}_{Fj} &= \sum_i L_{ij}\bar{\epsilon}_{ij} \\
\xi_i \tilde{L}_i &= \alpha\lambda_i^{1-\sigma} v^{\sigma-1}\beta Y \\
\lambda_i &= \tilde{\alpha}\frac{\xi_i^\alpha \delta_i^{1-\alpha}}{A_i}. \\
H_{Ri} &= (1-\beta)\frac{L_{Ri}\bar{y}_i}{\kappa_i} \\
\delta_i H_{Fi} &= (1-\alpha)\lambda_i^{1-\sigma} v^{\sigma-1}\beta Y \\
\bar{U} &\propto \left[\sum_i \left(u_i \bar{y}_i \kappa_i^{\beta-1}\right)^\theta\right]^{1/\theta} \\
\bar{y}_i &= \Phi_{Ri}^{1/\theta} - \frac{\theta}{\theta+1}\mu \\
\Phi_{Ri} &= \sum_j \left(\frac{\theta}{\theta+1}\xi_j/d_{ij}\right)^\theta
\end{aligned}$$

where $\bar{y}_i = \Phi_{Ri}^{1/\theta} - \frac{\theta}{\theta+1}\mu$ and $\Phi_{Ri} = \sum_j \left(\frac{\theta}{\theta+1}\xi_j/d_{ij}\right)^\theta$ are functions of these variables and the planner's multiplier on the residential feasibility constraint μ. This is the same set of equations as the decentralized equilibrium with $(\kappa_i, v, \xi_i, \delta_i, \lambda_i, \tilde{x}_{ij}, \mu) = (r_{Ri}, P, w_i, r_{Fi}, p_i, \frac{\theta}{1+\theta}w_j, -\frac{\theta}{1+\theta}e)$, i.e. when $t_{ij} = 1/(1+\theta)$. Therefore under this condition, any competitive equilibrium also solves the social planner's solution and is efficient.

Welfare Elasticity. Using the envelope theorem, the change in welfare to a change in commute costs is

$$\begin{aligned}
\frac{\partial \bar{U}}{\partial d_{ij}} &= -\frac{\theta\psi_{ij}L_{ij}}{d_{ij}} - \theta\frac{\tau_i\Phi_{Ri}}{d_{ij}}\frac{L_{ij}}{L_{Ri}} \\
\Rightarrow \frac{\partial \ln \bar{U}}{\partial \ln d_{ij}} &= -\frac{L_{ij}\left(\theta\psi_{ij} + \theta\frac{\tau_i\Phi_{Ri}}{L_{Ri}}\right)}{\bar{U}}
\end{aligned}$$

Combining the FOC for $\tilde{c}_{ij}$ and $\bar{\epsilon}_{ij}$ give $\xi_j\bar{\epsilon}_{ij} = \theta\psi_{ij} + \theta\tau_i\frac{\Phi_{Ri}}{L_{Ri}}$. Defining $w_{ij} \equiv \xi_j\bar{\epsilon}_{ij}$ to be average labor income for commuters along ij, this simplifies to

$$\frac{\partial \ln \bar{U}}{\partial \ln d_{ij}} = -\frac{w_{ij}L_{ij}}{\bar{U}}.$$

Substituting the expression for ρ_i into the FOC for $\bar{U}$ implies $\bar{U} = \sum_i L_{Ri}\bar{y}_i$. From the adding up condition we must have

$$\underbrace{\sum_{ij} L_{ij}\bar{y}_i}_{\text{Total Income}} = \underbrace{\sum_{ij} L_{ij}w_{ij}}_{\text{Labor Income}} + \underbrace{(1-\beta)\sum_{ij} L_{ij}\bar{y}_i}_{\text{Income from Res Floorspace}} + \underbrace{\beta(1-\alpha)\sum_{ij} L_{ij}\bar{y}_i}_{\text{Income from Comm Floorspace}} = \frac{1}{\alpha\beta}\sum_{ij} L_{ij}w_{ij}$$

Thus $\bar{U} = \frac{1}{\alpha\beta}\sum_{ij} L_{ij}w_{ij}$ and

$$\frac{\partial \ln \bar{U}}{\partial \ln d_{ij}} = -\alpha\beta \frac{w_{ij}L_{ij}}{\sum_{rs} w_{rs}L_{rs}}.$$

Adding up to compute the change in utility $d\ln\bar{U}$ to a vector of changes in commute costs $\{d\ln d_{ij}\}_{ij}$ gives the result in the proposition. Note that the parameters α, β account for the fact that some of the gains go to factors other than labor, but these equilibrium price effects do not matter to an infinitesimal change in commute costs and thus do not impact welfare. ■

C.8.3 Proof of Proposition 3

Part 1: Wages

To construct the system of equations used for solving for wages, I collect the expressions for supply and demand for workers. Labor supply $L_{Fjg} = w_{jg}^{\theta_g}\Phi_{Fjg}$ can be rearranged as

$$w_{jg} = L_{Fjg}^{\frac{1}{\theta_g}} \left[\sum_{i,a} \frac{L_{Riag}}{\sum_k w_{kg}^{\theta_g} d_{ika}^{-\theta_g}} d_{ija}^{-\theta_g} \right]^{-\frac{1}{\theta_g}}$$

This is a system of equations in w_{jg} given parameters and data $\{L_{Riag}, d_{ija}, L_{Fjg}\}$. The problem is that I do not observe employment by group, but only employment by industry L_{Fjs}. However, I can combine this data with the structure of the model to find employment by group for each location.

From CES demand for each group's labor, the share of any industry's (effective) employment by any group g is given by

$$\frac{\tilde{L}_{Fjgs}}{\tilde{L}_{Fjs}} = \frac{(w_{jg}/\alpha_{sg})^{-\sigma}}{\sum_h (w_{jh}/\alpha_{sh})^{-\sigma}}.$$

Summing this over industries yields total employment by group in a location

$$\tilde{L}_{Fjg} = \sum_s \frac{(w_{jg}/\alpha_{sg})^{-\sigma}}{\sum_h (w_{jh}/\alpha_{sh})^{-\sigma}} \tilde{L}_{Fjs}$$

It remains to express effective units of labor supply in terms of observed data and wages.

Start by decomposing $\tilde{L}_{Fjs}$ in terms of data and wages as follows. First, compute the average productivity of workers in j

$$\bar{\epsilon}_{jg} = E\left[\epsilon | g, \text{Choose } j\right] = \sum_{i,o} E\left[\epsilon | g, \text{Choose } j \text{ from}(i,o)\right] \Pr\left(i,o|j,g\right) = \sum_{i,o} \gamma_g \left(\frac{\tilde{T}_g}{\pi_{j|iog}}\right)^{\frac{1}{\theta_g}} \frac{1}{d_{ijo}} \Pr\left(i,o|j,g\right)$$

Next, break down the probability as

$$\Pr(i,o|j,g)=\pi_{io|jg}=\frac{\pi_{j|iog}\pi_{iog}}{\sum_{r,u}\pi_{j|rug}\pi_{rug}}=\frac{\pi_{j|iog}L_{Riog}}{\sum_{r,u}\pi_{j|rug}L_{Rrug}}$$

So

$$\bar{\epsilon}_{jg}=T_g\sum_{i,o}\pi_{j|iog}^{-\frac{1}{\theta_g}}\frac{1}{d_{ijo}}\frac{\pi_{j|iog}L_{Riog}}{\sum_{r,u}\pi_{j|rug}L_{Rrug}}$$

Next, note that

$$\bar{\epsilon}_{js}=\sum_g\bar{\epsilon}_{jg}\pi_{g|js}=\sum_g\bar{\epsilon}_{jg}\frac{L_{Fjgs}}{L_{Fjs}}=\sum_g\bar{\epsilon}_{jg}\frac{(w_{jg}/\alpha_{sg})^{-\sigma}/\bar{\epsilon}_{jg}}{\sum_h(w_{jh}/\alpha_{sh})^{-\sigma}/\bar{\epsilon}_{jh}}$$

Putting these results together, we have that

$$L_{Fjg}=\frac{\tilde{L}_{Fjg}}{\bar{\epsilon}_{jg}}=\sum_s\frac{(w_{jg}/\alpha_{sg})^{-\sigma}}{\sum_h(w_{jh}/\alpha_{sh})^{-\sigma}}\frac{\bar{\epsilon}_{js}}{\bar{\epsilon}_{jg}}L_{Fjs}$$

Substituting this result back into the expression for labor supply, we find that wages are the fixed point of the system $w_g=F_{wg}(w_g;L_{Rg},L_{Fs})$ where the operator F_{wg} is defined to have the j-th element

$$\begin{aligned}F_{wg}(w_g;L_{Fs},L_{Rg})_j&=\left[\sum_s\frac{(w_{jg}/\alpha_{sg})^{-\sigma}}{\sum_h(w_{jh}/\alpha_{sh})^{-\sigma}}\frac{\bar{\epsilon}_{js}}{\bar{\epsilon}_{jg}}L_{Fjs}\right]^{\frac{1}{\theta_g}}\left[\sum_{i,o}\frac{L_{Riog}}{\sum_k w_{kg}^{\theta_g}d_{iko}^{-\theta_g}}d_{ijo}^{-\theta_g}\right]^{-\frac{1}{\theta_g}}\\&=F_{1wg}(w_g;L_{Fs},L_{Rg})_jF_{2wg}(w_g;L_{Rg})_j\\\text{where }\bar{\epsilon}_{jg}&=T_g\sum_{i,o}\pi_{j|iog}^{-\frac{1}{\theta_g}}\frac{1}{d_{ijo}}\frac{\pi_{j|iog}L_{Riog}}{\sum_{r,u}\pi_{j|rug}L_{Rrug}}\\\bar{\epsilon}_{js}&=\sum_g\bar{\epsilon}_{jg}\frac{(w_{jg}/\alpha_{sg})^{-\sigma}/\bar{\epsilon}_{jg}}{\sum_h(w_{jh}/\alpha_{sh})^{-\sigma}/\bar{\epsilon}_{jh}}\end{aligned}$$

Note that the operator F_{wg} has the following properties:

- **Monotonicity**. Transform the system into log-space. From Euler's theorem since F_1 is homogenous of degree zero we know for any vector $d\ln w$ we have that

$$\sum_{k,h}\frac{\partial F_{1g}}{\partial\ln w_{kh}}=0$$

  so the total differential of F_{1g} to a vector of wage changes is zero. The second term is monotonic in w, which is a positive transformation of $\ln w$. Thus, the operator F_{wg} is a strictly increasing function of $\ln w$. By the chain rule, F_{wg} is a strictly increasing function of w.

- **Homogeneity**. Consider first F_{1wg}. The first part $\frac{(w_{jg}/\alpha_{sg})^{-\sigma}}{\sum_h(w_{jh}/\alpha_{sh})^{-\sigma}}$ is homogenous of degree zero in wages. From the definition of $\bar{\epsilon}_{js}$ and $\bar{\epsilon}_{jg}$ we see that these too are homogenous of degree zero in wages. Therefore F_{1wg} is homogenous of degree zero in wages. Next, we see that F_{2wg} is homogenous of degree one in wages, so that F_{wg} is homogenous of degree one.

Therefore, by the results in Fujimoto and Krause (1985) there exists a unique (to-scale) solution to the system $w_g = F_{wg}(w_g; L_{Fs}, L_{Rg})$.

Part 2: Remaining Unobservables

Given wages, Φ_{Riag}, W_{is} can be computed. The total wage bill is obtained from

$$\begin{aligned} W_{js}N_{js} &= \sum_g w_{jg}\tilde{L}_{Fjgs} \\ &= \sum_g w_{jg}\frac{(w_{jg}/\alpha_{sg})^{-\sigma}}{\sum_h (w_{jh}/\alpha_{sh})^{-\sigma}} L_{Fjs}\bar{\epsilon}_{js} \end{aligned}$$

This allow me to obtain sales from $\alpha_s X_{js} = W_{js}N_{js}$. With this in hand, productivity comes from

$$X_{js} = \left(\frac{W_{js}^{\alpha_s} r_{Fj}^{1-\alpha_s}}{A_{js}}\right)^{1-\varsigma} X$$

since X is also observed using Φ_{Riag}.

Lump sum income from the housing stock is recovered directly from $\pi = \bar{L}^{-1}\sum_i (r_{Ri}H_{Ri} + r_{Fi}H_{Fi})$. Amenities are retrieved from the resident supply condition

$$\begin{aligned} L_{Riag} &= \lambda_{Lg}\left(u_{iag}(T_g\Phi_{Riag}^{1/\theta} - \bar{h}r_{Ri} - p_a a + \pi)r_{Ri}^{\beta-1}\right)^{\eta_g} \\ \Rightarrow u_{iag} &= \frac{(L_{Riag}/\lambda_{Lg})^{1/\eta_g} r_{Ri}^{1-\beta}}{(T_g\Phi_{Riag}^{1/\theta} - \bar{h}r_{Ri} - p_a a + \pi)} \end{aligned}$$

To solve for unobservables on the housing side of the model, I need to introduce a new pair of location characteristics omitted in the main paper for notational brevity. In particular, the floorspace market clearing condition $r_{Ri} = \frac{E_i}{H_{Ri}}$ will not necessarily hold at the values for data and estimated wages (where E_i is total expenditure on housing from residents of i). I therefore introduce an additional unobservable so that $H_{Ri} = \tilde{H}_{Ri}\xi_{Ri}$, where $\tilde{H}_{Ri}$ are physical units of floorspace and ξ_{Ri} are effective units (or housing quality). These unobservables can be solved for from the housing market clearing condition $\xi_{Ri} = \frac{E_i}{\tilde{H}_{Ri}r_{Ri}}$. Similar residuals for effective units of commercial floorspace ξ_{Fi} are obtained from the commercial floorspace market clearing condition $\xi_{Fi} = \frac{\sum_s (1-\alpha_s)X_{is}}{\tilde{H}_{Fi}r_{Fi}}$, and total floorspace supplies are given by $H_{Ri} = \tilde{H}_{Ri}\xi_{Ri}$ and $H_{Fi} = \tilde{H}_{Fi}\xi_{Fi}$.

Finally, it remains to solve for the land use restrictions τ_i. These can be identified from

$$(1-\tau_i) = \frac{r_{Ri}\xi_{Ri}}{r_{Fi}\xi_{Fi}}$$

for locations with mixed land use. For locations with single land use, the wedges are not identified but these are rationalized by zero productivities (for all sectors) or zero amenities (for all worker groups) and thus will remain single use across counterfactuals.[62] ■

[62] These solutions are unique to scale. In practice, as discussed in Section D.3, I normalize the geometric mean of wages and floorspace prices to one. This affects the scale of unobservables such as productivities and amenities, but has no impact on relative differences in exogenous characteristics or endogenous variables across locations or counterfactuals.

C.8.4 Average Income in Single Group Model

Floorspace Market Clearing and Average Income. Average income of residents of i is

$$
\begin{aligned}
\bar{y}_i &= \sum_j \pi_{j|i}(w_j/d_{ij}) E\left[\epsilon_{ij}(\omega)|\omega \text{ chooses } (i,j)\right] = \frac{1}{\pi_i}\sum_j \pi_{ij}^{\frac{\theta-1}{\theta}}(w_j/d_{ij}) = \frac{1}{\pi_i}\bar{U}^{-(\theta-1)}\left(u_i r_{Ri}^{\beta-1}\right)^{\theta-1}\sum_j (w_j/d_{ij})^{\theta} \\
&= \bar{L}\bar{U}^{-(\theta-1)}\frac{\left(u_i r_{Ri}^{\beta-1}\right)^{\theta-1}\Phi_{Ri}}{L_{Ri}} \\
&= \bar{U}\frac{1}{u_i r_{Ri}^{\beta-1}} \\
&= \bar{U}\frac{1}{\left(L_{Ri}/\Phi_{Ri}\bar{L}\bar{U}^{-\theta}\right)^{\frac{1}{\theta}}} \\
&= \Phi_{Ri}^{1/\theta} L_{Ri}^{-1/\theta}.
\end{aligned}
$$

Total expenditure by residents in i is then simply $E_i = \bar{y}_i L_{Ri} = \Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}}$, the expression in the floorspace market clearing condition.

C.9 Bootstrap

To incorporate uncertainty from the parameter estimates into the welfare estimates, I bootstrap the quantification procedure 200 times.

For the single group sufficient statistics model, I draw values for the 10 estimated parameters $(\kappa, b_{Bus}, b_{Car}, b_{TM}, \lambda, \theta\kappa, \beta_{L_R}, \beta$ from normal distributions with means equal to the point estimates and standard deviations equal to the standard error of the estimates. I consider only draws which have non-negative commuting elasticities and reduced form elasticities otherwise the model has issues converging. This represents 95% of all drawn parameter vectors. I also disregard a small number of draws with an implausibly large value for the agglomeration elasticity ($\mu_A > 1$) since this can lead to non-sensical negative welfare estimates.[63] I then compute confidence intervals across the 200 bootstrap estimates. In Table 6 the estimated parameters are used for welfare estimates in the first two columns so confidence intervals are reported, and the non-parametric p-value for whether the fraction of welfare gains accounted for by VTTS is less than one is simply the fraction of the 200 draws for which this is not true.

For the multigroup model I repeat the same procedure for the 10 estimated parameters $(\kappa, b_{Bus}, b_{Car}, b_{TM}, \lambda, \theta_g\kappa, \eta_g, \mu_A, \mu_U^g)$

D Additional Model Results

D.1 Model Inversion

The model contains unobserved location characteristics, such as wages, productivities, amenities and land use wedges. While the presence of agglomeration forces allows for the possibility of multiple equilibria, I am able to recover unique values of composite productivities and amenities that rationalize the observed data as a model equilibrium.

There is a key difference in this process compared to recent quantitative urban models (e.g. Ahlfeldt et. al. 2015). In those models, there is one group of workers. It is straightforward to combine data on residence and employment

[63] Including these simulations widens the 90% and 95% confidence intervals to (-0.027,8.481) and [-1.035,10.754] respectively.

with the model structure provided by the gravity equation in commuting to solve for the unique vector of wages that rationalize the data. To replicate this in a model with multiple skill groups requires data on residence and employment by skill group. While the former are typically available in censuses, I am unaware of datasets that provide employment by skill group across small spatial units within cities. This is where the model's multiple industries become useful. The data contain employment by industry. Intuitively, given the differential demand for skills across industries, the relative employment by industries in a location should be informative about the relative employment across skill groups. The following proposition formalizes this intuition, and shows that a unique vector of group-specific wages can be recovered using data on residence by skill and employment by industry. Obtaining the remaining unobservables is straightforward.

Proposition 3. ***(i) Wages*** *Given data on residence by skill group L_{Rig}, employment by industries L_{Fjs}, commute costs d_{ija} and car ownership shares $\lambda_{a|ig}$ in addition to model parameters, there exists a unique vector of wages (to scale) that rationalizes the observed data as an equilibrium of the model.*

(ii) Remaining Unobservables *Given model parameters, wages and data $\{L_{Rig}, \pi_{a|iag}, L_{Fjs}, H_i, \vartheta_i, r_{Ri}, r_{Fi}\}$ there exists a unique vector of unobservables $\{u_{iag}, A_{js}, X_{js}, \tau_i, \pi\}$ (to scale) that rationalizes the observed data as an equilibrium of the model.*

The procedure to estimate the parameters of the model proceeds in four steps. First, a subset of parameters are calibrated and estimated without solving the full model. Second, wages are recovered using parameters from the first step. Third, the remaining elasticities are estimated via GMM using moments similar to those in the reduced form analysis. Fourth, with all parameters in hand the model is inverted to recover the remaining unobservables.

D.2 Calibrating α_{sg}

Under the CES aggregator for labor, the relative wage bill paid by firms to high-skill workers in location j and sector s defined as $\lambda_{jsH} \equiv w_{jH}\tilde{L}_{FjHs}/w_{jL}\tilde{L}_{FjLs}$ is

$$\lambda_{jsH} = \left(\frac{w_{jH}}{w_{jL}}\right)^{1-\sigma}\left(\frac{\alpha_{sH}}{\alpha_{sL}}\right)^{\sigma}.$$

Taking a double difference of this ratio in sector s relative to a reference sector s' gives

$$\lambda_{jsH}/\lambda_{js'H} = \left(\frac{\alpha_{sH}}{\alpha_{sL}}\right)^{\sigma} \Big/ \left(\frac{\alpha_{s'H}}{\alpha_{s'L}}\right)^{\sigma}$$

which holds for all workplace locations j. Using that $\alpha_{sL} = 1 - \alpha_{sH}$ yields

$$\alpha_{sH} = \frac{\frac{\alpha_{s'H}}{1-\alpha_{s'H}} E\left[\lambda_{jsH}/\lambda_{js'H}\right]^{1/\sigma}}{1 + \frac{\alpha_{s'H}}{1-\alpha_{s'H}} E\left[\lambda_{jsH}/\lambda_{js'H}\right]^{1/\sigma}},$$

where $E\left[\lambda_{jsH}/\lambda_{js'H}\right]$ are observed at the city-level in the ECH data. This allows identification of α_{sH} to scale (relative to the value of $\alpha_{s'H}$ in the reference sector). Using the manufacturing sector as the reference sector $s' = M$, I pin down α_{MH} with a departure from the spatial aspect of the model and use that under the CES aggregator the share of the wage bill paid to high-types is

$$\text{Share of Wage Bill to H}_M = \frac{w_H^{1-\sigma}\alpha_{MH}^{\sigma}}{w_H^{1-\sigma}\alpha_{MH}^{\sigma} + w_L^{1-\sigma}(1-\alpha_{MH})^{\sigma}}.$$

Plugging in the left hand side (observed at the city-level in the ECH data) along with the average wages w_H, w_L observed in the manufacturing sector in that data allows me to recover a value for α_{MH}.

The results are shown in Table A.9. The first column shows α_{Hs} while the second shows the relative wage bill of high-skill workers. We see a sensible and monotonic relationship, where industries such as Education and Financial Services have the highest weight on high-types and Domestic Services and Hotels & Restaurants have the lowest.

D.3 Model Solution

Calibrating $T_H, \bar{h}, p_a$ Given the parameter estimates in the previous section, for any value of T_g it is possible to solve for the full distribution of wages across the city. Since the vector T_g is not identified to scale, I normalize $T_L = 1$ and calibrate T_H so that the aggregate wage skill premium in the model matches that observed in the data. This involves jointly solving the system of equations for $\{T_H, w_{jg}\}$

$$\widehat{WP} = \frac{T_H \sum_{ia} \Phi_{RiaH}^{1/\theta_H} \lambda_{iaH}}{\sum_{ia} \Phi_{RiaL}^{1/\theta_L} \lambda_{iaL}}$$
$$w_g = F_g(w_g; L_{Fs}, L_{Rg}, T_H)$$

where $\widehat{WP} = 1.713$ is the wage premium observed in the data, the term next to it is the wage premium as predicted by the model (where λ_{iag} is the share of type-g workers in cell (i, a)), and the operator F_g is the system of equations used to solve for wages as a function of observables as given in Section C.8.3.

Next, having solved for wages the parameters $\bar{h}, p_a$ are set to exactly match the average expenditure share on housing and cars. In particular, they solve

$$1 - \beta + \bar{h} \sum_{i,a,g} \frac{r_{Ri} L_{Riag}}{E_{iag}} \lambda_{iag} = \hat{\omega}_H$$
$$\sum_{i,g} \lambda_{ig}^C \frac{p_a P}{T_g \Phi_{Riag}^{1/\theta_g}} = \hat{\omega}_C$$

where P is the aggregate price index,[64] $\hat{\omega}_H = 0.3075$ and $\hat{\omega}_C = 0.1513$ are the aggregate expenditure shares on housing and cars respectively from the GEIH, and λ_{iag} and λ_{ig}^C are the share of all individuals in cell (i, a, g) and the share of car owners in call (i, g) respectively.

I solve for these parameters to exactly match the observed data in each period. For example, for the post period I obtain $T_H = 2.016$, $\bar{h} = 1.2097$ and $p_a = 117.37$ (with 7).

Algorithm for Solving the Model The system of equations to be solved are provided in the proof of proposition 1. In this section, I outline the iterative algorithm used to solve for the equilibrium of the model

1. Guess a vector $w^0, \vartheta^0, r^0, u^0, A^0$

[64] This can be computed given calibrated wages and productivities, as well as observed commercial floorspace prices.

2. Given a wage vector $w^t, \vartheta^t, r^t, u^t, A^t$

   (a) Compute $H^t_{Ri} = \vartheta^t_i H_i$, $H^t_{Fi} = (1-\vartheta^t_i)H_i$, $\Phi^t_{Riag} = \sum_j (w^t_{jg}/d_{ija})^{\theta_g}$ and $W^t_{is} = \left(\sum_h \alpha^{\sigma_L}_{sh}(w^t_{ih})^{1-\sigma_L}\right)^{\frac{1}{1-\sigma_L}}$.

   (b) Compute $P_t = \left(\sum_{j,s}\left(((W^t_{js})^{\alpha}(r^t_{Fj})^{1-\alpha}/A_{js}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$, where $r^t_{Fj} = (1-\tau_i)r^t_i$

   (c) Compute L^t_R from

$$L^t_{Riag} = \bar{L}_g \frac{\left(u^t_{iag}(T_g\Phi^{1/\theta}_{Riag} - \bar{h}r^t_{Ri} - p^t_a a + \pi^t)r^{\beta-1}_{Ri}\right)^{\eta_g}}{\sum_{r,o}\left(u^t_{rog}(T_g\Phi^{1/\theta}_{Rrog} - \bar{h}r_{Rr} - p^t_o o + \pi^t)r^{\beta-1}_{Rr}\right)^{\eta_g}}$$

   where $p^t_a = p_a P^t$ and $\pi^t = \bar{L}^{-1}\left(\sum_i H_{Ri}r_{Ri} + H_{Fi}r_{Fi}\right)$.

   (d) Compute labor supply $\tilde{L}^t_{Fig} = (w^t_{jg})^{\theta_g-1}\Psi^t_{jg}$, where $\Psi^t_{jg} \equiv T_g \sum_{r,o}(\Phi^t_{Riog})^{-\frac{\theta_g-1}{\theta_g}} d^{-\theta_g}_{rjo} L^t_{Rrog}$.

   (e) Update the main variables

$$\begin{aligned}
\tilde{w}_{jg} &= \left[\frac{(P^t)^{\sigma-1}X^t\sum_s B_{isg}A^{\sigma-1}_{is}(W^t_{is})^{\sigma_L-(1+\alpha_s(\sigma-1))}(r^t_{Fi})^{-(1-\alpha_s)(\sigma-1)}}{\Psi^t_{jg}}\right]^{\frac{1}{\theta_g+\sigma_L-1}} \\
\tilde{r}_i &= \frac{E^t_i + (1-\alpha)Y^t_i}{H_i} \\
\tilde{\vartheta}_i &= \begin{cases} 1 & i \in \mathcal{D}_R \backslash \mathcal{D}_F \\ 0 & i \in \mathcal{D}_F \backslash \mathcal{D}_R \\ \frac{E^t_i}{E^t_i + (1-\alpha)Y^t_i} & i \in \mathcal{D}_F \cap \mathcal{D}_R \end{cases} \\
\tilde{A}_{js} &= \bar{A}_{js}(\tilde{L}^t_{Fj}/T_j)^{\mu_A} \\
\tilde{u}_{iag} &= \bar{u}_{iag}\left(L^t_{RiH}/L^t_{Ri}\right)^{\mu_U}
\end{aligned}$$

   where $X^t = \beta\sum_{i,g,a}(T_g(\Phi^t_{Riag})^{1/\theta_g} - p^t_a a - r^t_i\bar{h} + \pi^t)L_{Riag}$ is aggregate expenditure on goods, $Y^t_i = \sum_s (p^t_{is})^{1-\sigma}(A_{js}P^t)^{\sigma-1}X^t$ is firm sales in i and $E^t_i = r^t_i\bar{h}L^t_{Ri} + (1-\beta)\sum_{a,g}(T_g(\Phi^t_{Riag})^{1/\theta_g} - p^t_a a - r^t_i\bar{h} + \pi^t)L^t_{Riag}$ is expenditure on housing.

3. $||(\tilde{w}, \tilde{\vartheta}, \tilde{r}, \tilde{u}, \tilde{A}) - (w^t, \vartheta^t, r^t, u^t, A^t)||_\infty < \epsilon_{tol}$ then stop. Otherwise, set $(w^{t+1}, \vartheta^{t+1}, r^{t+1}, u^{t+1}, A^{t+1}) = \zeta(w^t, \vartheta^t, r^t, u^t, A^t) + (1-\zeta)(\tilde{w}, \tilde{\vartheta}, \tilde{r}, \tilde{u}, \tilde{A})$ for some $\zeta \in (0,1)$ and return to step 2.

Since the equilibrium system is only defined to scale (it is homogenous of degree zero), I normalize the geometric mean of wages to one. In order to keep the scale of different variables on the same order of magnitude, I also normalize the geometric mean of floorspace prices to one prior to solving for the model's unobservables. This affects the scale of unobservables such as productivities and amenities, but has no impact on relative differences in exogenous characteristics or endogenous variables across locations or counterfactuals.

D.4 Benchmarking the Amenity Spillovers

The estimated amenity spillovers can be benchmarked to Diamond (2016) who estimates a spillover of the form $u_{ig} = \bar{u}_{ig}(L_{Hi}/L_{Li})^{\mu_{U,g}}$ finds on average $\mu_U \approx 2.62$. To a first order, in this paper $u_{kig} \approx \bar{u}_{ikg}(L_{Hi}/L_{Li})^{\mu_{U,g}(1-\pi_H)}$ where π_H is the share of high-skill workers. Using $\pi_H = 0.3$ from 2005 (the midpoint of the period in question), the average estimate of 0.818 gives $E\left[\mu_{U,g}\right](1-\pi_H) = 0.572$, about one quarter of Diamond (2016).

E Model Extensions

E.1 Migration

The baseline model considers a closed city with a fixed population. This section relaxes this to allow for migration into the city from the rest of the country.

We assume that workers in Colombia face a choice to live in Bogotá or the rest of the country. Workers make their migration choice based on expected utility in the destination; their expected utility is $\bar{U}$ in Bogotá and $\bar{U}^{Rest}$ in the rest of the country. This latter term is an exogenous model parameter. Letting individuals have a multiplicative preference $\eta(\omega)$ for each choice distributed Frechet with shape parameter $\rho > 0$, the number of workers choosing to live in Bogotá is

$$\bar{L} = \bar{L}^{Col}\left(\frac{\bar{U}}{\bar{U}^{Rest}}\right)^{\rho},$$

where $\bar{L}^{Col}$ is the (exogenous) population of the entire country. In changes this yields

$$\hat{\bar{L}} = \frac{\hat{\bar{U}}^{\rho}}{\pi^{Bog}\hat{\bar{U}}^{\rho} + \pi^{Rest}},$$

where we have assumed that $\hat{\bar{U}}_{rest} = 1$ (i.e. average utility in the rest of Colombia is unaffected by TransMilenio), and π^{Bog}, π^{Rest} denote the share of Colombians living in Bogotá and the rest of Colombia respectively in the initial period. The remaining equations of the model are unchanged, this simply turns $\hat{\bar{L}}$ from a model parameter into an endogenous variable.

The change in welfare of Bogotanos is now now $\widehat{E\left[\bar{U}\eta(\omega)|\omega \text{ chose Bogotá}\right]} = \left[\pi^{Bog}\hat{\bar{U}}^{\rho} + \pi^{Rest}\right]^{1/\rho}$.

E.2 Congestion

Overview. This section develops an extension of the model that incorporates congestion. While the same system of equations will determine the equilibrium of economic activity in the city given a matrix of commute times, a separate system of equations will be added that determines commute times as a function of economic activity (through the number of commuters). These will then be solved jointly to quantify the response of the equilibrium to a change in infrastructure allowing for congestion.

The extension blends elements from Allen and Arkolakis (2021) and Gaduh et. al. (2022). Commuters travel along a network where census tracts are nodes and adjacent census tracts (in the network sense) are connected by edges. They choose a route between an origin and destination and for each edge in that route they pick a mode. This extension inherits elements from the nested logit model in the paper. If an individual travels using the public nest, they can choose between any mode in that nest (walking, bus, TransMilenio) for each node. However if they travel by the private nest (i.e. car), they travel by car along each edge. Individuals have route-specific Frechet shocks, yielding convenient expressions for d_{ij} as the expected cost over all the routes they might take between i and j. Travel time on roads by car is subject to within-mode congestion through a power function of the volume of car travel along that edge.

Congestion is incorporated by building off the working paper version of Allen and Arkolakis (2021).[65] Unlike the published version, I allow the elasticity with which commuters choose origin-destination pairs to differ from

[65] This version is Allen and Arkolakis (2019).

the elasticity with which they choose the particular route to get there. This choice is made for two reasons. First, it is restrictive to require commuters have the same heterogeneity in idiosyncratic preferences across pairs of neighborhoods to live and work as they do across potential routes to get between home and work. For example, commuters may by and large choose the fastest route between home and work (low dispersion in preferences over routes) but tend to choose quite different home and work locations all else equal (high dispersion in preferences over live-work pairs). Second, this choice keeps the economic and traffic modules of the model separate. In so doing, the reduced form elasticities $\boldsymbol{\beta}_R, \boldsymbol{\beta}_F$ from the baseline model continue to determine the response of economic activity to changes in travel times. The difference is that now, the change in travel times with respect to changes in infrastructure will depend on commuting choices through congestion. The extension borrows the idea from Gaduh et. al. (2022) to incorporate multiple travel modes by allowing commuters to choose routes between origins and destinations across alternative link-mode combinations, but allows for differential substitution patterns across modes when using public transit as opposed to driving.

Traffic Module. To construct a tractable way of incorporating congestion, I model the routing choice of commuters using the discrete choice framework from Allen and Arkolakis (2021). Between each pair of locations is an infrastructure matrix $\mathbf{T}(m) = [t_{kl}(m)]$ for mode $m \in$ {Walk,Bus,TransMilenio,Car}, where $t_{kl}(m) \geq 0$ is the minutes of travel between location k and l on mode m. If no direct link exists between k and l on the network of mode m, I set $t_{kl}(m) = \infty$. I also set $t_{kk}(m) = \infty$ to exclude self-loops.

The disutility of travel over link kl using mode m is simply $\exp\left(\kappa t_{kl}(m) + \tilde{b}(m)\right)$, where $\tilde{b}(m)$ is an amenity associated with each mode as in the baseline model. I assume these costs are multiplicative, so that if a commuter chooses a route $r = \{i = r_0, r_1, \ldots, r_K = j)$ of length K between i and j, the total cost is $\exp\left(\kappa t_{ijr} + \tilde{b}_r\right)$ where $t_{ijr} = \sum_{k=1}^{K} t_{r_{k-1},r_k}(m_{r_{k-1},r_k})$ and $\tilde{b}_r = \sum_{k=1}^{K} \tilde{b}(m_{r_{k-1},r_k})$. Note here that m_{r_{k-1},r_k} is the mode chosen on link r_{k-1}, r_k of the route. Lastly, I allow commuters to have an idiosyncratic multiplicative preference for a particular route $\exp(\nu_r(\omega))$, where $\nu_r(\omega)$ is distributed T1EV for minima with shape parameter $\lambda > 0$. Under the same structure of preferences from the baseline model, indirect utility from choice (i, j, r) is

$$U_{ijr}(\omega) = \frac{u_i w_j r_{Ri}^{\beta-1}}{\exp\left(\kappa t_{ijr} + \tilde{b}_r + \nu_r(\omega)\right)} \epsilon_{ij}(\omega).$$

Assuming that workers first choose where to live and work and then choose which route to commute with and solving this via backward induction, the route choice problem is simply

$$\min_{r \in P_K, K \geq 0} \left\{\exp\left(\kappa t_{ijr} + \tilde{b}_r + \nu_r(\omega)\right)\right\}.$$

Workers become car owners with probability ρ^{Car}. If they do not own a car, they choose between public modes only. Properties of the T1EV distribution imply that

$$E\left[\min_{r \in P_K^{Pub}, K \geq 0} \{\exp(\kappa t_{ijr} + \nu_r(\omega))\}\right] = \exp(-\kappa \bar{t}_{ij})$$

$$\text{where } \bar{t}_{ijPub} = -\frac{1}{\kappa\lambda} \ln \sum_{K=0}^{\infty} \sum_{r \in P_K^{Pub}} \exp\left(-\kappa\lambda \sum_{k=1}^{K} \left[t_{r_{k-1},r_k}(m_{r_{k-1},r_k}) + \tilde{b}(m_{r_{k-1},r_k})\right]\right),$$

where P_K^{Pub} are all paths of length K using the public transit network consisting of $m \in \{\text{Walk,Bus,TransMilenio}\}$. If a worker does own a car, they can also choose to to travel using the car alone with

$$\bar{t}_{ijCar} = -\frac{1}{\kappa\lambda} \ln \sum_{K=0}^{\infty} \sum_{r \in P_K^{Car}} \exp\left(-\kappa\lambda \sum_{k=1}^{K} \left[t_{r_{k-1},r_k}(m_{r_{k-1},r_k}) + \tilde{b}(m_{r_{k-1},r_k})\right]\right),$$

where P_K^{Car} are all paths of length K using the car network. If a worker owns a car, they decide whether or not to use it to commute and solve $\max\{\bar{t}_{ijPub} + \epsilon_{Pub}, \bar{t}_{ijCar} + \epsilon_{Car}\}$. Assuming the idiosyncratic preference draws $\epsilon_{Pub}, \epsilon_{Car}$ are drawn iid from a T1EV distribution, the probability of choosing to travel using the car conditional on owning a car is

$$P_{ijCar|Car} = \frac{\exp\left(-\kappa\bar{t}_{ijCar}\right)}{\exp\left(-\kappa\bar{t}_{ijCar}\right) + \exp\left(-\kappa\bar{t}_{ijPub}\right)}.$$

Note that overall expected utility is

$$\begin{aligned} E_a\left[\max_m \left\{U_{ijm|a}(\omega)\right\}\right] &= u_i w_j r_{Ri}^{\beta-1} \epsilon_{ij}(\omega) \times \Bigg[\rho_{car} \Bigg(E\Bigg[\min_{r \in P_K^{Pub}, K \geq 0} \left\{\exp\left(\kappa t_{ijr} + \nu_r(\omega)\right)\right\} + \epsilon_{Pub}, \min_{r \in P_K^{Car}, K \geq 0} \left\{\exp\left(\kappa t_{ijr} + \nu_r(\omega)\right)\right\} \\ &= u_i w_j r_{Ri}^{\beta-1} \epsilon_{ij}(\omega) \times \left[\rho_{car}\left(E \max\left\{\exp\left(\kappa\bar{t}_{ijPub} + \epsilon^{Pub}\right), \exp\left(\kappa\bar{t}_{ijCar} + \epsilon^{Car}\right)\right\}\right) + (1-\rho_{car}) \exp\left(\kappa\bar{t}_{ijCar}\right)\right] \\ &= u_i w_j r_{Ri}^{\beta-1} \epsilon_{ij}(\omega) \times \left[\rho_{car}\left(\exp\left(\kappa\bar{t}_{ijOwnCar}\right)\right) + (1-\rho_{car}) \exp\left(\kappa\bar{t}_{ijPub}\right)\right] \end{aligned}$$

where

$$\bar{t}_{ijOwnCar} \equiv -\frac{1}{\kappa} \ln\left[\exp\left(-\kappa\bar{t}_{ijPub}\right) + \exp\left(-\kappa\bar{t}_{ijCar}\right)\right].$$

So altogether

$$\begin{aligned} E_a\left[\max_m \left\{U_{ijm|a}(\omega)\right\}\right] &= \frac{u_i w_j r_{Ri}^{\beta-1} \epsilon_{ij}(\omega)}{\exp\left(\kappa\bar{t}_{ij}\right)} \\ \text{where } \bar{t}_{ij} &= -\frac{1}{\kappa} \ln\left[\rho_{Car} \exp\left(-\kappa\bar{t}_{ijOwnCar}\right) + (1-\rho_{car}) \exp\left(-\kappa\bar{t}_{ijPub}\right)\right]. \end{aligned}$$

This therefore is nested within the simple model of Appendix C.1, with a different formulation of commute costs d_{ij}.

Define $\mathbf{A}(m) \equiv \left[a_{kl}(m) \equiv \exp\left(t_{kl}(m) + \tilde{b}(m)\right)^{-\kappa\lambda}\right]$. As in Gaduh et. al. (2022), one can show via induction for the public transit network that

$$\begin{aligned} \exp\left(\bar{t}_{ij}\right)^{-\kappa\lambda} &= \sum_{K=0}^{\infty} \mathbf{A}_{ijPub}^K \\ \text{where } \mathbf{A}_{Pub} &= \sum_{m \in \mathcal{B}^{Pub}} \mathbf{A}(m) \end{aligned}$$

where $\mathbf{A}_{ijPub}^K$ is the ij element of the K matrix power of the matrix $\mathbf{A}_{Pub}$.[66] So long as the spectral radius of $\mathbf{A}_{Pub}$

[66] For $K = 1$, we simply have

$$\exp\left(\bar{t}_{ij,1}\right)^{-\kappa\rho} = \sum_m \exp(-\kappa\rho t_{ij}(m)) = \left[\sum_{m \in \mathcal{B}^{Pub}} \mathbf{A}(m)\right]_{ij} = \mathbf{A}_{ijPub}.$$

Now suppose that $\exp\left(\bar{t}_{ij,K}\right)^{-\kappa\rho} = \left[\mathbf{A}_{Pub}^K\right]_{ij}$. This is the sum of all weights along all paths between ij of length K. To compute is less than one, $\sum_{K=0}^{\infty} \mathbf{A}_{ijPub}^{K} = (\mathbf{I} - \mathbf{A})^{-1} \equiv \mathbf{B}_{Pub}$ and

$$\bar{t}_{ijPub} = -\frac{1}{\kappa\lambda} \ln b_{ijPub},$$

where b_{ijPub} is the ij element of $\mathbf{B}_{Pub}$. For the car network, as in Allen and Arkolakis (2021)

$$\bar{t}_{ijCar} = -\frac{1}{\kappa\lambda} \ln b_{ijCar},$$

where $\mathbf{B}_{Car} \equiv (\mathbf{I} - \mathbf{A}(car))^{-1}$.

To close this module, I need to define how travel costs on each link $t_{kl}(m)$ are determined. I allow these to depend on exogenous characteristics $e_{kl}(m)$ and, for the car network, the traffic using the link $\Xi_{kl}(m)$ [67] through the log-linear functional form

$$t_{kl}(m) = e_{kl}(m)\Xi_{kl}(m)^{\phi_m},$$

where $\phi_{Car} > 0$ and otherwise is zero. There are $L_{ijCar} = P_{ijCar|Car}\rho^{Car} L_{ij}$ commuters using the car network, and so the number of car trips using a link is therefore

$$\Xi_{kl}(Car) = \sum_{ij} \pi_{ij}^{kl}(Car) L_{ijCar}.$$

where $\pi_{ij}^{kl}(Car)$ is the number of times the average driver between i and j uses link kl. The results from Allen and Arkolakis (2021) imply

$$\pi_{ij}^{kl}(Car) = \frac{b_{ikCar} a_{kl}(m) b_{ljCar}}{b_{ijCar}}.$$

Letting $\mathbf{L}_{Car} \equiv [L_{ijCar}]$ denote the matrix of commute flows on the car network, this system can be written in matrix form as

$$\Xi(Car) = \mathbf{A}(Car) \odot [\mathbf{B}'_{Car}(\mathbf{L}_{Car} \oslash \mathbf{B}_{Car})\mathbf{B}'_{Car}]$$

where $\odot$ and $\oslash$ are Hadamard product and division operators respectively. This formulation reduces the size of the matrices that need to be stored, since $\mathbf{A}(Car), \mathbf{B}_{Car}, \mathbf{L}_{Car}$ are all $I \times I$ rather than $\{\pi_{ij}^{kl}(Car)\}$ which is $I^2 \times I^2$.

Lastly, I define the exogenous portion of travel costs in the same way as Allen and Arkolakis (2021). Assuming travel time is given by $t_{kl}(m) = (distance_{kl} \times speed_{kl}^{-1}(m))^{\delta_0}$ and inverse speed is given by $speed_{kl}^{-1}(m) = \gamma(m) \times \left(\frac{\Xi_{kl}(m)}{lanes_{kl}(m)}\right)^{\delta_1(m)} \times \epsilon_{kl}(m)$ where $\gamma(m)$ is a mode-specific shifter and $\epsilon_{kl}(m)$ is a link-mode-specific idiosyncratic term, then

$$t_{kl}(m) = \underbrace{\left[\frac{distance_{kl} \times \gamma(m) \times \epsilon_{kl}(m)}{lanes_{kl}(m)^{\delta_1(m)}}\right]^{\delta_0}}_{e_{kl}(m)} \times \Xi_{kl}(m)^{\phi_m},$$

---

the same for paths of length $K+1$, we simply multiply by the adjacency matrix and sum across all modes that could be taken next

$$\exp(\bar{t}_{ij,K+1})^{-\kappa\rho} = \sum_{m \in \mathcal{B}^{Pub}} \left[\mathbf{A}_{Pub}^{K}\mathbf{A}(m)\right]_{ij} = \left[\mathbf{A}_{Pub}^{K} \sum_{m \in \mathcal{B}^{Pub}} \mathbf{A}(m)\right]_{ij} = \left[\mathbf{A}_{Pub}^{K}\mathbf{A}_{Pub}\right]_{ij} = \left[\mathbf{A}_{Pub}^{K+1}\right]_{ij}.$$

This proves the conjecture.

[67] A previous version of the paper allowed for congestion on the bus and TransMilenio network too.

where $\phi_m \equiv \delta_0\delta_1(m)$.

Traffic Equilibrium. Collecting the previous results, a traffic equilibrium is a vector $\{t_{kl}(m), \Xi_{kl}(m), \bar{t}_{ij}\}$ that given commute flows L_{ij} and parameters $\delta_0, \delta_1(m), \tilde{b}(m), \gamma(m), lanes_{kl}(m), distance_{kl}$ satisfies the system

$$\begin{aligned}
t_{kl}(m) &= e_{kl}(m)\Xi_{kl}(m)^{\phi_m}\\
\Xi(Car) &= \mathbf{A}(Car) \odot [\mathbf{B}'_{Car}(\mathbf{L}_{Car} \oslash \mathbf{B}_{Car})\mathbf{B}'_{Car}]\\
\bar{t}_{ijPub} &= -\frac{1}{\kappa\lambda}\ln b_{ijPub}\\
\bar{t}_{ijCar} &= -\frac{1}{\kappa\lambda}\ln b_{ijCar}\\
\mathbf{A}(m) &= \left[\exp\left(t_{kl}(m) + \tilde{b}(m)\right)^{-\kappa\lambda}\right]_{kl}\\
\mathbf{A}_{Pub} &= \sum_{m\in\mathcal{B}^{Pub}} \mathbf{A}(m)\\
\mathbf{A}_{Car} &= \mathbf{A}(Car)\\
\mathbf{B}_{Pub} &= (\mathbf{I} - \mathbf{A}_{Pub})^{-1}\\
\mathbf{B}_{Car} &= (\mathbf{I} - \mathbf{A}_{Car})^{-1}\\
L_{ijCar} &= P_{ijCar|Car}\rho^{Car}L_{ij}\\
P_{ijCar|Car} &= \frac{\exp\left(-\kappa\bar{t}_{ijCar}\right)}{\exp\left(-\kappa\bar{t}_{ijCar}\right) + \exp\left(-\kappa\bar{t}_{ijPub}\right)}
\end{aligned}$$

The first three rows is a system of as many equations as unknowns, while the second three rows define the auxiliary variables of that system.

I refer to this as the traffic module of the model: it determines travel times $\bar{t}_{ij}$ given a matrix of commute flows $\mathbf{L}$. Recall that the baseline model pins down changes in economic activity $\{\hat{L}_{Ri}, \hat{L}_{Fi}, \hat{r}_{Ri}, \hat{r}_{Fi}, \hat{\Phi}_{Ri}, \hat{\Phi}_{Fi}, \hat{\tilde{\Phi}}_{Fi}, \hat{\tilde{L}}_{Fi}, \hat{\bar{U}}, \hat{E}\}$ given a change in travel times $\{\hat{d}_{ij}\}$. I therefore also need to express the change in travel times as a function of the change in commute flows. I will model changes in transit infrastructure as a change in the number of lanes on the mode in question, $\widehat{lanes}_{kl}(m)$. In particular, when simulating the removal of TransMilenio I will set $\widehat{lanes}_{kl}(m)$ to a very small number $\forall kl, m = TransMilenio$ so that $\hat{t}_{kl}(m) \to \infty$.[68] In response to this change in model parameters, the change in traffic equilibrium can be written as

$$\begin{aligned}
\hat{t}_{kl}(m) &= \widehat{lanes}_{kl}^{-\phi_m}(m)\hat{\Xi}_{kl}^{\phi_m}(m)\\
\hat{\Xi}_{kl}(Car)\Xi_{kl}(Car) &= \left[\mathbf{A}'(Car) \odot (\mathbf{B}'_{Car})' \left(\mathbf{L}'_{Car} \oslash \mathbf{B}'_{Car}\right)(\mathbf{B}'_{Car})'\right]\\
\bar{t}'_{ijk} - \bar{t}_{ijk} &= -\frac{1}{\kappa\lambda}\ln\left(b'_{ijk}/b_{ijk}\right) \quad k \in \{Pub, Car\}\\
\mathbf{A}'(m) &= \left[\exp\left(\hat{t}_{kl}(m)t_{kl}(m) + \tilde{b}(m)\right)^{-\kappa\lambda}\right]_{kl}\\
\mathbf{A}'_{Pub} &= \sum_{m\in\mathcal{B}^{Pub}} \mathbf{A}'(m)\\
\mathbf{B}'_{Pub} &= (\mathbf{I} - \mathbf{A}'_{Pub})^{-1}
\end{aligned}$$

[68] Since $\hat{e}_{kl}(m) = \widehat{lanes}_{kl}(m)^{-\phi_m}$, setting lanes equal to zero in the counterfactual would leave $\hat{e}_{kl}(m) = \infty$ and the new equilibrium would be undefined.

$$\mathbf{B}'_{Car} = (\mathbf{I} - \mathbf{A}'(Car))^{-1}$$

This provides a system that pins down $\{\hat{t}_{kl}(m), \hat{\Xi}_{kl}(Car), \bar{t}'_{ijk} - \bar{t}_{ijk}\}$ given data from the initial equilibrium $\{\Xi_{kl}(Car), \tilde{b}(m), t_{kl}\}$ and commute flows in the counterfactual equilibrium $\mathbf{L}' = \hat{\mathbf{L}} \odot \mathbf{L}$. Since $\hat{d}_{ij} = \exp\left(\kappa\left(\bar{t}'_{ij} - \bar{t}_{ij}\right)\right)$ with $\bar{t}_{ij}$ as defined above, this pins down the change in commute costs given the shock to infrastructure $\widehat{lanes}_{kl}(m)$ and the change in commute flows $\hat{\mathbf{L}}$. Combining the economic module of the model with the traffic module provides one large system of equations that jointly finds the distribution of changes in economic activity and traffic that is consistent with equilibrium in both modules of the model.

Calibrating the Model. To solve the model in changes, I require values for the parameters $\delta, \delta_1(m), \lambda, \tilde{b}(m), \gamma(m)$ and data $t_{kl}(m), \Xi_{kl}(m)$. Note that link-level traffic and travel times are unobserved, so these will be have to be calibrated along with the deep model parameters.[69]

Given a value for the parameters $\delta, \delta_1(m), \lambda$, I need to solve for the preference shifters $b(m)$ and speed shifters $\gamma(m)$. I estimate these to match average speed and choice shares for each mode. With these in hand, I can solve for $t_{kl}(m), \Xi_{kl}(m)$ which are consistent with the model and observed data given deep parameters $\delta, \delta_1(m), \lambda$.

Lastly, I calibrate these deep traffic parameters $\delta, \delta_1(m), \lambda$ to existing values from the literature. First, I set the routing elasticity $\lambda = 175$ from Allen and Arkolakis (2019). This implies highly elastic routing choices, so that commuters take close to the least cost route between origins and destinations. Second, as in Allen and Arkolakis (2021) I set $\delta_0 = 1/\theta$ to match a unit distance elasticity. Lastly, I calibrate $\delta_1(Car)$ to give a congestion elasticity $\phi_{Car} = \delta_0\delta_1(Car) = 0.06$, the average congestion elasticity estimated for Bogotá by Duranton and Akbar (2017). $\phi_m = 0$ for all other modes.

E.3 Endogenous Floorspace Use with Fixed Housing Supply

This section considers an extension of the baseline model in which total floorspace supply is fixed but the share used for commercial purpose ϑ_i is endogenous. To rationalize differences in commercial and residential floorspace prices, we allow for a tax equivalent of zoning regulations which mean that floorspace owners receive $(1-\tau_i)r_{Fi}$ for each unit of floorspace allocated to commercial use. Denoting $r_i = r_{Ri}$, no arbitrage across floorspace use implies $r_{Fi} = (1-\tau_i)r_i$. This implies that the share of floorspace used for commercial purpose and the floorspace price is pinned down by

$$\vartheta_i = \frac{H_{Fi}}{H_{Ri} + H_{Fi}} = \frac{(1-\alpha)\left(w_i^{\alpha}\left((1-\tau_i)r_i\right)^{1-\alpha}\right)^{1-\sigma} A_i^{\sigma-1} E}{(1-\beta)\,\Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}} + (1-\alpha)\left(w_i^{\alpha}\left((1-\tau_i)r_i\right)^{1-\alpha}\right)^{1-\sigma} A_i^{\sigma-1} E}$$

$$r_i = \frac{(1-\alpha)\left(w_i^{\alpha}\left((1-\tau_i)r_i\right)^{1-\alpha}\right)^{1-\sigma} A_i^{\sigma-1} E + (1-\beta)\,\Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}}}{H_i}.$$

[69] To construct these, I need values for $distance_{kl}$ and $lanes_{kl}(m)$. For walk, car and bus networks these are computed between adjacent census tracts. $distance_{kl}$ is the minimum distance on each mode's network between adjacent tract centroids along the network. For non-adjacent tracts or adjacent tracts not connected by a network, $distance_{kl} = \infty$. For the TransMilenio network, $distance_{kl}$ is finite for two census tracts that are directly connected via the network, i.e. there is no stop between them. The number of lanes is equal to one for any pair of connected tracts for the walk, bus and TransMilenio network. For the car network, I first construct dummies for whether a paid is connected via primary, secondary and tertiary connections (which are not mutually exclusive; a pair can be connected via multiple road types). I then assign 5 lanes to primary connections, 2 to secondary connections and 1 to tertiary connections to approximate the road widths documented in Google Earth, and then compute the total number of lanes between a pair as the sum all road type connections (i.e. the maximum number of lanes is 8).

These equations hold for mixed use locations with $\vartheta_i \in (0,1)$, which one can show to be locations where $\bar{u}_i > 0$ and $\bar{A}_i > 0$. If either exogenous amenities or productivities are zero in a location, that location becomes completely specialized in that type of floorspace.

Extending the equilibrium system to incorporate these new equations, and writing in changes assuming unobservables are constant across periods yields the system

$$
\begin{aligned}
\hat{L}_{Ri}^{1-\theta\mu_U}\hat{r}_i^{\theta(1-\beta)} &= \hat{\hat{L}}\hat{\hat{U}}^{-\theta}\hat{\Phi}_{Ri} \\
\hat{r}_i &= \frac{\vartheta_i\left(\hat{w}_i^{\alpha}\hat{r}_i^{1-\alpha}\right)^{1-\sigma}\hat{\hat{L}}_{Fi}^{\mu_A(\sigma-1)}\hat{E} + (1-\vartheta_i)\hat{\Phi}_{Ri}^{1/\theta}\hat{L}_{Ri}^{\frac{\theta-1}{\theta}}}{\hat{H}_i} \\
\hat{r}_i^{(\sigma-1)(1-\alpha)}\hat{\hat{L}}_{Fi}^{\frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1}} &= \left(\hat{\hat{L}}\hat{\hat{U}}^{-(\theta-1)}\right)^{-\frac{\alpha(\sigma-1)+1}{\theta-1}}\hat{E}\hat{\hat{\Phi}}_{Fi}^{\frac{\alpha(\sigma-1)+1}{\theta-1}} \\
\hat{\vartheta}_i &= \frac{\left(\hat{w}_i^{\alpha}\hat{r}_i^{1-\alpha}\right)^{1-\sigma}\hat{\hat{L}}_{Fi}^{\mu_A(\sigma-1)}\hat{E}}{(1-\vartheta_i)\hat{\Phi}_{Ri}^{1/\theta}\hat{L}_{Ri}^{\frac{\theta-1}{\theta}} + \vartheta_i\left(\hat{w}_i^{\alpha}\hat{r}_i^{1-\alpha}\right)^{1-\sigma}\hat{\hat{L}}_{Fi}^{\mu_A(\sigma-1)}\hat{E}} \\
\hat{w}_j &= \left(\left(\hat{\hat{L}}\hat{\hat{U}}^{-\theta}\right)^{\frac{\theta-1}{\theta}}\frac{\hat{\hat{L}}_{Fj}}{\hat{\hat{\Phi}}_{Fj}}\right)^{\frac{1}{\theta-1}}
\end{aligned}
$$

The two equations for $\hat{r}_i$ and ϑ_i are no longer log-linear. However, taking logs, differentiating the original system and substituting out for wages yields the following first order approximation of the system

$$
\begin{bmatrix}
1-\theta\mu_U & \theta(1-\beta) & 0 & 0 \\
-(1-\vartheta_i)\frac{\theta-1}{\theta} & (1+\vartheta_i(\sigma-1)(1-\alpha)) & \vartheta_i(\sigma-1)\left(\frac{\alpha}{\theta-1}-\mu_A\right) & 0 \\
0 & (\sigma-1)(1-\alpha) & \frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1} & 0 \\
\frac{(1-\vartheta_i)(\theta-1)}{\theta} & (\sigma-1)(1-\alpha)(1+\vartheta_i) & -(1-\vartheta_i)(\sigma-1)\left[\mu_A-\frac{\alpha}{\theta-1}\right] & 1
\end{bmatrix}
\begin{bmatrix}
\ln\hat{L}_{Ri} \\ \ln\hat{r}_i \\ \ln\hat{\hat{L}}_{Fi} \\ \ln\hat{\vartheta}_i
\end{bmatrix}
=
$$

$$
\begin{bmatrix}
1 \\ (1-\vartheta_i)\frac{1}{\theta} \\ 0 \\ -(1-\vartheta_i)\frac{1}{\theta}
\end{bmatrix}
\ln\hat{\Phi}_{Ri} +
\begin{bmatrix}
0 \\ \vartheta_i\frac{(\sigma-1)\alpha}{\theta-1} \\ \frac{1+\alpha(\sigma-1)}{\theta-1} \\ \frac{\alpha(\sigma-1)(1-\vartheta_i)}{\theta-1}
\end{bmatrix}
\ln\hat{\hat{\Phi}}_{Fi}
$$

$$
+\begin{bmatrix}
\theta\ln\hat{\bar{u}}_i + \ln\hat{\hat{L}} - \theta\ln\hat{\hat{U}} \\
-\vartheta_i(\sigma-1)\alpha\left(\frac{1}{\theta}\ln\hat{\hat{L}} - \ln\hat{\hat{U}}\right) - \vartheta_i\alpha(\sigma-1)\ln\hat{E} - \ln\hat{H}_i - \vartheta_i(1-\alpha)(\sigma-1)\ln(\widehat{1-\tau_i}) \\
(\sigma-1)\ln\hat{\bar{A}}_i - \frac{\alpha(\sigma-1)+1}{\theta}\left(\ln\hat{L} - \theta\ln\hat{\hat{U}}\right) + \ln\hat{E} \\
(\sigma-1)(1-\vartheta_i)\ln\hat{\bar{A}}_i - \vartheta_i\ln\hat{E} - \vartheta_i(1-\alpha)(\sigma-1)\ln(\widehat{1-\tau_i}) - \alpha(\sigma-1)(1-\vartheta_i)\left[\frac{1}{\theta}\ln\hat{\hat{L}} - \ln\hat{\hat{U}}\right]
\end{bmatrix}
$$

Now the A, b_R, b_F terms have data in them through the initial floorspace share terms ϑ_i. This system can once again be written as

$$
\ln\hat{\mathbf{y}}_i = A^{-1}b_R\ln\hat{\Phi}_{Ri} + A^{-1}b_F\ln\hat{\Phi}_{Ri} + \mathbf{e}_i
$$

$$
\text{where } A =
\begin{bmatrix}
(1-\theta\mu_U)I & \theta(1-\beta)I & \mathbf{0} & \mathbf{0} \\
-(1-\text{diag}(\vartheta_i))\frac{\theta-1}{\theta} & (1+\text{diag}(\vartheta_i)(\sigma-1)(1-\alpha)) & \text{diag}(\vartheta_i)(\sigma-1)\left(\frac{\alpha}{\theta-1}-\mu_A\right) & \mathbf{0} \\
\mathbf{0} & (\sigma-1)(1-\alpha)I & \frac{\theta+(\sigma-1)(\alpha-\mu_A(\theta-1))}{\theta-1}I & \mathbf{0} \\
\frac{(1-\text{diag}(\vartheta_i))(\theta-1)}{\theta} & (\sigma-1)(1-\alpha)(1+\text{diag}(\vartheta_i)) & -(1-\text{diag}(\vartheta_i))(\sigma-1)\left[\mu_A-\frac{\alpha}{\theta-1}\right] & I
\end{bmatrix}_{4I\times 4I}
$$

$$b_R = \begin{bmatrix} I \\ \frac{1}{\theta} I \\ \mathbf{0} \\ -(I - \text{diag}(\vartheta_i))\frac{1}{\theta} \end{bmatrix}_{4I \times I}$$

$$b_F = \begin{bmatrix} \mathbf{0} \\ \frac{\alpha(\sigma-1)}{\theta-1} I \\ \frac{1+\alpha(\sigma-1)}{\theta-1} I \\ \frac{\alpha(\sigma-1)}{\theta-1}(I - \text{diag}(\vartheta_i)) \end{bmatrix}_{4I \times I}$$

Two insights follow from comparing this system to the equilibrium in the baseline model. First, there are now heterogenous elasticities across locations since the A^{-1} matrix contains data on initial land shares which differ by location. Second, since A is no longer block diagonal, each outcome now depends on both RCMA and FCMA.

E.4 Housing Supply Adjustment and Land Value Capture

Housing Supply. This section outlines the extension of the model allowing for a housing supply response to the transit infrastructure. First, we consider a model where housing supply can freely adjust in each location, and floorspace use is endogenous as in Section E.3. Housing is produced according to the Cobb-Douglas technology $H_i = T_i^{1-\eta} K_i^{\eta}$. The price of capital is normalized to one. Defining the production function on one unit of land as $h_i = k_i^{\eta}$ where $k_i \equiv K_i/T_i$, developers solve the problem

$$\max_{k_i} k_i^{\eta} r_i - k_i - p_i$$

where p_i is the price of land in i. This yields the density of construction per unit of land of $k_i = (\eta r_i)^{\frac{1}{1-\eta}}$ and profits $\tilde{\eta} r_i^{\frac{1}{1-\eta}} - p_i$ were $\tilde{\eta} \equiv \eta^{\frac{\eta}{1-\eta}}$. The price of land adjusts so that developers earn zero profits $p_i = \tilde{\eta} r_i^{\frac{1}{1-\eta}}$. Total housing supply is then given by $H_i = T_i(\eta r_i)^{\frac{\eta}{1-\eta}}$. The share of floorspace allocated to commercial use ϑ_i is determined as in Section E.3. The remainder of the model equations are unchanged; this housing supply condition is simply added to them. To ensure this fits the data in the initial period, a residual $\zeta_i = H_i / T_i(\eta r_i)^{\frac{\eta}{1-\eta}}$ is introduced so that the effective units of land are actually $T_i \zeta_i$. This wedge can be interpreted as the quality of land.

Land Value Capture. In the Land Value Capture scheme, only a subset of locations are allowed to have their floorspace adjust.

Under the distance-based scheme, locations $i \in \mathcal{I}$ within 500m from a station are allocated a 30% increase in floorspace. Their floorspace is allowed to increase up to a maximum of 30%, but not decrease (which is the relevant case for a relatively short 16 year time horizon). That is,

$$\hat{H}_i = \begin{cases} \max\{1, (\hat{r}_i)^{\frac{\eta}{1-\eta}}\} & \text{if } i \in \mathcal{I} \text{ and } \max\{1, (\hat{r}_i)^{\frac{\eta}{1-\eta}}\} < 1.3 \\ 1.3 & \text{if } i \in \mathcal{I} \text{ and } \max\{1, (\hat{r}_i)^{\frac{\eta}{1-\eta}}\} > 1.3 \\ 1 & \text{if } i \notin \mathcal{I}. \end{cases}$$

Perfect competition ensures the price of the permits adjust so that that developers earn zero profits, so income from the scheme is $(H' - H_i) r_i'$ where prices are evaluated in the new equilibrium.

Under the CMA-based scheme, locations are allocated permits proportional to their change in CMA $\vartheta_i \Delta \ln \Phi_{Ri} +$

$(1-\vartheta_i)\Delta\ln\Phi_{Fi}$ so that the number of potential new permits (or, equivalently, the maximum amount of new floorspace created) is the same as under the distance-based scheme. Here ϑ_i is the commercial floorspace share in the initial equilibrium and the CMA changes are those using the baseline measure that hold population and employment fixed at their initial levels, so this is all information which the policy maker would have at the time of the intervention.

Parameterization. In the quantitative exercises, a conservative choice for the housing elasticity is made so that $\eta/(1-\eta) = 0.7$ to match the most inelastic cities in the US from Saiz (2010). This value corresponds to his value for Oakland, CA which is ranked the 6th most inelastic city, one position behind San Francisco and San Diego (3rd and 4th) and a couple ahead of New York and Chicago (9th and 12th). I also shut down spillovers for a conservative estimate, especially in the open city case where a large value for the amenity spillover can lead to larger changes in population than in the baseline model.

E.5 Employment in Domestic Services

This section outlines the extension of the model that incorporates employment in domestic services. I begin by noting the following facts. First, between 2000-2014 in the GEIH 7.3% of non-college educated Bogotanos worked as domestic helpers while almost no college educated workers did. Second, in the 2014 Multipurpose Survey I observe that 30.3% of college-educated households employ domestic services, compared to only 3.6% of non-college households. Third, conditional on employing domestic servants households spend on average 0.15 of their income on their wages, a fraction that remains constant with income. Unfortunately employment in domestic services by employment location is reported neither in the census nor in the CCB. Therefore, given that 90% of domestic servants are employed in college educated households, I impute domestic employment by assigning each worker equally to high skilled households and scaling up until the total matches the number observed in the GEIH.

These observations motivate the following extension of the model. I assume that only high-skilled households consume domestic services while only low-skilled workers are used in its production. I also assume domestic services enter the utility of the high skilled according to Cobb-Douglas preferences with an expenditure share of 0.045 (=0.303*0.15). That is, I assume the common component of utility is given by

$$U_H = C^{1-\beta_H-\beta_D}(H-\bar{h})^{\beta_H} D^{\beta_H^D}$$

In each location, a perfectly competitive firm produces domestic services under the linear technology $Y_{iD} = \tilde{L}_{FiL}$. The cost is therefore equal to the low-skill wage $p_i^D = w_{Li}$. Market clearing for domestic services therefore requires that

$$\beta_D E_{iH} = p_i^D D_i = \frac{w_{Li}\tilde{L}^D_{FiL}}{\bar{A}_{Di}}$$

where $\bar{A}_{Di}$ is a residual that ensures this condition holds and reflects factors that make i more or less easy to work in as a domestic servant.

The equilibrium equations of the model remain the same, apart from the labor demand equation which becomes

$$\tilde{L}_{Fig} = w_{ig}^{-\sigma_L} P^{\sigma-1} E \sum_s B_{isg} A_{is}^{\sigma-1} W_{is}^{\sigma_L-(1+\alpha_s(\sigma-1))} r_{Fi}^{-(1-\alpha_s)(\sigma-1)} + \mathbb{I}_{gL}\frac{\beta_H^D E_{iH}}{w_{Li}}$$

and the expression for residential populations for high skilled which becomes

$$L_{Riag} = \bar{L}_g \frac{\left(u_{iag}(T_g \Phi_{Riag}^{1/\theta} - \bar{h} r_{Ri} - p_a a) r_{Ri}^{\beta-1} w_{Li}^{\beta_g^D}\right)^{\eta_g}}{\sum_{r,o} \left(u_{rog}(T_g \Phi_{Rrog}^{1/\theta} - \bar{h} r_{Rr} - p_o o) r_{Rr}^{\beta-1} w_{Lr}^{\beta_g^D}\right)^{\eta_g}}, g = H.$$

The other ingredients of the model are unchanged. The procedure to solve the model and unobservables is unchanged, other than for wages. The system of equations is extended to include the domestic service sector:

$$D_{ig}(w) = w_{ig}^{\theta_g} \left[\sum_s \frac{L_{Rsg}}{\sum_k w_{kg}^{\theta_g} d_{sk}^{-\theta_g}} d_{si}^{-\theta_g}\right] - \left[\sum_s \frac{(w_{ig}/\alpha_{sg})^{-\sigma}}{\sum_h (w_{ih}/\alpha_{sh})^{-\sigma}} \frac{\bar{\epsilon}_{is}}{\bar{\epsilon}_{ig}} L_{Fis} + L_{FiD} \mathbb{I}_{gL}\right]$$

where $\mathbb{I}_{gL}$ is a dummy for whether g is L, and L_{FiD} is employment in domestic services as described above.

E.6 Home Ownership

This section outlines the extension of the model that allows for local home ownership across worker groups to match the ownership rates observed in the data. In the data, home ownership rates are 0.603 and 0.457 for college and non-college individuals respectively in 2015. Letting o_L and o_H be the shares of home owners in the data, I therefore assume that total income is given by

$$\frac{w_{jg}\epsilon_j(\omega)}{d_{ija}} + o_g \frac{E_i}{L_{Ri}}$$

where $E_i = \sum_{g,a} \left(r_{Ri}\bar{h} + (1-\beta)(\bar{y}_{iag} - p_a a - r_{Ri}\bar{h} + \pi_{ig})\right) L_{Riag}$ is total expenditure on housing by residents of i, L_{Ri} are total residents in i and $\pi_{ig} \equiv o_g \frac{E_i}{L_{Ri}}$ is income from home ownership. That is, the model is the same with one replacement of π with π_{ig}. The remaining equilibrium equations and procedure to solve for unobservables are easily extended to incorporate this change.

F Data Appendix

This section provides supplementary information on the data used in this paper.

F.1 Dataset Description

Population

The primary source of population data is DANE's General Census of 1993, 2005 and 2018. This contains the population in each block by education-level. I define "college" educated workers to be those with more than post-secondary education (defined by the level achieved during their last complete year of study). This contains both conventional universities and technical colleges, but the small size of the latter means the results are not sensitive to this grouping. My main results include adults 20 and older; the results are robust to including individuals of all ages.[70]

[70]The data provided to me by DANE provided population totals by education level and age across 10 year age bins.

Commuting

Commuting data comes from the city's Mobility Survey administered by the Department of Mobility and overseen by DANE. Conducted in 2005, 2011 and 2015, these are household surveys in which each member was asked to complete a travel diary for the previous day. For 1995, I obtained the Mobility Survey undertaken by the Japan International Cooperation Agency (JICA) to similar specifications as the DANE surveys. The samples sizes are similar across years, including 141,316 trips for 73,830 individuals in 20,002 households per round on average.[71] I include only trips that originate or end in municipal Bogotá in the analysis.[72] Sampling weights are also provided.

The survey reports the demographic information of each traveller and household, including age, education, gender, industry of occupation, car ownership and in some years income.[73] For each trip, the data report the departure time, arrival time, purpose of the trip, mode, as well as origin and destination UPZ.[74] Since all trips are reported, these include commutes (trips to work) as well as for other purposes (e.g. shopping, seeing friends). Reported modes are often quite detailed (e.g. 25 options in 2011); I often aggregate into car, bus, TransMilenio, and others (walking, bicycle, motorbike). Trips on TransMilenio trunk and feeder buses are reported separately, so I consider TransMilenio trips to be those involving at least one stage on a trunk bus (multiple modes can be reported in a single trip).

Housing

As described in the main text, the mission of the cadastre is to keep the city's geographical information up to date and thus 98.6% of the city's more than 2 million properties are included.[75] The city is recognized as a pioneer on the continent for the quality of its cadastre (Anselin and Lozano-Gracia 2012). In addition to having an updated record of the city's layout, the cadastre is important for the government due to its importance in city revenues: in 2008, for example, property taxes accounted for 19.8% of Bogotá's tax revenues (Uribe Sanchez 2010). These taxes depend on assessed property values. In developed countries, property valuations are typically determined using data on market transactions. However, Bogotá, like most developing cities, lacks comprehensive records of such data. The city circumvents this by assessing property prices as follows. First, they collect available data on transactions through outreach to the real estate sector (Uribe Sanchez 2010). Second, through a census-like process officials collect information on property sales announced through signs and local newspapers, survey these properties and then contact the owners pretending to be potential buyers. They negotiate to get as close as possible to an actual sales price and record the final value, under the premise of a cash payment (Anselin and Lozano-Gracia 2012). Third, the city hires teams of professional assessors to value at least one property in one of each of the city's "homogenous zones", which currently exceed 16,000 (Ruiz and Vallejo 2010).[76] The net effect of these efforts should be that a comprehensive record of property values which are less prone to under-reporting for tax avoidance.

[71]Minima-maxima across years are (i) 117,217-169,766 trips, (ii) 58,313-91,765 individuals and (iii) 15,519-28,213 households.

[72]Municipal Bogotá accounts for 85% of the residents of the Bogotá metropolitan area, and only 5% of employment in municipal Bogotá comes from outside the municipality (Akbar and Duranton 2017)

[73]The 1995 survey reports raw income, while in 2011 and 2015 eight income bin dummies are reported.

[74]In certain years more precise spatial information is reported, such as address of origin and destination in 2011, but UPZ are consistently reported across all years.

[75]I confirmed this comprehensive coverage by overlaying the shapefile of plots with data over satellite images.

[76]These zones are determined by employees of the cadastral office who physically walk around the city and classify each neighborhood into a zone of similar attributes based on observation and their knowledge of the city. Criteria used to define "homogeneity" include categories for main activities, access to public services, and dominant land use. This process is extremely cost intensive, representing around 73% of the total costs of estimating cadastral values (Anselin and Lozano-Gracia 2012).

The city then combines this data on actual and assessed valuations with building characteristics to construct assessed values for each property. By law, during every updating process each parcel must surveyed by enumerators using a "parcel form" that contains more than 60 questions about the property.

One concern is whether properties surveys and assessments are made very infrequently, with annual changes based solely on an aggregate inflation rate. While assessments are indeed inflated on a yearly basis, information for individual properties is frequently updated through visits: between 2000 and 2006 over 1,036,000 properties were updated, while a large push in 2008-2009 updated all of the city's 2 million properties (Ruiz and Vallejo 2010).[77] My primary focus on long-differences in housing market outcomes ensures that data for essentially all properties was updated.

To validate the valuations in the cadastre, I compare these assessed values per m2 in 2014 with purchase prices per room reported in DANE's 2014 Multipurpose Survey. This survey is a slightly more detailed version of the household survey discussed below. One question asks respondents to report the purchase price and year for their current home. I keep the 5,497 observations for which the purchase was made in the past 10 years,[78] and compute the average price per room within each locality (the smallest geographical unit in the survey). I merge these year-locality observations with the average price per m2 of residential floorspace in the cadastral database, and take weighted averages of both cadastral and reported unit prices across years where I weight by the number of observations in each year. Figure A.8 plots the average cadastral price against the reported purchase price, normalizing each variable to have unit mean. The measures have a high correlation coefficient of 0.947, with the majority of observations lying along the 45-degree line. Importantly, there appears to be no deviation of the relationship for expensive neighborhoods, which we would expect if cadastral values were systematically over- or under-valuing these properties.[79] Consistent with the city's efforts, it appears that property values in the cadastral data are fairly accurate representations of actual property prices throughout the city.

Finally, to construct comparable measures of floorspace prices by census tract I purge property prices driven by differences in building composition by regressing log floorspace prices per m2 on property characteristics (age bins, point bins) and a set of census tract fixed effects, and recover these fixed effects.

Employment (Firms)

The employment data used in this paper comes from two sources. The first is a census of the universe of establishments from DANE's 2005 General Census and 1990 Economic Census. Panel A of Table A.10 presents some summary statistics. There are many small firms in both years: while average firm size is close to 5 employees, the median firm only has 2 employees while firm size at the 90th percentile is between 6 and 7.

The second source is a database of all registered establishments from Bogotá's Chamber of Commerce (CCB by its Spanish acronym) in 2000 and 2015. The 2015 dataset contains the block of each establishment, its industry and, in many cases, the number of employees. Keeping only observations with non-missing values for all 3 variables leaves around 126,867 observations as reported in Panel B. In 2000 neither the number of employees nor the block are reported, but it does provide the address. Bogotá's clear grid system made it straightforward to geolocate the

[77] Updated assessments and property transaction records were conducted throughout, with assessments for each homogenous zone being updated during the 2008-2009 comprehensive update.

[78] The results are not sensitive to this choice.

[79] Of course, while it is possible that values in the Multipurpose survey themselves are biased, there is no strong reason to think this would be the case since DANE enumerators are well-trained in making clear that responses are anonymous and for statistical purposes only.

vast majority of these.[80] Retaining establishments with non-missing industry codes left 34,332 observations.

Two aspects of the CCB data need addressing. First, there is the absence of employment data for 2000. I therefore rely on establishment counts as a measure of employment when using the CCB in the main analysis. In the 2015 data, I compute the number of establishments in a locality as well as the mean employment and find a correlation of 0.033. In the 2005 census, the correlation is 0.09. Since average establishment size is fairly constant across the city, this suggests establishment counts are a fairly good proxy for employment.

Second, the coverage of establishments is much lower than in the census. While aggregate coverage gaps will not matter for the analysis, relative differences across the city will pose a problem since relative changes in employment in the CCB data may not be representative of actual changes (for example, if informal employment is more likely to be located in certain areas than others).[81] I diagnose the representativeness of the CCB dataset by comparing its spatial distribution of establishments with that reported in the 2005 census. Panels (a) and (b) Figure A.7 plots the density of establishments in each locality in the CCB dataset in each year on the y-axis against the density of establishments in the 2005 census on the x-axis, normalizing both variables to have unit geometric mean. Both figures show a reassuringly tight relationship, with correlations of 0.948 and 0.949 respectively. Importantly, the majority of localities lie along the 45-degree line regardless of whether they are poor (Ciudad Bolivar, Kennedy, Bosa, Tunjuelito) or rich (Chapinero, Usaquen), implying that the coverage is fairly uniform across different types of neighborhoods. Panel (c) confirms that the uniform coverage holds across smaller spatial units, by comparing establishment counts across 631 sectors.

Employment (Workers)

Worker-level employment data comes from DANE's Continuing Household Survey (ECH) between 2000 and 2005, and its extension into the Integrated Household Survey (GEIH) for the 2008-2014. These are monthly labor market surveys covering approximately 10,000 households in Bogotá each year. In the external processing room of DANE's offices in Bogotá, I was able to access versions of these datasets with the block of each household provided.[82] The sampling scheme is a repeated cross-section, and so while it is possible to document changes within geographic areas it is not possible to track individuals over time. The survey includes questions pertaining to individual and household characteristics, as well details on employment such as income, hours worked and industry of occupation across primary and secondary jobs.

Maps and other Datasets

The city provides a geodatabase for use in ArcMap containing spatial datasets on the features of Bogotá. From the road layer I extract shapefiles for primary, secondary and tertiary roads. Walk routes consist of the union of the road network in addition to some smaller pedestrian-only paths. The routes of the bus official bus system (which was integrated towards the end of 2012) are also provided. Given that the aim of the government was to bring the provision of existing routes under one integrated system, I use these current routes to measure the location of the

[80] The success rate was around 95%. Addresses in Bogotá are of the form C26#52-18 which stands for the 26th street (Calle in Spanish) and 52nd avenue, 18 meters from the intersection.

[81] Note that I also require the coverage of the CCB to be representative of overall employment across 1-digit industries used in the analysis, too. I find this indeed to be the case, the correlation between the share of establishments in each 1 digit industry in the CCB data vs the 2005 census is 0.991 in 2015 and 0.984 in 2000.

[82] Public versions provide no additional geographic information within the city

bus network throughout the period.[83] Since buses tended to ignore posted bus stops, I create random bus stops every 250m along each route. The database also includes TransMilenio stations and routes, as well as the routes of feeder buses (which I create stops for in the same way as for normal buses). Finally, I use the topographical layer to compute the slope of land across the city in the computation of the least cost construction path instrument.

In all datasets above, the spatial units are either defined through the Cadastre or DANE's classification. The city's geodatabase provides a map of the geography used by the Cadastre (down to the property-level), while DANE provides a shapefile for their map at the block-level. Luckily, these spatial units remained essentially constant during my period of study.[84] I merge the Cadastre's map to DANE's to use as consistently across analyses, and compute the distance from each tract centroid to particular features (CBD, nearest main road, nearest TransMilenio station in each year) in ArcMap. I place the central business district at the center of the high employment density area in the center-east of the city. This is the historical center of the city cited in the literature; when including this variable in regressions I will allow for a different coefficient depending on whether a tract is in the North, West or South of the city in order to account for the different types of neighborhoods in each axis of the city.

Geographic units referred to in the paper consist of localities (19), UPZs (113), sectors (631), census tracts or sections (2,799) and blocks (43,672).

Lastly, data on crime come from the Bogotá police department, and report the GPS location of all reported violent, property and sexual crimes between 2007 and 2013.

F.2 Computing Commute Times

I compute commute times using the Network Analyst toolbox in ArcMap. This accepts as inputs a set of points to be used as origins and destinations (census tract centroids in my setting), as well as a network consisting of a set of edges and nodes at which these edges can be traversed. Each edge of the network is assigned a cost to travel along it; the toolbox then uses Djikstra's algorithm to compute the least cost paths connecting any origin-destination pair.

In my setting, the networks are defined separately for each mode of transit. The walk network consists of single layer of pedestrian paths. The car network consists of the union of primary, secondary and tertiary roads, that can be joined at any intersection, each of which is associated with a different speed. The bus network is comprised of bus routes described above as well as the walk network; the two intersect only at bus stops which are placed randomly every 250m. The TransMilenio network consists of the trunk network (which can only be entered at stations), the feeder bus network (which can be entered at stops placed in the same was as for buses), and the walk network.[85] In order to compute the time cost to traverse each edge of these networks, it remains to assign a speed to each mode.

While Section G provided evidence that speeds were not changing on routes affected by TransMilenio relative to other locations, Table A.12 shows that aggregate speeds were not quite constant over the period. There was a significant reduction in speeds between 1995 and 2005 (a period of city expansion), which remained relatively constant thereafter. I therefore seek to assign two sets of speeds to match the distribution of observed commute times in the "pre" and "post" periods. In the main results, I use the average of both but provide evidence in

[83] While I acknowledge this might introduce measurement error in the bus network location for early years, the strong association between predicted times and those observed in the 1995 Mobility Survey suggests this is a fairly good approximation.

[84] For the cadastre, while old properties were partitioned and new ones created, the underlying block structure and "barrios" remained unchanged (up to new ones being added as the city grew). Similarly, existing blocks and census tracts DANE's map were kept in almost all instances unchanged, again up to new blocks being added between 2005 and 1993.

[85] From the commuting data, I observe that the majority of trips taken by TransMilenio do not involve other buses (other than feeders). Therefore I exclude the bus network in the construction of the baseline TransMilenio.

robustness checks that the results are similar if either set of times is used separately. Finally, note that average speeds reflect the net effect of traveling on different road types (for cars), modes (for buses and TransMilenio) as well as wait times incurred at transfers.

I set speeds to match travel times observed in the data for commutes to and from work during rush hours in the Mobility Surveys (departing between 5-8am and 4-6pm). I set walk speeds to 5km/h in all years (Ahlfeldt et. al. 2015). Car speeds were reportedly as high as 27 km/h (Steiner and Vallejo 2010) in early years, while the Department of Mobility reports average speeds along main roads of 24 km/h from 2010-2015. To allow for additional time spent parking and slower speeds during rush hours, I set speeds of 20 km/h, 14 km/h and 10 km/h on primary, secondary and tertiary roads respectively for the pre-period, and 14 km/h, 10 km/h and 8km/h for each type during the post-period. Buses were reported to travel at 10 km/h during rush hour before TransMilenio, with some estimates as low as 5 km/h (ESMAP 2009; Muller 2014). I set bus speeds of 13 km/h and 11 km/h for the pre- and post-period respectively, and set transfer times of 4 minutes to enter or exit the network by foot implying a total of 8 minutes spent waiting on each trip. Finally, most reports cite system speeds of 26.2km/h for trunk service on TransMilenio routes (Cracknell 2003; Transportation Research Board 2003). However, this was for earlier years and reports suggest speeds may have slowed later on. I therefore set speeds of 26 km/h for the pre-period and 20 km/h for the post-period. I set the speed of feeder buses equal to those of regular buses, and again impose a 4 minute transfer time to enter or exit each network.[86]

Figure A.10 explores how these predicted times compare with those observed in the data. I construct observed times for each mode using those reported in the Mobility survey for rush hour trips to and from work, and create an average for each origin-destination UPZ pair. I construct the predicted time for the same trip by taking an area-weighted average of the commute times calculated in Arc between each census tract pair within the UPZ pair. I use 1995 as the pre-period for each mode other than TransMilenio for which I use 2005, and 2015 as the post-period. For each mode, the times are highly correlated with the majority of observations lying close to the 45-degree line.

In the main results, I use the average of the pre- and post-period calibrated commute times from ArcMap. In columns (1)-(3) of Table A.14, I run difference in difference specifications to formally test whether the coefficient from a regression of log observed times on log (average) predicted times changes over time. The difference in slopes in the third row are insignificant for cars and TransMilenio, but is positive for the case of buses. However, inspection of Figure A.10 suggests this is driven by a drop in the intercept for 2015 caused b y movements in the left tail: overall the majority of points lie along the 45-degree line in both years.[87] Finally, the last column examines whether the relationship between predicted and observed times is constant across modes within a year. The insignificant coefficients in rows 4-8 confirm this to be the case.

F.3 Constructing the Instruments

Least Cost Construction Path From Transportation Research Board (2007), I obtain engineering estimates for building BRT on different types of land. Their estimates suggest it costs $4mm to build a mile of BRT by converting a median arterial busway, $25mn to build a new bus lane on vacant land, $50mn to build an elevated lane and

[86] I decided on these times to balance the reported speeds in the literature and matching those in the data. Unfortunately, there was not a simple way to automate the procedure to choose speeds that matched the fit with the data since each creation of a Network dataset in ArcMap must be done manually.

[87] Attempts to shift the intercept by varying the fixed time cost within reasonable bounds had negligible effects on this specification.

$200mn to build a tunnel.[88] The maximum grade of BRT is 10% for short runs (Barr et. al. 2010), so I assume tunnels are built on land steeper than that. I assume that building over developed land costs twice as much as vacant land.[89] I then digitize a land use map of the city in 1980 produced by the United States Defense Mapping Agency (Figure A.11, panel (a)) and clean the image into vacant, arterial road, water and developed land use categories. I infill the medians that can be seen in between a handful of large main roads throughout the city, so that these are also coded as arterial. I then compute the share of each land use category in each 20m by 20m pixel, and use a topographical shapefile to compute the average slope in each pixel. Multiplying the share of each land use type by the prior cost estimates yields a cost to build BRT on each pixel. Panel (b) of Figure A.11 shows the results, with lighter shades representing higher cost.

I read this cost raster into Matlab, and use the Fast Marching Method to compute the least cost routes between portals and the CBD. Panel (c) of Figure A.11 shows the resulting paths. We see that for the majority of cases, the actual lines follow the least cost routes suggesting that conditional on the locations of origin and destinations the costs were a large driver of actual placement. To construct the final input for ArcMap, I create stops every 700m to match the spacing of TransMilenio stations. I add instruments for the Feeder routes by placing a 2km radius disk around each portal connecting the two with 8 "spokes", and create stops every 250m.

Tram System From Morrison (2007), I obtained an image of the city's tram system that was last placed in 1921 and stopped operating in 1951.[90] Since the city was far smaller at that time, I digitize the shapefile and extend the routes to the edge of the city in present day. This might reduce concerns about the direct effects of the tram instrument, since the large portions of it were not built. Panel (d) of Figure A.11 shows the extended lines (as well as the originals). As before, I create stops every 700m and construct the least cost commute times in ArcMap using the same speed of travel as trunk lines.

Instrument Construction These procedures provide counterfactual TransMilenio networks. To construct the pairwise travel times under each instrument, I take the modern street and transit network and then replace the Transmilenio with the networks implied by the two instruments. I then recalculate travel times for each pair over the counterfactual networks.

F.4 Cost-Benefit Calculations

This section presents some of the calculations behind the cost figures in Table A.7. Phase 1 of the system cost $5.85mm per km to build in 2003 dollars.[91] This was financed through local fuel taxes (46%), national government grants (20%), a World Bank loan (6%) and other local funds (28%). Phase 2 was more expensive at $13.29mm per km in 2003 dollars, with funding coming from the national government (66%) and a local fuel surcharge (34%).

[88] These numbers are close to the costs of $8mn per mile in 2003 USD reported by the first phase of TransMilenio (Transportation Research Board 2003).

[89] All figures are in 2004 USD and are per mile of construction. Since I have less guidance over the cost of building on developed land, I experimented with higher values and found the routes were unchanged.

[90] The chief of the Liberal Party was assassinated during an international conference in Bogota in 1948, after which riots led to the destruction of one quarter of the city's trams. Combined with the demand for higher capacity transit, this led to the retiring of the trams and their replacement with buses. While trams operated on rail lines, the buses that followed shared roads with cars.

[91] All figures from Baltes et. al. (2006), except the cost per km for phase 3 which is from https://www.esci-ksp.org/archives/project/bogota-brt-colombia.

The higher costs were due to road widening, increased investment in public space and associated infrastructure improvements. Phase 3 continued the trend costing \$20mm per km in 2014 dollars. Averaging over the 41km of lines in phase 1 and 2 and 21km of lines in phase 3, the average construction cost for the whole 103km network as of phase 3 was \$14.08mm in 2016 dollars.

Operating costs are recovered at the farebox by private operators; the cost to transport a passenger is close to the fare (Transportation Review Board 2003). Using the figure of 565mm rides in 2013 from BRT Data (2017) and the fare of \$0.66 in 2016 dollars yields an operational cost of \$372.97mm per year.

GDP in Bogotá in 2016 from DANE[92] is equal to 221,456 bn 2016 Colombian Pesos, equivalent to \$72.57bn in 2016 dollars.

G Supplementary Empirical Results

G.1 TransMilenio Trip Characteristics

Table A.11 presents some descriptives of trips taken in Bogotá in 2015. Three points are worth emphasizing. First, TransMilenio is an important mode of transit constituting 16% of all trips, exceeding the 13.7% taken by cars but less than the roughly 34% taken by bus and walking. Second, the average TransMilenio trip is 10.5km, far longer than the 6.6km and 6.1km average trips taken by other motorized transport. The fixed costs involved in reaching and entering stations make the benefits of BRT pronounced for longer journeys. Third, when compared to other modes we see that TransMilenio is primarily used for trips to work and business. These constitute around 40% of trips on the system. For private matters or shopping, walking is by far the dominant mode, reflecting that these trips tend to be shorter and closer to home. TransMilenio's outsized role in commuting motivates the focus on its effects on access to jobs emphasized in this paper.

Table A.12 examines how each mode's role in commuting has evolved over time. Panel A shows the changes in each mode's share of commutes to work. TransMilenio's rise has been primarily at the expense of a reduction in bus trips. Panel B shows that TransMilenio is on average 26.7% faster than buses and roughly the same speed as trips taken by cars.[93] TransMilenio speeds have fallen over time as the system has become congested with greater use over time. Changes in aggregate speeds on cars and buses appears not so correlated with TransMilenio ridership: speeds fall significantly between 1995 and 2005 (a period of significant population growth of over 29%) while stabilizing between 2005 and 2015. This highlights the role of external aggregate shocks, such as urbanization lead by the country's civil war, that motivates the more local analysis pursued in this paper. Panel C reports a mild fall in the share of car owners consistent with its decreased role in commuting. However, the persistently higher proportion of car owners vs car commuters reflects the importance of cars for other trip purposes.

[92]Source: https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-departamentales/cuentas-nacionales-departamentales-pib-trimestral-bogota-d-c

[93]Note that these are observed door-to-door speeds rather than system speeds: TransMilenio buses are reported to operate faster than the results in Table A.12 suggest, but queueing at stations and time taken to walk between stations and final destinations decrease average observed speeds. Average speeds are also conflated by the different nature of trips taken across modes (such as TransMilenio being used for longer trips, which are typically faster). Section F.2 compares speeds across modes controlling for trip characteristics and composition, and reports that while the relative performance of TransMilenio is more muted it remains a substantive improvement over existing buses.

G.2 Impact on Other Mode Speeds

BRT may affect equilibrium speeds through impacts on travel mode and route choices, and the number of lanes available for other traffic. In Bogotá, the number of lanes available for other traffic was left unchanged: one might then expect TransMilenio to have reduced congestion faced by cars and other buses. To examine the impact of TransMilenio on car and bus speeds, I run regressions of the form

$$\ln \text{Speed}_{ijkt} = \alpha_{ij} + \beta \text{TM Route}_{ij} \times \text{Post}_t + \gamma_t' X_{ijkt} + \epsilon_{ijt}$$

separately for each mode. Here (i, j) indexes a UPZ origin-destination pair, k indexes an individual, Post_t is a dummy equal to one in 2015 and zero in 1995,[94] and X_{ijkt} is a vector of control variables containing individual and trip characteristics, which are allowed to have time-varying effects on speeds. All specifications include a gender dummy, hour of departure dummies and age quantile dummies, origin and destination locality fixed effects, each interacted with the Post dummy. Certain specifications additionally control for log trip distance interacted with the Post dummy.

The variable TM Route_{ij} captures whether the trip from i to j has been "treated" by TransMilenio. To define this measure, I compute the routes for the least cost commutes between each pair of UPZ origin and destination in ArcGIS separately for cars and buses. I then intersect this route with the TransMilenio network (within a 100m tolerance) to compute the share of a trip that lies along a TransMilenio line. With this in hand, I create two treatment measures. The first is simply the share of a trip that lies along a TransMilenio line. The second is a dummy for whether more than 75% of the trip is adjacent to TransMilenio, allowing for a non-linear effect on speed.

Table A.13 presents the results. Once the composition of trips is properly controlled for (columns 2 and 4, since trips intersecting with TransMilenio are more likely to be longer going from the outskirts to the city center), TransMilenio has no impact on neither car nor bus speeds. Note this only identifies relative changes in speeds: any aggregate effect TransMilenio had on the overall level of speeds would be absorbed into the intercept. Consistent with a small congestion elasticity, Akbar and Duranton (2017) find the elasticity of speed with respect to the number of travelers is only 0.06 during peak hours in Bogotá, while Akbar et. al. (2021) find that only 15% of differences in driving speeds in Indian cities are due to congestion.

G.3 Impact on Housing Supply

Table A.15 provides evidence that TransMilenio had no significant impact housing development. The outcome variable is the growth of total floorspace in a census tract between 2000 and 2018.[95] The specification is otherwise the same as from the baseline specification. Columns 1 and 2 show no significant impact of either CMA term on floorspace supply. Column 3 provides a robustness check regressing floorspace supply on log distance to each phase of the system, confirming the previous results. It does appear more development may be happening around the third phase (the negative coefficient on ln Distance F3), but the effect is insignificant. Column 4 interacts distance to each phase with a dummy for whether a tract is above the median tract distance from the CBD, to test if more development is occurring near TransMilenio at the periphery. This does not appear to the case as all the interactions are insignificant.

[94] Results are similar when intermediate years are included, and are omitted for clarity.

[95] I use the Davis-Haltiwanger growth rate $g_i = (X_{it} - X_{it-1})/(0.5 \times (X_{it} + X_{it-1}))$ which allows me to incorporate tracts with no development in 2000.

Figure A.6 repeats the main event study from Figure 3 with floorspace area as the outcome, and shows no significant effect either before or after on property development. While there is a noisy increase in development in 8 to 4 years before line opening, this is neither significant nor enough to show up in the aggregate numbers in Table A.15.

Overall, there was no significant new development close to TransMilenio stations. Reports suggest that constraints to re-development restricted the supply response (Cervero et. al. 2013), in large part due to no significant change in zoning regulations that remained unchanged over the period.

G.4 Impact on Wages and Sorting

Table A.16 examines the impact of market access on income by place of residence. It runs a difference-in-difference specification similar to (16) to examine the effect of improved RCMA on log average weekly labor income reported by full-time workers between 18 and 55 across UPZs. Since the survey is a sample survey, there are not many observations in each census tract in each period and so the variation in RCMA is aggregated to the UPZ-level. Standard errors are clustered by UPZ and Post-period pair in Panel A, and by UPZ in Panel B.

Column (1) shows a strong association between improved access to jobs and incomes over the period. However, column (2) controls for the changing educational composition of workers and shows that about half of the relationship is explained by re-sorting of workers by skill. The result is qualitatively unchanged when controlling for hours worked in column (3) (i.e. when looking at the wage). While my cross-sectional data do not allow me to control for individual fixed effects, that wages rise even when controlling for changing worker characteristics supports the idea that CMA reflects accessibility to high-paid jobs. The last row also reports the results from a test of whether the coefficient on log RCMA equals $1/\theta$, and in both panels this cannot be rejected.

Table A.17 examines TransMilenio's impact on the educational composition of residents. The outcome is the change in a tract's share of college-educated residents between 2018 and 1993. In 1993 this is measured within all adults 18 or older, and in 2018 this is measured within adults 40 and older. This is to try to look within a cohort, since the overall college share grew substantially over this period. Results are not sensitive to this choice. Column 1 shows a semi-elasticity of 0.05 of the change in the college share to the change in RCMA. Column 2 examines whether this is mechanical: if the change in RCMA is correlated with the initial college share there may be mechanically more or less room for the share to increase in exposed locations. Controlling for the initial college share has little qualitative effect on the coefficient, increasing it slightly. These results suggest the college educated tended to move into neighborhoods with improved accessibility due to TransMilenio. This is consistent both with the results in Table A.16, as well as the sorting channel in the model whereby the rich are more likely to move into neighborhoods with appreciating house prices since they spend a smaller fraction of income on housing.

G.5 Impact of Both Types of CMA

The baseline model predicts no impact of FCMA on residential outcomes and no impact of RCMA on commercial outcomes (see proof of Proposition 1). Table A.18 extends the baseline specification to include both types of CMA separately in the regressions. In general, the results are noisy: conditional on the set of controls, there does not appear to be a huge amount of residual variation in RCMA conditional on FCMA within a locality and vice versa. For five out of seven outcomes (residential population, commercial prices, commercial floorspace share and census employment) the basic prediction that RCMA affects residential outcomes and FCMA affects commercial outcomes holds in the data, although many of these specifications are noisy. For residential floorspace prices the effect is similar for both types of CMA, although the effect is noisy and neither coefficient can be distinguished from zero. For establishment counts in the CCB, the impact is positive only for RCMA.[96]

Even the fact that floorspace shares are observed to change to TransMilenio already suggests some basic assumptions from the simple model are not borne out in the data, since it assumes floorspace use shares are fixed. Appendix E.3 extends the model to include this, and shows that a weighted average of CMA types will now matter for outcomes in each location, where the weights depend on the initial floorspace shares across residential and commercial uses. In fact, this would be the correct regression framework to use to fully test the model given that changes in floorspace use shares are observed in the data. However the log-linear reduced form no longer holds and the constant CMA elasticities are replaced with a more complex matrix of elasticities (where a location's weight on each change in CMA depends on its initial floorspace use share). Given the parsimony of the basic model, I focus on this for the main results. The full model allows for endogenous floorspace use.

G.6 Main Results: Robustness

Table A.1 assesses the robustness of the main results to a number of alternative specifications. First, I use alternative ways to aggregate mode-specific commute times and alternative travel speeds on each mode (columns 2 to 4). Second, I vary the commute elasticity θ to 1.5 and 0.5 times its estimated value (columns 5 and 6). Third, I consider only tracts within 3km a TransMilenio station to ensure the results are not driven by outliers at implausible distances from the network (column 7). Fourth, I use heteroscedasticity robust standard errors and standard errors clustered at the sector level (560 administrative units above the census tract) in columns 8 and 9. Fifth, I exclude tracts within 1km of a portal (compared to the 500m exclusion in Table 2) to further ensure the results are not driven by the targeting of these neighborhoods (column 10). Sixth, I control for distance to a tract's closest TransMilenio station interacted with distance to the CBD (column 11). This assesses whether the CMA effect is simply due to heterogeneity of the distance effect at different distances from the CBD (a possibility given the trends in Figure 1). Reassuringly, the results are robust to this, highlighting how the key source of identifying variation is local changes in RCMA within localities.

Seventh, I run an unweighted regression for the change in establishments which is weighted by the initial share of establishments in a tract in the main results (Table A.5). The unweighted results are significant as controls are added, but become noisy and insignificant in the full specification in column 3 (p-value of 0.15). I use the weighted regressions in the main results for two reasons. First, we might expect noise in the CCB data which is a database of establishments registered with the city's chamber of commerce rather than a census. Weighting by initial shares places more weights on tracts where establishment growth is more precisely estimated. Second, I document sharp positive impacts of CMA on the share of floorspace used for commercial purposes (another measure of the changing allocation of real production activity). Taken together, these suggest employment is indeed responding to TransMilenio.

[96] Digging into exactly why this result is occurring did not lead to clear conclusions. I interpret this as due to the finite sample nature of the data whereby running enough specifications will lead to some unexpected results in a finite dataset.