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

Nick Tsivanidis†
University of California, Berkeley

June, 2019

Abstract

This paper estimates the effects of improving transit infrastructure on city structure and welfare. It derives a new reduced form framework from a class of general equilibrium urban models to examine how they capture Bogotá's response to the construction of the world's largest Bus Rapid Transit system. To quantify the system's distributional impacts, it extends these models to incorporate low- and high-skilled workers with non-homothetic preferences over neighborhoods and transit modes. Relative to valuing benefits based on time savings alone, welfare gains are 20-40% larger and 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, Fernando Alvarez, Costas Arkolakis, Jonathan Dingel, Ben Faber, Rick Hornbeck, Ralph Ossa, Andrés Rodríguez-Clare, Chris Snyder, Nancy Stokey, Daniel Sturm, Felix Tintelnot, Owen Zidar, my discussants Treb Allen, David Atkin and Cecile Gaubert, and seminar participants at Berkeley (ARE/Econ/Haas), Boston College, Brown, Chicago, Chicago Fed, CMU Tepper, Darmouth, Harvard/MIT, IGC Cities, IADB, Johns Hopkins SAIS, LBS, LSE, Minneapolis Fed, NBER SI (ITI/DEV), NBER Public, NYU, Penn State, Stanford (Econ/GSB), St Louis Fed, UMD College Park, Universidad de los Andes, Universidad de Javeriana, Urban Economics Association, USC, Wharton, World Bank and Yale for helpful comments. I also thank 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 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. Any 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 and Department of Economics, University of California at Berkeley. Email: ntsivanidis@berkeley.edu. First version: November 2017.

1 Introduction

How large are the economic gains to improving public transit systems within cities and how are they shared between low- and high-skilled workers? With 2.5 billion people predicted to move into mostly developing country cities by 2050, governments will spend vast sums on mass transit to reduce congestion associated with this rapid urban growth.[1] The reliance of poor, low-skilled individuals on public transit suggests they may benefit the most. While existing approaches focus on the value of travel time saved (VTTS),[2] measuring the benefits of these systems is challenging: individuals' decisions of where to live and work will change as new alternatives become attractive, and land and labor markets will adjust. The lack of detailed intra-city data in less developed countries coinciding with the construction of large transit systems makes the task of evaluating their causal impact even more daunting.

This paper exploits uniquely detailed spatial data before and after the opening of the world's largest Bus Rapid Transit (BRT) system–TransMilenio–in Bogotá, Colombia to make three contributions to our understanding of the impact of urban transit infrastructure on cities. First, it shows that a wide class of quantitative urban models deliver a sufficient statistic–"commuter market access" (CMA)–that summarizes the impact of the entire transit network on equilibrium outcomes in any location. For individuals this reflects access to jobs while for firms it reflects access to workers. These models deliver log-linear reduced form relationships linking population, employment and house prices to CMA which is used to guide the empirical analysis. Second, motivated by the heterogeneous response across worker groups, it develops a richer model where low- and high-skilled workers with non-homothetic preferences sort over commutes and car ownership to quantify the system's distributional effects. Third, it estimates the model and uses it to quantify the welfare gains from TransMilenio. It then compares these with the VTTS approach to isolate the importance of reallocation and general equilibrium effects.

The paper presents three main findings. First, changes in CMA parsimoniously fit the rich patterns of adjustment of population, employment and housing markets to TransMilenio. Second, the system led to large aggregate gains for the city, increasing average welfare by 1.49% and output by 1.09% (net of construction and operating costs) at the most conservative estimates. Reallocation and general equilibrium effects account for between 20-40% of these gains, with the remainder captured by VTTS. This suggests that focussing on VTTS alone–which is precisely the first order welfare effect in an efficient case of the full general equilibrium model–misses a substantial portion of the benefits from new infrastructure. Third, high-skilled workers benefitted slightly more, which is surprising given the reliance of the low-skilled on public transit.

To build intuition, I find the incidence of improving public transit depends not only on who uses it most (favoring the low-skilled), but also on how easily individuals substitute between em-

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

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

ployment and residential locations, whether the system connects workers with employment opportunities, and equilibrium adjustment of housing and labor markets (favoring the high-skilled). This contrasts with the VTTS approach which delivers a large reduction in welfare inequality by focussing on mode choice alone without accounting for the reorganization or general equilibrium effects that are important for large shocks. This suggests improving public transit is a less precise way to target welfare gains for the poor than is implied by time savings alone.

Opened in 2000, TransMilenio is the world's most used BRT system with a daily volume of over 2.2mn trips. The system operates more like a subway than the informal bus system that preceded it: 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] I collect new sources of data covering 2,800 census tracts on residence, employment, commuting patterns, and land markets spanning the system's construction.

The paper uses the variation provided by TransMilenio's construction in four parts. It begins by empirically evaluating its impact on city structure. A large literature estimates treatment effects of transit based on distance to stations. In contrast I show that in a wide class of models that feature a gravity equation for commute flows, the total impact of the transit network on firms and workers in a location is summarized by its CMA. This captures a rich heterogeneity in treatment effects separately for firms and workers, and is easily computed using data on residence and employment. A class of models with log-linear demand for residents and workers across locations deliver reduced forms in which equilibrium outcomes are log-linear functions of CMA. These are isomorphic to a host of alternative assumptions over production technologies, housing supply and worker preferences. The implied regression framework then guides the empirical analysis.

The empirical strategy examines the impact of changes in CMA induced by TransMilenio's construction on the growth of outcomes such as population, employment and house prices. Given the potential endogeneity of route placement, identification relies on predicting TransMilenio's location using (i) a historical tram system built by 1921 and (ii) a least-cost construction route connecting the end points with the central business district (CBD) as was the intent of the government. A threat to identification is that features that make a location cheaper to build BRT, such as proximity to a main road, can have direct effects on outcomes. Relative to distance-based analyses, a key advantage of the CMA approach is that I can control for the distance to these features and use only residual variation in the instrumented change in CMA for identification. To provide additional evidence of causality, I (i) run falsification tests exploiting the timing of station openings, (ii) use residual variation in market access conditional on distance to stations and (iii) leverage changes in CMA to locations further than 1.5km from each tract.

Changes in CMA parsimoniously capture the heterogeneous response of population, employ-

[3]For example, 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).

ment and land markets to TransMilenio: the log-linear predictions predicted by the model are borne out in the data. Increased residential CMA drove growth in commute distances and wages, supporting the intuition that it measures access to jobs. These effects are heterogeneous across workers: while the low-skilled increased their commute distances the most, it was the high-skilled who enjoyed greater increases in wages. The system also caused a re-sorting of workers: the high-skilled moved into high-amenity, expensive neighborhoods in the North while the low-skilled moved into poorer neighborhoods in the South.

The second part of the paper develops a quantitative urban model to understand the implications of improving public transit on worker welfare. The model is motivated by (i) the rich heterogeneous responses across skill groups in the reduced form analysis that the simpler models do not capture and (ii) that prior to TransMilenio low-skilled workers relied on a network of informal buses which were on average 30% slower than cars. The key ingredients are multiple worker skill groups with non-homothetic preferences over residential locations and transit modes. These non-homotheticities mean that rich, high-skilled workers are more likely to live in high amenity neighborhoods and own cars. Individuals work in different locations due in part to differential demand for skills from firms in different industries across the city. Individuals differ in their match-productivity with firms in each location and their preference to live in each neighborhood. Together, these determine the sensitivity of commute flows to commute costs. Differences in residential locations, commuting elasticities and the relative demand for worker skills turn out to be crucial in determining the distributional effects from improving transit.

The third part of the paper structurally estimates the model. The parameters are identified using the same variation as the reduced form analysis. The commuting elasticity is identified from the responsiveness of changes in commute flows to (the instruments for) changes in commute times. Some parameters, such as spillovers in productivities and amenities, are challenging to estimate in cross-sectional data. For example, a location's productivity may be a cause or consequence of the number of workers employed there. Since the supply of workers and residents are functions of CMA, the instruments provide exogenous variation in the number of individuals living and working across the city and permit identification of these key elasticities through a Generalized Method of Moments (GMM) procedure.

The estimates of productivity and amenity spillovers are some of the first from within a developing country city. The agglomeration elasticity is roughly three times the size of median estimates in the US, but close to other studies using experimental approaches. I estimate a substantial elasticity of amenities to the college share of residents, reflecting the endogeneity of neighborhood characteristics like crime prevalent in such settings. The model also performs well in matching a number of non-targeted moments such as income, employment and commute flows by skill group, and the change in residential segregation. Amenities and productivities recovered from the model correlate well with observable proxies like local homicide rates and the slope of land.

The final part of the paper uses the estimated model to quantify the welfare effects of the new infrastructure. The analysis begins by using two first order welfare approximations as bench-

marks. First, an approximation to an arbitrary indirect utility function relates data on commute shares, the change in CMA and the reduced form elasticities to changes in welfare. Second, an application of the envelope theorem to the social planner's problem in an efficient economy yields a welfare elasticity 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 small only the direct effects of time saved matter. These approaches require only (i) cross-sectional data on commute shares by skill and transport mode, (ii) changes in commute times and (iii) estimated elasticities that relate changes in time to changes in welfare and are therefore widely applicable. Both expressions lead to similar increases in average welfare that benefit the low-skilled the most.

It then examines how these welfare approximations compare with the full general equilibrium benefits. The VTTS account for between 20-40% of the total gains depending on the variant of the model used (20% at the baseline). Intuitively, allowing for more margins of adjustment increases the return to new infrastructure.

The sign of the distributional effects reverses once the spatial reorganization and general equilibrium effects ignored by the first order approaches are incorporated. While the low-skilled use public transit the most (which drives the reduction in inequality in the first order approaches that rely on commute share data alone), two factors act against them. First, low-skilled workers have a larger elasticity of commuting decisions to commute costs.[4] In the presence of high commute costs, low-skilled workers are better able to substitute to less costly commutes and thus benefit less when costs fall. Second, low-skilled wages fall in response to the greater shift in labor supply amongst public transit users. In contrast, high-skilled wages are partially shielded from this supply shock when skills are imperfect substitutes in production. The net effect of these three forces is that low- and high-skilled workers benefit about the same (welfare inequality rises by only 0.08% vs a 0.23% fall under VTTS).

These results are robust to a host of different parameter values and modeling assumptions. I incorporate different models of home ownership, alternative timing assumptions and employment of domestic workers. I use the extreme assumptions of either zero or infinite mobility costs between Bogotá and the rest of Colombia to bound the impact on welfare, population, land rents and output.

Lastly, three sets of counterfactuals are run to draw additional insights. First, I compute the effect of constructing alternative TransMilenio networks. The effect of different network segments is heterogeneous: lines serving poor (rich) neighborhoods disproportionately benefit the low- (high)-skilled. The conclusion that the low-skilled benefit less than implied by time savings alone remains generalizable, since existing evidence suggests that the key elasticities that vary

[4] In the model this depends on the heterogeneity of match productivity with firms. For example, a high-skilled IT worker may be more willing to incur a costly commute to an especially well-paid position. A low-skilled cleaner who receives similar wages wherever they work may instead substitute towards other alternatives. While one might expect rich high-skilled workers to be more sensitive to commute costs since their value of time is higher, choices of where to work are made based on *relative* differences in net incomes.

across groups have similar relative magnitudes in other countries. Second, I document large gains to improving service delivery that increases the amenity value of riding TransMilenio. Third, I evaluate the impact of a "Land Value Capture" (LVC) scheme under which development rights to increase building densities near stations are sold by the government to developers. While similar schemes have been used with great success in Asian cities such as Hong Kong and Tokyo, one of the main criticisms of TransMilenio was that the city experienced such a large change in transit without any adjustment of zoning laws to allow housing supply to respond. I find that a well-targeted LVC scheme would have increased the welfare gains from TransMilenio by around 19%, while government revenues cover 8-40% of construction costs depending on the migration response from the rest of Colombia.

This paper contributes to several literatures. Most closely related is the body of work that examines the impact of transportation infrastructure on economic activity. A first strand examines the impact of new transit infrastructure and typically measures changes in population and property prices as a function of distance to the CBD (Baum-Snow 2007; Gonzalez-Navarro and Turner 2016; Baum-Snow et. al. 2017) or distance to stations (Gibbons and Machin 2005; Glaeser et. al. 2008; Billings 2011).[5] However, when spatial units are interlinked spillovers across treatment and control locations confound causal inference from such comparisons. 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 may not be externally valid in another. This paper derives a measure from a class of commuting models that explicitly captures the full direct and indirect effects of changes in the transit network between connected locations, allowing for a causal identification of transit connections that captures heterogeneous responses as a function of city geography.

A second strand of this literature explores the effect of infrastructure between regions on economic development through goods market access (Redding and Sturm 2008; Donaldson forthcoming; Bartelme 2015; Donaldson and Hornbeck 2015; Alder 2019). However these models contain no notion of commuting within cities and are silent on the effects of transit infrastructure. This paper considers a different class of urban commuting models where individuals can live and work in separate locations, and shows that model-derived exact reduced form relationships between outcomes and accessibility measures explain the change in city structure in response to a real world change in transit infrastructure.

This paper also contributes to the growing body of work on quantitative spatial models (Ahlfeldt et. al. 2015; Allen et. al. 2015; Fajgelbaum and Schaal 2017; Monte et. al. 2017; Owens et. al. 2017; Severen 2017; Bryan and Morten 2018; Heblich et. al. 2018; Adao et. al. 2019; Allen and Arkolakis 2019). First, the model features multiple types of workers, firms and transit modes, necessary to assess the distributional impacts of improvements in particular modes of transit. Workers have non-homothetic preferences for residential locations and car ownership. This has important implications for matching the sorting response observed in the data as well as for the estimates of

[5] At the city-level, Duranton and Turner (2012) measure population growth as a function of the stock of roads.

amenity spillovers. Incorporating multiple firm types me to use the model to solve for unobserved wages that vary by census tract and skill group, which provide a better fit to the data and have important equity implications. Second, the paper shows these models share a common measure that summarizes the effect of transit on the supply of residents and workers across locations. A class of these models admits an exact reduced form representation in terms of these accessibility measures that can be used to evaluate the impact of changes in transit infrastructure. Third, it leverages the construction of the world's largest BRT in a validation exercise to show the regression framework delivered by these models performs well in explaining the change in city structure.

Lastly, this paper relates to work in transportation economics that measures the benefits of improved transportation through the value of travel time saved (Train and McFadden 1978; Small and Verhoef 2007). While this is precisely the first order welfare effect in an efficient equilibrium of a quantitative urban model, I show it misses a sizable portion of aggregate welfare gains and delivers opposite implications for distributional consequences. This paper also connects with an extensive literature on agglomeration spillovers, providing intra-city estimates of productivity and amenity spillovers within a developing country city, identified using an expansion in the transit network that separately shifts the supply of labor and residents across the city.[6]

The rest of the paper proceeds as follows. Section 2 discusses the context of Bogotá and TransMilenio as well as the data. Section 3 presents the reduced form framework and its results. Section 4 develops the model while Section 5 estimates it. Section 6 quantifies the impact of TransMilenio, Section 7 evaluates the effects of counterfactual policies and Section 8 concludes.

2 Background and Data

Bogotá is the political and economic center of Colombia, accounting for 16% and 25% of the country's population and GDP respectively. Its population of eight million inhabitants makes it the world's ninth densest, with a stark divide between rich and poor.[7] This section provides background on the city and its transit system.

2.1 Structure of Bogotá

Residence and Employment Bogotá is characterized by a high degree of residential segregation between the rich and poor. Defining high-skill workers as individuals who have completed some post-secondary education, panel (a) in Figure 2 plots the share of college residents within a census tract in 1993.[8] The high-skilled are more likely to live in the North, with the low-skilled workers

[6]See Rosenthal and Strange (2004) for a review. Other papers using potentially exogenous sources of variation in the density of employment include Combes et. al. (2010), Greenstone et. al. (2010), Kline and Moretti (2014), Ahlfeldt et. al. (2015). Those examining how amenities depend on the composition of local residents include Bayer, Ferreira and McMillan (2007), Guerrieri, Hartley and Hurst (2013), Diamond (2016), Giannone (2018).

[7]Colombia is the eleventh most unequal country in the world according to the ranking of Gini coefficients from the World Bank. The income distribution in Bogotá had a slightly higher Gini than the country as a whole in 2014.

[8]Datasets are described in Section 2.3. In this section, population data is from the 1993 census, employment location data uses the 1990 economic census, other employment data is from DANE's GEIH and ECH and commuting data is

living in the city's South and periphery. Panel (b) shows that these poorer neighborhoods have a much higher population density, reflecting the smaller per capita housing consumption.

High- and low-skilled residents work in different industries and neighborhoods. Table 1 shows the share of workers employed in each one-digit industry with post-secondary education. Workers in domestic services, hotels and restaurants, manufacturing and retail are relatively unskilled, while those in real estate, education and financial services tend to be high-skilled. These jobs are located in different parts of the city. Defining high-skill intensive industries as those with college employment shares above the median, Figure 3 shows that while overall employment is concentrated along two bands to the west and north of the city center, high-skill intensive industries are located more towards the North.

Commuting Prior to TransMilenio In 1995 the average trip to work in Bogotá 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%).[9] Despite its importance, public transportation in the city was highly inefficient due in large part to its industrial organization. The government allocated the administration of routes to companies called "afiliadoras" which acted as intermediaries between the government and bus companies. Afiliadoras sold slots to run their routes to bus operators. Since their profits depended only on the number of buses the result was a huge over-supply of vehicles. Low enforcement meant that up to half of the city's bus fleet operated illegally (Cracknell 2003).[10] Disregard of bus stops promoted boarding and alighting along curbs, further reducing traffic flows.

The result was that while the crowding of Bogotá's streets slowed traffic overall, buses were much slower than cars. Table 2 compares speeds between buses and cars in 1995. Column (1) shows that commutes by car were around 35% faster than by bus. This is robust to controlling for differences in trip composition with trip origin-destination fixed effects in column (2). However, columns (3) and (4) show that low-skill Bogotanos were about 29% more likely to use buses than cars. The burden of slow public transit therefore fell disproportionately on the low-skilled.

2.2 TransMilenio: The World's Most Used BRT System

Background 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, its first phase opening a mere 21 months later adding 42 km along Avenida Caracas and Calle 80, two arteries of the city.[11] Phases 2 and 3 added an additional 70km in 2006 and 2011, creating a network

from DANE's mobility surveys.

[9] Bicycles and motos account for the remaining 1% of commutes. For comparison, the average commute in US cities was 21 minutes in 1980 to 26 minutes in 2015.

[10] The Department of Mobility estimated the number to be more than double the amount actually required. A typical practice through which bus companies avoided government controls was duplication of license plates and vehicle documentation.

[11] 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 2005). Two years prior to TransMilenio, Peñalosa implemented

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 patronized system of its kind in the world (Cervero 2013).[12] Its average operational speed of 26.2kmh reported during phase one is on par with that of the New York subway (Cracknell 2003; Johnson 2010), and provided a pronounced improvement on reported bus speeds of 10kmh 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. In contrast to the informal network that preceded it, 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, as well as 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 while the government sets fares close to those offered by existing buses.[13]

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 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). The city conducted a planning study to reconfirm these suggested routes and identify new ones based on (i) current and future demand level and (ii) expected capital costs. The result was a plan that aimed to connect 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.[14]

a "pico y placa" driving restriction which restricted cars to 3 days of peak hour weekday road based on their license plate endings (this was later extended to all day in 2009). While the main change occurred before the period of interest, my controls for locality fixed effects and distance to CBD capture potential trends in the benefit of access to public transport that vary across space induced by the policy. The policy did not have the intended consequence of reduced car use: rising pollutants suggest increased purchases of old vehicles (Lawell et. al. 2016) and I report a mild increase in car ownership over the period in Appendix Table A.11. This matches the experience of Mexico City (Davis 2008).

[12] A map of each system component and their opening date is provided in Appendix Figure A.1. 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.

[13] For example, in 2011 (the only year where fare information is reported in the Mobility Survey), the average bus fare is 1400 COP compared to the 1700 COP fare on TransMilenio. While the fare difference of 21.4% is non-trivial, this does not reflect the free transfers across trunk and feeder lines not offered by the existing bus network.

[14] See Cracknell (2003) for discussion. This was confirmed through inspection of satellite images. Since road widening is not always possible (e.g. Jakarta, Delhi), an interesting extension would be to assume car and bus speeds fell along TransMilenio routes to assess the impacts in these contexts. That certain routes already contained median busways did not mean that there was efficient bus transit available along them (e.g. Avenida Caracas). Within a few years of their opening in 1990 the busways "became anarchic as, for example, (i) buses competed for passengers and this, together with little effective stop regulations, resulted in bus stop congestion and hazardous operating conditions, (ii) buses without a license to operate on Av. Caracas were attracted to the busway seeking passengers" (Cracknell 2003).

Three features make TransMilenio an attractive context for empirical analysis. First, having identified neighborhoods towards the city's periphery to be connected with the center, final routes were chosen to a large extent by the desire to minimize construction cost. Second, lines were placed along wide arterial roads that were cheaper to convert and determined by the the city's historical evolution. I leverage both in constructing instruments for the system's layout. Third, TransMilenio was was rolled out so quickly primarily to complete a portion of the system 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 series variation used in the analysis.

Finally, one central criticism of TransMilenio was its singular focus on improving urban mobility without coordinated changes in land use regulation (Bocajero et. al. 2013). As a result, Appendix Section F.2 shows that housing supply did not respond to the system's construction. An integrated land use and transit policy tailored towards increasing housing densities near stations promotes a more efficient urban structure where many residents can take advantage of improved commuting infrastructure, and sales of development rights can finance construction. In counterfactuals, I assess the impact of TransMilenio had Bogotá pursued a such a policy.

Trip Characteristics Appendix Section E provides additional details on the way in which TransMilenio is used which are summarized here. First, TransMilenio is a quantitatively important mode of transit used for longer trips than other modes. Second, TransMilenio provides an improvement in door-to-door speeds of around 17% over existing buses, but remains around 8.1% slower than cars. Third, the system is more likely to be used for commutes to work rather than leisure trips compared to other modes, motivating the focus on access to jobs in this paper. Fourth, TransMilenio use appears to have come primarily from substitution away from buses. Fifth, conditional on car ownership the rich and poor are equally likely to use TransMilenio, consistent with the similar fares as traditional buses.

Impact on Congestion 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. Appendix Section F.3 shows that there were in fact no significant changes in car and bus speeds along routes most affected by TransMilenio. This could be explained by substitution across modes and routes arbitraging any initial speed differences caused by the BRT, or a small elasticity of driving speeds to vehicles volumes at high levels of rush hour traffic.[15] While incorporating congestion in the model would be an interesting exten-

[15]The former is consistent with the "fundamental law of road congestion": Duranton and Turner (2012) find that vehicle-kilometers travelled (VKT) increase one for one with roadway lane kilometers, and find no evidence that the provision of public transportation affects VKT. Akbar and Duranton (2017) estimate congestion in Bogotá and find the elasticity of speed with respect to the number of travelers is only 0.06 during peak hours, while Akbar et. al. (2017) find that only 15% of differences in driving speeds in Indian cities are due to congestion.

sion, these moments suggest that my abstraction from the effects of TransMilenio on other mode speeds appears a reasonable approximation to reality.[16]

2.3 Data

This section provides an overview of the datasets used in the analysis. Additional details are provided in Appendix Section D.

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 Census of 1993 and 2005. This provides the residential population of each block by education level. College-educated individuals are defined as those with some post-secondary education. In 2015, DANE provides population totals at the UPZ. I combine this with the share of college-educated workers in each UPZ from the GEIH survey in that year (described below) to construct population by skill group. Combined with the census, this provides the growth rate of college and non-college residents in each UPZ between 2005 and 2015. Population by census tract in 2015 is then calculated by inflating the 2005 totals by these growth rates.

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. In 2015 this contains the location, industry and employment of each establishment, but in 2000 employment is not provided. I therefore use variation in establishment counts to proxy for employment in the CCB data, 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 employment may be different from that of total employment. Appendix Figure A.8 shows that the employment and establishment densities in both years of the CCB data are highly correlated with that from the 2005 census. Importantly, coverage is even across rich and poor neighborhoods, suggesting both that the CCB data is fairly representative of overall employment.

[16] In the presence of congestion, constant speeds may also simply reflect an increase in the number of trips taken. Recent work has begun to incorporate congestion in trade models (Fajgelbaum and Schaal 2017; Allen and Arkolakis 2019). Doing so in this context would mean the results likely underestimate the welfare gains: when I simulate the counterfactual of removing TransMilenio from the present day equilibrium, the times on other buses and cars are kept unchanged and the gains are measured as the difference in welfare across equilibria. In a world with congestion and/or endogenous number of trips, moving passengers from TransMilenio to other modes would slow speeds (and/or reduce the number of trips) further increasing the welfare gap. Note the empirical results only speak to relative changes in speeds and are silent on the overall effect of TransMilenio. In the data, aggregate speeds for cars and (non-TransMilenio) buses are uncorrelated with the system's 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 led by the country's civil war, that motivates the more local analysis pursued in this paper.

Housing market data between 2000 and 2012 comes from Bogotá's Cadastre. Its mission is to keep the city's geographical information up to date; all parcels, formal or informal, are included with the result that the dataset covers 98.6% of the city's more than 2 million properties (Ruiz and Vallejo 2015).[17] It reports the use, floorspace and land area, 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 2015).[18] As a result, Appendix Figure A.7 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 member was asked to complete a travel diary for the previous day. 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. For each trip, the data report the departure time, arrival time, purpose of the trip, mode, as well as origin and destination UPZ.

Employment data at the worker level 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 each year. They report individual and household characteristics, as well details on employment such as income, hours worked and industry of occupation across primary and secondary jobs. I was able to access versions of these datasets with the block of each household reported.

Commute times between more than 7.8mm pairs of census tracts by each mode are computed in ArcMap. I obtain the shape of each mode's network by combining spatial datasets provided by the city. To construct the time to traverse each edge of the network, I assign speeds in order to

[17] High coverage was confirmed by overlaying the shapefile for available properties over satellite images of the city. Underlining the importance of property tax revenues, in 2008 they accounted for 19.8% of Bogotá's tax revenues (Uribe Sanchez 2015).

[18] Surveyors are sent out to update the characteristics of each property every couple of years. Since the primary data informative about prices is not necessarily updated each year, I focus on long-differences in my analysis.

match both reported values in the literature as well as the distribution of commute times observed in the Mobility Surveys. Appendix Figure A.9 shows the computed times correlate well with observed door-to-door times from these surveys.

Finally, I measure the distance of tracts to various spatial features provided by the city. I also use a land use map of the city in 1980 provided by the US Defense Mapping Agency and a Tramway map from Morrison (2007).

3 Reduced Form Results: Using Theory to Guide Empirical Work

This section shows that in a wide class of urban models, a single measure—CMA—summarizes the effect of a city's entire transit network on any location. A subset of these models admit a log-linear reduced form representation where endogenous outcomes such as population, employment and floorspace prices can be written as log-linear functions of CMA. I take this regression framework to the data as a validation exercise to examine whether the city responds to the change in infrastructure as predicted by this class of models. The elasticities are also used directly in the first order welfare approximations presented in Section 6.1.

3.1 Commuter Market Access: Using Theory to Measure Treatment Effects of Transit

Model Overview I outline a benchmark quantitative urban model based on Ahlfeldt et. al. (2015) and Allen et. al. (2015). Full details are provided in Appendix Section C.1. The city is comprised of a large number of discrete locations $i \in I$ that differ in their exogenous amenities $\bar{u}_i$, productivities $\bar{A}_i$, residential/commercial housing supplies H_{Ri}, H_{Fi} and the time t_{ij} it takes to commute to any other location. A continuum of workers with unit mass choose where to live and work and have Cobb-Douglas preferences over a freely-traded numeraire good and housing. Indirect utility from living in i and working in j is given by

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

where $\epsilon_{ij}(\omega)$ is an idiosyncratic productivity if worker ω chooses commute (i, j), $d_{ij} = \exp(\kappa t_{ij})$ converts commute times into commute costs and $u_i = \bar{u}_i L_{Ri}^{\mu_U}$ allows for spillovers in residential amenities.[19] Assuming the shocks are drawn iid from a Frechet distribution with shape parameter θ, the supply of residents and labor to each location can be written as

$$L_{Ri} \propto \left(u_i r_{Ri}^{\beta-1}\right)^{\theta} \Phi_{Ri} \tag{1}$$

$$L_{Fj} \propto w_j^{\theta} \Phi_{Fj}. \tag{2}$$

[19] This functional form delivers a semi-log equation for commute flows in commute times which enjoys a long empirical support (e.g. Fortheringham and O'Kelly 1989; McDonald and McMillen 2010). Appendix Figure A.2 shows this fits the data in Bogotá too.

I refer to $\Phi_{Ri} = \sum_j (w_j/d_{ij})^\theta$ as Residential Commuter Market Access (RCMA) since it reflects access to well-paid jobs. I refer to $\Phi_{Fj} = \sum_i (u_i r_{Ri}^{\beta-1}/d_{ij})^\theta$ as Firm Commuter Market Access (FCMA) since it reflects access to workers (i.e. locations with high amenities or cheap housing).

Firms produce using the Cobb-Douglas technology $Y_i = A_i \tilde{L}_{Fi}^{\alpha} H_{Fi}^{1-\alpha}$ where $\tilde{L}_{Fi}$ is effective labor. There are spillovers in productivity which depend on a location's employment $A_i = \bar{A}_i \tilde{L}_{Fi}^{\mu_U}$. Solving firms' profit maximization problem delivers expressions for labor and housing demand that depend on prices (wages and commercial floorspace prices) and location characteristics (productivity and commercial floorspace supply). Equating supply and demand of labor and both types of floorspace pins down prices, while imposing the closed city condition that the population must sum to one pins down the level of aggregate welfare.

Commuter Market Access In this model the transit network only matters for equilibrium outcomes through two variables, RCMA and FCMA. Substituting (1) and (2) into their definitions allows them to be expressed (to scale) as the solution to the following system of equations

$$\Phi_{Ri} = \sum_j d_{ij}^{-\theta} \frac{L_{Fj}}{\Phi_{Fj}} \tag{3}$$

$$\Phi_{Fj} = \sum_i d_{ij}^{-\theta} \frac{L_{Ri}}{\Phi_{Ri}}. \tag{4}$$

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 wage 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 population, particularly so when these other locations lack access to jobs (lowering the wage individuals are willing to work there for). Proposition A.1 shows the solution to this system of equations exists and is unique (to scale), so that the market access measures are easily computed using data on population, employment and commute costs as well as a value for the commuting elasticity.

Figure 1 plots the distribution of changes in commuter access across the city induced by the construction of the first two phases of the system.[20] The system increases access to jobs much more for tracts in the outskirts of the city, which were far from the high-employment densities towards the center. Firms' access to workers rose more in the center, since these locations were best positioned to take advantage of increased labor supply along all spokes of the network.[21]

[20] To compute these CMA terms I use the average θ_g and κ estimates from Section 5. I compute commute times t_{ij} by averaging over t_{ija} according to the share of car owners in the data. The baseline approach computes the indices t_{ija} using those implied by the logit model in Appendix Section G.1. It assumes individuals take the quickest mode of public transit available; Appendix Table A.2 shows robustness to alternative aggregations. The figure plots the change in CMA induced by holding population and employment fixed at their initial level in 1993 and 1990 respectively (from the population and economic census) and changing only commute costs to isolate graphically the change due only to TransMilenio (i.e. the instrument from Section 3.2).

[21] Firm CMA increases toward the center-North due to the high density of (low-skill) workers in the South.

Regression Framework Appendix Section C.1 shows that this benchmark model has the following reduced form representation

$$\Delta \ln Y_{Ri} = \beta_R \Delta \ln \Phi_{Ri} + e_{Ri} \tag{5}$$

$$\Delta \ln Y_{Fi} = \beta_F \Delta \ln \Phi_{Fi} + e_{Fi}. \tag{6}$$

where $\Delta \ln Y_{Ri} = \begin{bmatrix} \Delta \ln L_{Ri} & \Delta \ln r_{Ri} \end{bmatrix}'$ are changes in residential populations and floorspace prices, $\Delta \ln Y_{Fi} = \begin{bmatrix} \Delta \ln \tilde{L}_{Fi} & \Delta \ln r_{Fi} \end{bmatrix}'$ are changes in employment and commercial floorspace prices, and e_{Ri} and e_{Fi} are structural residuals that reflect changing location fundamentals (such as productivities and amenities).[22] The reduced form coefficients β_R and β_F reflect both the direct and indirect effect of improving CMA as it filters through land and labor markets. These regressions form the foundation of the empirical analysis, allowing me to assess whether these relationships predicted by the model explain the city's response to the new transit infrastructure.

Isomorphisms The CMA measures and regression specifications are robust to a host of alternative modeling assumptions. The first part of Proposition A.1 shows that the only structure required to recover the CMA measures is a gravity equation for commute flows. This enjoys wide empirical support and is contained in the vast majority of recent quantitative urban models. The second part shows that for a class of models with log-linear demand for residents and labor, equilibrium population and employment can be written as log-linear functions of CMA. Appendix Section C.4 shows this accommodates iso-elastic housing supply, separate shocks and timing assumptions over workplace and residential choices, alternative production technologies (e.g. Eaton and Kortum 2002, and individual entrepreneurs who sort over where to produce) and worker preferences (such as utility over leisure). The regression framework I take to the data is therefore robust to a host of alternative modeling assumptions.

Relation to Market Access Literature Relative to the market access literature in trade and economic geography, this framework allows individuals to live and work in different locations. This delivers measures of accessibility of residents (firms) to jobs (workers). CMA can also be recovered from observable data using less model structure than is typically used in these literatures. The only structure used to derive the system (3) and (4) is the gravity equation for commuting. In economic geography settings, additional assumptions such as symmetric trade costs, balanced trade and goods market clearing are often made to recover market access measures from the data. In fact, one can show that it is precisely the absence of balanced trade in commuters that delivers separate notions of resident and firm CMA in this paper. This distinction is important given that

[22] In fact, a slight variant of $\ln \Phi_{Ri}$ appears on the right hand side of (6) adjusted to reflect the effective units of labor supplied to locations differ from the number of workers in the presence of idiosyncratic productivity shocks across workers. Formally, the two FCMA terms are log-proportional around the point $d_{ij}^{-\theta}$ and have a 0.98 correlation in the data.

changes in firm and resident CMA capture very different sources of variation.[23]

3.2 Identification

The baseline specification runs (5) and (6) with locality fixed effects and a set of variables to partially control for changes in observables that may be correlated with CMA growth. The elasticities of outcomes to CMA is then identified from variation in CMA within census tracts over time, comparing tracts within a locality with similar observable characteristics which experienced different changes in market access.

There are two key threats to identification. First, changes in CMA contain population and employment in both periods. Since productivity and amenity shocks are in the error term, it will be mechanically correlated with changes in CMA. I therefore instrument for the change in CMA when population and employment are fixed at their initial values in the system (3) and (4). This isolates the variation in CMA due only to changing commute costs through TransMilenio's construction.[24]

Second, CMA growth may be correlated with the error if TransMilenio routes targeted neighborhoods with differential trends in productivities or amenities. For example, the government may have wanted to support growing neighborhoods or to stimulate lagging ones. I therefore construct two instruments for TransMilenio routes, which in turn imply two instruments for the change in CMA.[25] The first instrument 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. To do this, I first digitize a land use map of Bogotá in 1980 to measure the different types of land use on small pixels across the city (e.g. arterial roads, vacant, developed etc). Using engineering estimates for the cost to build BRT on different types of land use, this provides a construction cost raster for the city based on the share of land use in each pixel. This allows me to solve for the least-cost paths connecting portals with the CBD. This will be 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).

[23] While balanced trade seems appropriate in a trade setting, it makes less sense in an urban model where it would require the number of workers in a location to equal the number of residents, which is counterfactual. The amount of additional model structure these papers impose is inversely related to the granularity of the data. Redding and Venables (2004) impose none of these additional restrictions but have much stronger data requirements: they need to observe trade flows between each geographic unit in their data. Bartelme (2015) only requires symmetric trade costs and balanced trade, while Donaldson and Hornbeck (2015) also impose goods market clearing since they only observe population rather than expenditure in each location.

[24] I exclude the location itself when calculating its predicted change in CMA due to the potential correlation between initial residence/employment and unobserved shocks (and exclude tracts within a 1.5km band in robustness checks).

[25] Additional details can be found in Appendix Section D.3. To compute the instruments, I first calculate the commute times had the system been built along each instrument. Plugging these into (3) and (4) 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 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 2015, Alder 2017).

The second instrument exploits the location of a tram system opened in 1884, which was last extended in 1921 and stopped operating in 1951. I extend the 1921 lines to the edge of the city in present day, 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 last well after its construction, which I capture by including historical controls.[26] 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 worry 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. This is much harder to implement in distance-based specifications where there is likely to be little residual variation in distance to one particular least cost road conditional on distance to main roads. To provide further evidence in support of my identification assumption, I check the stability of IV point estimates as controls are added and test that both instruments yield similar coefficients. I also run a host of robustness checks described below.

3.3 Results: Main Outcomes

Main Outcomes Table 3 presents the main results. In all specifications, only tracts further than 500m from a portal and the CBD are included in order to keep a constant sample across specifications.[27] Columns (1) and (2) report the OLS results where the change in CMA is measured using (3) and (4). Basic controls included in the first column are then extended to capture a richer set of land market and demographic characteristics, and historical conditions.[28] On average these have little impact; the attenuation observed in some rows reflects the positive correlation between accessibility improvements and observable characteristics associated with faster outcome growth.

Columns (3) and (4) run the baseline IV specification, which instrument for the total change in CMA holding employment and population fixed at their initial levels. The point estimates

[26] These include 1918 population and distance to main roads in 1933. The extension of the tram lines to the city's edge should also reduce concerns over direct effects on outcome growth, since much of the instrument was never built.

[27] One limitation of my data is that variables do not line up over time periods and each specification may therefore rely on changes over different periods. However, I always use changes in market access constructed between the two waves in question and measure CMA using the values for population and employment in each period. Population regressions using differences from 1993 to 2005 measure changes in market access induced by phase one (opened between 2000 and 2003, with 47% and 73% of the phase opened by 2000 and 2001 respectively). Land market and employment regressions using differences between 2000-2012 and 2000-2015 respectively measure changes in market access induced by phase one and two (opened between 2005-2006). Employment and population regressions are weighted by initial establishment counts and population respectively to increase precision, but in robustness checks I show the results also hold in unweighted regressions. I also restrict the sample to tracts within 3km of stations for main specifications to ensure the results are not being driven by changes in CMA in very distant tracts, but include all tracts in robustness checks.

[28] Appendix Table A.1 adds each control sequentially. Controls are described in full in the note to Table 3 .

tend to fall slightly, reflecting the positive mechanical correlation previously discussed. Columns (5) and (6) instrument for the change in CMA by both holding initial employment and population constant and computing the change in commute times had TransMilenio been built along the least-cost path instrument. For residential outcomes, the point estimates are larger than columns (3) and (4). While this could be (partially) due to measurement error, the difference suggests a negative correlation between TransMilenio placement and growth in unobserved amenities and productivities. This seems plausible, given that the system was built to serve areas of the city that had been growing during the 1990s and may have slowed down during the 2000s as they became congested. Commercial outcomes are more noisy, but the overall pattern is that the IV estimates are slightly higher than the previous estimates. That the estimates are stable as additional controls are added provides additional evidence in support of the exclusion restriction. Finally, columns (7) and (8) use both the tram and LCP instruments. The coefficients remain stable compared to using the LCP instrument alone, and all but one case fail to reject validity of the overidentification restrictions.

Heterogeneous Effects of Transit Figure 4 plots the non-parametric relationship between (residual) growth in outcomes and (residual) changes in CMA. The relationship appears approximately log-linear for each outcome, as predicted by the model. This suggests the model performs well in fitting the heterogeneous effects observed in the data: tracts that experience large improvements in market access report large changes in outcomes.

Robustness The Online Appendix reports a number of additional results which I summarize here. Results are reported in the last table mentioned.

First, I run falsification tests to check that changes in CMA induced by particular lines are not associated with growth in outcomes before they open (Appendix Table A.4). Second, I use less model-dependent measures of resident and firm CMA (Appendix Table A.1). These are commute-time weighted sums of employment and residence respectively, and recall the "market potential" discussed by Harris (1954) and alluded to in the discussion of accessibility in Hansen (1959).[29] Third, I condition on distance to stations to show the effects are driven by changes in accessibility rather than features of stations (e.g. changes in foot traffic, pollution or complementary infrastructure). Fourth, I show the results are robust to measuring changes in market access to distant locations more than 1.5km away, reducing the potential for bias resulting from local unobservables. Both approaches are not possible with a distance-based specification. Fifth, I use alternative speeds to compute the commute times for each mode and alternative methods of aggregating across them (Appendix Table A.2). Sixth, I vary the commute elasticity θ to 1.5 and 0.5 times its estimated value. Seventh, I run unweighted regressions for specifications weighted in the main results. Eighth, I use Conley (1999) HAC standard errors (compared to the baseline estimates which

[29] In particular, I define $RMP_i = \sum_j t_{ij}^{-1} L_{Fj}$ and $FMP_i = \sum_j t_{ij}^{-1} L_{Rj}$ as resident and firm market potential respectively.

cluster by census tract) to allow for arbitrary spatial correlation of errors across tracts within 500m of each other. Ninth, I include all census tracts in the analysis, rather than those within 3km of a station (Appendix Table A.3). That my results are robust across these alternatives provides additional evidence in support of the causal effect of TransMilenio through improvements in CMA.

Comparison with Distance Band-Based Predictions One key benefit of the CMA approach is that by capturing the specific geography of a city and the change in its transit network, estimates from one context are more likely to port to others than those based on distance to stations. Another is the rich heterogeneity in treatment effects shown in Figure 1. Appendix Figure A.10 compares the predictions for residential house price growth in the CMA model with those from a distance-based regression on two dummies for being <750m and 750-1500m from a station (relative to the omitted tracts between 1.5-3km away). The dissimilarity index for the predicted changes is 0.631, with appreciation over- (under-)predicted in the center (outskirts).[30]

3.4 Results: Additional Outcomes

Commute Distance Table 4 examines whether TransMilenio led to changes in commute distances. Column (1) shows that changes in market access caused by TransMilenio were indeed associated with greater probability of using the system in 2015, providing reassurance that the measure captures changes in commuting opportunities. Columns (2)-(4) run difference-in-difference specifications similar to (5) exploring how changes in market access affected commute distances within residential locations (UPZs) between 1995 and 2015. Throughout the OLS and IV specifications, improvements in RCMA led to increases in commute distances, suggesting the system made employment in distant locations more attractive. Finally, column (5) tests for heterogeneous effects across workers and finds the effect on commute distances is mildly greater for low-skill workers. This could reflect either a greater reliance on public transit or a greater sensitivity of commute flows to commute costs, a topic revisited in the next section.

College Share A central question surrounding the effects of public transit is whether it leads to a re-sorting of worker groups. In the US, investments in transit have typically been followed by reductions in the share of rich residents (e.g. Glaeser et. al. 2008) although there is evidence this effect varies across different types of neighborhoods (Heilman 2017). The evidence in developing countries is far sparser.

Table 5 explores how the share of college residents in a census tract responds to changes in RCMA. Column (1) shows that on average there was no significant effect on demographic composition. However, this may mask underlying heterogeneity. Columns (2) to (4) test whether the response differed by tracts according to the college share of the surrounding neighborhood.[31] The

[30] For two variables X_i, Y_i this is defined as $\frac{1}{2}\sum_i |\frac{X_i}{\sum_k X_k} - \frac{Y_i}{\sum_k Y_k}|$. It varies between zero and one, with zero indicating identical distributions across locations.

[31] I measure a tract's surrounding college share using the share of college residents within a 1km disk around each

college share did increase in response to an increase in market access, but only in neighborhoods with an initially high college share. In other words, the high-skilled were only willing to pay for improved transit access in "nicer" neighborhoods and not in low amenity, poor neighborhoods in the South. In contrast, the low-skilled were more likely to move into poorer neighborhoods with a lower initial college share. Overall, this shows that TransMilenio increased residential segregation between the low- and high-skilled.

Wages Table 6 examines the impact of market access on wages by place of residence. It runs a difference-in-difference specification similar to (5) to examine the effect of improved RCMA on log average hourly wages reported by full-time workers between 18 and 55 across UPZs. Column (1) shows a strong association between improved access to jobs and wages 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 using the IVs in columns (3) and (4). Finally, column (5) shows that the effect of RCMA on wages is greater for high-skilled individuals. 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. That the effect is greater for high-skilled workers suggests they benefitted more through this channel.[32]

4 A Quantitative Model of a City with Heterogeneous Skills

The evidence provided in the previous section highlights the challenge of assessing TransMilenio's welfare impact. Improved speeds led to travel time savings, but also drove a large reorganization of the city with residence, employment and land markets adjusting in response. While the low-skilled use public transit the most, they experienced smaller wage gains and may have been "pushed out" and replaced by the high-skilled in high amenity neighborhoods where accessibility improved. The poor spend a greater fraction of their income on housing and may have been hurt more by the associated house price appreciation.[33] The models underpinning the reduced form framework in (5) and (6) lack the heterogeneity across workers and transit modes to connect with these empirical patterns. This section develops a quantitative model rich enough to parse the aggregate and distributional effects of new infrastructure through all of these channels.[34]

tract centroid in 1993 (excluding the tract itself). I then define a high college dummy equal to one for tracts in the top two terciles of its distribution. The results are robust to using own-tract college share (subject to mechanical bias as a lagged dependent variable).

[32]Effects on other labor market outcomes are available upon request. I find a mild fall in hours worked and the probability an individual is employed at a small establishment with less than 5 workers, but the vast majority of these effects are driven by changing skill-composition of residents.

[33]See Appendix Section F.4 for evidence on the Engel curve for housing. Section 6.3 examines how different assumptions over home ownership impact the distributional implications.

[34]Appendix Section C.4.2 shows a simplified version of this model has the exact same reduced form (5) and (6). A concern is whether the results from Section 3 really serve as validation of this model since it abstracts from the

4.1 Setup

Locations $i \in I$ differ in their commute times to every other location, their housing floorspace as well as their amenities and productivities as before.[35] High- and low-skilled 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 willing to pay to own a car have the option to drive. Firms from multiple industries are located across the city and produce using labor and commercial floorspace. Industries differ in their demand for skills: for example, hotels and restaurants demand more low-skilled workers while financial services rely more on the high-skilled. Demand for skill therefore varies across the city based on the productivity of each industry in each location. Landowners choose how to allocate the fixed amount of floorspace across residential or commercial use. In equilibrium, the price of floorspace, the share allocated to each use and wages adjust to clear land and labor markets. This setup differs from recent quantitative urban models (e.g. Ahlfeldt et. al. 2015) by incorporating multiple skill groups of workers, commute modes and industries, where workers have non-homothetic demand for cars and residential amenities.

4.2 Workers

The city is populated by different worker skill 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, and whether or not to own a car denoted by $a \in \{0, 1\}$. Individuals derive utility from consumption of a freely traded numeraire good ($C_i(\omega)$); consumption of residential floorspace ($H_{Ri}(\omega)$); an amenity reflecting the average preference of each group to live in i under car ownership a (u_{iag}); and have a disutility from commuting that reduces their productivity at work ($d_{ija} \geq 1$). Workers are heterogeneous in their match-productivity with firms in each location ($\epsilon_j(\omega)$) and their preference for each residence-car ownership pair ($\nu_{ia}(\omega)$). All land is owned by residents and rents are redistributed lump sum through payment π.[36]

Commute costs differ by car ownership because car owners can choose between commuting by car or public transit (such as walking, bus or TransMilenio), whereas individuals without cars can only choose between public modes.[37] Cars also provide an amenity reflecting improved leisure

heterogeneity I claim to be in the data generating process. Appendix Section H.8 presents a Monte Carlo exercise in which I run the same regressions on data simulated from building TransMilenio in the full model. The non-parametric relationships between outcomes and CMA in the simplified regression model are still log-linear. Since log-linearizing the full model would deliver similar regression equations but with heterogeneous effects (e.g. the change in residential house prices would depend on a weighted average of the RCMA shocks for each skill group and car ownership cell), an interpretation is that the simplified regression is approximating the average effect from the full model.

[35] The choice to keep the total supply of floorspace fixed is motivated by the result that this is mostly unaffected by TransMilenio as documented in the Appendix Section F.2. Section 7 explores the impact of allowing floorspace supply to respond.

[36] Specifically $\pi = \bar{L}^{-1} \sum_i (r_{Ri} H_{Ri} + r_{Fi} H_{Fi})$. This choice ensures that all the gains are accounted for within the model, and avoids inefficiencies introduces by absentee landlords that impact the application of Proposition 2. However, in Section 6.3 I consider alternative home ownership assumptions where (i) all workers rent and (ii) a share of individuals own their home, calibrated to match the home ownership rates of each skill group in the data.

[37] Appendix Section G.1 derives this from a mode choice problem in a third stage where having made their live-work-car ownership decision, individuals decide how to commute (walk, bus, TransMilenio, car). Modes differ by commute

benefits, but come at a fixed cost of ownership $p_a > 0$.

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

$$\max_{C_i(\omega), H_{Ri}(\omega)} u_{iag} C_i(\omega)^{\beta} (H_{Ri}(\omega) - \bar{h})^{1-\beta} \nu_{ia}(\omega)$$
$$\text{subject to } C_i(\omega) + r_{Ri} H_{Ri}(\omega) + p_a a = \frac{w_{jg} \epsilon_j(\omega)}{d_{ija}} + \pi$$

Solving for optimal demands yields the following expression for indirect utility from choice (i, j, a)

$$U_{ijag}(\omega) = u_{iag} \left(\frac{w_{jg} \epsilon_j(\omega)}{d_{ija}} - p_a a - r_{Ri} \bar{h} + \pi \right) r_{Ri}^{\beta-1} \nu_{ia}(\omega) \tag{7}$$

where the iceberg commute cost $d_{ija} = \exp(\kappa t_{ija})$ increases with the time t_{ija} it takes to commute between i and j under car ownership a. The parameter $\kappa > 0$ controls the size of these commute costs.

In contrast to models with homothetic preferences, the fixed expenditures on cars and housing allows me to match the Engel curves I document for car ownership and housing expenditure and drive sorting of workers over car ownership and residential neighborhoods by income.[38] When cars are quicker than public transit, the rich are more willing to pay the fixed cost since their value of time is higher. Similarly, the fixed expenditure on subsistence housing means that the poor spend a greater share of income on housing and are attracted to low amenity neighborhoods where it is cheap.

Timing Workers first choose where to live and whether or not to own a car, and then choose where to work. I solve their problem by backward induction. This simplifies the model's estimation, but Section 6.3 shows the results are qualitatively similar if all choices are made simultaneously.[39]

times and amenity values (e.g. comfort). Individuals have idiosyncratic preferences over modes from a Generalized Extreme Value (GEV) distribution. The result is that average commute times differ by car and non-car owners which is reported here in t_{ij0} and t_{ij1}.

[38] See Appendix Section F.4 and Appendix Figure A.5.

[39] Appendix Section H.4 lays out the model with a joint decision over residence, employment and car ownership. This no longer delivers a log-linear gravity equation for commute flows, complicating the estimation of the commute elasticity θ. Instead, I attempted to estimate θ via an indirect inference procedure in conjunction with the spillover parameters. However, this rendered the estimation procedure computationally infeasible. For each value of θ, I needed to solve for wages pre- and post- to generate commute flows, estimating the same regression on the simulated data as in Table 8, and searching for the θ that minimized the distance to the observed coefficients (the spillover moments were unchanged). It takes about 20 minutes to solve for wages pre- and post-, which needs to happen hundreds of time in one optimization search and then across one hundred bootstrap replications. In contrast, the formulation of the model in the paper allows me to estimate θ directly from commute flow data, solve for wages once, and then minimize the moment conditions for the spillover parameters without re-solving the model.

4.2.1 Employment Decisions

Having chosen where to live i and whether or not to own a car a, individuals draw a vector of match-productivities with firms across the city iid from a Frechet distribution $F(\epsilon_j) = \exp\left(-\tilde{T}_g \epsilon_j^{-\theta_g}\right)$.[40] The parameter θ_g measures the dispersion of productivities for type-g workers, with a higher θ_g corresponding to a smaller dispersion. The scalar $\tilde{T}_g$ controls the overall level of productivities for workers in a particular group.

With these draws in hand, linearity of (7) means that workers choose to work in the location that offers the highest income net of commute costs $\max_j\{w_{jg}\epsilon_j(\omega)/d_{ija}\}$. Properties of the Frechet distribution imply that the probability a worker of type g who has made choice (i, a) decides to work in j is given by

$$\pi_{j|iag} = \frac{(w_{jg}/d_{ija})^{\theta_g}}{\sum_s (w_{sg}/d_{isa})^{\theta_g}} \equiv \frac{(w_{jg}/d_{ija})^{\theta_g}}{\Phi_{Riag}}. \tag{8}$$

Individuals are more likely to commute to a location when it pays a high wage net of commute costs (the numerator) relative to those in all other locations (the denominator). The sensitivity of employment decisions to commute costs is governed by the dispersion of productivity. When workers have similar matches with firms in different locations (high θ_g), choices are more sensitive to commute costs. Differences in productivity heterogeneity across skill groups will important in determining the incidence of commute costs, 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 directly related to the denominator in (8) through

$$\bar{y}_{iag} = T_g \Phi_{Riag}^{1/\theta_g}, \tag{9}$$

where T_g is a transformation of the location parameter of the Frechet distribution.[41] Φ_{Riag} is RCMA that is group- and car ownership-specific in this model. Intuitively, in locations with better access to jobs workers earn higher expected income.

4.2.2 Residential Location and Car Ownership Decisions

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

$$L_{Riag} = \lambda_{U,g}\left(u_{iag}\tilde{y}_{iag} r_{Ri}^{\beta-1}\right)^{\eta_g} \tag{10}$$

[40] The iid assumption can be relaxed to allow for within-person correlation in productivity draws; the model's equations are isomorphic. This Frechet distribution can be microfounded by a process of undirected job search where workers and firms meet according to a poisson process with match-productivity learned after each meeting.

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

where $\tilde{y}_{iag} \equiv \bar{y}_{iag} - p_a a - r_{Ri}\bar{h} + \pi$ is income net of fixed expenditures and $\lambda_{U,g}$ is an equilibrium constant. Workers are attracted to locations with high amenities, high net incomes and low house prices, with an elasticity determined by the dispersion of their idiosyncratic preferences η_g.[42]

4.2.3 Aggregation

Firm Commuter Market Access and Labor Supply Using the commuting probabilities (8), 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

$$L_{Fjg} = w_{jg}^{\theta_g} \Phi_{Fjg} \tag{11}$$
$$\text{where } \Phi_{Fjg} = \sum_{i,a} d_{ija}^{-\theta_g} \frac{L_{Riag}}{\Phi_{Riag}}$$

Labor supply in the model takes a log-linear form that depends on two forces. First, more workers commute to destinations paying higher wages. Second, firms attract workers when they have better access to them through the commuting network, captured through the term Φ_{Fjg}. This is because individuals care about wages net of commute costs. Φ_{Fjg} is FCMA that is group-specific in this model. Total effective labor supply to location is given by $\tilde{L}_{Fjg} = \bar{\epsilon}_{jg} L_{Fjg}$, where $\bar{\epsilon}_{jg}$ is the average productivity of type-g workers who decide to work in j.[43]

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

$$\bar{U}_g = \gamma_{\eta,g} \left[\sum_{i,a} \left(u_{iag} \tilde{y}_{iag} r_{Ri}^{\beta-1} \right)^{\eta_g} \right]^{1/\eta_g} \tag{12}$$

4.3 Firms

Technology There are $s \in \{1, \ldots, S\}$ industries which produce varieties differentiated by location under perfect competition. Output is freely traded, and consumers have CES preferences over each variety with elasticity of substitution $\sigma_D > 1$.[44] Firms produce using a Cobb-Douglas

[42] 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 (and net income, the expression in brackets in (7) must be positive for all workers). This requirement is satisfied in the data once I solve the model in Section 6 for both equal and local home ownership assumptions (it is not when all individuals rent since $\pi = 0$). The model with preference rather than productivity shocks over workplaces in Appendix Section H.4 does not require this condition, but is infeasible to estimate as discussed above.

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

[44] This is the numeraire good introduced in the consumer's problem. Appendix Section C.4 shows this Armington assumption is isomorphic to more realistic setups.

technology over labor and commercial floorspace

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

where 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 they take as given.

Industries differ in the intensity in which they use different types of workers α_{sg}. All else equal, industries such as real estate and financial services require a higher share of high-skill 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 industries 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[45]

$$\tilde{L}_{Fjgs} = \left(\frac{w_{jg}}{\alpha_{sg} W_{js}} \right)^{-\sigma} N_{js} \tag{13}$$
$$H_{Fjs} = (1 - \alpha_s) \frac{X_{js}}{r_{Fj}}. \tag{14}$$

4.4 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 that the supply of residential floorspace $H_{Ri} = \vartheta_i H_i$ equals demand:

$$r_{Ri} = (1 - \beta) \frac{E_i}{H_{Ri} - \beta \bar{h} L_{Ri}} \tag{15}$$

[45] From CES demand $X_{js} = p_{js}^{1-\sigma_D} X$ 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.

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 demanded by firms:

$$r_{Fj} = \frac{\sum_s (1 - \alpha_s) \left(W_{js}^{\alpha_s} r_{Fj}^{1-\alpha_s} / A_{js} \right)^{1-\varsigma} X}{H_{Fj}}. \tag{16}$$

Floorspace Use Allocation Landowners choose the fraction ϑ_i of floorspace allocated to residential use to maximize profits. They receive r_{Ri} per unit of floorspace allocated to residential use, but land use regulations limit the return to each unit allocated to commercial use to $(1 - \tau_i)r_{Fi}$. Landowners allocate floorspace to its most profitable use so that

$$\begin{aligned} \vartheta_i &= 1 \text{ if } r_{Ri} > (1 - \tau_i)r_{Fi} \\ (1 - \tau_i)r_{Fi} &= r_{Ri} \;\; \forall \{i : \vartheta_i \in (0, 1)\} \\ \vartheta_i &= 0 \text{ if } (1 - \tau_i)r_{Fi} > r_{Ri} \end{aligned} \tag{17}$$

4.5 Externalities

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

$$A_{js} = \bar{A}_{js} \left(\tilde{L}_{Fj} / T_j \right)^{\mu_A}, \tag{18}$$

where $\tilde{L}_{Fj} = \sum_s \tilde{L}_{Fjs}$ is the total effective labor supplied to that location and T_j is the total units of land. The strength of agglomeration externalities is governed by the parameter μ_A.[46]

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

$$u_{iag} = \bar{u}_{iag} \left(L_{RiH} / L_{Ri} \right)^{\mu_{U,g}}. \tag{19}$$

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

[46] Given the 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 approach can incorporate such spillovers.

additional force towards residential segregation, since the high-skilled are more willing to pay to live in high-amenity neighborhoods and by doing so increase amenities even more. While sorting could be driven by the subsistence housing requirement alone, I allow the strength of residential externalities $\mu_{U,g}$ to differ across groups and let the data speak to the relative strength of these forces in estimation.[47]

4.6 Equilibrium

I now define general equilibrium in the city.[48]

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 (11) is consistent with demand for labor by firms (13),*

2. **Floorspace Market Clearing** *The market for residential floorspace clears (15) and its price is consistent with residential populations (10), the market for commercial floorspace clears (16) and floorspace shares are consistent with land owner optimality (17),*

3. **Closed City** *Populations add up to the city total, i.e.* $\bar{L}_g = \sum_{i,a} L_{Riag}\ \forall g$.

5 Structural Estimation

This section structurally estimates the model from Section 4. It first describes how the model can be inverted to obtain the unobservable wages, amenities and productivities that rationalize the observed data as an equilibrium of the model. It then outlines the procedure to estimate the model's parameters, and presents the estimation results and model diagnostics.

5.1 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

[47] Apart from improving accessibility, TransMilenio may have had direct impacts on locations productivities $\bar{A}_{js}$ and amenities $\bar{u}_{iag}$ (e.g. through street improvements, effects on crime or pollution). In Section 5 I allow for this possibility but find no economically significant effects, motivating their exclusion from the model.

[48] A previous version of this paper established existence of equilibrium, and uniqueness in a special case of the model without non-homotheticities or heterogeneity across groups, firms and transit modes. In that case, bounds can be provided on the strength of spillovers such that the equilibrium is unique when externalities are sufficiently small.

combine data on residence and employment 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 1. ***(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.*

5.2 Parameter Estimation

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.

5.2.1 Parameters Calibrated to Exogenous Values

The parameters $\{\sigma, \sigma_D, \alpha_s\}$ are calibrated to existing values from the literature. I set the elasticity of substitution between labor skill groups to $\sigma = 1/0.7$ based on the review in Card (2009). I set the cost share of commercial floorspace to the estimates from Greenwood, Hercowitz, and Krusell (1997) who measure the share of labor, structures and equipment in value added for the US to be 70, 13, and 17 respectively. A floorspace share of $1 - \alpha_s = 0.156$ corresponds to their estimates renormalized to exclude equipment which is absent from my model. This is set to be equal across industries. The elasticity of substitution of demand is set to $\sigma_D = 6$ close to median estimates from Feenstra et. al. (2014). I vary both elasticities of substitution in robustness checks.

I now discuss estimation of $\{\beta, \alpha_{sg}, \kappa, \theta_g, T_g\}$ using relationships from the model.

5.2.2 Parameters Estimated without Solving the Model

Share Parameters I estimate $1 - \beta = 0.24$ to match the long-run housing expenditure share in Bogotá.[49] The labor shares α_{sg} are estimated by industry using the average share of the wage bill paid to college and non-college educated workers in Colombia between 2000 and 2014 in all cities other than Bogotá. Assuming that firms outside Bogotá aggregate labor using Cobb-Douglas technology, these labor cost shares identify α_{sg}.

Commute Costs Appendix Section G.1 outlines how commute times for car and non-car owners are constructed using averages of the time on each available mode implied by a discrete choice model. In the third stage of the model, having made their live-work-car ownership decision, individuals decide how to commute (walk, bus, TransMilenio, car) given their idiosyncratic preference for each mode. These are drawn from a GEV distribution allowing for a nested preference structure across public and private nests. There are three sets of parameters to estimate: (i) preference shifters for each mode, (ii) the disutility of commuting κ and (iii) the degree of correlation of draws within the public mode nest λ. I estimate the mode choice model using the 2015 Mobility Survey by Maximum Likelihood. The elasticity κ is therefore identified from the sensitivity of individuals' mode choices to differences in times across modes within particular commutes.

Table 7 reports the results. The estimate of $\kappa = 0.012$ is very close to that of 0.01 reported in Ahlfeldt et. al. (2015). Rows (2)-(4) report the preference shifters for each mode relative to walking. These are identified off differences in choice shares conditional on observed travel times. Intuitively, cars are most attractive followed by buses and TransMilenio. That TransMilenio is least desirable likely reflects high crowds on the system as well as the inconvenience of having to walk between stations and final origins and destinations.

Commute Elasticity Taking logs and first differences of the expression for commute flows (8) combined with the specification of commute costs $d_{ija} = \exp(\kappa t_{ija})$ provides a gravity equation relating the change in commute flows to changes in times

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

where γ_{iag} and δ_{jg} are fixed effects and ε_{ijag} is an unobserved component of commute costs. This equation is estimated at the locality-level using the commuting data from the 1995 and 2015 Mobility Surveys. Given the estimate of κ from the previous step, θ_g is identified off the sensitivity of changes in commute flows to changes in commute times induced by TransMilenio. As before, the change in commute times is instrumented with its value had the BRT been built along the predicted routes to address the endogeneity of its placement.

Table 8 reports the results. Column (1) reports the baseline specification where the change in times is instrumented using both LCP and Tram instruments. Low-skilled workers are more

[49] See the Engel curves presented in Appendix Section F.4.

sensitive to changes in commute times than high-skilled workers. These correspond to $\theta_H = 2.724$ and $\theta_L = 3.299$. The overall magnitude and fact that more educated workers are estimated to have a greater dispersion of match-productivities lines up with existing estimates (e.g. Lee 2015; Hsieh et. al. 2016; Galle et. al. 2017). Column (2) reports the OLS result, which is similar but attenuated. Columns (3) and (4) control for other factors that might affect commute costs other than time (a route's average crime, average house price, and road type). These variables are insignificant determinants of commute flows and the time coefficients are fairly stable after their inclusion.[50] This motivates the focus on commute times as primary determinants of flows in this paper. Columns (5) and (6) follow the trade literature (e.g. Santos Silva and Tenrayo 2006) in estimating the model in a single year via PPML to deal with the zeros in the commute data.[51] This yields somewhat smaller estimates, but with a larger difference across skill groups. I use the estimates from the time-differenced model in column (1) as the baseline values for θ_g, and explore sensitivity of the results to using the values from the PPML model in robustness checks.

5.2.3 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 Section H.1 shows that given prior parameter estimates there is a unique 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. These are solved in the process of recovering the model's unobservables to exactly match these moments in each year of data. I now turn to estimating the residential supply elasticity η_g and spillover parameters $\mu_A, \mu_{U,g}$ by exploiting the fact that changes in market access induced by TransMilenio provide a shock to the supply of labor and residents across the city.

Amenities Moment Taking logs of the expression for residential populations in (10) delivers the following expression for residential population growth across skill groups

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

where $\Delta \ln V_{iag} \equiv \Delta \ln \tilde{y}_{iag} - (1-\beta)\Delta \ln r_{Ri}$ is the change in indirect utility from living in (i, a) net of changes in amenities, γ_ℓ and Controls_i are locality fixed effects and tract characteristics (to partially control for changing fundamentals) and $\Delta \ln \epsilon_{Riag}$ reflects residual variation in unobserved amenity growth. Identification of η_g requires a source of exogenous variation in the common component of utility from living in a location $\Delta \ln V_{iag}$. Identification of the spillovers $\mu_{U,g}$ requires a separate source of exogenous variation in the college share of residents $\Delta \ln L_{RiH}/L_{Ri}$.

The change in indirect utility is instrumented using the instruments for the change in RCMA.

[50] Coefficients for controls not reported, available on request.

[51] The IV-PPML model failed to converge in the two-period specification (columns 1-4): work has shown the Poisson model is not subject to an incidental parameter problem in the case with two fixed effects (e.g. Fernandez-Val and Weidner 2016), but I am not aware of results for the case with three fixed effects that applies in the two-period specification.

This is adjusted for the fixed expenditures and transfers that are not location-specific to improve the fit over the curvature these introduce, denoted $\Delta \ln \tilde{\Phi}^k_{Riag}$ for $k \in \{LCP, Tram\}$.[52] Two additional instruments provide separate variation in the share of college residents living in a tract. First, tracts that experience a greater growth in CMA to high-skill jobs relative to low-skill jobs should experience a larger increase in the share of college residents. This is captured by the instruments $Z^k_{Diff,i} = \Delta \ln \bar{\tilde{\Phi}}_{RiH} - \Delta \ln \bar{\tilde{\Phi}}_{RiL}$ where $\bar{X}_i = \sum_a X_{ia}$. Second, tracts with expensive housing where accessibility improves should see a greater increase in the college share than cheap neighborhoods. This comes directly from log-linearizing the expression for residential populations (10): 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.[53] I capture this by interacting the change for high-skilled residents with the house price in the initial period $Z^k_{Rents,i} = \Delta \ln \bar{\tilde{\Phi}}_{RiH} \times \ln r^{2000}_{Ri}$. The moment conditions used to identify η_g and $\mu_{U,g}$ are therefore[54]

$$E\left[\Delta \ln \epsilon_{Riag} Z_{Riag}\right] = 0, \qquad Z_{Riag} \in \begin{Bmatrix} \Delta \ln \tilde{\Phi}^{LCP}_{Riag} & Z^{LCP}_{Diff,i} & Z^{LCP}_{Rents,i} \\ \Delta \ln \tilde{\Phi}^{Tram}_{Riag} & Z^{Tram}_{Diff,i} & Z^{Tram}_{Rents,i} \end{Bmatrix}.$$

Productivity Moment Recall that firm sales are given by $X_{js} \propto \left(W^{\alpha_s}_{js} r^{1-\alpha_s}_{Fj}\right)^{1-\sigma_D} A^{\sigma_D - 1}_{js}$. Commercial floorspace prices are observed. Wages are recovered from model inversion in Proposition 1 using data on employment, residence and commute costs. These define the labor cost index W_{js}. Lastly, the model implies that firm sales are proportional to the wage bill through $\alpha_s X_{js} = \sum_g w_{jg} \tilde{L}_{Fjgs}$. Since effective labor is obtained using data on employment and model-implied wages, this allows me to recover firm sales X_{js}.

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. 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). Using data before and after TransMilenio's construction provides two values for composite productivities in each location. Taking logs of (18) and including a set of control variables to (partially) capture

[52] In particular, letting $t-1$ and t reference the pre- and post-TM periods respectively, adjusted RCMA is defined as $\tilde{\Phi}_{Riag,t-1} \equiv T_{g,t-1} \Phi_{Riag,t-1}{}^{1/\theta_g} - p_{a,t-1} a + \pi_{t-1}$ and $\tilde{\Phi}_{Riag,t} \equiv T_{g,t} (\Phi^k_{Riag,t})^{1/\theta_g} - p_{a,t} a + \pi_t$. The change is simply $\Delta \ln \tilde{\Phi}_{Riag} = \ln \tilde{\Phi}_{Riag,t} - \ln \tilde{\Phi}_{Riag,t-1}$.

[53] Log-linearizing the expression for residential populations (10) 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 $\mu^L_{iag} \equiv T_g \Phi^{1/\theta_g}_{Ri} / \tilde{y}_{iag}$ and $\mu^R_{iag} = r_{Ri} \bar{h} / \tilde{y}_{iag}$ are the share of labor income and fixed housing expenditure of total net income. Note that μ^R_{iag} is greater for poor individuals in expensive neighborhoods. Thus, poor low-skilled workers are more sensitive to house price appreciation in expensive neighborhoods and are less willing to pay for improved CMA there than the high-skilled.

[54] Orthogonality conditions with each control variable are also included. The baseline specification measures changes in outcomes between 2000 and 2015 and uses the change in transit network due to the first phase of the system since the raw population data at the tract level comes from 2005 (before using the 2015 UPZ totals to inflate to that year).

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. The identification challenge is clear: locations may become more productive because more people work there, or locations whose productivity is growing may attract more workers. Guided by the reduced form results, I exploit the fact that labor supply in the model is a log-linear function of FCMA. TransMilenio therefore provides a shock to labor supply in each location through the commuting network, and my instruments isolate the portion of this variation orthogonal to changes in location fundamentals. The moment condition used to identify μ_A is therefore

$$E\left[\Delta \ln \epsilon_{Fis} Z_{Fig}\right] = 0, \qquad Z_{Fig} \in \begin{Bmatrix} \Delta \ln \bar{\Phi}^{LCP}_{FiL} & \Delta \ln \bar{\Phi}^{LCP}_{FiH} \\ \Delta \ln \bar{\Phi}^{Tram}_{FiL} & \Delta \ln \bar{\Phi}^{Tram}_{FiH} \end{Bmatrix}$$

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.[55]

5.2.4 GMM Results

Table 9 presents the main results. Three comments are in order. First, the estimate of the productivity externality of 0.212 is larger than the Ahlfeldt et. al. (2015) estimate of 0.07 in Berlin, but within the bounds established in experimental approaches in the US with estimates as high as 0.12 and 0.2 (Greenstone, Hornbeck, and Moretti 2010; Kline and Moretti 2014). The returns to agglomeration may be higher in developing countries due to factors such as weak infrastructure or high crime. To my knowledge, this is the first intra-city estimate of agglomeration in a less developed country using quasi-experimental variation. Second, the residential population elasticity is slightly larger for the high-skilled than the low-skilled. The difference is around half the size than the difference between commute elasticities. Third, the spillovers for residential amenities are 0.419 and 0.576 for the low- and high-skilled respectively. The share of college-educated residents in a tract increases the amenities from living there, and the high-skilled value living around each other more than the low-skilled.

Appendix Table 9 checks the robustness of these estimates. First, column (2) assesses whether TransMilenio impacted amenities and productivities directly (e.g. through street improvements, effects on crime or pollution) by controlling for log distance to the closest TransMilenio station (instrumented using the log distance to the instruments). I find no effect on amenities, and an

[55] Bootstrapping is necessary since units of observation vary across moment conditions, rendering the standard asymptotic variance formulas inapplicable.

economically small reduction in productivity.[56] This suggests the primary effect of TransMilenio was indeed through improved accessibility. Columns (3) to (7) vary the elasticity of substitution of demand (from 4 to 9) and elasticity of substitution between skill groups (from 1/0.7 to 2.5).[57] The residential supply elasticity is slightly sensitive to the substitution elasticity of labor, but qualitatively the results are similar. The agglomeration point estimate is mechanically related to the demand elasticity since both affect returns to scale. While my preferred estimate lies in the middle of the observed range, I later examine the robustness of my quantitative results to these parameters. Lastly, column (8) shows that excluding non-homothetic housing demand shrinks the estimates of the amenity spillover by about one quarter with implications for efficiency.

5.3 Non-targeted Moments: Model vs Data

This section evaluates the model's performance by assessing its ability to match moments not targeted in estimation.

Wages Figure 5a compares the average wage earned by residents of each locality with that observed in 2014 in the GEIH data. The latter was not used in estimation. The two variables are highly correlated with values of 0.537 for non-college and 0.601 for college workers. Most observations lie along the 45-degree line for low-skilled workers, but there is noticeable deviation for the richest localities amongst high-skill workers.[58] While the model is unable to capture all factors that drive differences in average income, the high correlation suggests that the spatial forces perform well in explaining income differences across the city.

Amenities and Productivities Amenities and productivities represent characteristics that make locations more or less desirable to individuals and firms. Appendix Table A.6 shows that (i) amenities are higher in neighborhoods with less crime and (ii) productivities are higher in tracts with less crime, a flatter slope and a higher density of roads. Overall, the model performs well at capturing features that affect the desirability of locations.

Commute Flows I use the model to compute commute flows between origin, destination and car ownership pairs according to the gravity equation (8). The model exactly matches total residence, employment and the share of car owners, but information over the bilateral commute shares was not used in estimation. Figure 5b compares these model-implied commute shares with those observed for each locality origin-destination pair in the 2015 Mobility Survey. The model matches commute flows across skill and car ownership groups well. The fit is even across college

[56] In additional results available on request, I show directly there was no significant impact of distance to TransMilenio on changes in crime between 2007 and 2012.

[57] The value of 2.5 comes at the upper end of estimates in Card (2009) for skill groups using regional data in the US.

[58] The model loads all spatial variation in employment unexplained by bilateral commute costs to residential locations into wages. There may be amenities such as safety, access to food and retail establishments, that determine how attractive it is to work in a location. This could explain the relatively poor fit for high-skilled workers if they value these attributes more. So long as these amenities are fixed in the counterfactuals they will not affect the quantitative results.

and non-college workers, showing the method to back out wages by skill group using the location of employment by industries performs well in predicting commute flows.

Employment By Skill Group Figure 6 compares the skill employment ratio $\ln(L_{FiH}/L_{FiL})$ within each UPZ in the model with that implied by trips to work in the 2015 Mobility Survey. To show the importance of the ingredients in the model, panel (a) plots the results from a simplified version in which labor skill groups are perfect substitutes and share the same commute elasticity (set to the average value) as in Ahlfeldt et. al. (2015). In this model, relative employment by skill group has an oddly smooth pattern that slowly declines as one moves further south in the city. This is because low- and high-skilled workers receive the same relative wage in each location and have the same sensitivity to commute costs, so differences in commuting behavior are solely due to differences in residential locations. Thus, the supply of high-skilled workers is much greater in the North close to where they live, and vice versa for the poor who live in the South. This is clearly counterfactual to the distribution in the data shown in panel (c). By contrast, my model performs better in matching this spatial distribution of the employment of relative skills (panel (b)): the correlation between the skill share in the data and in the baseline model of 0.408 is mildly higher than the 0.358 in the simplified model, and qualitatively the pattern appears more realistic.

Heterogeneous Sorting Response Appendix Table A.7 shows the model can match the sorting response to TransMilenio despite not being targeted in estimation. Column (2) shows the model comes remarkably close to matching the differential response of the college share to the system by initial surrounding college composition (interaction term of 0.091 vs 0.095 in column 1). Column (3) shows the model with homothetic demand for residential housing cannot match this moment so well (interaction term of 0.067).

6 The Welfare Effects of TransMilenio

This section quantifies the impact of TransMilenio by simulating its removal to ask what Bogotá would look like in 2012 had the system not been built. It begins with first-order approximations to the welfare change that are easy to implement using readily available travel survey data. It then compares these with the results from the model that capture the global, general equilibrium responses and decomposes the differences.[59]

[59] I refer to the "2012 equilibrium" as the post-TransMilenio equilibrium. Population and employment data come from 2015, land market data come from 2012, and the TransMilenio network is taken to include phases 1 and 2 of the system. There may be multiple equilibria in the presence of spillovers. 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.

6.1 First Order 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 the model without spillovers and no preference heterogeneity, the competitive equilibrium is efficient. The elasticity of welfare to a change in commute costs is*

$$d \ln \bar{U}_g = -\beta\kappa \sum_{i,j,a} \frac{w_{ijga} L_{ijga}}{\sum_{r,s,o} w_{rsgo} L_{rsgo}} dt_{ija}, \tag{21}$$

where $w_{ijga} \equiv T_g \Phi_{Riag}^{1/\theta_g}$ is average labor income of type-g individuals along commute (i,j) using mode a.

A straightforward application of the envelope theorem shows that, in an efficient economy, only the direct effect of new infrastructure matters for welfare. This is proportional to a weighted average of commute time reductions, with weights determined by the value of each commute. The constant of proportionality reflects the relation between commute time and costs (through κ), and that income from transfers is unchanged since there are no house price changes to a first order (through β).

Appendix Section G.2 provides a method to map the reduced form elasticities of Section 3 to welfare using readily available data. I provide a first order approximation to the change in an arbitrary indirect utility function that depends on house prices and income. These in turn depend on RCMA, which allows me to link the change in RCMA to changes in welfare through

$$d \ln \bar{U}_g^{FOA} = \sum_{i,a} \frac{L_{Riag}}{\sum_{r,o} L_{Rrog}} \left(\varepsilon^Y - \beta_{iag} \varepsilon^R\right) d \ln \Phi_{Riag}. \tag{22}$$

Here $d \ln \Phi_{Rigm}$ is a simple approximation to the change in RCMA using initial employment levels and the change in travel times, $\varepsilon^X \equiv \frac{\partial \ln X}{\partial \ln \Phi_{Ri}}$ are the reduced form elasticity estimates above, and β_{iag} is the expenditure share on housing (approximated using each group's citywide average).[60]

Results Table 10 reports the change in average welfare and welfare inequality (defined as $\bar{U}_H / \bar{U}_L$) under the first order approaches and compares them with the full general equilibrium response. I do so by first using the model to simulate what Bogotá would look like in 2012 without TransMilenio. I then compute the welfare change from adding TransMilenio under the different ap-

[60] Specifically, I approximate $d \ln \Phi_{Rigm} \approx \ln \left(\sum_j (d'_{ijm})^{-\theta} L_{Fjg}\right) - \ln \left(\sum_j (d_{ijm})^{-\theta} L_{Fjg}\right)$ which ignores the FCMA term in the denominator and holds employment fixed at its initial level. A similar first order approach is used in Atkin et. al. (2018). Ignoring the indirect effects of changing commute costs on employment and FCMA yields $d \ln \Phi_{Rigm} \approx -\theta\kappa \sum_j \pi_{j|img} dt_{ijm}$. Thus, this approximation is also proportional the average time savings with constant of proportionality $\theta\kappa \left(\varepsilon^Y - \beta_g \varepsilon^R\right)$.

proaches.[61]

The first row shows the CMA-based approximation in (22) delivers an increase in average welfare of 0.94% that reduces inequality by 0.23%. The second row shows a larger increase in welfare of 1.31% using the VTTS approach from Proposition 2 with inequality falling by a similar 0.17%. Both approaches deliver large welfare gains that disproportionately benefit the low-skilled. This reflects their greater reliance on public transit relayed through the commute shares in (21) and (22). Lastly, the third row shows the full general equilibrium response in the model delivers an average welfare gain of 1.63% that *increases* inequality by 0.08%. So while around 80% of the welfare gains are captured by the VTTS approximation, it delivers the opposite implication for inequality with the model implying that the high-skilled benefit mildly more than the low-skilled.

Two main factors may drive the differences in the approaches' implications for equity. First, the approximation holds all discrete choices fixed while the GE model incorporates the substitution elasticities that determine the ability of agents to reorganize spatially. This grows in importance for larger changes in commute costs. Second, equilibrium responses may now be important due to the presence of externalities and the large size of the shock. I now turn to decomposing the sources of these differences.

6.2 Unpacking the General Equilibrium Response

Aggregate Effects Table 11 presents the effect of TransMilenio on GDP, total rents and welfare.[62] Panel A presents the closed city results, in which the population of the city remains constant and utility adjusts in equilibrium. The effects on all outcomes are large, independent of whether spillovers are included: TransMilenio increases city GDP between 1.47%-1.82%, total city rents by 1.77%-1.91% and worker welfare by 1.49-1.60%, the higher number referring to the case with spillovers. Panel B shows the open city results, where population rather than utility adjusts (which is fixed to the reservation level in the wider economy). TransMilenio drives a large population increase of 4.73-5.78%, which in turn delivers a much larger impact on GDP of 5.27-7.70% and rents of 7.42-8.91%. These two extreme cases help bound TransMilenio's aggregate impact.

Figure A.3 plots the changes in employment and population in each tract by each variable's initial level. Panel (a) shows that tracts with the largest employment lose the most when TransMilenio is removed. By enabling productive locations to "import" more workers, the system allowed a reallocation of employment that increased aggregate productivity. Panel (b) shows similar

[61] An alternative would be to simulate adding TransMilenio to the pre-TM equilibrium in which I could use observed commute data from the 1995 travel survey to compute the first order approaches. I use this method instead to maintain consistency with the remaining counterfactuals which ask what Bogotá would look like in 2012 without pre-TM (i.e. using the 2012 unobservables to simulate going back to pre-TM commute times). This is the relevant counterfactual given the other forces changing unobservables over the period.

[62] The exercise asks: what Bogotá would look like today had TransMilenio not been built? I answer this by simulating the effect of removing TransMilenio from the 2012 equilibrium (i.e. reverting to the Pre-TM commute times) while holding unobservables fixed at their 2012 levels. I report the absolute value of the percentage change in each variable under this counterfactual change, i.e. $100 \times (X^{NoTM}/X^{TM} - 1)$ for any variable X. Numbers may therefore differ from the first order approaches in Table 10 which inverts the ordering by using the equilibrium without TransMilenio as the base.

but more muted patterns for residence.

Costs vs Benefits How did the output gains from TransMilenio compare with its costs? Panel A of Table 13 provides a breakdown of the system's costs and benefits (see Appendix Section E for details on cost calculations). Even using the most conservative estimate in column (1), the net present value of the net increase on GDP was about $20bn, or a net increase of 1.09% in the steady-state level of GDP. This suggests the system was a highly profitable investment for the city.

Distributional Effects Why does the model deliver the opposite implication for equity than the first order approach? Table 12 decomposes the welfare gains, starting with a simplified case of the model and slowly adding its ingredients to isolate their impact.

Row (1) considers the model where workers share the same (average) value for η and θ and are perfect substitutes in production. Relative wages are therefore equal across employment locations. Low-skilled workers benefit the most with inequality falling by 0.35%. Row (2) allows worker groups to differ in their commute elasticities set to the estimated values. In the first order approximation results, discrete choices over where to work are fixed. In the full model, individuals can substitute between commutes in response to the removal of TransMilenio with elasticity of substitution θ_g. Since low-skilled workers have a greater commuting elasticity than the high-skilled, they are more able to substitute to less costly commutes when transit is slow and therefore benefit less when it improves. Intuitively, the group with the lower elasticity bears a greater incidence of slow transit infrastructure through this channel. This shifts the gains towards the high-skilled with welfare inequality now falling by only 0.14%. Row (3) incorporates differences in residential choice elasticities, and the result is qualitatively similar.

Finally, row (4) considers the full model with imperfect substitution in production. Relative wages now differ across employment locations, based on demand and supply in each tract's labor market. This has two effects. First, what matters is whether workers are connected to locations where demand for their specific skill (and hence their wage) is 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 (Figures 2 and 3). Residence and employment for the low-skilled is more dispersed, so TransMilenio connected a smaller fraction of these workers with high-wage locations. Second, since skill groups are imperfect substitutes in production, 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. Accounting for these forces means that TransMilenio ultimately increased welfare inequality between the low- and high-skilled by 0.08%.

Taking Stock The welfare gains from reallocation and general equilibrium adjustments are quantitatively important. Appendix Table A.8 shows these represent 20-40% of the total benefits de-

pending on the precise model used, with gains from time savings representing the remainder.[63] In terms of the distributional consequences, the incidence of improving public transit depends not only on how much each group uses it, but also how willing each group is to bear high commute costs to work at a particular location, whether the system connects workers with high-wage locations and the general equilibrium response of wages and house prices. I find these additional channels favor the high-skilled most and welfare inequality is left roughly unchanged as a result. While there is no universal answer to who benefits most from transit infrastructure (which depends on the geography of the city and the transit improvements), comparing the conclusions of these approaches suggests that accounting for spatial reorganization of the city and general equilibrium adjustment of prices implies that investments in public transit are a less precise way to target welfare improvement for the poor than is implied by travel time savings alone.

6.3 Robustness and Model Extensions

Appendix Table A.9 explores the robustness of the quantitative results to alternative parameter values. The effects on output, rents and welfare are qualitatively similar across specifications.[64]

Table 14 explores the results' sensitivity to a number of model extensions. These are outlined in Appendix Section H. First, row (2) considers a model where individuals make a joint decision over each residence-employment-car ownership choice, and where shocks by employment locations affect utility rather than productivity. Since commute costs do not affect time available for work in this model, wages do not fall as much (there is no mechanical increase in labor supply) and welfare gains from the system rise. For the same reason, the effect on output falls by more than two thirds but this difference is eliminated by the increase in labor supply from population growth in the open city model.

Second, I extend the model to allow for employment in domestic services. 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 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 benefit from this increased labor supply which lowers the cost of hiring domestic workers. The model extension incorporates both of these possibilities and row (3) shows they tend to balance out: allowing for employment

[63] Relative to Table 10, Appendix Table A.8 uses expressions for the exact change in commute costs through a DEK-like expression shown in Appendix Section H.3. The first order results approximate these with log changes introducing approximation error. The table also highlights which features shape the importance of general equilibrium effects. E.g. column (4) shows the fraction of gains accounted for by these effects is only 20% in the model with imperfect substitutes versus 40% under perfect substitution. Since firms need access to both types of workers, for any configuration which affects the supply of one group more than the other firms will benefit less under imperfect substitution.

[64] These include (i) increasing θ and η by one third (in case my estimates reflect medium- rather than long-run responses), (ii) using alternative values of θ estimated via IV-PPML in Table 8, (iii) setting spillovers to one third of their estimated values (to match the magnitude of productivity spillovers in Ahlfeldt et. al. 2015), (iv) using a larger elasticity of substitution across labor skill groups $\sigma_L = 2.5$, (v) measuring the distribution of employment using the 2005 census rather than the 2015 CCB (to address whether missing informal establishments impacts the results) and (iv) using alternative values of the elasticity of demand $\sigma = 4, 9$.

in domestic services has little effect on inequality.

Third, I incorporate different assumptions over home ownership.[65] I consider both a model where all individuals are renters, as well as one where workers own a share of the local residential housing stock to match the home ownership rates in the data (0.46 and 0.60 for the low- and high-skilled respectively). The results remain fairly invariant across these alternatives, showing that the particular assumption over home ownership does not qualitatively affect the results.

7 Policy Counterfactuals

This section presents results from Table 15 that use the model to draw further insights from a number of policy-relevant counterfactuals.

Impact of Different Lines The first panel of Table 15 analyzes the impact of alternative network configurations. Rows (1) and (2) simulate the effect of removing lines H and A that connect the South and North of the city with the CBD. Both lines delivered large welfare gains with welfare 0.68% and 0.54% higher due the lines to the South and North respectively. However, the benefits accrue disproportionately to the group who live in neighborhoods reached by each line: for example, the high-skilled benefit most from the line to the North (inequality would be lower without it) which the low-skilled benefit most from the line to the South. This underlines that who benefits from new commuting infrastructure depends on the geography of the city and transit network in question.

Row (3) shows that the feeder system, which connects outlying areas with portals using buses that run on existing roadways, increases welfare more than any other line of the network. This underscores the large benefits to providing cheap, complementary services that reach residents in outlying but dense residential areas, thereby reducing the last-mile problem of traveling between stations and final destinations.

Improving Service Quality A frequent complaint of TransMilenio is that the system is congested, leading to queues for buses and cramped conditions on board. How large are the potential gains to improving service delivery? Recalling from Table 7 that the amenity for traveling on TransMilenio was the lowest of all modes, row (4) sets the amenities to be constant across all transit modes. Row (5) allows individuals to prefer cars more than public transit, but assumes that individuals take the quickest mode of public transit (conditional on choosing within that nest). The welfare gains are large from either improvement, suggesting sizable (gross) gains to improving TransMilenio service.

[65] The welfare effect in these two models cannot be properly compared with the baseline model due to accounting differences. In the baseline model, gains reflected through floorspace price appreciation are accounted for in the welfare calculation since workers own the housing stock. The initial welfare level is higher than the model with absentee landlords, so the same percentage welfare change lead to more total gains in the baseline model. One would need to separately account for gains to absentee landlords in these alternative model to make them comparable.

Land Value Capture One main criticism of TransMilenio was that it was not accompanied by an adjustment of zoning laws to allow housing supply to respond where it was needed. Appendix Section F.2 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 (see Hong et. al. 2015 for a review). These policies achieve the dual aim of increasing housing supply and raising revenue to finance the infrastructure's construction.

I evaluate the impact of TransMilenio if housing supply had responded to the opening of the system. In the most extreme case, I assume housing supply freely adjusts. This provides a useful upper bound on the welfare gains from facilitating housing supply response. 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 Asian, European and American cities.[66] 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.[67] Details on these model extensions are provided in Appendix Section H.7. I compare the two equilibria from first removing TransMilenio (without housing adjustment) and then adding it back under each housing supply model.

The third panel of Table 15 presents the results. Under free adjustment in row (6), welfare would have been 0.30% higher that it is today (or, the gains would have been around 19% higher). Under the LVC schemes in rows (7) and (8), welfare would have been 0.08% and 0.29% higher for under the CMA and distance-band policies respectively (5% and 18% larger gains). While this exercise does not factor in potential adjustment costs associated with demolitions and displacement of prior residents, it highlights the large potential benefits left on the table and suggests well-targeted zoning adjustments that allocate permits towards where they are most needed deliver bigger benefits. Lastly, panel B in Table 13 shows the fiscal benefits of LVC schemes. The distance-based instrument recoups between 4-18% of construction costs, while the CMA-based scheme covers around twice as much a (8-41% of 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

[66] These schemes have a number of benefits over property taxes. They are likely to incur less opposition from stakeholders, are less distortionary, are more likely to work in settings with weak property tax systems, and provide additional benefits such as new residential and commercial units. See Hong et. al. (2015) and Salon (2014) for further details. My choice of parameters for this policy is motivated by the example of Nanchang, China, where floor area ratios were increased by a uniform amount within 500m of stations. While the precise increase is hard to find, revenues from the scheme covered 20.5% of costs. In examples covered in Salon (2014) between 14-88% of costs are covered. The 30% increase in permitted densities I choose therefore results in similar revenues. Of course, the revenue raised varies across alternative candidate policies.

[67] In particular, I let the change in permitted FAR be proportional to $\vartheta_i \Delta \ln \Phi_{Ri} + (1 - \vartheta_i) \Delta \ln \Phi_{Fi}$ where ϑ_i are the residential floorspace shares in the initial equilibrium and $\Delta \ln \Phi$ are the instruments for the change in CMA holding population and employment at their initial values. Each of these values is based only on information the government has at the time of the policy change.

transit, and targeting zoning adjustment based on where demand for housing will increase the most delivers the largest benefits.

8 Conclusion

This paper makes three contributions to our understanding of the aggregate and distributional effects of urban transit systems. First, it shows a wide class of models delivers a reduced form framework to evaluate the effects of transit based on "commuter market access". It takes this to the data in the context of the construction of the world's largest BRT system in Bogotá, Colombia. Second, it develops a quantitative urban model in which low- and high-skill workers with non-homothetic preferences sort over where to live, where to work, and whether or not to own a car. Third, it estimates the model and uses it to quantify the welfare gains and how they compare with the traditional travel time savings approach.

It finds the reduced form representation fits the heterogeneous adjustment of population, employment and housing markets to TransMilenio. The system led to large welfare and GDP gains that were more than worth its costs. Between 20-40% of the benefits accumulated through reallocation and general equilibrium effects. This suggests focusing on time savings alone misses a significant portion of transit's impact. The gains in the full model accumulate disproportionately to the high-skilled. The opposite distributional impact is predicted by the time savings approach that ignores spatial reorganization of the city and general equilibrium adjustment of prices that occur from large changes in infrastructure. These results therefore imply that investments in public transit are a less precise way to target welfare improvements for the poor than is implied by the typical framework. The paper also finds the welfare gains would have been around one fifth larger had the government implemented a more accommodative zoning policy, underscoring 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.

AHLFELDT, G. (2011), "If Alonso was Right: Modeling Accessibility and Explaining the Residential Land Gradient", *Journal of Regional Science*, 51(2): 318-338.

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.

AHLFELDT, G.M., REDDING, S.J., AND D.M STURM (2016), "A Quantitative Framework for Evaluating the Impact of Urban Transport Improvements," Working Paper.

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 (2017), "Accessibility and Mobility in Urban India", Working Paper.

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

ALLEN, T., AND C. ARKOLAKIS (2019), "The Welfare Effects of Transportation Infrastructure Improvements," Working Paper.

ALLEN, T., ARKOLAKIS, C. AND X. LI (2015), "Optimal City Structure," Working Paper.

ALLEN, T., ARKOLAKIS, C. AND Y. TAKAHASHI (2017), "Universal Gravity," Working Paper.

ALONSO, W. (1964), *Location and Land Use*. Harvard, Cambridge MA.

ATKIN, D., FABER, B., AND GONZALEZ-NAVARRO, M. (2018), "Retail globalization and household welfare: Evidence from Mexico", *Journal of Political Economy*, 161(1): 1-73.

ANAS, A., ARNOTT, R., AND K. SMALL (1998), "Urban Spatial Structure", *Journal of Economic Literature*, 36(3): 1426–64.

BARTELME, D. (2015), "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., 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.

BOCAJERO, 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.

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

BRYAN, G., AND M. MORTEN (2018): "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," Journal of Political Economy, forthcoming.

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

CERVERO, R. (2013), "Bus Rapid Transit (BRT): An Efficient and Competitive Mode of Public Transport", *Berkeley Institute of Urban and Regional Development Working Paper* 2013-01.

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.

CICCONE, A., AND R. HALL (1996), "Productivity and the Density of Economic Activity," *American Economic Review*, 86(1), 54–70.

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.

COUTURE, V. (2016), "Valuing the Consumption Benefits of Urban Density", Working Paper.

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

DAVIS, L. (2008), "The Effect of Driving Restrictions on Air Quality in Mexico City", *Journal of Political Economy*, 116(1):38-81.

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. (Forthcoming) "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, forthcoming.

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

DONALDSON, D., AND R. HORNBECK (2015), "Railroads and American Economic Growth: A Market Access Approach", *Quarterly Journal of Economics*, 131(2): 799-858.

DURANTON, G., AND M. TURNER (2012), "Urban Growth and Transportation", *Review of Economic Studies*, 79: 1407-1440.

DURANTON, G., AND M. TURNER (2017), "Urban form and driving: Evidence from US cities", Working Paper.

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

FAJGELBAUM, P. D., AND E. SCHAAL (2017), "Optimal Transport Networks in Spatial Equilibrium," NBER Working Paper, 23200.

FEENSTRA, R., P. LUCK, M. OBSTFELD AND K. RUSS (2014), "In Search of the Armington Elasticity," NBER Working Paper No. 20063.

FERNANDEZ-VAL, I. AND M. WEIDNER (2016), "Individual and Time Effects in Nonlinear Panel Data Models with Large N, T," *Journal of Econometrics*, 196: 291-312

FORTHERINGHAM, S., AND M. O'KELLY (1989), *Spatial Interaction Models: Formulations and Applications*, Kluwer, Dordrecht.

GADUH, A., GRACNER, T., AND A ROTHENBERG (2018) "Improving Mobility in Developing Country Cities: Evaluating Bus Rapid Transit and Other Policies in Jakarta", Working Paper.

GALLE, S., RODRIGUEZ-CLARE, A., AND M. YI (2017), "Slicing the Pie: Quantifying the Aggregate and Distributional Consequences of Trade", Working Paper.

GIANNONE, E. (2018), "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 (2016), "Subways and urban growth: Evidence from Earth", Working Paper.

GUERRIERI, G., HARTLEY, D., AND E. HURST (2013), "Endogenous Gentrification and Housing Price

Dynamics" *Journal of Public Economics*, 100:45-60.

GRAFTIEAUX, P., AND D. HIDALGO (2005), "A Critical Look at Major Bus Improvements in Latin America and Asia: Case Study TransMilenio, Bogotá, Colombia", Working Paper.

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.

GREENWOOD, J., HERCOWITZ, Z., AND P. KRUSELL (1997): "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, 87(3), 342-362.

HANDY, S., AND D. NIEMEIER (1997), "Measuring Accessibility: An Exploration of Issues and Alternatives" *Environment and Planning A*, 29(7): 1175— 94.

HANSEN, W. (1959), "How Accessibility Shapes Land Use," *Journal of the American Institute of Planners*, 25: 73–76.

HARRIS, C. (1954), "The Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 44(4), 315–348.

HEILMANN, K. (2016), "Transit Access and Neighborhood Segregation:A Study of the Dallas Light Rail System", Working Paper.

HONG, Y-H., MURAKAMI, J., SUZUKI, H., AND B. TAMAYOSE (2015), *Financing Transit-Oriented Development with Land Values*, Washinton, DC: The World Bank.

HEBLICH, S., REDDING, S., AND D. STURM (2018): "The Making of the Modern Metropolis: Evidence from London," Working Paper.

HSIEH, C-T., HURST, E., JONES, C., AND P. KLENOW (2016), "The Allocation of Talent and US Economic Growth", Working Paper.

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., (2016), "The Effects of License Plate-Based Driving Restrictions on Air Quality: Theory and Empirical Evidence", Working Paper.

LEE, E. (2015), "Trade, Inequality, and the Endogenous Sorting of Heterogeneous Workers", Working Paper.

LLERAS, G. (2003), "Bus Rapid Transit: Impacts on Travel Behavior in Bogota", MIT Masters Thesis.

LUCAS , R., AND E. ROSSI-HANSBERG (2002), "On the Internal Structure of Cities," *Econometrica*, 70(4), 1445–1476.

MACKIE, P., J. NELLTHORP, AND J. LAIRD (2005), "Notes on the Economic Evaluation of Transport Projects", World Bank Transport Note No. TRN-15.

MAJID, H., A. MALIK, AND K. VYBORNY (2018): "Infrastructure Investments, Public Transport Use and Sustain- ability: Evidence from Lahore, Pakistan," Working Paper.

MCARTHUR, D., L. OSLAND, AND I. THORSEN (2012), "Spatial Transferability of Hedonic House Price Functions", *Regional Studies*, 46(5): 597-610.

MCDONALD, J., AND D. MCMILLEN (2010), *Urban Economics and Real Estate: Theory and Policy*, John Wiley & Sons, Hoboken, NJ.

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.

MILLS, E. (1967), "An Aggregative Model of Resource Allocation in a Metropolitan Centre," American Economic Review, 57(2), 197–210.

MONTE, F., REDDING, S., AND E. ROSSI-HANSBERG (2017), "Commuting, Migration and Local Employment Elasticities", NBER Working Paper 21706.

MUTH, R. (1969), *Cities and Housing*, Chicago: University of Chicago Press.

OWENS, R., E. ROSSI-HANSBERG, AND P-D. SARTE (2017), "Rethinking Detroit", Working Paper.

REDDING, S. (2010), "The Empirics of New Economic Geography", *Journal of Regional Science,* 50(1), 297-311.

REDDING, S., AND A. VENABLES (2004), "Economic Geography and International Inequality," *Journal of International Economics*, 62, 53–82.

REDDING, S., AND D. STURM (2008), "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98(5), 1766–1797.

REDDING, S., AND D. STURM (2016), "Estimating Neighborhood Effects: Evidence from War-time Destruction in London", Working Paper.

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.

RUIZ, F., AND G. VALLEJO (2015), "Using Land Registration as a Tool to Generate Municipal Revenue: Lessons from Bogota", Cadastre of Bogota Working Paper.

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

SANTOS SILVA, J.M, AND S. TENRAYO (2006), "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

SEVEREN, C. (2016), "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification", Working Paper.

SMALL, K., AND E. VERHOEF (2007), *The Economics of Urban Transportation*, New York: Routledge.

URIBE SANCHEZ, M.C. (2015), "Land Information Updating, a De Facto Tax Reform: Bringing up to Date the Cadastral Database of Bogota", Cadastre of Bogota Working Paper.

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

Tables

Table 1: College-Employment Shares by Industry

Industry	College Share	Employment Share
Domestic Services	0.085	0.050
Construction	0.181	0.052
Hotels & Restaurants	0.235	0.057
Wholesale, Retail, Repair	0.300	0.222
Manufacturing	0.315	0.173
Transport, Storage, Communications	0.341	0.089
Other Community, Social, Personal Serv	0.380	0.050
Real Estate	0.556	0.120
Social & Health Services	0.634	0.053
Public Administration	0.707	0.038
Education	0.810	0.052
Financial Services	0.827	0.028

Note: Data is an average over 2000-2014 and comes from the GEIH and ECH. The first column shows the share of workers which have post-secondary education within each one-digit industry. The second column shows the industry's share of total city employment. Only industries accounting for at least 1% of employment reported.

Table 2: Commuting in 1995

	lnSpeed	lnSpeed	Bus	Bus
Bus	-0.353***	-0.305***		
	(0.021)	(0.016)		
Low-Skill			0.287***	0.163***
			(0.010)	(0.011)
R^2	0.06	0.76	0.18	0.47
N	14,841	12,877	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. Data is from 1995. 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 3: IV Results: Main Outcomes

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV-LCP	(6) IV-LCP	(7) IV All	(8) IV All
Panel A: Residents								
ln(Res Floorspace Price)	0.386***	0.257***	0.135*	0.228***	0.350***	0.405***	0.333***	0.401***
	(0.067)	(0.055)	(0.073)	(0.060)	(0.119)	(0.093)	(0.125)	(0.096)
N	1,943	1,943	1,943	1,943	1,943	1,943	1,943	1,943
F-Stat					685.05	690.36	329.64	337.54
Over-ID p-value							0.88	0.09
ln(Residential Pop)	0.213**	0.222**	0.140	0.144	0.240*	0.258*	0.224*	0.243*
	(0.102)	(0.104)	(0.103)	(0.104)	(0.133)	(0.138)	(0.135)	(0.139)
N	1,997	1,997	1,997	1,997	1,997	1,997	1,997	1,997
F-Stat					1,608.38	1,562.65	788.22	769.49
Over-ID p-value							0.84	0.95
Panel B: Firms								
ln(Comm Floorspace Price)	0.211**	0.211**	0.222**	0.238**	0.216	0.234	0.250*	0.245*
	(0.097)	(0.102)	(0.100)	(0.105)	(0.135)	(0.143)	(0.134)	(0.142)
N	1,884	1,884	1,884	1,884	1,884	1,884	1,884	1,884
F-Stat					881.67	721.25	656.04	535.07
Over-ID p-value							0.66	0.97
Comm Floorspace Share	0.146***	0.143***	0.144***	0.145***	0.107**	0.097*	0.118**	0.109**
	(0.034)	(0.036)	(0.036)	(0.038)	(0.049)	(0.055)	(0.051)	(0.055)
N	1,981	1,981	1,981	1,981	1,981	1,981	1,981	1,981
F-Stat					939.03	757.93	693.88	555.68
Over-ID p-value							0.35	0.45
ln(Establishments)	0.939***	0.552*	0.819***	0.523*	1.187**	1.150**	0.964*	0.742
	(0.295)	(0.284)	(0.306)	(0.293)	(0.521)	(0.529)	(0.503)	(0.505)
N	1,724	1,724	1,724	1,724	1,724	1,724	1,724	1,724
F-Stat					157.03	178.69	252.31	208.62
Over-ID p-value							0.66	0.04
Locality Fixed Effects	X	X	X	X	X	X	X	X
CBD X Region Controls	X	X	X	X	X	X	X	X
Basic Tract Controls	X	X	X	X	X	X	X	X
Historical Controls		X		X		X		X
Init. Land Controls		X		X		X		X
Init. Demographic Controls		X		X		X		X
Distance to Tram Controls							X	X

Note: Observation is a census tract. Each entry reports the coefficient from a regression of the variable in each row on firm or residential commuter market access in first differences. Each column corresponds to a specification. Land market regressions use changes in outcomes between 2000 and 2012, measuring the change in CMA induced by phases 1 and 2 of TransMilenio. Establishment regressions combine changes between 2000 and 2015 with the same CMA variation induced by phases 1 and 2, and are weighted by number of establishments in 2000. Population regressions use changes between 1993 and 2005 measuring the change in CMA induced by phase 1 and are weighted by 1993 population. Only tracts further than 500m from a portal and the CBD (and less than 3km from a station) are included. 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 are log area and log distance to main road. Historical controls are dummies for quartile of 1918 population and a dummy for whether a tract is closer than 500m to main road in 1933. Initial land controls are the share of land developed, share of floorspace that is commercial, floor area ratio and log value of floorspace per square meter in 2000. Initial demographic controls are log population density and college share in 1993. Land market and demographic controls that represent initial values of outcome variable are excluded in each regression. Distance to tram is a dummy for whether a tract is closer than 500m from the historical tram line. Columns (1) and (2) run an OLS specification. Columns (3) and (4) instrument for the change in CMA holding residence and employment fixed at their initial levels and changing only commute costs, excluding the census tract itself from the variable construction. Kleinberg-Paap F-statistics are high and not reported for brevity. Columns (5) and (6) instrument using the change in CMA induced by the LCP route, while (7) and (8) include both the LCP and tram instrument. Standard errors cluestered by census tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4: Commute Distance

Outcome	(1) OLS UseTM	(2) OLS lnDist	(3) IV lnDist	(4) IV-LCP lnDist	(5) IV All lnDist	(6) IV All lnDist
ln(RCMA)	0.834***	0.459**	0.317	0.782**	0.827**	0.241
	(0.188)	(0.210)	(0.251)	(0.345)	(0.344)	(0.399)
ln(RCMA) X High Skill						-0.125**
						(0.049)
N	9,088	22,119	22,119	17,212	17,212	19,920
R^2	0.07	0.10				
F-Stat				99.39	70.56	17.27
Over-ID p-value				.	0.52	0.47
UPZ FE	X	X	X	X	X	X
Locality FE X Post FE	X	X	X	X	X	X
Log Dist CBD X Region FE X Post FE	X	X	X	X	X	X
Trip Controls X Post FE	X	X	X	X	X	X
Tract Controls X Post FE	X	X	X	X	X	X
Historical Controls X Post FE	X	X	X	X	X	X
Educ X Post FE						X

Note: Observation is a trip, only trips to work for working age adults (18-65) included. Column (1) reports coefficients from a regression of a dummy for whether an individual uses TransMilenio in 2015 on the change in lnRCMA in the origin UPZ. The other columns run difference-in-difference specifications using data from 2015 (Post) and 1995 (Pre), examining how changes in commute distances vary with changes in RCMA. RCMA is measured at the UPZ level using the pre-TM network in the pre-period and the 2006 network in the post period. IV specifications instrument for CMA using each instrument in the post-period. Trip controls include hour of departure dummies and demographic characistics (sex, log age, hh head dummy, occupation dummies). Tract controls include log area, log distance to a main road and log population density in 1993. Historical controls include quartile dummies of 1918 population, dummy for whether closer than 500m to main road in 1933, and when the tram instrument is used a dummy for whether a tract is closer than 500m from the historical tram line. Last column includes education level dummies interacted with Post FE. Standard errors clustered by origin UPZ are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 5: College Share

Outcome: Change in College Share	(1) OLS	(2) OLS	(3) IV	(4) IV LCP	(5) IV All
Δ lnRCMA	-0.012	-0.040	-0.032	-0.054	-0.052
	(0.026)	(0.025)	(0.025)	(0.042)	(0.040)
Δ lnRCMA X HighColl		0.043*	0.053*	0.091**	0.095**
		(0.023)	(0.028)	(0.043)	(0.040)
N	1,886	1,886	1,886	1,886	1,886
R^2	0.27	0.27			
F-Stat				89.38	122.14
Over-ID p-value					0.54
Locality FE	X	X	X	X	X
HighColl FE		X	X	X	X
Log Dist CBD X Region FE	X	X	X	X	X
Tract Controls	X	X	X	X	X
Historical Controls	X	X	X	X	X

Note: Outcome is the change in a census tract's share of residents older than 20 with post-secondary education between 1993 and 2005. Dependent variable is change in RCMA between these years using the pre-TM and phase 1 of the system to measure commute times, interacted with a dummy for whether a tract is high college. The high college measure is constructed by first computing the share of college residents within a 1km disk around each tract centroid in 1993 (excluding the tract itself) and then setting high college dummy equal to one for tracts in the top two terciles of its distribution. Specifications with interactions include an intercept to allow growth to differ across low and high college tracts (HighColl FE). Tract controls include log area, log distance to a main road and log population density in 1993; all other controls are as described in previous tables. Final column includes additional control for whether tract is closer than 500m from historical tram route. Columns (1) and (2) run OLS. Column (3) instruments for the change in CMA holding residence and employment fixed at their initial levels. Column (4) instruments using the change in CMA induced by the LCP route, while column (5) additionally includes the tram instrument. Only tracts further than 500m from a portal and the CBD (and less than 3km from a station) are included. Standard errors clustered by tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6: Wages

Outcome: lnWage	(1) OLS	(2) OLS	(3) IV	(4) IV-All	(5) IV-All
lnRCMA	0.479***	0.202*	0.282**	0.221	0.185
	(0.162)	(0.108)	(0.129)	(0.221)	(0.236)
lnRCMA X College					0.298***
					(0.054)
N	75,981	75,981	75,981	75,981	75,981
R^2	0.35	0.47	0.47	0.47	0.47
F-Stat				30.94	16.41
Over-ID p-value				0.94	0.64
UPZ FE	X	X	X	X	X
Region X Post FE	X	X	X	X	X
Log Dist CBD X Region FE X Post FE	X	X	X	X	X
Tract Controls X Post FE	X	X	X	X	X
Worker Controls X Post FE	X	X	X	X	X
College FE X Post FE		X	X	X	X
Historical Controls X Post FE				X	X

Note: Dependent variable is the log hourly wage 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 2009-2014 in the post period. RCMA is measured at the UPZ-level using the pre-TM network in the pre-period, and using the 2006 network in the post-period. IV specification uses both the LCP and Tram instruments. Region are dummies for the North, West and South of the city. College is a dummy for having post-secondary education. Worker controls include gender and log age. Remaining controls are as described in previous tables. Standard errors are clustered by UPZ and period. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 7: Mode Choice Model Estimates

	(1)
Time	-0.012**
	(0.005)
Bus	-0.086*
	(0.050)
Car	0.837***
	(0.292)
TM	-0.216**
	(0.105)
λ	0.140**
	(0.064)
Time of Day Controls	X
Demographic Controls	X

Note: Table shows estimation from nested logit regression on trip-level data from the 2015 Mobility Survey. λ is the correlation parameter for the public nest. Demographic controls include a sex dummy as well as dummies for quintiles of the age distribution, time of day controls include dummies for the hour of trip departure. Each have choice-varying coefficients. Only trips during rush hour (hour of departure 5-8am,4-6pm) to and from work are included. Robust standard errors are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 8: Gravity Regression

	(1)	(2)	(3)	(4)	(5)	(6)
HighSkill X Commute Time	-0.0319***	-0.0256**	-0.0283**	-0.0249**	-0.0166***	-0.0165***
	(0.0123)	(0.0121)	(0.0124)	(0.0122)	(0.0025)	(0.0025)
LowSkill X Commute Time	-0.0387***	-0.0281**	-0.0357***	-0.0284**	-0.0298***	-0.0284***
	(0.0131)	(0.0114)	(0.0132)	(0.0117)	(0.0025)	(0.0024)
N	1,778	1,778	1,778	1,778	1,444	1,444
Method	IV	OLS	IV	OLS	PPML	IV-PPML
Years	1995,2015	1995,2015	1995,2015	1995,2015	2015	2015
Orig-Dest-Skill-Car Own FE	X	X	X	X		
Dest-Skill-Year FE	X	X	X	X	X	X
Orig-Skill-Car Own-Year FE	X	X	X	X	X	X
Crime, House Price, Main Road Ctrls			X	X		

Note: Outcome is the log conditional commute shares (columns 1-4) and conditional commute shares (columnes 5-6). 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 heads of households included. Columns 1-4 estimate fixed effect model using variation within cells between 1995 and 2015. Columns 1 and 2 estimate the baseline model using IV and OLS, where the IV model using the times computed for both car and non-car owners under the LCP and Tram to instrument for times computed using the observed network in the post-TM year. Columns 3 and 4 include controls for (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, interacted with year FE. Columns 5 and 6 estimate the same specification using a PPML model using data from 2015 only (the PPML model with 3 sets of fixed effects did not converge), where column 6 uses the same instruments for transit times. Standard errors are clustered at the origin-destination locality. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 9: GMM Results

Parameter	Estimate
Panel A: Firms	
μ_A	0.212**
	(0.093)
Panel B: Workers	
η_L	2.959***
	(0.861)
η_H	3.329***
	(0.862)
μ_U^L	0.419***
	(0.126)
μ_U^H	0.576***
	(0.144)

Note: Estimates are from joint GMM procedure as described in text. Controls include locality fixed effects, historical controls (dummies for quartile of 1918 population, dummies for whether a tract is closer than 500m from the historical tram line in 1921 and main roads in 1933), log distance to main road, commercial floorspace share in 2000, and log population density and college share in 1993 for employment moment conditions (so as to not include initial values of outcome variables). Only tracts within 3km of the network and those more than 500m from portals and the CBD are included. Standard errors clustered by tract obtained from 100 block-bootstrapped replications resampled at the tract-level.* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 10: Welfare Approximations

	Average Welfare	**Inequality**
First Order Approximation (CMA)	0.937	-0.230
First Order Approximation (VTTS-Model)	1.308	-0.172
General Equilibrium	1.628	0.085

Note: Table shows the percentage change in average welfare and inequality (the ratio of high- to low-skill welfare) from adding TransMilenio to the equilibrium without it. The equilibrium without TransMilenio is computed using the model to simulate its removal. The first row is the first order welfare approximation using the CMA regression elasticities. The second is the VTTS approximation from Proposition 2. The third line is the full general equilibrium response.

Table 11: Effect of Removing Phases 1 and 2 of TransMilenio

	No Spillovers	**Spillovers**
Panel A: Closed City		
GDP	1.470	1.820
Rents	1.769	1.905
Welfare Low	1.492	1.573
Welfare High	1.481	1.656
Inequality	-0.011	0.085
Panel B: Open City		
GDP	5.273	7.701
Rents	7.423	8.910
Population Low	4.729	5.665
Population High	4.726	5.996
Relative Population High	-0.003	0.351

Note: Table shows the (negative of the) value of the percentage change in each variable from removing phases 1 and 2 of the TransMilenio network from the 2012 equilibrium, with and without spillovers.

Table 12: Welfare Effects of TransMilenio: Decomposing the Channels

Model	**Average Welfare**	**Inequality**
Same η, θ, Perf Sub	1.853	-0.346
Diff θ, same η, Perf Sub	1.892	-0.137
Diff θ, η, Perf Sub	2.051	-0.108
Diff θ, η, Imperf Sub	1.602	0.085

Note: Table shows the percentage change in welfare and inequality from TransMilenio under each model. 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. Rows 1 considers a simplified model where worker groups share the same value for η, θ and are perfect substitutes in production. Rows 2 to 4 then compute the full GE effects where η, θ and σ are slowly turned to their estimated or calibrated values.

Table 13: Cost vs. Benefits of TransMilenio

	Closed City		Open City	
	No Spillovers	**Spillovers**	**No Spillovers**	**Spillovers**
Panel A: Costs & Benefits				
NPV Increase GDP (mm)	27,394	33,909	98,234	143,461
Capital Costs (mm)	1,137	1,137	1,137	1,137
NPV Operating Costs (mm)	5,963	5,963	5,963	5,963
NPV Total Costs (mm)	7,101	7,101	7,101	7,101
NPV Net Increase GDP (mm)	20,293	26,808	91,134	136,360
Annual Net Increase GDP	1.09%	1.44%	4.89%	7.32%
Panel B: Land Value Capture				
LVC Band Revenue (mm)	46	65	145	203
As share of capital costs	4.01%	5.72%	12.71%	17.82%
LVC CMA Revenue (mm)	90	116	352	467
As share of capital costs	7.88%	10.21%	30.95%	41.07%

Note: All numbers in millions of 2016 USD. NPV calculate over a 50 year time horizon with a 5% discount rate. Each column refers to a different model. Row (1) reports the increase in NPV GDP from phases 1 and 2 of the TransMilenio network from the baseline equilibrium in 2012 (calculated as the fall in GDP from its removal). Row (2) reports the capital costs of constructing the system, averaging 12.23mm per km over 93km of lines. Row (3) reports the NPV of operating costs, defined conservatively as farebox revenue in 2012. Row (4) reports the NPV of total costs, while row (5) reports the difference between row (1) and row (4). Row (6) reports this difference as a percent of the NPV of GDP in 2012. Row (7) reports the government revenue from the distance band-based land value capture scheme as described in the text, while row (8) reports this as a percentage of capital costs. Rows (9) and (10) report the same figures for the commuter market access-based LVC scheme.

Table 14: Model Extensions

	Closed City			Open City
	Avg. Welfare	**Inequality**	**Output**	**Output**
Baseline	1.602	0.085	1.820	7.701
Alternative Models				
Preference Shocks & Joint Decision	1.866	0.093	0.674	6.570
Domestic Services	1.542	0.091	1.800	-
Alternative Home Ownership				
All Renters	1.667	0.060	1.716	-
Local Home Ownership	1.626	0.054	1.758	-

Note: Table shows the absolute value of the percentage change in welfare and GDP from removing phases 1 and 2 of the TransMilenio network from the 2012 equilibrium across different models. Columns (1)-(3) report values from the closed city model; column (4) reports the change in city output in the open city model where the value from the closed city model varies significantly from the baseline model. Row 1 reports results from the baseline model. Row 3 reports the model where individuals make a joint decision over locations of residence and employment simultaneously and have idiosyncratic preference rather than productivity shocks across locations. Rows 4 and 5 show results from models with alternative home ownership assumptions. In row 4 all individuals are renters, while in row 5 individuals own their homes with probabilities set to home ownership rates of 0.46 and 0.60 for low- and high-skill workers respectively.

Table 15: Policy Counterfactuals

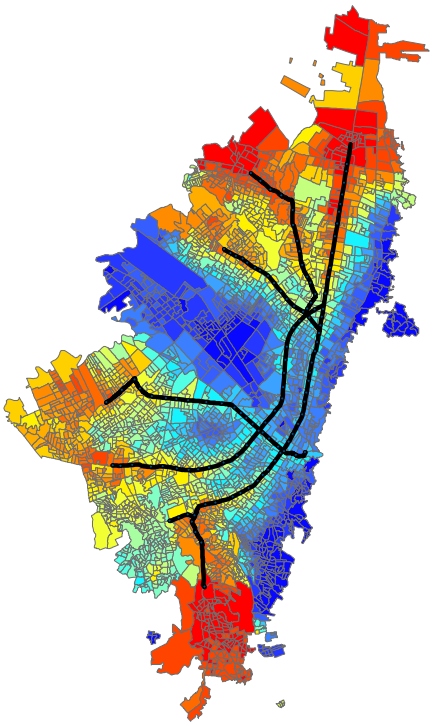
	Avg. Welfare	Inequality	Output
Alternative Networks			
Remove Line South	-0.684	0.003	-0.544
Remove Line North	-0.540	-0.025	-0.676
Remove Feeders	-0.994	-0.028	-0.904
Alternative Time Aggregation			
Equal Preference Weights	0.537	0.038	0.901
Quickest Mode of Public Transit	1.543	-0.031	1.793
Alternative Housing Adjustment			
Free Adjustment	0.299	0.013	0.210
LVC, Bands	0.080	0.028	0.145
LVC, CMA	0.295	0.001	0.184

Note: The first 3 rows suppose TransMilenio had been built without line H in the south, without line A in the North, and without the feeder system respectively. I simulate the effect of moving from the observed equilibrium to each counterfactual one, and report the percentage change in each variable relative to the observed equilibrium. Rows 4 and 5 consider alternative methods of aggregating mode-specific times in time indices. Row 4 supposes the preference shifters are equal across all transit modes, and row 5 assumes individuals take the quickest mode of public transit (i.e. there is no preference heterogeneity within the public nest). The last 3 rows consider alternative housing supply models. I first solve for the counterfactual equilibrium without TransMilenio using the unobservables recovered in the post-period. I then compute the equilibrium returning to the TransMilenio network under each housing supply model, and report the percentage change in each variable relative to the observed equilibrium. Row 6 is the case with freely adjusting housing. Row 7 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. Row 6 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.

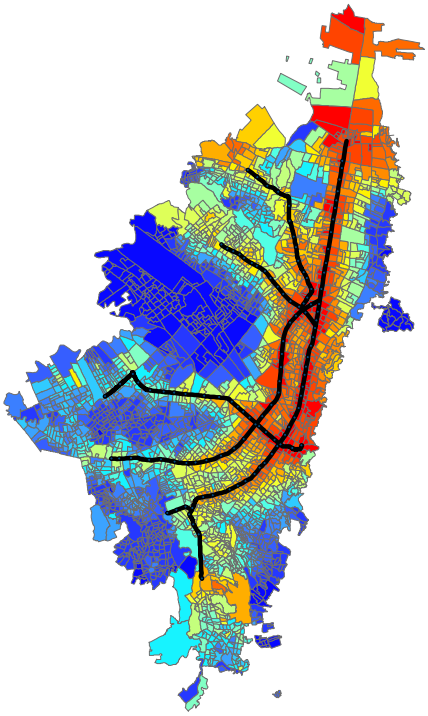
Figures

Figure 1: Change in Commuter Market Access from TransMilenio

(a) Resident CMA



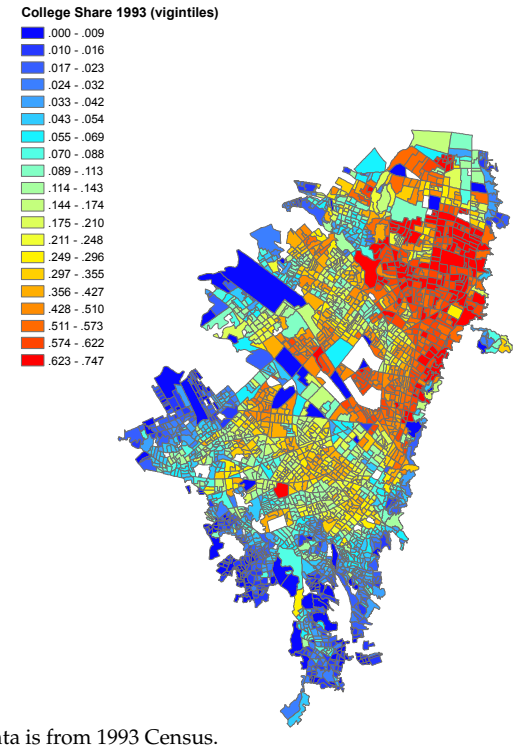
(b) Firm CMA



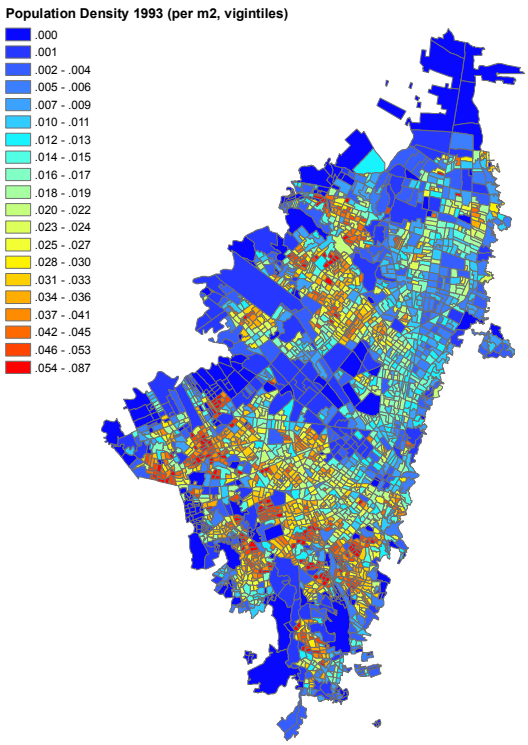
Note: Plot shows the baseline instrument for the change in CMA induced by holding population and employment fixed at their initial level and changing only commute costs. Tracts are grouped into deciles based on the the change in CMA, with warmer colors indicating a larger increase in CMA. Black line shows the TransMilenio routes as of 2006. See Section 3.2 for full discussion.

Figure 2: Population Density and Demographic Composition in 1993

(a) College Share



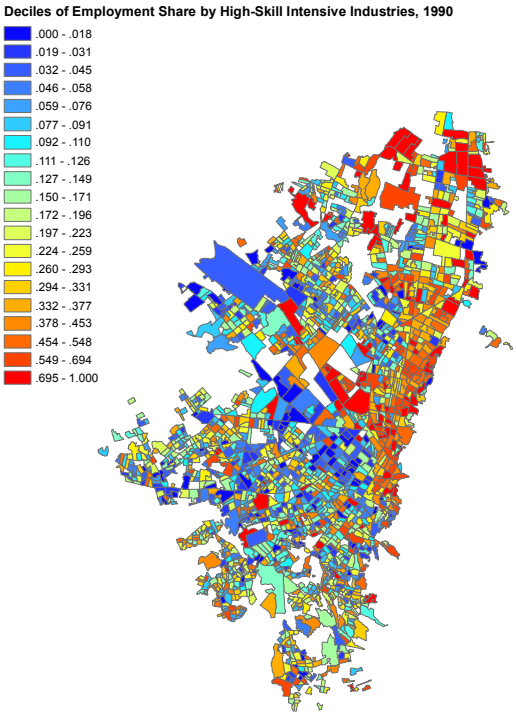
(b) Population Density



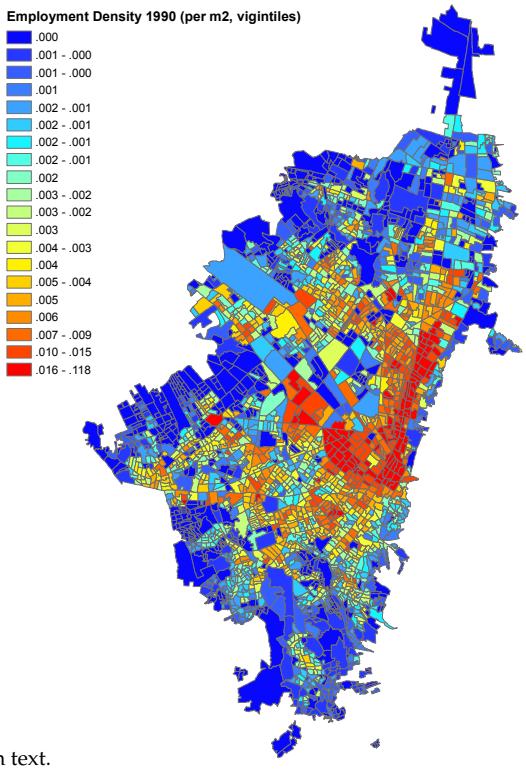
Note: Data is from 1993 Census.

Figure 3: Employment Density and Industry Composition in 1990

(a) High-Skill Industry Share

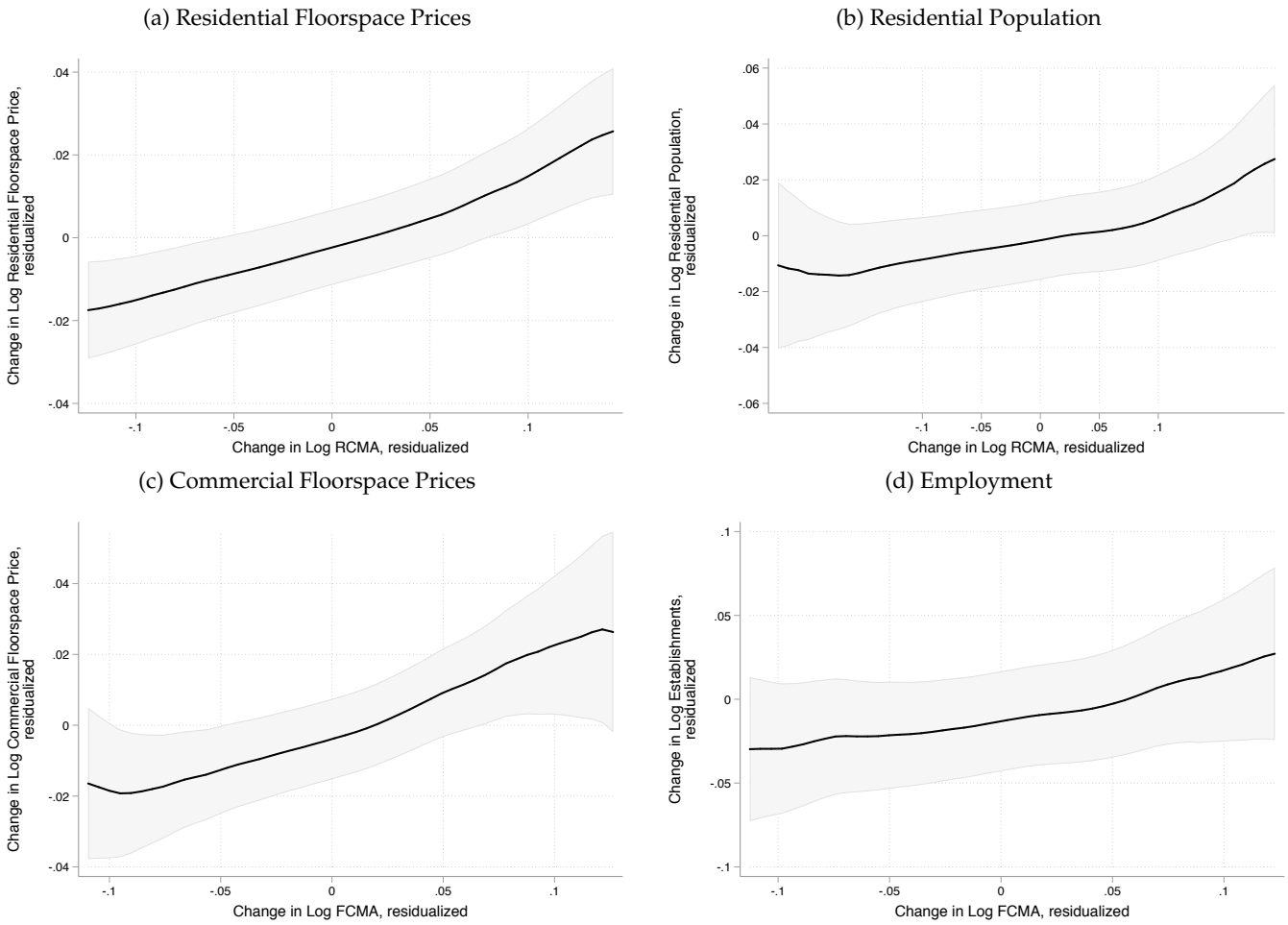


(b) Employment Density



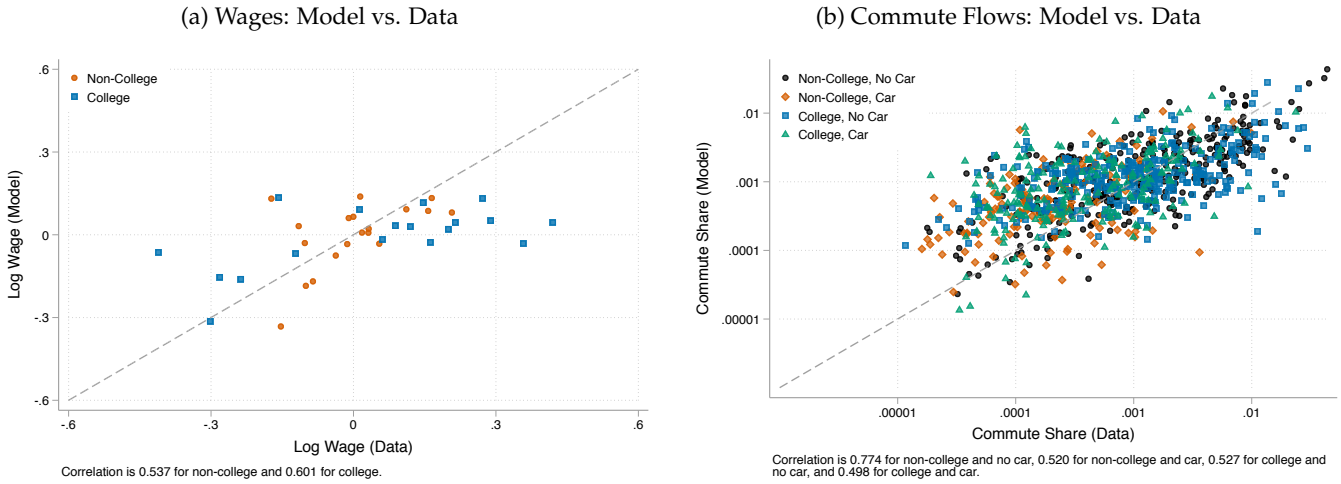
Note: Data is from 1990 Economic Census. High-skill industries defined in text.

Figure 4: 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 (4) of Table 3 in which CMA is measured holding population and employment fixed at their initial levels, with the full set of baseline controls included, and is regressed directly on outcomes.

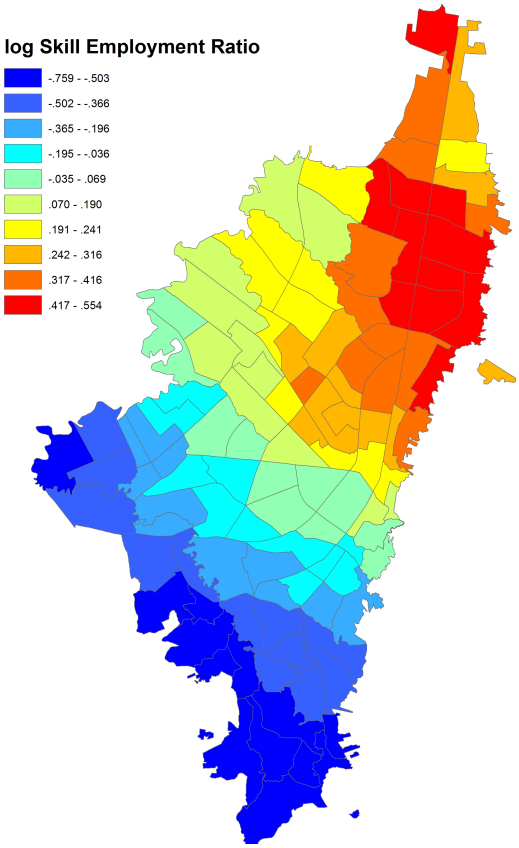
Figure 5: Wages and Commute Flows: Model vs. Data



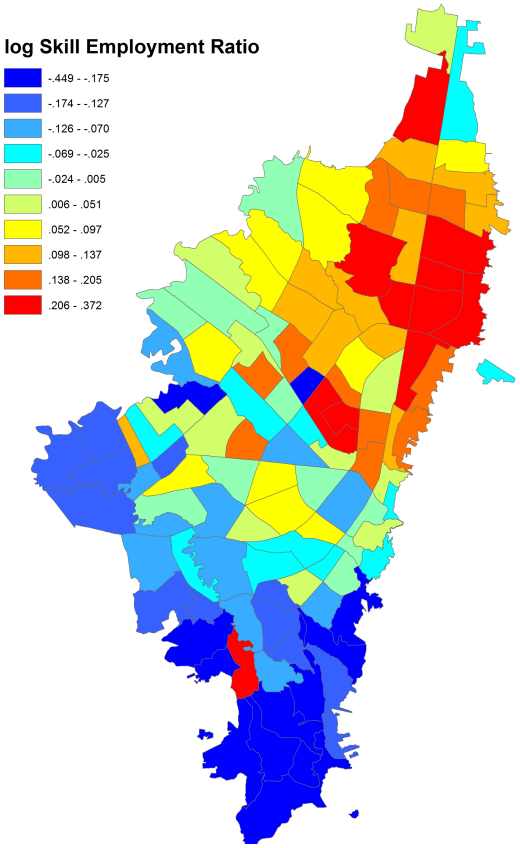
Note: Panel (a) compares the average wage by skill group in each locality as predicted by the model with that observed in the GEIH data (not used in estimation). In panel (b), a observation is a locality origin-destination pair, skill group and car ownership combination. Plot shows relationship between share of commuters choosing each (i, j, a) pair in the model vs those doing so in the 2015 Mobility Survey.

Figure 6: Relative Employment by Skill by UPZ: Model vs Data

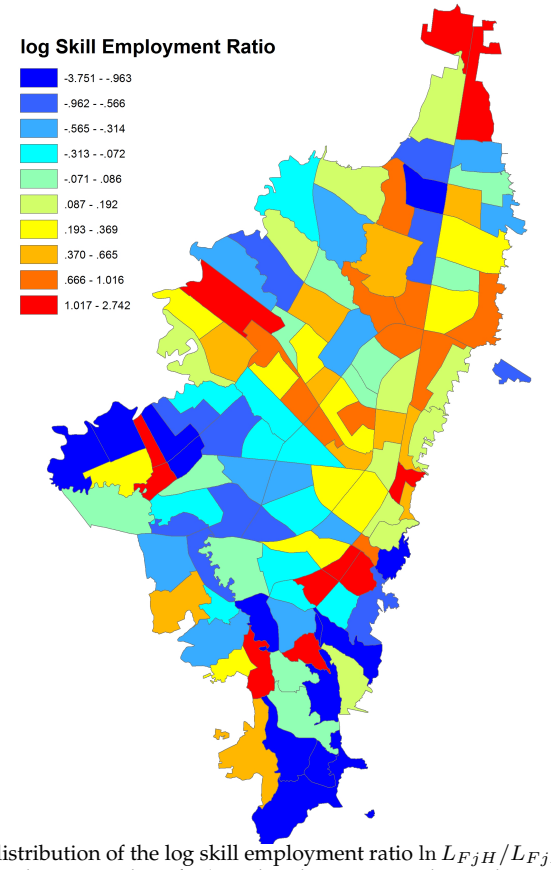
(a) Model: Perfect Substitutes & Same θ



(b) Model: Baseline Estimates



(c) Data



Note: Panel (a) shows the deciles of the distribution of the log skill employment ratio $\ln L_{FjH}/L_{FjL}$ by UPZ in the model when skill groups are perfect substitutes in production and have the same value of θ (equal to the average value in the population. Panel (b) shows the distribution for the baseline model. Panel (c) shows the distribution in the 2015 Mobility Survey. Correlation between data in panel (a) and (c) is 0.358, while that between panel (b) and (c) is 0.408.

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

Contents

A Additional Tables **1**

B Additional Figures **13**

C Reduced Forms from Quantitative Urban Models **22**
- C.1 Baseline Model . . . 22
- C.2 Reduced Form Representation . . . 23
- C.3 Generalizing the Framework . . . 24
- C.4 Mapping Common Modeling Assumptions to the Framework . . . 25

D Data Appendix **29**
- D.1 Dataset Description . . . 29
- D.2 Computing Commute Times . . . 34
- D.3 Constructing the Instruments . . . 36

E Additional Information on Commuting and TransMilenio **36**

F Supplementary Empirical Results **38**
- F.1 Regressions of Relative Speed over Time . . . 38
- F.2 Effect of TransMilenio on Growth in Floorspace . . . 38
- F.3 Effect of TransMilenio on Other Mode Speeds . . . 39
- F.4 Engel Curve for Housing . . . 40

G Supplementary Theoretical Results **40**
- G.1 Mode Choice Problem . . . 40
- G.2 First Order Welfare Approximations . . . 42

H Supplementary Quantitative Results **42**
- H.1 Calibrating $T_H, \bar{h}, p_a$ . . . 42
- H.2 Algorithm for Solving The Model . . . 43
- H.3 Welfare Decomposition . . . 44
- H.4 Model with Preference Shocks and Joint Live-Work Decision . . . 45
- H.5 Model with Employment in Domestic Services . . . 46
- H.6 Model with Home Ownership . . . 47
- H.7 Model with Variable Housing Supply . . . 47
- H.8 Monte Carlo: Single-Group Regressions on Multiple-Group Model . . . 48

I Proofs **48**

A Additional Tables

Table A.1: Additional OLS Results

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Residents								
ln(Res Floorspace Price)		0.362***	0.120	0.151*	0.249***	0.295***	0.241***	0.311***
		(0.075)	(0.078)	(0.078)	(0.065)	(0.081)	(0.057)	(0.078)
	N	1,975	1,975	1,975	1,975	1,975	1,975	1,785
R^2		0.36	0.40	0.43	0.55	0.55	0.55	0.50
ln(Res Population)		0.253***	0.127	0.151	0.174*	0.228**	0.148*	0.299**
		(0.087)	(0.095)	(0.095)	(0.094)	(0.111)	(0.086)	(0.148)
	N	2,028	2,028	2,028	2,028	2,028	2,028	1,756
	R^2	0.11	0.11	0.14	0.16	0.16	0.14	0.15
Panel B: Firms								
ln(Comm Floorspace Price)		0.146	0.220**	0.215**	0.245**	0.358***	0.191**	0.211**
		(0.089)	(0.098)	(0.101)	(0.105)	(0.130)	(0.095)	(0.100)
	N	1,914	1,914	1,914	1,914	1,914	1,914	1,755
	R^2	0.09	0.11	0.13	0.15	0.15	0.15	0.17
Comm Floorspace Share		0.181***	0.160***	0.160***	0.146***	0.143***	0.134***	0.105***
		(0.033)	(0.035)	(0.036)	(0.037)	(0.045)	(0.033)	(0.035)
	N	2,013	2,013	2,013	2,013	2,013	2,013	1,818
	R^2	0.06	0.06	0.07	0.08	0.08	0.08	0.07
ln(Establishments)		0.187	0.816***	0.696**	0.580*	0.668*	0.580**	0.764***
		(0.292)	(0.305)	(0.308)	(0.296)	(0.370)	(0.265)	(0.284)
	N	1,753	1,753	1,753	1,753	1,753	1,753	1,750
	R^2	0.19	0.21	0.25	0.30	0.30	0.30	0.31
Locality FE		X	X	X	X	X	X	X
Log Dist CBD X Region FE			X	X	X	X	X	X
Basic Tract Controls				X	X	X	X	X
Historical Controls				X	X	X	X	X
Tract Demographic Controls					X	X	X	X
Land Market Controls					X	X	X	X
Distance to TM Controls						X		
Exclude Band							1.5km	
Accessibility Measure								X

Note: Observation is a census tract. Each entry reports the coefficient from a regression of the variable in each row on firm or residential commuter market access in first differences. Each column corresponds to a specification. Land market regressions use changes between 2000 and 2012, measuring the change in CMA induced by phases 1 and 2 of TransMilenio holding population and employment fixed at their initial values. Establishment regressions use changes between 2000 and 2015 and are weighted by number of establishments in 2000. Population regressions use changes between 1993 and 2005 measuring the change in CMA induced by phase 1 and are weighted by 1993 population. Only tracts within 3km of a station in the respective phases are included. Column (1) includes locality fixed effects. Column (2) includes log distance to the CBD, interacted with dummies for whether the locality is in the North, West or South of the city. Column (3) includes basic tract controls (log area, log distance to main road) and historical controls (quartile dummies of 1918 population, dummy for whether closer than 500m to main road in 1933). Column (4) includes tract demographic controls (1993 college share in all specifications, and 1993 log population density for outcomes other than population) and initial land market characteristics in 2000 for outcomes other than population (average floor area ratio, share of land developed, share of floorspace used for commercial purposes, log value of floorspace per square meter) since 2000 is not the initial year for these outcomes. Land market controls that represent initial values of outcome variable (i.e. value of floorspace in rows 1 and 3, commercial floorspace share in row 4) are excluded in each specification. Column (5) includes dummy for whether tract is closer than 500m from any TransMilenio station for each respective phase. Column (6) computes the change in market access to tracts further than 1.5km. Column (7) uses the Hansen (1969) accessibility measure as the measure of change in market access (see text). Standard errors cluestered by census tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.2: OLS Results: Robustness

		(1) Baseline	(2) Alt Times	(3) Alt Times	(4) Slow	(5) Fast	(6) High θ	(7) Low θ	(8) No Cutoff	(9) Unweighted	(10) Conley SE
Panel A: Residents											
ln(Res Floorspace Price)		0.249***	0.442***	0.270***	0.240***	0.255***	0.166***	0.464***	0.183***		0.249***
		(0.065)	(0.117)	(0.070)	(0.065)	(0.063)	(0.051)	(0.104)	(0.061)		(0.093)
	N	1,975	1,975	1,975	1,975	1,975	1,975	1,975	2,231		1,975
	R^2	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.58		0.55
ln(Res Population)		0.174*	0.285*	0.189*	0.158*	0.198**	0.113	0.318*	0.302***	0.326**	0.174
		(0.094)	(0.165)	(0.103)	(0.093)	(0.093)	(0.070)	(0.162)	(0.093)	(0.137)	(0.151)
	N	2,028	2,028	2,028	2,028	2,028	2,028	2,028	2,426	2,372	2,028
	R^2	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.18	0.20	0.16
Panel B: Firms											
ln(Comm Floorspace Price)		0.245**	0.472**	0.264**	0.263**	0.221**	0.198**	0.319**	0.247**		0.245*
		(0.105)	(0.189)	(0.114)	(0.109)	(0.099)	(0.082)	(0.159)	(0.102)		(0.126)
	N	1,914	1,914	1,914	1,914	1,914	1,914	1,914	2,145		1,914
	R^2	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15		0.15
Comm Floorspace Share		0.161***	0.289***	0.175***	0.168***	0.150***	0.121***	0.225***	0.148***		0.161***
		(0.037)	(0.064)	(0.040)	(0.038)	(0.034)	(0.030)	(0.054)	(0.036)		(0.044)
	N	2,071	2,071	2,071	2,071	2,071	2,071	2,071	2,341		2,071
	R^2	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06		0.07
ln(Establishments)		0.580*	0.795	0.638**	0.535*	0.576**	0.351	1.137**	0.622**	0.206	0.580
		(0.296)	(0.532)	(0.321)	(0.308)	(0.279)	(0.236)	(0.446)	(0.291)	(0.265)	(0.357)
	N	1,753	1,753	1,753	1,753	1,753	1,753	1,753	1,948	1,880	1,753
	R^2	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.28	0.27	0.30
Time Specification			Pref. Wgt	Min. All	Slow	Fast					
Locality Fixed Effects		X	X	X	X	X	X	X	X	X	X
CBD X Region Controls		X	X	X	X	X	X	X	X	X	X
Tract Controls		X	X	X	X	X	X	X	X	X	X
Historical Controls		X	X	X	X	X	X	X	X	X	X

Note: Each entry reports the coefficient from a regression of the variable in each row on firm or residential commuter market access in first differences. Each column corresponds to a specification. Column (1) reports the baseline specification including census tracts within 3km of a station in the corresponding phase of the system. Controls are full set of those described in text and previous tables. Column (2) measures average commute costs weighting by the preference parameters from the logit model. Column (3) uses the minimum time across all modes. Column (4) uses slower commute times that best match average speeds in the post-period. Column (5) uses faster ones that do the same for the pre-period. Column (6) uses a larger value of θ equal to 1.5 times its baseline estimate, while column (7) scales it down by the same factor. Column (8) relaxes the 3km distance cutoff and includes all census tracts, while column (9) provides unweighted estimates for outcomes where weighted regressions are presented in the main tables. Tracts within 4km of a station are included. Column (10) provides spatial HAC standard errors (Conley 1999), using a spatial weight of 1 for tracts within 500m of each other and zero otherwise. Standard errors clustered by tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.3: IV Robustness, Alternate LCP Cutoffs

	(1) OLS	(2) IV	(3) IV-LCP	(4) IV All	(5) OLS	(6) IV	(7) IV-LCP	(8) IV All
Panel A: Residents								
ln(Res Floorspace Price)	0.266***	0.244***	0.356***	0.351***	0.247***	0.228***	0.397***	0.392***
	(0.058)	(0.064)	(0.101)	(0.104)	(0.053)	(0.059)	(0.090)	(0.094)
N	1,823	1,823	1,823	1,823	1,975	1,975	1,975	1,975
F-Stat			585.73	288.89			764.33	374.25
Over-ID p-value				0.49				0.04
ln(Res Population)	0.309***	0.227**	0.292*	0.276*	0.272***	0.192*	0.311**	0.297**
	(0.110)	(0.111)	(0.159)	(0.160)	(0.102)	(0.103)	(0.136)	(0.138)
N	1,883	1,883	1,883	1,883	2,028	2,028	2,028	2,028
F-Stat			1,230.41	612.98			1,753.08	862.34
Over-ID p-value				0.82				0.92
Panel B: Firms								
ln(Comm Floorspace Price)	0.218**	0.244**	0.222	0.231	0.209**	0.244**	0.223	0.245*
	(0.105)	(0.107)	(0.144)	(0.142)	(0.101)	(0.104)	(0.144)	(0.142)
N	1,766	1,766	1,766	1,766	1,914	1,914	1,914	1,914
F-Stat			691.58	519.81			700.68	533.20
Over-ID p-value				0.96				0.76
Comm Floorspace Share	0.129***	0.134***	0.086	0.098*	0.159***	0.160***	0.119**	0.131**
	(0.037)	(0.039)	(0.056)	(0.056)	(0.036)	(0.038)	(0.055)	(0.055)
N	1,861	1,861	1,861	1,861	2,013	2,013	2,013	2,013
F-Stat			719.92	537.66			737.69	554.23
Over-ID p-value				0.46				0.43
ln(Establishments)	0.524*	0.527*	1.267**	0.823*	0.600**	0.578**	1.061*	0.669
	(0.293)	(0.299)	(0.502)	(0.490)	(0.282)	(0.292)	(0.553)	(0.521)
N	1,616	1,616	1,616	1,616	1,753	1,753	1,753	1,753
F-Stat			178.75	198.64			135.50	181.08
Over-ID p-value				0.02				0.07
LCP Cutoff	1km	1km	1km	1km	None	None	None	None
Locality Fixed Effects	X	X	X	X	X	X	X	X
CBD X Region Controls	X	X	X	X	X	X	X	X
Full Tract Controls	X	X	X	X	X	X	X	X

Note: Observation is a census tract. Each entry reports the coefficient from a regression of the variable in each row on firm or residential commuter market access in first differences. Each column corresponds to a specification. Controls are full set of those described in main IV table. Columns (1)-(4) repeat the main IV specifications on a subsample that drops all tracts within 1km of a portal and the CBD vs the 500m cutoff reported in the main table. Columns (5)-(8) repeat each specification including all census tracts. Standard errors clustered by tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.4: Falsification Tests

		(1) ResPr	(2) ResPr	(3) ResPop	(4) ResPop	(5)	(6)
Panel A: Residents							
ln(RCMA)		0.156***	0.160***	0.182*	0.134		
		(0.054)	(0.054)	(0.095)	(0.093)		
ln(RCMA) Later Phase			0.075		0.037		
			(0.112)		(0.099)		
	N	1,980	2,084	2,028	2,327		
	R^2	0.46	0.46	0.16	0.15		
		(1) CommPr	(2) CommPr	(3) CommSh	(4) CommSh	(5) Estb	(6) Estb
Panel B: Firms							
ln(FCMA)		0.222**	0.211**	0.090***	0.089***	0.521*	0.490
		(0.099)	(0.099)	(0.027)	(0.027)	(0.301)	(0.300)
ln(FCMA) Later Phase			0.345		0.088		1.785
			(0.365)		(0.093)		(1.316)
	N	1,910	2,011	2,013	2,119	1,753	1,845
	R^2	0.06	0.06	0.09	0.09	0.30	0.29
Locality FE		X	X	X	X	X	X
Log Dist CBD X Region FE		X	X	X	X	X	X
Basic Tract Controls		X	X	X	X	X	X
Historical Controls		X	X	X	X	X	X
Tract Demographic Controls		X	X	X	X	X	X
Land Market Controls		X	X	X	X	X	X

Note: Each column reports coefficients from a regression of the growth in the outcome on a commuter market access measure. For each outcome, the first column reports the baseline specification where CMA growth is measured using the change in commute costs induced by TransMilenio holding residence and employment fixed at their initial levels (i.e. the reduced form). The second column adds an additional variable containing the growth in this CMA measure induced by opening later phases of the system. In Panel A, columns (1) and (2) report the change in log residential floorspace price per m2 between 2000 and 2007 (as opposed to 2012) as the outcome variable. Column (1) reports the coefficient on the change in RCMA induced by phase 1 and 2 of the system. Column (2) adds the change in RCMA going from phase 2 to 3 (opened in 2011). Tracts closer than 3km from a station of the phases in question are included in all specifications, explaining the expanding sample size. Columns (3) and (4) repeat the exercise for growth of log residential populations (where growth is measured between 1993 and 2005), where RCMA is the change induced by phase 1 and the later phase refers to the change from adding phases 2 and 3 (opened in 2005/2011). In Panel B, the outcomes are the growth in log commercial floorspace price per m2, commercial floorspace share (both between 2000 and 2007) and log number of establishments (between 2000 and 2015) correspond to columns (1)-(2), (3)-(4) and (5)-(6) respectively. For all 3 outcomes, the change in FCMA in row (1) is that induced by phase 1 and 2 while the later phase refers to the change induced by phase 3. For the establishment counts, the later phase is of course open before the post-period. There is still less of an effect than for the earlier phases and, although this is noisy, it confirms there wasn't significant employment growth before the opening of the third phase. Standard errors clustered by tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.5: GMM Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Firms								
μ_A	0.212**	0.359***	0.326***	0.144*	0.333***	0.231***	0.187***	0.230**
	(0.093)	(0.115)	(0.098)	(0.087)	(0.084)	(0.071)	(0.063)	(0.097)
log(Dist TM)		0.042**						
		(0.021)						
Panel B: Workers								
η_L	2.959***	2.913***	2.859***	2.707**	1.507*	1.506*	1.506*	3.352***
	(0.861)	(1.008)	(0.809)	(1.052)	(0.905)	(0.905)	(0.905)	(1.150)
η_H	3.329***	3.289***	3.244***	3.086***	2.318**	2.317**	2.317**	3.715***
	(0.862)	(1.011)	(0.809)	(1.054)	(0.914)	(0.914)	(0.914)	(1.137)
log(Dist TM)		-0.014						
		(0.033)						
μ_U^L	0.419***	0.416***	0.416**	0.392***	0.300***	0.301***	0.301***	0.322***
	(0.127)	(0.111)	(0.165)	(0.096)	(0.071)	(0.071)	(0.071)	(0.097)
μ_U^H	0.576***	0.574***	0.569***	0.552***	0.485***	0.485***	0.485***	0.479***
	(0.144)	(0.139)	(0.162)	(0.129)	(0.119)	(0.119)	(0.119)	(0.123)
Labor Demand Elasticity	1.3	1.3	1.3	1.3	2.5	2.5	2.5	1.3
Demand Elasticity	6	6	4	9	4	6	9	6
$\bar{h} = 0$								X

Note: Estimates are from the specification used in the joint GMM procedure as in the baseline specification and include full controls from it. Column (1) is the baseline specification. Column (2) adds a control for log distance to the nearest TransMilenio station, instrumented with the distance to the LCP instrument. Columns (3) and (4) use alternate values for the demand elasticity, while columns (5)-(7) use alternate values for the labor demand elasticity and demand elasticity. Column (8) sets $\bar{h} = 0$. Only tracts within 3km of the network and those more than 500m from portals and the CBD are included. Standard errors clustered by tract obtained from 100 block-bootstrapped replications resampled at the tract-level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.6: Amenities and Productivities: Model vs Data

Panel A: Amenities	(1) ln Crime Density	(2) ln Crime Density	
Amenity Elasticity	-0.190***	-0.221***	
	(0.015)	(0.016)	
Skill	College	Non-College	
R^2	0.28	0.27	
N	503	501	
Panel B: Productivities	(1) ln Crime Density	(2) ln Slope	(3) ln Roads
Productivity Elasticity	-0.062	-0.336***	0.192***
	(0.047)	(0.022)	(0.025)
R^2	0.01	0.28	0.10
N	504	615	615

Note: Estimates show coefficients from regressions of log (composite) productivities and amenities on variable given in each column. Observation is a sector. Crime is measured as total homicides in a sector between 2007 and 2012. In column (2) of Panel B, the dependent variable is log of the average slope of land. In column (3), the dependent variable is log of 1 plus the kilometers of primary and secondary roads within a disk of 1.5km radius around the sector centroid. Standard errors clustered by sector reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.7: Non-Targeted Moment: Sorting Response to TransMilenio

Outcome: Change in College Share	(1)	(2)	(3)
Δ lnRCMA	-0.052	-0.023	-0.045
	(0.040)	(0.029)	(0.028)
Δ lnRCMA X HighColl	0.095**	0.091***	0.067**
	(0.040)	(0.033)	(0.031)
Model	Data	Non-Homothetic	Homothetic

Note: Column (1) replicates regression from Table 5 in the data. Columns (2) and (3) run the same regression on the data generated from the model, where the post-period data is observed and the pre-period data is the counterfactual data generated from the model when TransMilenio is removed. Column (2) reports the results from the model with non-homotheticities (and local home ownership), while column (3) reports the results from the model without non-homotheticities. Standard errors clustered by census tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.8: Welfare Gains Accounted for by Time Savings Across Models

	Same θ, η, Perf Sub	**Diff θ, same η, Perf Sub**	**Diff θ, η, Perf Sub**	**Diff θ, η, Imperf Sub**
Welfare Gain from Time Savings	1.198	1.228	1.220	1.268
Full Welfare Gain	1.888	1.929	2.094	1.628
Fraction Accounted by Time Savings	0.635	0.637	0.583	0.779

Note: Table shows the percentage change in average welfare and inequality from adding TransMilenio to the 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 the first row all choices are held fixed so that all the gains come from time savings. The expression uses the DEK-like term derived in the appendix. The second row reports the welfare change in the full GE model, while the third reports the fraction of the total gains accounted for by time savings alone.

Table A.9: Effect of TransMilenio: Robustness

	Avg. Welfare	**Inequality**	**Output**
Baseline	1.602	0.085	1.820
Larger θ	1.072	0.045	1.360
Larger η	1.482	0.150	1.801
Alt θ	2.617	0.318	2.452
Smaller Spillovers	1.552	0.015	1.582
$\sigma_L = 2.5$	1.697	0.033	1.796
Census Employment	1.605	0.084	1.819
$\sigma = 4$	1.484	0.144	1.854
$\sigma = 9$	1.712	0.027	1.788

Note: Table shows the (negative of the) value of the percentage change in welfare from removing phases 1 and 2 of the TransMilenio network from the 2012 equilibrium across different models, with spillovers set to their estimated values. Row (1) reports the values from the baseline model. Rows (2) and (3) sets θ_g and η_g to 1.33 times their estimated values respectively. Row (4) uses alternative values of theta from the PPML regression. Row (5) sets spillovers equal to one third of their estimated values. Row (6) uses a larger elasticity of substitution across labor groups. Row (7) uses census employment measured in 2005 instead of the CCB employment measured in 2015 as the measure of employment in the baseline equilibrium. Rows (8) and (9) use alternate values of the elasticity of demand.

Table A.10: 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.11: 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 (ownership is at the household level).

Table A.12: TransMilenio Use and Income

	TM	TM	TM	TM
Bottom Income Tercile	0.119***	0.055	0.091*	0.030
	(0.040)	(0.045)	(0.049)	(0.050)
Middle Income Tercile	0.052*	0.006	0.074**	0.033
	(0.028)	(0.031)	(0.037)	(0.036)
R^2	0.03	0.04	0.66	0.67
N	4,299	4,299	2,813	2,813
Own Car		X		X
UPZ O-D FE			X	X
Time of day Controls	X	X	X	X
Demographic Controls	X	X	X	X

Note: Standard errors clustered at upz origin-destination pair. TM is a dummy for whether TransMilenio is used during a commute, relative to the omitted categories of car and buses. Data is from 2015. Car is a dummy for whether household owns a 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. Trips to work included. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.13: Relative TransMilenio Speeds over Time

	lnSpeed	lnSpeed	lnSpeed	lnSpeed	lnSpeed	lnSpeed
Bus	-0.167***	-0.205***	-0.157***	-0.233***	-0.239***	-0.219***
	(0.017)	(0.034)	(0.029)	(0.015)	(0.040)	(0.033)
TM	0.093***	0.011	0.082***	-0.045**	-0.105**	-0.094***
	(0.020)	(0.037)	(0.031)	(0.020)	(0.042)	(0.036)
R^2	0.04	0.07	0.06	0.53	0.68	0.65
N	15,209	5,486	9,106	13,199	3,524	7,154
Year	2005	2011	2015	2005	2011	2015
UPZ O-D FE				X	X	X
Time of day Controls	X	X	X	X	X	X
Demographic Controls	X	X	X	X	X	X

Note: Standard errors clustered at upz origin-destination pair. Bus is a dummy for whether bus is used during a commute, relative to the omitted category of car. Data is from 1995. 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. Trips to work during rush hour included. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.14: Effect of TransMilenio on Growth in Floorspace

Outcome: Floorspace Growth	(1)	(2)	(3)	(4)	(5)	(6)
log Distance F1	0.099***	0.053***	0.086***	0.052***		
	(0.013)	(0.010)	(0.011)	(0.010)		
log Distance F2	0.093***	0.014	0.080***	0.018		
	(0.016)	(0.013)	(0.013)	(0.012)		
log Distance F3	0.083***	-0.018	0.097***	0.003		
	(0.019)	(0.024)	(0.016)	(0.019)		
Vacant Pre			4.299***	4.360***		
			(0.243)	(0.458)		
Vacant Pre X log Distance F1			-0.097***	-0.119***		
			(0.016)	(0.036)		
Vacant Pre X log Distance F2			-0.075***	-0.152***		
			(0.013)	(0.038)		
Vacant Pre X log Distance F3			-0.158***	-0.095***		
			(0.017)	(0.028)		
log RCMA					-0.084	
					(0.119)	
log FCMA						-0.009
						(0.148)
N	27,209	23,058	27,209	23,058	2,015	2,015
R^2	0.14	0.38	0.31	0.48	0.43	0.43
Locality FE	X	X	X	X	X	X
Block Controls		X		X	X	X
Dist. CBD X Region Controls		X		X	X	X

Note: Standard errors are clustered at the census tract. Observation is a block in columns 1-4, census tract in 5 and 6. Outcome is growth in floorspace between 2012 and 2000 measured using the Davis-Haltiwanger measure. Only observations closer than 3km from station included. Block controls are log distance to nearest main road, log population density 1993, and floor-area-ratio in 2000. CBD controls is log distance to CBD interacted with a dummy for whether the block is in the North, West or South of the city. F1 indicates distance to closest station in fase 1, the same applies to F2 and F3. Vacant pre is dummy equal to one if the block was vacant in 2000. Columns (1)-(4) measure to distance to closest station, while columns (5)-(6) use the change in CMA from holding employment and residence fixed at their initial level and changing the commute network from pre-TM to phases 1 and 2 of the system (i.e. the baseline instrument from the housing market specifications in the paper). *p<0.1; ** $p < 0.05$; *** $p < 0.01$

Table A.15: Effect of TransMilenio on other Mode Speeds

Outcome: lnSpeed	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Car Trips						
TM Route X Post	-0.160**	-0.107	-0.060	-0.043	0.014	0.052
	(0.079)	(0.086)	(0.089)	(0.062)	(0.064)	(0.065)
R^2	0.79	0.80	0.80	0.79	0.80	0.80
N	9,916	9,916	9,916	9,916	9,916	9,916
Panel B: Bus Trips						
TM Route X Post	-0.096**	-0.164***	-0.074	-0.015	-0.064	-0.020
	(0.042)	(0.046)	(0.047)	(0.041)	(0.041)	(0.040)
R^2	0.71	0.72	0.72	0.71	0.72	0.72
N	38,616	38,616	38,616	38,616	38,616	38,616
Route Measure	Share TM	Share TM	Share TM	TM>75%	TM>75%	TM>75%
Baseline Controls	X	X	X	X	X	X
Origin-Destination FE	X	X	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 Mobility Survey. 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 during rush hours included. Panel A includes only trips by car, while panel B includes only those by bus. $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.16: 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: Column (1) 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.17: 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

B Additional Figures

Figure A.1: TransMilenio Routes

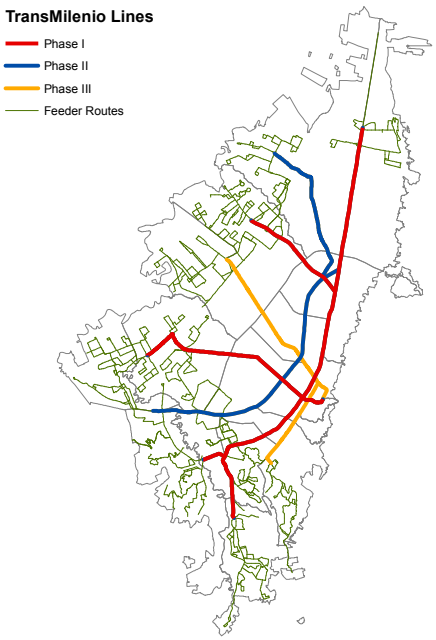
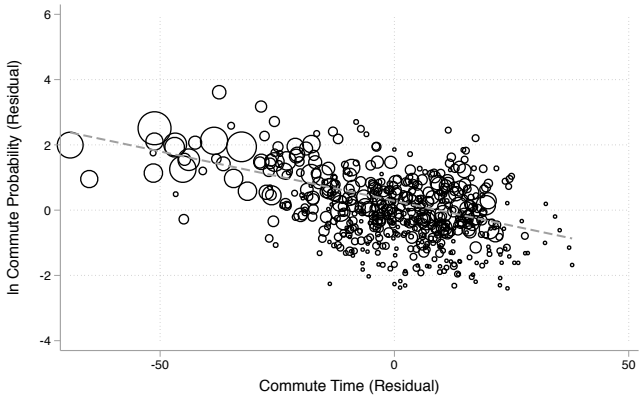


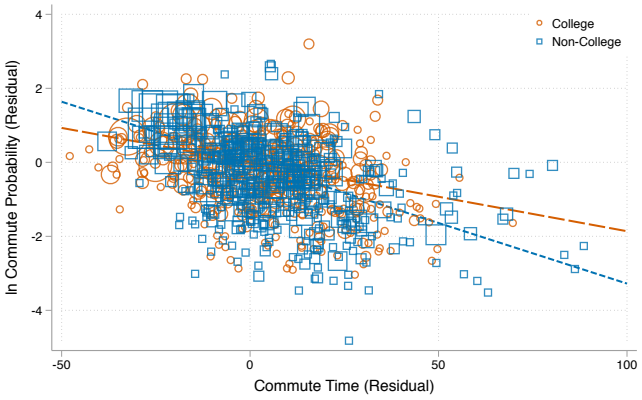
Figure A.2: Fit of Gravity Commuting Model

(a) Overall



Markers proportional to number of observations at the origin-car-destination level. R2 is 0.426

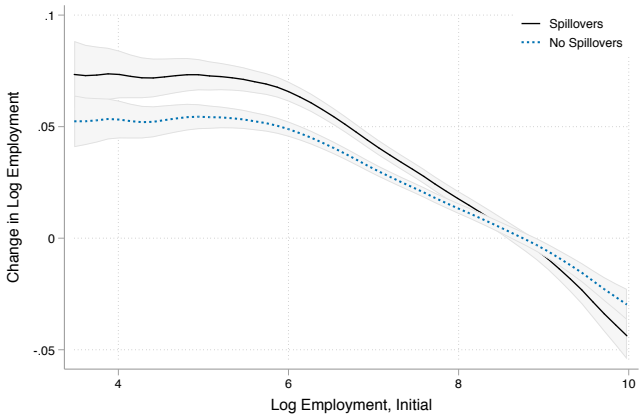
(b) By Group



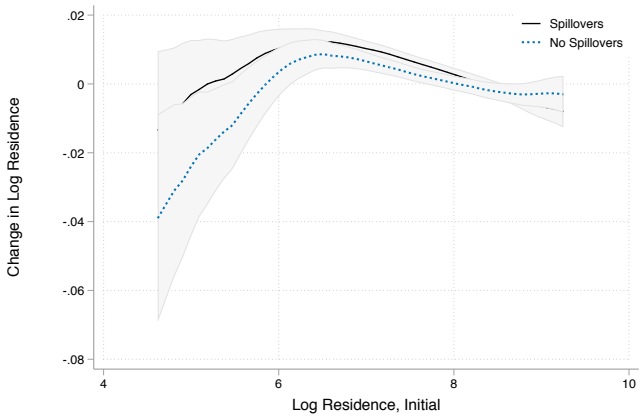
Markers proportional to number of observations at the origin-car-destination level.

Figure A.3: Simulated Changes in Outcomes

(a) Employment



(b) Residential Population



Note: Panels (a) and (b) plot the change in employment and population in each tract when TransMilenio is removed by each variable's initial level in the equilibrium with the system.

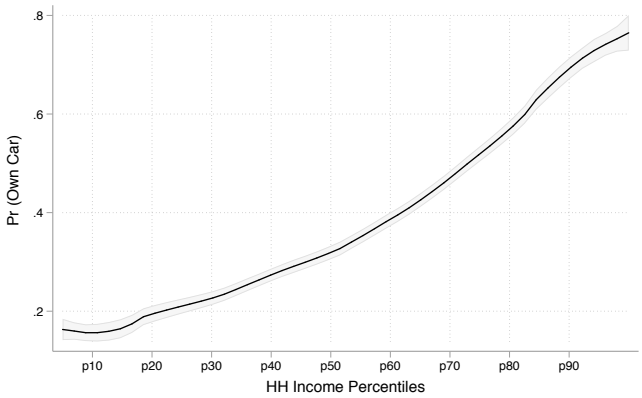
Figure A.4: TransMilenio

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

(b) TransMilenio Station, Avenida Caracas (Norte)

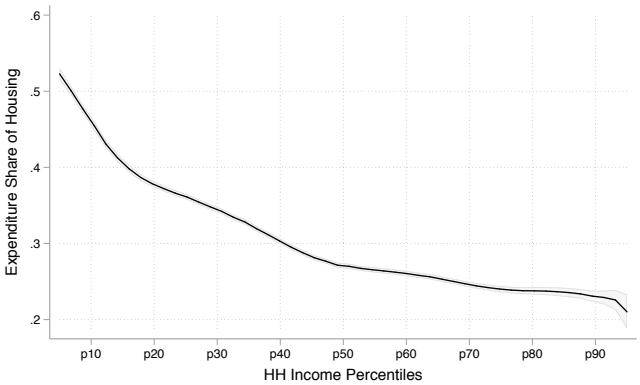
Figure A.5: Engel Curves for Car Ownership and Housing

(a) Car Ownership



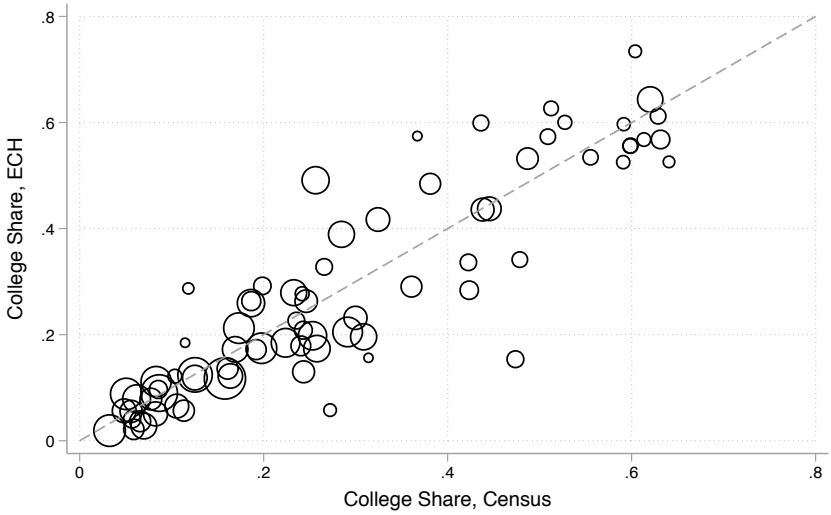
Data is from 1995 Mobility Survey.

(b) Housing Expenditure



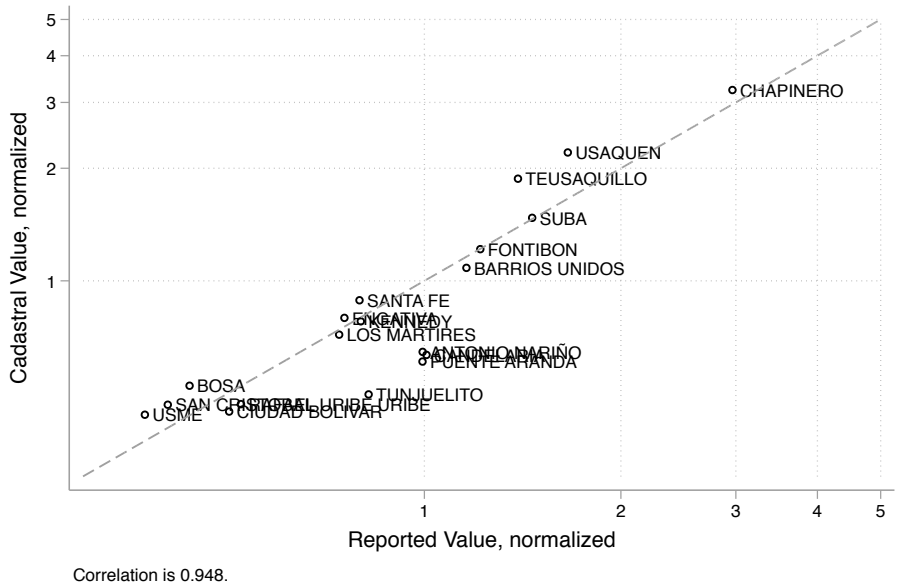
Data covers 2005-2014. Income is predicted from a regression of age bins interacted with (i) education, (ii) occupation and (iii) gender dummies. Housing expenditure includes rents; only renters included.

Figure A.6: College Share in Census vs ECH, 2005



Data covers 2005, observation is a UPZ. Correlation is 0.896

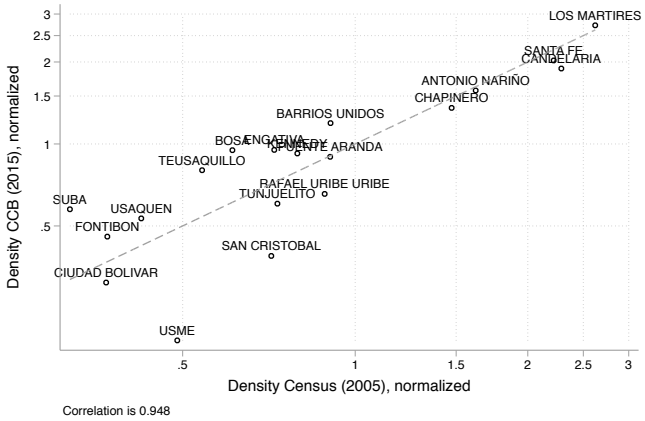
Figure A.7: Cadastral vs Reported Property Values



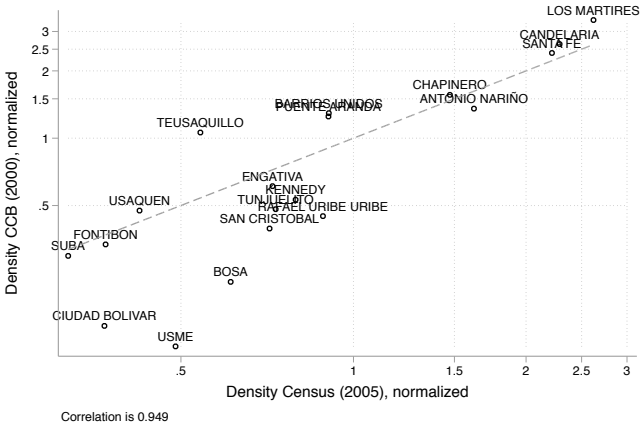
Note: Reported value is the reported purchase price per room as observed in the Multipurpose survey in 2014, for properties bought after 2005 (both the purchase price and year are reported). The cadastral value is the average residential property value per m2 in the locality in that year. Prices are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

Figure A.8: Employment in Chamber of Commerce vs Census

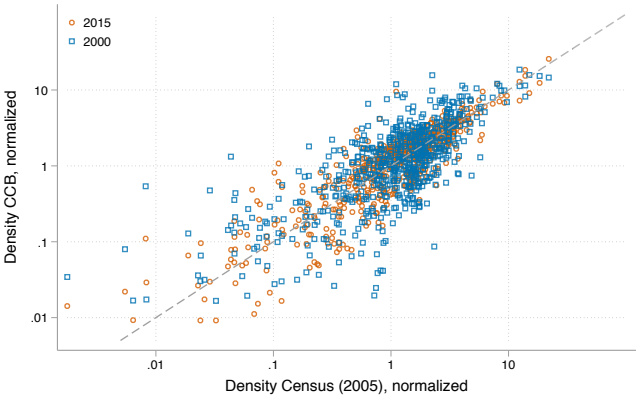
(a) 2015 Establishment Comparison by Locality



(b) 2000 Establishment Comparison by Locality



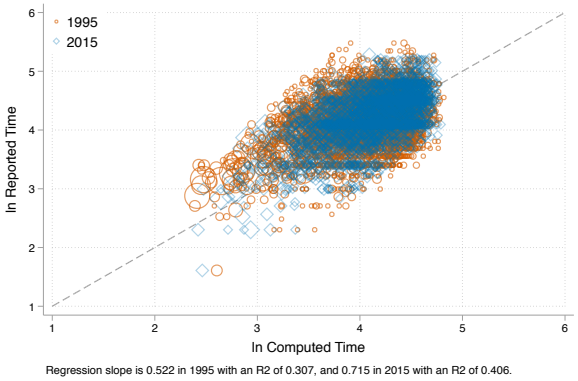
(c) Establishment Comparison by Sector



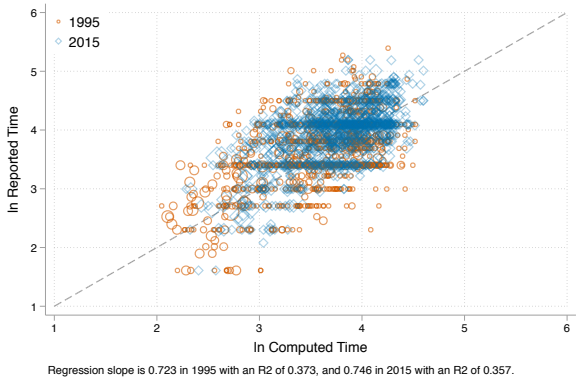
Correlation is 0.901 in 2015, 0.745 in 2000.

Figure A.9: Computed vs Observed Commute Times

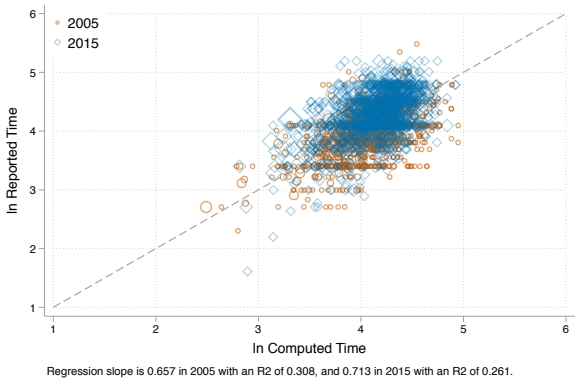
(a) Buses



(b) Cars

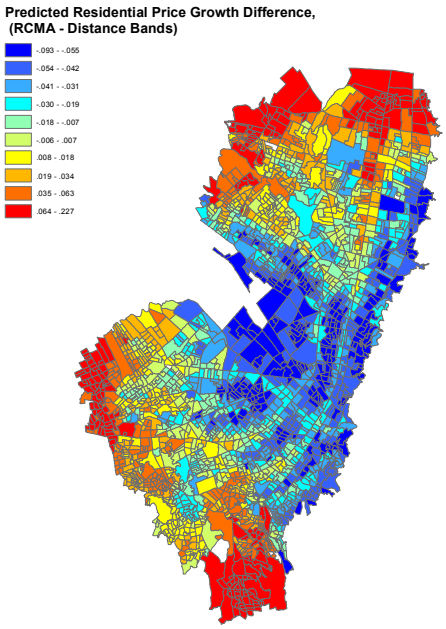


(c) TransMilenio



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).

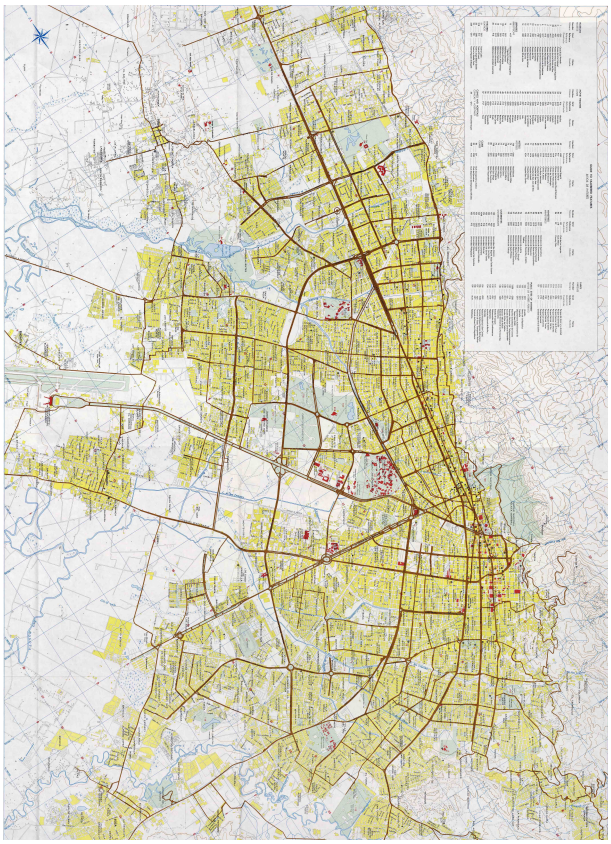
Figure A.10: CMA vs Distance Band Predictions For Floorspace Values



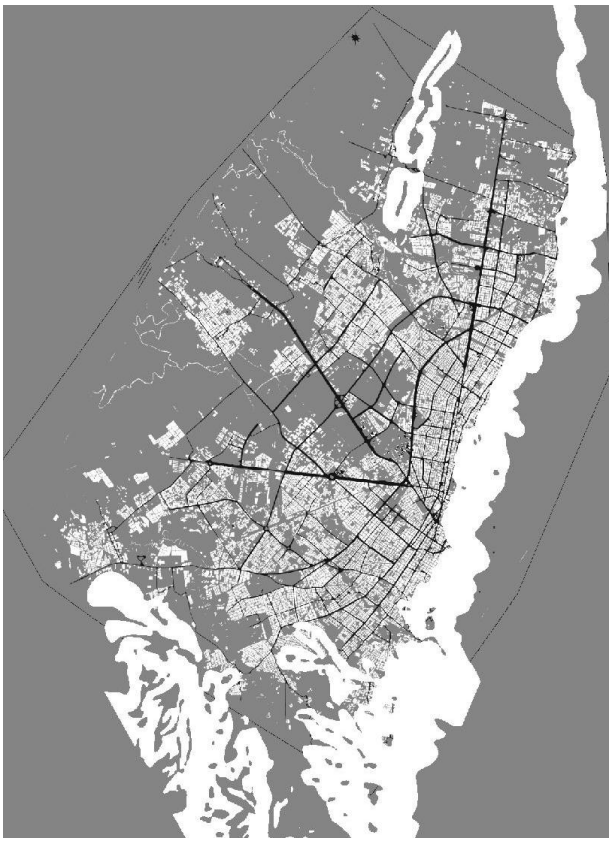
Note: Plot shows the log difference in predicted residential floorspace price growth between the commuter market access specification and the distance band based model. Only tracts within 3km of a TransMilenio station are plotted. The dissimilarity index between the predicted changes, which varies between 0 an 1 with 0 indicating the changes are identical in each location, takes on a value of 0.671.

Figure A.11: Instruments

(a) Raw Land Use Map 1980



(b) Cost Raster



(c) LCP Instrument

(d) Tram Instrument

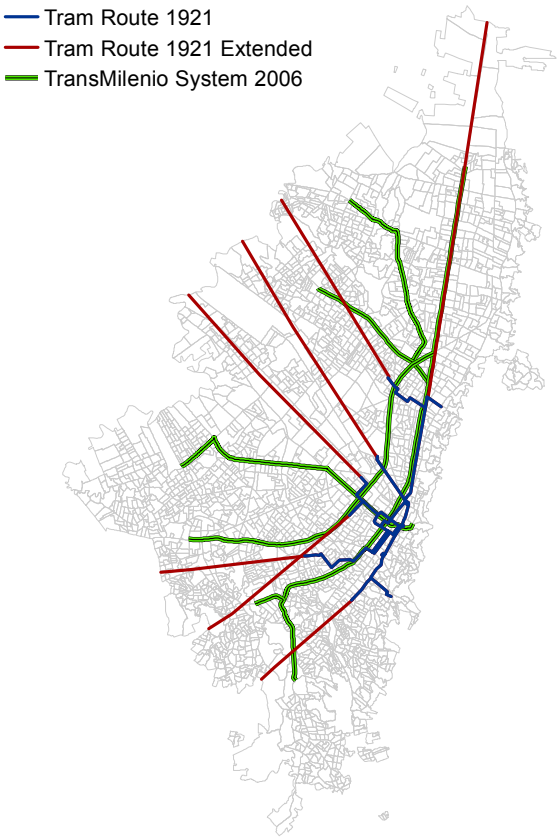
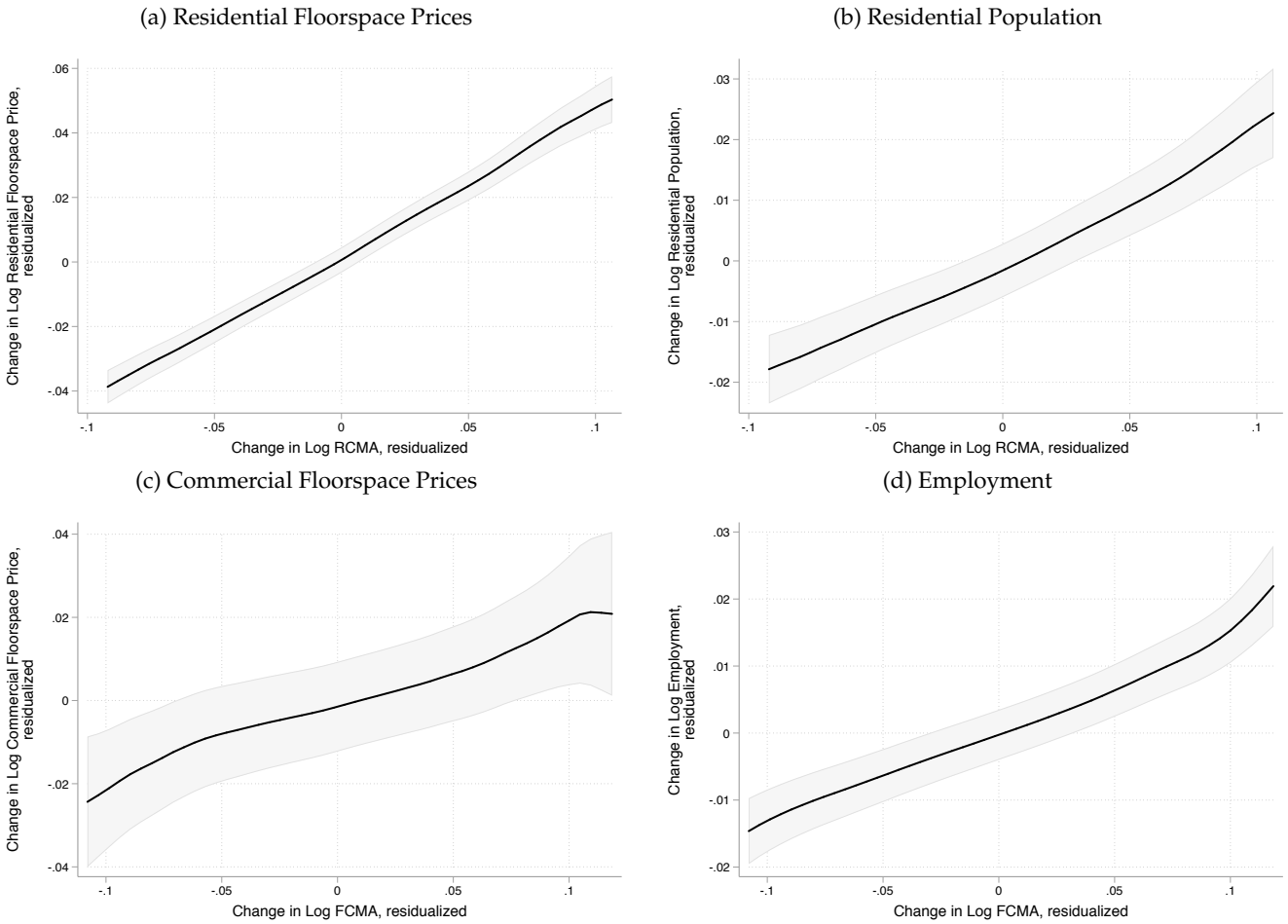


Figure A.12: Monte Carlo: Non-Parametric Relationship Between Outcomes and Commuter Market Access



Note: Plot shows the non-parametric relationship between outcomes and CMA on data simulated from full model with multiple skill groups, industries and transit modes as discussed in Section H.8.

C Reduced Forms from Quantitative Urban Models

This section derives the baseline reduced form from a simple quantitative urban model, and shows this representation is isomorphic to a wider class of models.

C.1 Baseline Model

I now setup a simple model based on Ahlfelt et. al. (2015) and Allen et. al. (2015). Appendix Section C.4.2 shows this is isomorphic to a special case of the model in Section 4 with (i) no non-homotheticities, (ii) no homeownership, (iii) fixed supplies of residential and commercial housing and (iv) one group of workers, firms and transit modes. Notations are extended from the main paper.

Setup. There are $i \in I$ locations that differ in their exogenous amenities $\bar{u}_i$, productivities $\bar{A}_i$, residential/commercial housing supplies H_{Ri}, H_{Fi} and the time t_{ij} it takes to commute to any other location. A continuum of workers with unit mass choose where to live and work and have Cobb-Douglas preferences over a freely-traded numeraire good and housing. Commuting reduces productivity 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

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

where $\epsilon_{ij}(\omega)$ is an idiosyncratic productivity for worker ω on commute (i, j). Assuming these are drawn iid from Frechet distributions with shape parameter θ, the supply of residents and workers to locations is given by

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

$$L_{Fj} = \lambda_U w_j^{\theta} \Phi_{Fj} \tag{24}$$

where $\Phi_{Ri} = \sum_j (w_j/d_{ij})^{\theta}$ is RCMA, $\Phi_{Fj} = \sum_i (u_i r_{Ri}^{\beta-1}/d_{ij})^{\theta}$ is FCMA and λ_U is an equilibrium constant. The Frechet distribution implies that average productivity of workers who have chosen (i, j) is $\bar{\epsilon}_{ij} \propto \pi_{ij}^{-1/\theta}$. The total effective units of labor supplied to a location is therefore $\tilde{L}_{Fj} = \sum_j \pi_{ij}^{\frac{\theta-1}{\theta}}$, which simplifies to

$$\tilde{L}_{Fj} = \lambda_F w_j^{\theta-1} \tilde{\Phi}_{Fj} \tag{25}$$

where $\tilde{\Phi}_{Fj} = \sum_i (u_i r_{Ri}^{\beta-1}/d_{ij})^{\theta-1}$ is adjusted FCMA capturing access to effective units of labor and $\lambda_F = \lambda_U^{\frac{\theta-1}{\theta}}$. There are also spillovers in amenities so that $u_i = \bar{u}_i L_{Ri}^{\mu_U}$.

CMA Measures. Substituting (24) and (23) into the definitions of RCMA and FCMA respectively yields

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

This is the expression from the main text. Appendix Section C.3 establishes existence and uniqueness (to scale) of the system's solution.

Firms. Firms produce using the Cobb-Douglas technology $Y_i = A_i \tilde{L}_{Fi}^{\alpha} H_{Fi}^{1-\alpha}$. Solving their profit maximization problem delivers labor demand

$$\tilde{L}_{Fi} = \frac{1}{\alpha} w_i^{\alpha(1-\sigma)-1} A_i^{\sigma-1} r_{Fi}^{(1-\sigma)(1-\alpha)} E \tag{26}$$

where E is aggregate expenditure. There are spillovers in productivity so that $A_i = \bar{A}_i L_{Fi}^{\mu_U}$.

Market Clearing. Residential and commercial housing market clearing imply that supply must equal demand[68]

$$r_{Ri} = \frac{1-\beta}{H_{Ri}} \Phi_{Ri}^{1/\theta} L_{Ri}^{\frac{\theta-1}{\theta}} \tag{27}$$

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

Labor market clearing pins down the wage by equating (25) and (26). The closed city condition implies the population must sum to one, which pins down the expression for average utility $\bar{U} \propto \left[\sum_{ij} \left(u_i w_j r_{Ri}^{\beta-1} / d_{ij} \right)^{\theta} \right]^{1/\theta}$ (which determines the equilibrium constant λ_U).

C.2 Reduced Form Representation

Stacking equations (23)-(28), taking logs, substituting out for wages and considering long-differences between two time periods (ignoring differences in constants λ_U, λ_F) yields

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

[68]The first line uses properties of the Frechet distribution to derive average income as $\Phi_{Ri}^{1/\theta} L_{Ri}^{-1/\theta}$.

$$= \underbrace{\begin{bmatrix} 1 \\ \frac{1}{\theta} \\ 0 \\ 0 \end{bmatrix}}_{B_R} \Delta \ln \Phi_{Ri} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \frac{\alpha(\sigma-1)}{\theta-1} \\ \frac{1+\alpha(\sigma-1)}{\theta-1} \end{bmatrix}}_{B_F} \Delta \ln \tilde{\Phi}_{Fi} + \underbrace{\begin{bmatrix} \theta \Delta \ln \bar{u}_i \\ -\Delta \ln H_{Ri} \\ (\sigma-1)\Delta \ln \bar{A}_i - \Delta \ln H_{Fi} \\ (\sigma-1)\Delta \ln \bar{A}_i \end{bmatrix}}_{e_i}$$

Pre-multiplying by A^{-1} yields the system

$$\Delta \ln Y_{Ri} = \beta_R \Delta \ln \Phi_{Ri} + e_{Ri} \tag{29}$$

$$\Delta \ln Y_{Fi} = \beta_F \Delta \ln \tilde{\Phi}_{Fi} + e_{Fi}. \tag{30}$$

where β_R is the first two elements of $A^{-1}B_R$ and β_F is the last two elements of $A^{-1}B_F$. The empirical part of the paper approximates $\tilde{\Phi}_{Fj}$ with Φ_{Fj}: formally approximating around $d_{ij}^{-\theta}$ yields $d \ln \tilde{\Phi}_{Fj} = \frac{\theta-1}{\theta} d \ln \Phi_{Fj}$ so that the same log-linear form of (30) is retained when substituting $\tilde{\Phi}_{Fi}$ with Φ_{Fi}. In the data, the log correlation between $\tilde{\Phi}_{Fi}$ and Φ_{Fi} is 0.98 so the results are robust to using either measure.

C.3 Generalizing the Framework

This section shows this reduced form representation and the ability to retrieve measures of market access using only the gravity equation for commuting is shared by a wide class of gravity models.

Proposition A.1. *Consider a model where commute flows are of the "gravity" form*

$$L_{ij} = \gamma_i \delta_j \kappa_{ij}$$

where $\gamma_i, \delta_j > 0$ are endogenous and $\kappa_{ij} > 0$ is exogenous. Then

(i) **Measuring CMA** *The supply of residents and workers to locations are given by $L_{Ri} = \gamma_i \Phi_{Ri}$ and $L_{Fi} = \delta_i \Phi_{Fi}$. Given data $\{L_{Ri}, L_{Fi}\}$ and parameters $\{\kappa_{ij}\}$, the commuter market access terms Φ_{Ri}, Φ_{Fi} are the unique solution to the system*

$$\Phi_{Ri} = \sum_j \frac{L_{Fj}}{\Phi_{Fj}} \kappa_{ij} \quad \text{and} \quad \Phi_{Fi} = \sum_j \frac{L_{Rj}}{\Phi_{Rj}} \kappa_{ji}$$

(ii) **General Gravity Model** *When there is log-linear demand for labor and residents of the form $\tilde{L}_{Fj} = A_j \delta_j^{\alpha}$ and $L_{Ri} = B_i \gamma_i^{\beta} \Phi_{Ri}^{\gamma}$ where $A_i, B_i > 0$ are exogenous constants and the supply of labor (potentially different from the number of workers) is given by $\tilde{L}_{Fj} = \delta_j^{\delta} \tilde{\Phi}_{Fj}$, where $\tilde{\Phi}_{Fj} = \sum_i \gamma_i^{\epsilon} \kappa_{ij}^{\delta} \Phi_{Ri}^{\zeta}$ then an equilibrium always exists and is unique whenever $|\epsilon| + |(\beta-1)\zeta - \epsilon\gamma| \leq |\beta-1||\alpha-1|$. Moreover, the economy has a reduced form representation*

$$\Delta \ln Y_i = B \Delta \ln \Phi_i + e_i$$

where $\Delta \ln Y_i = \begin{bmatrix} \Delta \ln L_{Ri} \\ \Delta \ln \tilde{L}_{Fi} \end{bmatrix}, \Delta \ln \Phi_i = \begin{bmatrix} \Delta \ln \Phi_{Ri} \\ \Delta \ln \tilde{\Phi}_{Fi} \end{bmatrix}, B = \begin{bmatrix} \frac{\gamma-\beta}{1-\beta} & 0 \\ 0 & \frac{\alpha}{\alpha-\delta} \end{bmatrix}$ and e_i is a structural error term containing changes in the exogenous constants.

The gravity equation for commuting that determines the supply side of the model enjoys wide empirical support and is used in the vast majority of recent quantitative urban models.[69] The first part of

[69] See McDonald and McMillen (2010) for a review of the evidence in support of gravity in commute flows; these include all

the proposition shows that unique values of market access can be computed using data on the location of residence and employment, as well as a parameterization of commute costs, using only the supply side of the model through the gravity equation for commute flows. The second part shows that for a class of models with log-linear demand for residents and labor, equilibrium population and employment can be written as log-linear functions of CMA. Note that once the CMA terms have been recovered, the supply curves permit recovery of origin and destination fixed effects which in turn allow computation of $\tilde{\Phi}_{Fj}$.

C.4 Mapping Common Modeling Assumptions to the Framework

This final section maps some common modeling assumptions to the framework from Proposition A.1. After establishing the baseline model from Appendix Section C.1 falls within this class, it considers various extensions and shows the equations that have changed still satisfy Proposition A.1.

C.4.1 Baseline Model

In the model above,

$$L_{ij} = \underbrace{w_j^{\theta}}_{\delta_j} \underbrace{\left(u_i r_{Ri}^{\beta-1}/\bar{U}\right)^{\theta}}_{\gamma_i} \underbrace{d_{ij}^{-\theta}}_{\kappa_{ij}}$$

so that commute flows are of the gravity form. Summing over origins and destinations implies $L_{Ri} = \gamma_i \sum_j \delta_j \kappa_{ij} = \gamma_i \Phi_{Ri}$ and $L_{Fj} = \delta_j \sum_i \gamma_i \kappa_{ij} = \delta_j \Phi_{Fj}$. Effective labor supply is given by $\tilde{L}_{Fj} = \sum_i L_{ij}\bar{\epsilon}_{ij} = w_j^{\theta-1} \sum_i (u_i r_{Ri}^{\beta-1}/\bar{U})^{\theta-1} d_{ij}^{-(\theta-1)}$ which satisfies the restriction with $\epsilon = \delta = \frac{\theta-1}{\theta}$ and $\zeta = 0$.

Substituting the commercial floorspace market clearing condition into the expression for labor demand delivers

$$\tilde{L}_{Fj} = \underbrace{\kappa_3 \bar{A}_j^{\frac{\sigma-1}{1+(\sigma-1)(1-\alpha-\mu_A)}} H_{Fj}^{\frac{(\sigma-1)(1-\alpha)}{1+(\sigma-1)(1-\alpha-\mu_A)}}}_{A_j} \underbrace{w_i^{-\frac{\sigma}{1+(\sigma-1)(1-\alpha-\mu_A)}}}_{\delta_j^{\alpha}}$$

where $\alpha = -\frac{\sigma}{\theta(1+(\sigma-1)(1-\alpha-\mu_A))}$.

Substituting the definition of γ_i into the residential floorspace market clearing condition yields an expression for the demand for residents

$$L_{Ri} = \underbrace{\left(\left(\bar{u}_i/\bar{U}\right)^{\frac{1}{1-\beta}} \frac{H_{Ri}}{1-\beta}\right)^{\frac{\theta(1-\beta)}{(\theta-1)(1-\beta)-\mu_U\theta}}}_{B_i} \underbrace{\gamma_i^{-\frac{1}{(\theta-1)(1-\beta)-\mu_U\theta}}}_{\gamma_i^{\tilde{\beta}}} \underbrace{\Phi_{Ri}^{-\frac{1-\beta}{(\theta-1)(1-\beta)-\mu_U\theta}}}_{\Phi_{Ri}^{\gamma}}$$

Relabelling β as $\tilde{\beta}$ in the proposition, we have $\tilde{\beta} = -\frac{1}{(\theta-1)(1-\beta)-\mu_U\theta}$ and $\gamma = -\frac{1-\beta}{(\theta-1)(1-\beta)-\mu_U\theta}$.

C.4.2 Baseline Model with Separate Resident and Employment Shocks

Consider a special case of the model in Section 4 with one group of workers, firms and transit modes and no non-homotheticities. This is the same as the baseline model above, just allowing for separate

the quantitative models in the literature review.

preference shocks by residential location and productivity shocks by workplace location. In particular,

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

I begin by showing this model yields exactly the same reduced form in terms of the full set of four variables. I then show it satisfies the conditions of Proposition A.1.

Residential and labor supply are given by

$$\begin{aligned} L_{Ri} &= \lambda_U \left(u_i \Phi_{Ri}^{1/\theta} r_{Ri}^{\beta-1}\right)^{\eta} \\ L_{Fj} &= w_j^{\theta} \Phi_{Fj} \end{aligned} \tag{31}$$

where RCMA is as before and $\Phi_{Fj} = \sum_i d_{ij}^{-\theta} \frac{L_{Ri}}{\Phi_{Ri}}$ is FCMA. The expression for average productivity that arises from the Frechet distribution $\bar{\epsilon}_{j|i} \propto \pi_{j|i}^{-1/\theta}$ delivers the total effective units of labor supplied to a location

$$\tilde{L}_{Fj} = w_j^{\theta-1} \tilde{\Phi}_{Fj}$$

where $\tilde{\Phi}_{Fj} = \sum_i d_{ij}^{-(\theta-1)} \frac{L_{Ri}}{\Phi_{Ri}^{\frac{\theta-1}{\theta}}}$ is adjusted FCMA capturing access to effective units of labor. Residential market clearing is given by

$$r_{Ri} = \frac{1-\beta}{H_{Ri}} \Phi_{Ri}^{1/\theta} L_{Ri}. \tag{32}$$

The remaining model equations are unchanged. Thus, stacking (31) and (32) with (26) and (28) and taking first differences yields exactly the same reduced form (29) and (30).

Next I show this model satisfies the conditions of Proposition A.1. Using $L_{ij} = \pi_{j|i} L_{Ri}$ delivers

$$L_{ij} = \underbrace{w_j^{\theta}}_{\delta_j} \underbrace{\lambda_U \left(u_i r_{Ri}^{\beta-1}\right)^{\eta} \Phi_{Ri}^{(\eta-\theta)/\theta}}_{\gamma_i} \underbrace{d_{ij}^{-\theta}}_{\kappa_{ij}}$$

so that commute flows are of the gravity form. Note that summing over origins and destinations implies $L_{Ri} = \gamma_i \sum_j \delta_j \kappa_{ij} = \gamma_i \Phi_{Ri}$ and $L_{Fj} = \delta_j \sum_i \gamma_i \kappa_{ij} = \delta_j \Phi_{Fj}$. Effective labor supply is given by $\tilde{L}_{Fj} = w_j^{\theta-1} \sum_i \frac{L_{Ri}}{\Phi_{Ri}^{\frac{\theta-1}{\theta}}} d_{ij}^{-(\theta-1)}$ which satisfies the restriction with $\delta = \frac{\theta-1}{\theta}$, $\epsilon = 1$ and $\zeta = \frac{1}{\theta}$.

The floorspace market clearing condition delivers the exact same relation as the previous section, so $\alpha = -\frac{\sigma}{\theta(1+(\sigma-1)(1-\alpha-\mu_A))}$.

Finally, substituting the definition of γ_i into the residential floorspace market clearing condition yields an expression for the demand for residents

$$L_{Ri} = \underbrace{\left(\frac{H_{Ri}}{(1-\beta)}\right)^{\frac{1-\beta}{1-\beta-\mu_U}} (\lambda_U \bar{u}_i^{\eta})^{\frac{1}{\eta(1-\beta-\mu_U)}}}_{B_i} \underbrace{\gamma_i^{-\frac{1}{\eta(1-\beta-\mu_U)}}}_{\gamma_i^{\tilde{\beta}}} \underbrace{\Phi_{Ri}^{\frac{\eta\beta-\theta}{\theta\eta(1-\beta-\mu_U)}}}_{\Phi_{Ri}^{\gamma}}$$

Relabelling β as $\tilde{\beta}$ in the proposition, we have $\tilde{\beta} = -\frac{1}{\eta(1-\beta-\mu_U)}$ and $\gamma = \frac{\eta\beta-\theta}{\theta\eta(1-\beta-\mu_U)}$.

C.4.3 Ahlfeldt et. al. (2015)

The model in Ahlfeldt et. al. (2015) with no spillovers across locations, fixed supplies of residential and commercial floorspace (as in Allen et. al. 2015) and productivity rather than preference shocks is the same as the baseline model with $\sigma \to \infty$. I show below the reduced form framework also holds with preference shocks.

C.4.4 Alternate Production Technologies

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 $\theta_F > 1$. As in their model, I assume only labor is used in production so $\alpha = 1$. There is perfect competition so that $p_j(\omega) = w_j/z_j(\omega)$. Goods market clearing implies that sales are given by

$$X_j = \sum_i \frac{A_j w_j^{-\theta_F}}{\sum_s A_s w_s^{-\theta_F}} E_i = A_j w_j^{-\theta_F} P^{\theta_F} E$$

where E_i and P are expenditure from residents and the price index respectively, and $E = \sum_i E$ is aggregate expenditure. Labor market clearing implies all payments are made to workers, so that

$$\tilde{L}_{Fj} = A_j w_j^{-\theta_F - 1} P^{\theta_F} E$$

which is of the form in Proposition A.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. These entrepreneurs have idiosyncratic preferences for producing in each block so that the return from locating in j is given by

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

where $\bar{\sigma} \equiv \sigma/(\sigma - 1)$ is the optimal markup 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

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

Each firm demands the same amount of labor in a location, and CES demand and no fixed costs implies that profits are a constant share of sales $\pi_j = \frac{1}{\sigma} r_j$, which since all costs are paid to labor implies that the wage bill is also proportional to sales

$$w_j \ell_j = \frac{\sigma - 1}{\sigma} r_j.$$

Since total labor demand is simply $\tilde{L}_{Fj} = N_j \ell_j$, we find that

$$\tilde{L}_{Fj} = \lambda_F A_j^{(1+\theta_F)(\sigma-1)} w_j^{-(\sigma+\theta_F(\sigma-1))}$$

where λ_F is an equilibrium constant. Here the mass of firms is fixed and there are positive profits. Equivalently, we could allow for free entry which would ensure zero profits and an endogenous mass of firms. However these are aggregates and thus absorbed into λ_F.

C.4.5 Endogenous Housing Supply

I now show the framework applies to a model with log-linear housing supply. For simplicity, suppose $\alpha = 1$ so that only residents consume land. Suppose that housing is produced using $H_i = T_i^{1-\zeta} K_i^{\zeta}$, where capital is freely traded and land is owned by atomistic land owners. Then each owner, owning one unit of land, produces using $k_i = (\zeta r_{Ri})^{\frac{1}{1-\zeta}}$ and so $h_i = (\zeta r_{Ri})^{\frac{\zeta}{1-\zeta}}$ is housing supply per unit of land and therefore housing supply is given by

$$H_i = T_i (\zeta r_{Ri})^{\frac{\zeta}{1-\zeta}}$$

Equating this with housing demand $(1-\beta)\frac{\Phi_{Ri}^{1/\theta} L_{Ri}}{r_{Ri}}$, and using that $u_i = \bar{u}_i L_{Ri}^{\mu_U}$ and $\gamma_i = \lambda_U \left(u_i r_{Ri}^{\beta-1}\right)^{\eta} \Phi_{Ri}^{(\eta-\theta)/\theta}$, this simplifies to

$$L_{Ri} = \left(\frac{T_{Ri}\zeta^{\frac{\zeta}{1-\zeta}} \bar{u}_i^{\frac{1}{(1-\beta)(1-\gamma)}}}{(1-\beta)}\right)^{\frac{(1-\beta)(1-\zeta)}{(1-\beta)(1-\zeta)-\mu_U}} \gamma_i^{-\frac{1}{\theta((1-\beta)(1-\zeta)-\mu_U)}} \Phi_{Ri}^{\left(\frac{\eta-\theta}{\theta\eta} - \frac{(1-\beta)(1-\zeta)}{\theta}\right)\frac{1}{(1-\beta)(1-\zeta)-\mu_U}}$$

which is of the required form. Intuitively, it is now land rather than housing that acts as a shifter of the resident demand equation.

C.4.6 Leisure

Suppose individuals have Cobb-Douglas preferences over goods, housing and leisure. Allowing for a labor-leisure decision and a joint location decision (for simplicity) yields the problem

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

Solving for commute flows yields

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

where $d_{ij} \equiv \frac{1}{1-t_{ij}}$. Commute flows are therefore still of the gravity form. Agents spend a constant fraction of post-commuting time $1 - t_{ij}$ working, so the rest of the equations are the same (up to scale) as in the baseline model.

C.4.7 Preference Shocks

The baseline model assumes agents have productivity shocks at different workplace locations. If these are preference shocks instead, the exact reduced form system (29) and (30) and conditions of Proposition A.1 no longer hold. However a very similar approximation applies.

Suppose (i) workers have an idiosyncratic preference for each origin-destination pair with shape parameter θ and (ii) commute costs affect utility rather than productivity. The equilibrium equations are given by

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

where $\bar{y}_i = \sum_j \pi_{j|i}(w_j/d_{ij})$ is average income.[70] Income is no longer log proportional to RCMA, which means there are two forcing variables on the residential since after taking logs and stacking as before:

$$A\Delta \ln Y_i = B_{R1}\Delta \ln \Phi_{Ri} + B_{R2}\Delta \ln \bar{y}_i + B_F \Delta \ln \Phi_{Fi} + e_i$$

Approximating $\Delta \ln \bar{y}_i$ around the point $d_{ij}^{-\theta} = 0$ yields $\Delta \ln \bar{y}_i \approx \frac{1}{\theta}\ln \Phi_{Ri}$ and the system returns to

$$A\Delta \ln Y_i = B_R \Delta \ln \Phi_{Ri} + B_F \Delta \ln \Phi_{Fi} + e_i$$

Empirically, I find the log correlation between $\bar{y}_i$ and Φ_{Ri} to be 0.98 so the approximation seems reasonable.

D Data Appendix

This section provides supplementary information on the data used in this paper.

D.1 Dataset Description

Population

The primary source of population data is DANE's General Census of 1993 and 2005. 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 all age groups; the results are robust to considering individuals 18 and older only.[71]

[70] The commuting equation in this model $L_{ij} \propto (u_i w_j r_{Ri}^{\beta-1}/d_{ij})^{\theta}$ retains the gravity form, so the CMA terms can be recovered using the same system of equations as in the paper. Wages and then $\bar{y}_i$ are then solved from $w_j = (L_{Fj}/\Phi_{Fj})^{1/\theta}$.

[71] In the model with spillovers, a requirement is that tracts with positive non-college population have positive college population. If not, amenities would be zero which contradicts the positive non-college population. In the data, this is satisfied in all but 11 of 2799 census tracts in 2005. In quantitative exercises, I add to these tracts enough workers so that their college share is 5% (equivalent to the 3rd percentile). The results are robust to dropping these observations. In estimation, I only use tracts with positive population of both skill groups.

In quantitative exercises, I use data on population in 2015. DANE provides population projections in this year at the UPZ level. To obtain population total by education group at the census tract level in 2015, I merge this with the college share of the UPZ from the GEIH survey. To increase accuracy, I pool GEIH data between 2010 and 2014.[72] In combination with the 2005 census data, this enables me to compute the compute the population growth rate g_u^g of skill group g in UPZ u between 2005 and 2015. I then assume that within each UPZ the growth of high-skilled workers across census tracts is constant so that $L_{i,2015}^H = (1 + g_u^H)L_{i,2005}^H$, where $L_{i,2005}^H$ is a tract's college population total in the 2005 census. The same applies for the calculation of low-skill population.

Comparing the college share by UPZ in the 2005 census with those in the ECH survey (the GEIH's predecessor) suggests this dataset reflects the true demographic composition of each UPZ. Figure A.6 plots the college share from the UPZ in the census (x-axis) with that in the ECH (y-axis): the observations are highly correlated (correlation coefficient 0.896) and lie along the 45-degree line. Importantly, the coverage appears stable across low- and high-college share neighborhoods, as well as across low and high population UPZs (reflected through the size of each marker).

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.[73] I include only trips that originate or end in municipal Bogotá in the analysis.[74] 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.[75] For each trip, the data report the departure time, arrival time, purpose of the trip, mode, as well as origin and destination UPZ.[76] 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).

[72] Of the 112 UPZs with positive population in 2015, 19 are not contained in the GEIH data. These account for only 6.04% of total city population. I assign a college share to these UPZs by taking an average of adjacent UPZs.

[73] Minima-maxima across years are (i) 117,217-169,766 trips, (ii) 58,313-91,765 individuals and (iii) 15,519-28,213 households.

[74] 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)

[75] The 1995 survey reports raw income, while in 2011 and 2015 eight income bin dummies are reported.

[76] 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.

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.[77] 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 2015). 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 2015). 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 2015).[78] 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.

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 (Forero et. al. 2008 ; Ruiz and Vallejo 2015).[79] 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 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,[80] 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.7 plots the av-

[77] I confirmed this comprehensive coverage by overlaying the shapefile of plots with data over satellite images.

[78] 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).

[79] Updated assessments and property transaction records were conducted throughout, with assessments for each homogenous zone being updated during the 2008-2009 comprehensive update.

[80] The results are not sensitive to this choice.

erage 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.[81] 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.16 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 vast majority of these.[82] 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).[83] 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.8 plots the density of establishments in each locality in

[81] 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.

[82] 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.

[83] 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.

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.

One final issue is that of household employment of domestic services. Employment in domestic services, such as maids, cooks and cleaners, is an important sector for low-skilled in Bogotá: between 2000-2014, 7.3% of non-college educated Bogotanos worked as domestic helpers while almost no college educated workers did.[84] However, employment of domestic help by households is not captured in either the census' or CCB data since, in contrast to other types of (often informal) household enterprises, this was not interpreted by DANE as constituting the household as an economic establishment.[85] The 2014 Multipurpose Survey reports whether households employ domestic services: I find that 30.% of college-educated households do, compared to only 3.6% of non-college households. I therefore impute the spatial distribution of domestic services employment by assuming that the total employment of domestic services in a given year (observed from the worker-level ECH/GEIH datasets) is spread evenly over college-educated households. I only include this data in the counterfactual with domestic employment.

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.[86] 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

[84] As seen in Table 1 in the paper, 91.% of domestic helpers are low-skill.

[85] I confirmed this by comparing the share of city-wide for each 1-digit industry as reported by workers in the 2005 ECH with those reported by establishments in the 2005 census. While the two were very highly correlated for all other industries (with a correlation coefficient of 0.918), there was no employment in domestic services reported in the census.

[86] Public versions provide no additional geographic information within the city

current routes to measure the location of the bus network throughout the period.[87] 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.[88] 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.

D.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.[89] 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 F.3 provided evidence that speeds were not changing on routes affected by TransMilenio relative to other locations, Table A.11 shows that aggregate speeds were not quite constant over the

[87] 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.

[88] 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.

[89] 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.

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 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 Review Board 2002). 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.[90]

Figure A.9 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.17, 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.9 suggests this is driven by a drop in the intercept for 2015 caused by movements in the left tail: overall the majority of points lie along the 45-degree line in both years.[91] 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.

[90] 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.

[91] Attempts to shift the intercept by varying the fixed time cost within reasonable bounds had negligible effects on this specification.

D.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 $4mn 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 $200mn to build a tunnel.[92] The maximum grade of BRT is 10% for short runs (American Public Transportation Association 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.[93] 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. With the resulting shapefile, I then compute in ArcMap the least cost times to commute via this instrument by assigning the same speeds to trunk and feeder lines as in the main calculations.

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.[94] 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.

E Additional Information on Commuting and TransMilenio

Trip Characteristics Table A.10 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,

[92] 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).

[93] 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.

[94] 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.

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 compared which far exceeds the 6.6km and 6.1km average trips taken by other motorized transport. Given the fixed costs involved in reaching and entering stations, the benefits of BRT are particularly pronounced for longer journeys. Third, when compared to other modes we see that TransMilenio is primarily used for trips to work - constituting 21.5% of commutes - than for more leisure-related activities such as trips for private matters or shopping. For these purposes, 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.11 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. It appears 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.[95] Interestingly, aggregate speeds on cars and buses is uncorrelated with the 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.

Finally, one might wonder whether given TransMilenio is more likely to be used by the poor given the sorting across transit modes previously documents. Table A.12 shows that while the poorest Bogotanos are significantly more likely to use TransMilenio than the rich, the difference is entirely explained by the fact that they are less likely to own cars. Consistent with the similar fares charged by TransMilenio and traditional buses, the principal monetary trade-off across modes remains between cars and public transit.

Construction and Operating Costs Phase 1 of the system cost $5.8mm per km to build in 2005 dollars. 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 $13.23mm per km, with funding coming from the national government (66%) and a local fuel surcharge (34%). The higher costs were due to road widening, increased investment in public space and associated infrastructure improvements.[96] Overall, the average cost to construct both phases was therefore $12.23mm in 2016 dollars across 93km of lines.

Operating costs are recovered at the farebox by private operators; the cost to transport a passenger is close to the fare (Transportation Review Board 2002). Using the figure of 565mn rides in 2013 from BRT Data (2017) and the fare of $0.66 in 2016 dollars yields an operational cost of $372.97mn per year. Deflating this by the share of the network accounted for by phases 1 and 2 gives a final operational cost of $309.69mn per year in 2016 dollars.

[95] 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.11 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.1 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.

[96] All figures from Cain et. al. (2006).

F Supplementary Empirical Results

F.1 Regressions of Relative Speed over Time

Table A.13 compares speeds on buses and TransMilenio versus cars in each year of the Mobility Survey (where all modes are available, see paper for comparison in 1995) controlling for the characteristics and composition of trips. Columns (1)-(3) control for hour of departure fixed effects and demographic controls, and for the most part are qualitatively similar to the average results in Table A.11. However, it is possible a portion of these differences are due to the different composition of commutes across modes. Columns (4)-(6) therefore include origin-destination pair fixed effects. While the speed difference for buses vs cars is relatively unchanged, the relative speed of TransMilenio drops a lot. This reflects the fact that the nature of TransMilenio is indeed very different - they are much longer trips, which tend to be faster - and once we control for this TransMilenio trips appear on average 8.1% slower than car trips across the sample. While still a pronounced improvement over buses, which are on average 25.3% slower than cars (including observations from 1995), the difference is substantially less than the aggregate figures imply.

F.2 Effect of TransMilenio on Growth in Floorspace

In Table A.14, I provide evidence of TransMilenio's muted effect on new housing development. To begin with, I regress the growth in a block's floorspace between 2000 and 2013 on the distance to the closest TransMilenio station.[97] I include separate measures for each phase of the TransMilenio system, to explore whether the effect was different across phases. In all specifications, I include locality fixed effects to control flexible for trends in construction across different areas of the city. I also use the CMA measure to repeat the baseline specifications from the main paper for this outcome.

Column (1) presents the baseline result. We see that the growth in building floorspace is greater far from TransMilenio stations. Of course TransMilenio stations were placed in dense, built-up areas, so column (2) controls for a number of block characteristics such as its population density in 1993, initial floor area ratio, distance to the nearest main road as well as distance to the CBD (which is allowed to have different effects based on whether the block is in the North, West or South of the city). The effects of proximity to TransMilenio become for the most part insignificant. Overall, there was not much new development close to TransMilenio stations.

Reports suggest that constraints to re-development restricted the supply response, but vacant parcels close to stations were in fact more likely to get developed (Cervero et. al. 2013). Columns (3) and (4) tests this by examining whether the effect of proximity to TransMilenio on the growth in floorspace was heterogeneous across vacant and non-vacant blocks. We can see that this was in fact that case: while vacant tracts were more likely to experience a growth in floorspace overall, vacant blocks close to TransMilenio were much more likely to get developed than those far away. For the most part, this effect was stronger towards the later phases of the system. However, since only a small proportion of land near stations was vacant, this suggests that the overall effect of TransMilenio on new construction was small.

[97] 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 blocks with no development in 2000, although the results are similar if I measure the log change in floorspace (adding a small number to include blocks with no construction).

Finally, columns (5) and (6) repeat the baseline specifications from the paper and show there is no effect of changing CMA on the supply of floorspace.

F.3 Effect of TransMilenio on Other Mode Speeds

In this section, I provide evidence that perhaps surprisingly TransMilenio seemed to not have significant effects on the speeds of other modes. To do so, 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 commutes by car and by bus, where (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,[98] and X_{ijkt} is a vector of control variables containing individual and trip characteristics, which are allowed to have time-varying effects on speeds. In all specifications these controls include a gender dummy, hour of departure dummies and age quantile dummies, each interacted with the Post dummy. In certain columns, these include origin locality fixed effects, destination locality fixed effects, and log trip distance, all interacted with the Post dummy.

The variable TM Route_{ij} captures whether the trip from i to j has been "treated" by TransMilenio and is defined in two ways. To construct this I compute the routes for the least cost commutes between each pair of UPZ origin and destination in ArcMap 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.

Panel A in Table A.15 presents the results for car trips. In column 1, we see that increasing the share of a trip lying along TransMilenio from 0 to 1 reduces car speed by 16%. However, this may well reflect differences in trip composition given that TransMilenio trips are longer and typically go through the outskirts to the city center. Column 2 includes locality origin and locality destination fixed effects (interacted with the Post dummy) to control for trends in speeds for different types of trips, and we see that the point estimate falls by about 40% and is no longer significant. When we control for the fact that speeds for long trips may have been trending differently than slow trips in column 3, the coefficient halves once more. All in all, there is no significant effect on driving speeds once we control for differences in trip composition on TransMilenio routes. Finally, columns 4-6 repeat the exercise with the treatment measure equal to one if more than 75% of the trip is adjacent to TransMilenio, and there is no significant difference in any specification. It even looks like speeds may have increased slightly when using this measure.

Panel B repeats the same set of regressions for trips taken by bus. The results are qualitatively similar to those for cars.[99]

[98] Results are similar when intermediate years are included, and are omitted for clarity.

[99] Finally, note that this only identifies whether speeds along treated routes changed relative to those in other locations in the city. As with as difference-in-difference analysis, it cannot identify whether TransMilenio had an effect on the overall level of speeds in the city. Indeed, if we think of removing the system as a whole, there is reason to think speeds might slow since the system constitutes more than 2.2 trips per day. This would suggest my structural results underestimate the true effect of TransMilenio, since in the baseline estimates I keep the speeds on other routes constant.

F.4 Engel Curve for Housing

In Figure A.5, I plot the relationship between the share of income spent on housing and average labor income on average between 2005-2014. Data comes from the GEIH. Income is defined as all labor income received by the household in the month of the survey, and housing expenditure is defined as monthly rents (only renters are included). This relationship might display a mechanical negative relationship were I to compare the raw variables if monthly household income is volatile. To address this, I predict worker income from a regression of log income on age bin dummies interacted with (i) education, (ii) occupation and (iii) gender dummies, as well as year fixed effects, and construct predicted household income by summing the fitted values over working household members. The adjusted relationship in Figure A.5 is indeed flatter than that in the raw data (which ranges from over 0.6 at the 10th percentile to below 0.2 at the 90th), but still displays a significant non-homotheticity in house expenditures: Bogotanos at the bottom 10th percentile of the income distribution spend over 50% of income on housing, whereas those at the 90th percentile spend just under one quarter.

Two comments are in order. First, the figure considers renters only. The survey also asks homeowners to report (i) monthly home payments (i.e. mortgages) and (ii) estimated rents. When I include home owners and produce separate plots for (i) renters plus home payments by home owners and (ii) renters plus estimated rents by home owners, the resultant Engel curve is essentially unchanged. Second, I use labor income rather than total income. But since non-labor income comprises a greater share of wealth for the rich, accounting for this would only increase the slope of the Engel curve.

While this result may seem at odds with evidence of constant expenditure shares in the US based on inter-city data (e.g. Davis and Ortalo-Magné 2011), recent evidence using within-city data has documented a downward sloping Engel curve (Ganong and Shoag 2014). Moreover, the middle-income country setting of Colombia may well be different than the US.

G Supplementary Theoretical Results

G.1 Mode Choice Problem

In the paper, car owners and non-car owners face a different distribution of commute times. In this section, I show how this is derived from a discrete choice problem in which individuals decide which mode to use to commute to work having already decided where to live, where to work and whether or not to own a car.

For a worker of type-g, conditional on having made the choice (i, j, a), in the third stage they choose which mode to use to commute to work $m \in \mathcal{M}_a$ to maximize utility

$$U_{ijamg}(\omega) = u_{iag}\left(\frac{w_{jg}\epsilon_j(\omega)}{d_{ijm}(\omega)} - p_a a - r_{Ri}\bar{h} + \pi\right) r_{Ri}^{\beta-1}\nu_i(\omega) \text{ where } d_{ijm}(\omega) = \exp\left(\kappa t_{ijm} + \upsilon_{ijm}(\omega)\right).$$

Linearity of this expression implies that the mode choice problem reduces to $\min_m \{d_{ijm}(\omega)\}$.

Following the precedent in the transportation literature (e.g. McFadden (1974)), I assume that transit modes are contained within two nests: $\mathcal{B}_{Pub} \equiv \{\text{Walk}, \text{Bus}, \text{TransMilenio}\}$ is the nest of public modes while $\mathcal{B}_{Priv} \equiv \{\text{Car}\}$ is the nest of private modes. Omitting the i, j subscript for brevity, individual ω has

idiosyncratic preferences over each mode $\upsilon_m(\omega)$ drawn from a GEV distribution[100]

$$F(\upsilon_1, \ldots, \upsilon_N) = 1 - \exp\left(-\sum_k \left(\sum_{m \in \mathcal{B}_k} \exp\left((\upsilon_{m\omega} - \tilde{b}_m)/\lambda_k\right)\right)^{\lambda_k}\right) \quad \text{where } k \in \{\text{Public}, \text{Private}\}$$

This is a Gumbel distribution for minima allowing for correlation of preference shocks within nests, with $\lambda_k \to 0$ being the case of perfect correlation. Note that $\lambda_{Priv} = 1$ by virtue of there being only one mode within the nest of private modes. The parameters $\tilde{b}_m$ control the mean preference for mode m, reflecting that all else equal some modes may be more pleasant to use to commute than others. Finally, note that the choice between nests only applies to individuals who own cars: those without cars can only choose between public modes of transit.

Standard results imply that the choice probabilities are given by

$$\begin{aligned}
\pi_{m|ija} &= \pi_{k|ija} \times \pi_{m|ijka} \\
&= \frac{\left(\sum_{n \in \mathcal{B}_k} \exp\left(b_n - \frac{\kappa}{\lambda_k} t_{ijn}\right)\right)^{\lambda_k}}{\sum_{k'} \left(\sum_{n \in \mathcal{B}_{k'}} \exp\left(b_n - \frac{\kappa}{\lambda_{k'}} t_{ijn}\right)\right)^{\lambda_{k'}}} \times \frac{\exp\left(b_m - \frac{\kappa}{\lambda_k} t_{ijm}\right)}{\sum_{n \in \mathcal{B}_k} \exp\left(b_n - \frac{\kappa}{\lambda_k} t_{ijn}\right)}
\end{aligned} \tag{33}$$

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.

As for the employment and residential preferences, I assume that the mode-specific preference shocks are only realized after other choices have been made. To compute expected utility prior to drawing these shocks, it remains to solve for $E\left[\max_{m \in \mathcal{M}_a} \{1/d_{ijm}(\omega)\}\right] \equiv 1/\bar{d}_{ija}$. Calculating this expectation using the GEV distribution above yields

$$\begin{aligned}
\bar{d}_{ija} &= \exp\left(\kappa \bar{t}_{ija}\right) \\
\text{where } \bar{t}_{ij0} &= -\frac{\lambda}{\kappa} \ln \sum_{m \in \mathcal{B}_{\text{Public}}} \exp\left(b_m - \frac{\kappa}{\lambda} t_{ijm}\right) \\
\bar{t}_{ij1} &= -\frac{1}{\kappa} \ln\left(\exp(b_{car} - \kappa t_{ij\text{Car}}) + \exp\left(\kappa \bar{t}_{ij0}\right)\right)
\end{aligned} \tag{34}$$

Intuitively, the expected commute cost can be expressed as the commute costs of an average commute time across the modes available to the individual. Expected utility is therefore

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

which is the expression in the paper.[101]

[100] An alternative interpretation is that there are a continuum of days for which each individual decides how to commute.

[101] For each car ownership $a \in \{0, 1\}$ before and after TransMilenio is introduced, I normalize the set of shifters so that there is no time cost to commuting to the same origin-destination (i.e. $\bar{t}_{iia} = 0 \; \forall i, a$). This ensures the option value of a larger choice set is not baked into car ownership through the greater cardinality of choices available.

G.2 First Order Welfare Approximations

This section derives the welfare approximations used in Section 9.1 of the paper, other than the result established in Proposition 2.

VTTS In the travel mode choice models used in the transportation literature (e.g. Train and McFadden 1978, reviewed in Small and Verhoef 2007) an individual n chooses across modes m for a trip with indirect utility

$$V_{nm} = \alpha \frac{c_m}{w_n} + \kappa t_m + \delta_m + \epsilon_{mn}$$

When the ϵ_{mn} are T1EV and workers have a constant wage, then aggregating across commutes, in response to a system of small changes $\{dt_m\}$ the change in consumer surplus is given by $dCS = \sum_{ijm} L_{ijm} \times \frac{1}{\lambda} dE\left[\max\{V_m\}\right] = -\frac{\kappa}{\alpha} \sum_m w L_{ijm} dt_m$ where λ is the marginal utility of income. This is the value of travel time savings result. This is measured in dollars; to convert into percentage changes in utility for comparison with my model, use that this model implies the change in average utility is[102]

$$d \ln \bar{U} = \kappa \sum_m \pi_{ijm} dt_m.$$

This is very similar to the expression used in Proposition 2.

First Order Approximation Suppose individuals of group g living in i have indirect utility $V_g(r_{Ri}, y_{igm})$ over house prices and income. Define aggregate welfare to be a population-weighted average of these, $V_g \equiv \sum_{i,m} \pi_{igm} V_g(r_{Ri}, y_{igm})$. Then a first order approximation to a shock to RCMA (ignoring indirect effects of shocks to other locations) yields

$$\begin{aligned}
dV_g &= \sum_{i,m} \pi_{igm} \left[\frac{\partial V_g}{\partial y_{igm}} \frac{\partial y_{igm}}{\partial \ln \Phi_{Rigm}} + \frac{\partial V_g}{\partial r_{Ri}} \frac{\partial r_{Ri}}{\partial \ln \Phi_{Rigm}} \right] d \ln \Phi_{Rigm} \\
\Leftrightarrow d \ln V_g &= \sum_{i,m} \pi_{igm} \left[\frac{\partial \ln V_g}{\partial \ln y_{igm}} \frac{\partial \ln y_{igm}}{\partial \ln \Phi_{Rigm}} - \frac{\partial \ln V_g}{\partial \ln y_{igm}} \beta_{img} \frac{\partial \ln r_{Ri}}{\partial \ln \Phi_{Rigm}} \right] d \ln \Phi_{Rigm} \\
\Leftrightarrow d \ln \bar{U}_g^{FPA} \equiv \frac{d \ln V_g}{\lambda_g} &= \sum_{i,m} \pi_{igm} \left[\varepsilon^Y - \beta_{img} \varepsilon^R \right] d \ln \Phi_{Rigm}
\end{aligned}$$

where the second line uses Shephard's lemma, $\beta_{img} \equiv r_{Ri} h_{img} / y_{igm}$ is the expenditure share on housing and $\lambda_g \equiv \frac{\partial \ln V_g}{\partial \ln y_{igm}}$ is the marginal utility of income.

H Supplementary Quantitative Results

H.1 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.

[102] To derive this, take the monotonic transformation $U_{nm} = \exp(V_{nm})$ and differentiate $\bar{U} = E\left[\max\{U_m\}\right]$ to the small shocks $\{dt_m\}$

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 I.

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,[103] $\hat{\omega}_H = 0.3075$ 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 in 2012 I obtain $T_H = 2.016$, $\bar{h} = 1.2097$ and $p_a = 117.37$ (with $P = 0.025$).

H.2 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 $w^0, \vartheta^0, r^0, u^0, A^0$
2. Given a wage vector $w^t, \vartheta^t, r^t, u^t, A^t$
   (a) Compute $H_{Ri}^t = \vartheta_i^t H_i$, $H_{Fi}^t = (1-\vartheta_i^t) H_i$, $\Phi_{Riag}^t = \sum_j (w_{jg}^t / d_{ija})^{\theta_g}$ and $W_{is}^t = \left(\sum_h \alpha_{sh}^{\sigma_L} (w_{ih}^t)^{1-\sigma_L}\right)^{\frac{1}{1-\sigma_L}}$.
   (b) Compute $P_t = \left(\sum_{j,s} \left(((W_{js}^t)^\alpha (r_{Fj}^t)^{1-\alpha} / A_{js}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$, where $r_{Fj}^t = (1 - \tau_i) r_i^t$
   (c) Compute L_R^t from

$$L_{Riag}^t = \bar{L}_g \frac{\left(u_{iag}^t (T_g \Phi_{Riag}^{1/\theta} - \bar{h} r_{Ri}^t - p_a^t a + \pi^t) r_{Ri}^{\beta-1}\right)^{\eta_g}}{\sum_{r,o} \left(u_{rog}^t (T_g \Phi_{Rrog}^{1/\theta} - \bar{h} r_{Rr} - p_o^t o + \pi^t) r_{Rr}^{\beta-1}\right)^{\eta_g}}$$

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

   (d) Compute labor supply $\tilde{L}_{Fig}^t = (w_{jg}^t)^{\theta_g - 1} \Psi_{jg}^t$, where $\Psi_{jg}^t \equiv T_g \sum_{r,o} (\Phi_{Riog}^t)^{-\frac{\theta_g - 1}{\theta_g}} d_{rjo}^{-\theta_g} L_{Rrog}^t$.

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

(e) Update the main variables

$$\tilde{w}_{jg} = \left[\frac{(P^t)^{\sigma-1} X^t \sum_s B_{isg} A_{is}^{\sigma-1} (W_{is}^t)^{\sigma_L - (1+\alpha_s(\sigma-1))} (r_{Fi}^t)^{-(1-\alpha_s)(\sigma-1)}}{\Psi_{jg}^t} \right]^{\frac{1}{\theta_g + \sigma_L - 1}}$$

$$\tilde{r}_i = \frac{E_i^t + (1-\alpha) Y_i^t}{H_i}$$

$$\tilde{\vartheta}_i = \begin{cases} 1 & i \in \mathcal{D}_R \backslash \mathcal{D}_F \\ 0 & i \in \mathcal{D}_F \backslash \mathcal{D}_R \\ \frac{E_i^t}{E_i^t + (1-\alpha) Y_i^t} & i \in \mathcal{D}_F \cap \mathcal{D}_R \end{cases}$$

$$\tilde{A}_{js} = \bar{A}_{js} (\tilde{L}_{Fj}^t / T_j)^{\mu_A}$$

$$\tilde{u}_{iag} = \bar{u}_{iag} \left(L_{RiH}^t / L_{Ri}^t \right)^{\mu_U}$$

where $X^t = \beta \sum_{i,g,a} (T_g (\Phi_{Riag}^t)^{1/\theta_g} - p_a^t a - r_i^t \bar{h} + \pi^t) L_{Riag}$ is aggregate expenditure on goods, $Y_i^t = \sum_s (p_{is}^t)^{1-\sigma} (A_{js} P^t)^{\sigma-1} X^t$ is firm sales in i and $E_i^t = r_i^t \bar{h} L_{Ri}^t + (1-\beta) \sum_{a,g} (T_g (\Phi_{Riag}^t)^{1/\theta_g} - p_a^t a - r_i^t \bar{h} + \pi^t) L_{Riag}^t$ 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.

H.3 Welfare Decomposition

This section provides a description of the DEK method used to compute the welfare from time savings in Appendix Table A.8. Expected income of workers of type-g commuting to j from (i,a) is given by $\tilde{y}_{iag} = \bar{y}_{iag} - r_i \bar{h} - p_a a + \pi$. Holding all choices and prices fixed, following the change in commute costs $\hat{d}_{ija} \equiv d'_{ija}/d_{ija}$ the new gross income for commuters to j is $\bar{y}_{iag} \frac{1}{\hat{d}_{ija}}$. The new gross income for residents of i is then

$$\bar{y}_{iag} \frac{1}{\hat{\bar{d}}_{ia}}, \quad \text{where } \frac{1}{\hat{\bar{d}}_{ia}} \equiv \sum_j \pi_{j|iag} \frac{1}{\hat{d}_{ija}}$$

The change in net income for residents of i is then

$$\hat{\tilde{y}}_{iag} = \frac{\bar{y}_{iag} \frac{1}{\hat{\bar{d}}_{ia}} - r_i \bar{h} - p_a a + \pi}{\tilde{y}_{iag}}$$

The change in average utility is then $\hat{\tilde{U}} = \sum_{i,a} \pi_{ia|g} \hat{\tilde{y}}_{iag}$.

H.4 Model with Preference Shocks and Joint Live-Work Decision

This section outlines the extension of the model where workers have preference rather than productivity shocks across employment locations, and make all choices simultaneously.

Workers solve the problem

$$\max_{(i,j,a),C_i(\omega),H_{Ri}(\omega)} \frac{u_{iag}C_i(\omega)^{\beta}(H_{Ri}(\omega)-\bar{h})^{1-\beta}}{d_{ija}}\epsilon_{ija}(\omega)$$
$$\text{subject to } \; C_i(\omega)+r_{Ri}H_{Ri}(\omega)+p_a a = w_{jg}$$

In contrast to the baseline model, workers have a joint preference shock over each choice (i,j,a). Substituting in goods and housing demand, this reduces to

$$\max_{(i,j,a)} U_{ijag}(\omega) = \frac{u_{iag}\tilde{y}_{ijag}r_{Ri}^{\beta-1}}{d_{ija}}\epsilon_{ija}(\omega)$$
$$\text{where } \tilde{y}_{ijag} \equiv w_{jg}-p_a a - r_{Ri}\bar{h}$$

Define $\mathcal{A}_g = \left\{(i,j,a) : \bar{u}_{iag} > 0,\ r_{Ri} < (w_{jg}-p_a a)/\bar{h}\right\}$ to be the set of active commutes that are both desirable and affordable. Then assuming ϵ_{ija} are drawn from a Frechet distribution with shape θ_g,

$$L_{ijag} = \begin{cases} \lambda_{U,g}\left(u_{iag}\tilde{y}_{ijag}r_{Ri}^{\beta-1}/d_{ija}\right)^{\theta_g} & \forall (i,j,a)\in\mathcal{A}_{Rg} \\ 0 & \text{otherwise} \end{cases}$$

We then get that resident and labor supply are given by

$$L_{Riag} = \begin{cases} \lambda_{U,g}\left(u_{iag}\Phi_{Riag}^{1/\theta_g}r_{Ri}^{\beta-1}\right)^{\theta_g} & \forall (i,a) : (i,a)\in\mathcal{A}_g \\ 0 & \text{otherwise} \end{cases}$$
$$L_{Fjg} = \lambda_{U,g}w_{jg}^{\theta}\Phi_{Fjg}$$

where

$$\Phi_{Riag} = \sum_{j:(i,j,a)\in\mathcal{A}_g} (\tilde{y}_{ijag}/d_{ija})^{\theta_g}$$
$$\Phi_{Fjg} = \sum_{ia\in\mathcal{A}_{Rg}} d_{ija}^{-\theta}\frac{L_{Riag}}{\Phi_{Riag}}\left(\frac{\tilde{y}_{ijag}}{w_{jg}}\right)^{\theta_g}$$

are RCMA and FCMA respectively. L_{Fjg} is the number of workers who commute to j; $\tilde{L}_{Fjg} = T_g L_{Fjg}$ is total effective labor supplied. Average income is

$$\bar{y}_{iag} = \sum_j \pi_{j|iag}w_{jg} \;\text{ where }\; \pi_{j|iag} = \frac{(\tilde{y}_{ijag}/d_{ija})^{\theta_g}}{\sum_s(\tilde{y}_{isag}/d_{isa})^{\theta_g}}.$$

Worker welfare is $\bar{U}_g = \gamma_{\theta,g}\left[\sum_{i,a}\left(u_{iag}\tilde{y}_{iag}r_{Ri}^{\beta-1}\right)^{\theta_g}\right]^{1/\theta_g}$. The remaining expressions in the model are

unchanged.

H.5 Model with 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

$$\beta_D E_{iH} = p_i^D D_i = \frac{w_{Li}\tilde{L}_{FiL}^D}{\bar{A}_{Di}}$$

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.

H.6 Model with 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.

H.7 Model with Variable Housing Supply

This section outlines the extension of the model allowing for a housing supply response to the transit infrastructure.

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 $p_i = \tilde{\eta} r_i^{\frac{1}{1-\eta}}$.

These results imply that total housing supply is given by $H_i = T_i(\eta r_i)^{\frac{\eta}{1-\eta}}$. The system of equations in this model is identical to that used in the paper, with one additional equation determining the supply of housing given its price in each location. 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 either as quality of land, or a distortion faced by developers (so that revenues are $\zeta_i^{\eta/(1-\eta)} r_i$).

In the Land Value Capture scheme, there are constraints on building densities in each location so that

$$H_i^S = \begin{cases} T_i(\eta r_i)^{\frac{\eta}{1-\eta}} & \text{if } T_i(\eta r_i)^{\frac{\eta}{1-\eta}} \leq \bar{H}_i \\ \bar{H}_i & \text{otherwise} \end{cases}$$

The government increases $\bar{H}_i$ in some locations to $\bar{H}'$, shifting this supply curve. Perfect competition ensures the price of the permits adjust so that that developers earn zero profits, so income from the scheme is $(\bar{H}' - \bar{H}_i)r_i'$ where prices are evaluated in the new equilibrium. When considering the counterfactual, I assume that ζ_i wedges remain the same so that changes in housing supply are due only to the change in transit.

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).

H.8 Monte Carlo: Single-Group Regressions on Multiple-Group Model

In this section, I provide evidence the reduced form regressions in the paper derived from the special case of the model with one group of workers, firms and commute modes are consistent with the full model with multiple layers of heterogeneity.

The key benefit to the regression framework used as a model validation exercise in the paper is the transparent manner in which it can be taken to the data. An alternative would be to log-linearize the equilibrium equations from the full model, but this would deliver more complicated specifications.[104] To show the regression framework from the simple model is consistent with the full model, I run a Monte Carlo exercise in which I simulate data from the full model, run the regression specifications from the simple model on this simulated data, and show the log-linear non-parametric relationships hold. To construct the simulated data, I first use the data and unobservables from the post equilibrium in 2012. I then remove TransMilenio, and scale all unobservables by a log-normal variable with mean 0 and standard deviation 0.1 (so that the log change in each unobservable is normal). I then run the same specifications as in the reduced form section of the paper, and plot the non-parametric relationship between CMA and each outcome.

Figure A.12 plots the results. Each relationship displays a tight, log-linear relationship. This suggests that the reduced form regressions in the paper are consistent as a model validation exercise for the full model.

I Proofs

Proof of Proposition 1

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

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

[104] For example, house price growth would depend on a weighted average of each group's change in CMA with heterogeneous coefficients, where the weights and coefficients reflect residential composition of the tract in consideration.

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

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

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

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

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

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

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.[105] ■

Proof of Proposition 2

To simplify notation, I assume that worker groups share the same values of θ, omit subscripts for travel modes and industries, and assume floorspace is only used for residential purposes (i.e. $\alpha = 1$). I begin by setting up the planner's problem. The planner knows the distribution of individual productivities, but not their specific draws. She therefore chooses the number of people living in any location, announces a policy where total consumption for someone who works in j with productivity ϵ is

$$c_{ijg}(\epsilon) = \tilde{c}_{ijg}\frac{\epsilon}{d_{ij}} + \bar{c}_{ig}$$
$$h_{ijg}(\epsilon) = \tilde{h}_{ijg}\frac{\epsilon}{d_{ij}} + \bar{h}_{ig} + \bar{h}$$

and then individuals decide where to commute. Since utility $U_{ijg}(\epsilon) = u_{ig}\left(\frac{\tilde{c}_{ijg}}{d_{ij}}\epsilon + \bar{c}_{ig}\right)^{\beta}\left(\frac{\tilde{h}_{ijg}}{d_{ij}}\epsilon + \bar{h}_{ig}\right)^{1-\beta}$ is non-linear in ϵ, to make progress I further assume the planner is constrained to policies of the form $\tilde{h}_{ijg} = \iota_{ig}\tilde{c}_{ijg}$ and $\bar{c}_{ig} = \iota_{ig}\bar{h}_{ig}$.[106] Then $U_{ijg}(\epsilon) = u_{ig}\iota_{ig}^{1-\beta}c_{ijg}(\epsilon)$. Average welfare for residents in i is then $E\left[\max_j\left\{u_{ig}\iota_{ig}^{1-\beta}c_{ijg}(\epsilon)\right\}\right] = u_{ig}\iota_{ig}^{1-\beta}\left(\gamma\left[\sum_j\left(\tilde{c}_{ijg}/d_{ij}\right)^{\theta}\right]^{1/\theta} + \bar{c}_{ig}\right)$ and commute flows are $L_{ijg} = \frac{(\tilde{c}_{ijg}/d_{ij})^{\theta}}{\sum_s(\tilde{c}_{isg}/d_{is})^{\theta}}L_{Rig}$.

The planner maximizes the sum of worker welfare subject to the following constraints:

- Housing Feasibility: $H_i = \sum_{jg} L_{ijg}\left(\tilde{h}_{ijg}\bar{\epsilon}_{j|ig} + \bar{h}_{ig} + \bar{h}\right)$

[105] These solutions are unique to scale. In practice, as discussed in Section H.2, 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.

[106] This is rationalized as follows. Suppose the planner knows the shock for an individual and maximizes utility subject to consumption adding up to some exogenous endowments C_{ijg} and H_{ijg}. The FOC require the ratio of marginal utilities to be equal to a constant. Since $U_{ig} = u_{ig}c_{ijg}(\epsilon)^{\beta}h_{ijg}(\epsilon)^{1-\beta}$ implies this ratio is $\frac{\beta}{1-\beta}\frac{h_{ijg}(\epsilon)}{c_{ijg}(\epsilon)}$, I get that $h_{ijg}(\epsilon) \propto c_{ijg}(\epsilon)$. Since the multipliers on goods and housing will only vary by i, I constrain the planner's announced policy to do so too.

- Goods Feasibility: $\left(\sum_k c_{kijg}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} = L_{ijg}\left(\tilde{c}_{ijg}\bar{\epsilon}_{j|ig} + \bar{c}_{ig}\right)$
- Residential Mobility: $\bar{U}_g = u_{ig}\iota_{ig}^{1-\beta}\left(\Phi_{Rig} + \bar{c}_{ig}\right)$
- Commuting Mobility: $L_{ijg} = \frac{(\tilde{c}_{ijg}/d_{ij})^\theta}{\sum_s(\tilde{c}_{isg}/d_{is})^\theta} L_{Rig}$.
- Goods Market Clearing: $A_iN_i = \sum_{rsg} c_{irsg}$
- Local and National Labor Market Clearing: $\tilde{L}_{Fjg} = \sum_i \bar{\epsilon}_{j|ig}L_{ijg}$, $\sum_i L_{Rig} = \bar{L}_g$
- Auxiliary variables: $N_j = \left[\sum_g \alpha_g \tilde{L}_{Fjg}^{\frac{\sigma_F-1}{\sigma_F}}\right]^{\frac{\sigma_F}{\sigma_F-1}}$, $\Phi_{Rig} = \gamma\left[\sum_j\left(\tilde{c}_{ijg}/d_{ij}\right)^\theta\right]^{1/\theta}$, $\bar{\epsilon}_{j|ig} = \gamma\frac{d_{ij}\Phi_{Rig}}{\tilde{c}_{isg}}\frac{1}{d_{ij}}$.

The lagrangian is therefore

$$
\begin{aligned}
\mathcal{L} = &\sum_g \bar{U}_g \\
&+ \sum_{ijg} v_{ijg}\left(\left(\sum_k c_{kijg}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - L_{ijg}\left(\bar{c}_{ig} + \tilde{c}_{ijg}\bar{\epsilon}_{j|ig}\right)\right) + \sum_i \kappa_i\left(H_i - \sum_{jg} L_{ijg}\left(\iota_{ig}\left(\tilde{c}_{ijg}\bar{\epsilon}_{j|ig} + \bar{c}_{ig}\right) + \bar{h}\right)\right) \\
&+ \sum_i \lambda_i\left(A_iN_i - \sum_{rsg} c_{irsg}\right) + \sum_j \xi_j\left(\left[\sum_g \alpha_g \tilde{L}_{Fjg}^{\frac{\sigma_F-1}{\sigma_F}}\right]^{\frac{\sigma_F}{\sigma_F-1}} - N_j\right) \\
&+ \sum_{jg}\delta_{jg}\left[\sum_i \bar{\epsilon}_{j|ig}L_{ijg} - \tilde{L}_{Fjg}\right] + \sum_{ig}\rho_{ig}\left(u_{ig}\iota_{ig}^{1-\beta}\left(\Phi_{Rig} + \bar{c}_{ig}\right) - \bar{U}_g\right) \\
&+ \sum_{ijg}\psi_{ijg}\left(\left(\frac{\tilde{c}_{ijg}/d_{ij}}{\Phi_{Rig}}\right)^\theta L_{Rig} - L_{ijg}\right) + \sum_{ig}\tau_{ig}\left(\left(\sum_s\left(\tilde{c}_{ijg}/d_{ij}\right)^\theta\right)^{1/\theta} - \Phi_{Rig}\right) \\
&+ \sum_{ijg}\phi_{ijg}\left(\gamma\frac{d_{ij}\Phi_{Rig}}{\tilde{c}_{isg}}\frac{1}{d_{ij}} - \bar{\epsilon}_{j|ig}\right) + \mu_g\left(\bar{L}_g - \sum_i L_{Rig}\right)
\end{aligned}
$$

The FOC are

$$
\begin{aligned}
\left(v_{ijg} + \kappa_i\iota_{ig}\right)\bar{\epsilon}_{j|ig}L_{ijg} &= \frac{L_{ijg}}{\tilde{c}_{ijg}}\left(\theta\psi_{ijg} + \frac{\tau_{ig}\Phi_{Rig}}{L_{Rig}} - \phi_{ijg}\frac{\bar{\epsilon}_{j|ig}}{L_{ijg}}\right) && (\tilde{c}_{ijg}) \\
\sum_j \kappa_i L_{ijg}\left(\tilde{c}_{ijg}\bar{\epsilon}_{j|ig} + \bar{c}_{ig}\right) &= (1-\beta)\frac{\rho_{ig}\bar{U}_g}{\iota_{ig}} && (\iota_{ig}) \\
\rho_{ig}u_{ig}\iota_{ig}^{1-\beta} &= \sum_j\left(v_{ijg} + \iota_{ig}\kappa_i\right)L_{ijg} && (\bar{c}_{ig}) \\
\lambda_k &= \nu_{ijg}\left(\frac{c_{kijg}}{C_{ijg}}\right)^{-\frac{1}{\sigma}} && (c_{kijg}) \\
\frac{\tilde{L}_{Fjg}}{N_j} &= \left(\frac{\delta_{jg}}{\alpha_g\xi_j}\right)^{-\sigma_F} && (\tilde{L}_{Fjg}) \\
\delta_{jg}\bar{\epsilon}_{j|ig} &= v_{ijg}\left(\bar{c}_{ig} + \tilde{c}_{ijg}\bar{\epsilon}_{j|ig}\right) + \kappa_i\left(\bar{h}_{ig} + \tilde{h}_{ijg}\bar{\epsilon}_{j|ig} + \bar{h}\right) + \psi_{ijg} && (L_{ijg}) \\
\sum_j \pi_{j|ig}\psi_{ijg} &= \mu_g && (L_{Rig})
\end{aligned}
$$

$$\tau_{ig} = \rho_{ig}\iota_{ig}^{1-\beta}u_{ig} + \sum_j \phi_{ijg}\frac{\bar{\epsilon}_{j|ig}}{\Phi_{Rig}} - \theta\sum_j \frac{\psi_{ijg}\pi_{j|ig}L_{Rig}}{\Phi_{Rig}} \tag{Φ_{Rig}}$$

$$\phi_{ijg} + \nu_{ijg}L_{ijg}\tilde{c}_{ijg} + \kappa_i L_{ijg}\tilde{h}_{ijg} = \delta_{jg}L_{ijg} \tag{$\bar{\epsilon}_{j|ig}$}$$

$$\xi_i = \lambda_i A_i \tag{N_i}$$

$$1 = \sum_i \rho_{ig} \tag{$\bar{U}_g$}$$

Part 1: Efficiency of Equilibrium

Consumption, Housing and Welfare. Define $\tilde{x}_{ijg} \equiv v\tilde{c}_{ijg} + \kappa_i\tilde{h}_{ijg}$ as the cost of providing one unit of the consumption aggregate per unit of effective labor, where I have conjectured that $v_{ijg} = v\ \forall ijg$ as will be established below. Using $\tilde{h}_{ijg} = \iota_{ig}\tilde{c}_{ijg}$, this implies that $\tilde{c}_{ijg} = \frac{\tilde{x}_{ijg}}{v+\kappa_i\iota_{ig}}$ and thus $\Phi_{Rig} = \frac{\gamma\left(\sum_s(\tilde{x}_{ijg}/d_{ij})^\theta\right)^{1/\theta}}{v+\kappa_i\iota_{ig}}$. Likewise, define expenditure on fixed goods to be $y_{ig}^e \equiv v\bar{c}_{ig} + \kappa_i\left(\bar{h}_{ig} + \bar{h}\right)$ so that $\bar{c}_{ig} = \frac{y_{ig}^e - \kappa_i\bar{h}}{v_{ijg}+\kappa_i\iota_{ig}}$. Putting these together yields

$$\bar{U}_g = u_{ig}\frac{\iota_{ig}^{1-\beta}}{v+\kappa\iota_{ig}}\left(\left(\sum_s(\tilde{x}_{ijg}/d_{ij})^\theta\right)^{1/\theta} + y_{ig}^e - \kappa_i\bar{h}\right).$$

To make progress, I need to solve for ι_{ig}. First, note that from the definition of $\bar{\epsilon}_{j|ig}$

$$\bar{\epsilon}_{j|ig} = \gamma\frac{\Phi_{Rig}}{\tilde{c}_{ijg}} = \gamma\frac{\left(\sum_s(\tilde{x}_{isg}/d_{is})^\theta\right)^{1/\theta}}{\tilde{x}_{ijg}} \equiv \gamma\frac{\tilde{\Phi}_{Rig}^{1/\theta}}{\tilde{x}_{ijg}}$$

which implies that $\tilde{x}_{ijg}\bar{\epsilon}_{j|ig} \equiv x_{ijg} = x_{ig} = \gamma\tilde{\Phi}_{Rig}^{1/\theta}\ \forall j$. The FOC for ι_{ig} implies

$$\begin{aligned}(1-\beta)\frac{\rho_{ig}\bar{U}_g}{\iota_{ig}} &= \sum_j \kappa_i L_{ijg}\left(\tilde{c}_{ijg}\bar{\epsilon}_{j|ig} + \bar{c}_{ig}\right)\\ &= \frac{\kappa_i}{v+\kappa\iota_{ig}}L_{Rig}\left(\gamma\tilde{\Phi}_{Rig}^{1/\theta} + y_{ig}^e - \kappa_i\bar{h}\right)\\ \Leftrightarrow \kappa_i\iota_{ig}\frac{L_{Rig}\bar{y}_i}{\rho_{ig}\bar{U}_g} &= (1-\beta)\left(v+\kappa\iota_{ig}\right)\end{aligned}$$

where I have defined $\bar{y}_{ig} \equiv \gamma\tilde{\Phi}_{Rig}^{1/\theta} + y_{ig}^e - \kappa_i\bar{h}$.

To simplify, I need to solve for ρ_{ig}. Condition (L_{Rig}) implies $\psi_{ijg} = \mu_g\forall ij$, and from $(\bar{\epsilon}_{j|ig})$ we have $\phi_{ij} = (\mu + y_i^e)L_{ij}/\bar{\epsilon}_{ij}$. Notice also that $(\tilde{c}_{ijg})$ implies $\tilde{x}_{ijg}\bar{\epsilon}_{j|ig} = \theta\psi_{ijg} + \frac{\tau_{ig}\Phi_{Rig}}{L_{Rig}} - \phi_{ijg}\frac{\bar{\epsilon}_{j|ig}}{L_{ijg}}$. Substituting these results into (L_{ijg}) and (Φ_{Rig}), the system of equations characterizing the multipliers is then

$$\begin{aligned}\delta_{jg}\bar{\epsilon}_{j|ig} &= \tilde{x}_{ijg}\bar{\epsilon}_{j|ig} + y_{ig}^e + \mu_g\\ \tilde{x}_{ijg}\bar{\epsilon}_{j|ig} &= \theta\mu_g + \frac{\tau_{ig}\Phi_{Rig}}{L_{Rig}} - (\mu + y_i^e)\\ \frac{\tau_{ig}\Phi_{Rig}}{L_{Rig}} &= \frac{\rho_{ig}u_{ig}\iota_{ig}^{1-\beta}\Phi_{Rig}}{L_{Rig}} + (\mu + y_i^e) - \theta\mu_g\end{aligned}$$

The first line implies $y_{ig}^e = -\mu_g\ \forall i$ and thus $\tilde{x}_{ijg} = \delta_{jg}$. Solving for the remaining multipliers delivers $\frac{\tau_{ig}\Phi_{Rig}}{L_{Rig}} = x_{ijg} - \theta\mu_g$ and $\rho_{ig} = \frac{\bar{y}_{ig}L_{Rig}}{\bar{U}_g}$. Substituting this value for ρ_{ig} back into the FOC for ι_{ig} yields

$\iota_{ig} = \frac{1-\beta}{\beta}\frac{v}{\kappa_i}$. Plugging $\frac{\iota_{ig}^{1-\beta}}{v+\kappa\iota_{ig}} = \frac{(1-\beta)^{1-\beta}\beta^{\beta}}{v^{\beta}\kappa_i^{1-\beta}}$ into the expression for average utility gives

$$\bar{U}_g \propto u_{ig}\bar{y}_{ig}v^{-\beta}\kappa_i^{-(1-\beta)}.$$

The value for ι_{ig} also implies $\tilde{c}_{ijg} = \beta\frac{\tilde{x}_{ijg}}{v}$, $\tilde{h}_{ijg} = (1-\beta)\frac{\tilde{x}_{ijg}}{\kappa_i}$, $\bar{c}_{ig} = \beta\frac{y_g^e - \kappa_i\bar{h}}{v}$ and $\bar{h}_{ig} = (1-\beta)\frac{y_g^e-\kappa_i\bar{h}}{\kappa_i}$.

Housing Market. Housing market clearing requires that $H_i = \sum_{jg} L_{ijg}\left(\tilde{h}_{ijg}\bar{\epsilon}_{j|ig} + \bar{h}_{ig} + \bar{h}\right)$, which is equivalent to

$$\kappa_i = (1-\beta)\frac{\sum_g L_{Ri}e_{ig}}{H_i - \beta\bar{h}L_{Ri}}$$

where $e_{ig} \equiv x_{ijg} + y_{ig}^e$ is the cost of providing goods agents get positive utility from (i.e. everything other than $\bar{h}$).

Goods Market. The FOC for each consumption variety (c_{kijg}) implies $c_{kijg} = \left(\frac{\lambda_k}{v_{ijg}}\right)^{-\sigma} c_{ijg}$. Using that $c_{ijg} + \bar{c}_{ig} = \beta L_{ijg}\bar{y}_{ig}/v$, the goods market clearing condition implies

$$\begin{aligned} A_k N_k &= \sum_{ijg}\left(\frac{\lambda_k}{v}\right)^{-\sigma}\frac{\beta L_{ijg}\bar{y}_{ig}}{v} \\ \Leftrightarrow N_k &= \xi_k^{-\sigma}v^{\sigma-1}A_k^{\sigma-1}X \end{aligned}$$

where $X \equiv \beta\sum_{ig} L_{Rig}\bar{y}_{ig}$ is total expenditure on goods, and I used $\lambda_k = \xi_k/A_k$ from (N_i). Note also that plugging in the expression for c_{kijg} into the definition of $c_{ijg} = \left(\sum_k c_{kijg}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ yields

$$v_{ijg} = v = \left(\sum_k (\xi_k/A_k)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \quad \forall i,j,g$$

confirming the conjecture that the multipliers are constant.

Labor Market. The FOC for labor demand imply $\lambda_j = \xi_j/A_j$ and $\frac{\tilde{L}_{Fjg}}{N_j} = \left(\frac{\delta_{jg}}{\alpha_g\xi_j}\right)^{-\sigma_F}$, while labor market clearing requires $\tilde{L}_{Fjg} = \sum_i \bar{\epsilon}_{j|ig}L_{ijg} = \sum_i \gamma\pi_{j|ig}^{-1/\theta}\frac{L_{ijg}}{d_{ij}}$. Finally, we obtain an expression for labor supply from $\left(\frac{\tilde{c}_{ijg}/d_{ij}}{\Phi_{Rig}}\right)^{\theta} = \frac{(\tilde{x}_{ijg}/d_{ij})^{\theta}}{\sum_s(\tilde{x}_{isg}/d_{is})^{\theta}}$ which implies $L_{ijg} = \frac{(\tilde{x}_{ijg}/d_{ij})^{\theta}}{\sum_s(\tilde{x}_{isg}/d_{is})^{\theta}}L_{Rig}$.

Taking Stock. Putting these results together yields the system

$$\begin{aligned} \bar{U}_g &= u_{ig}\bar{y}_{ig}v^{-\beta}\kappa_i^{\beta-1} \\ \bar{y}_{ig} &= \gamma\tilde{\Phi}_{Rig}^{1/\theta} + y_g^e - \kappa_i\bar{h} \\ \kappa_i &= (1-\beta)\frac{\sum_g L_{Ri}e_{ig}}{H_i - \beta\bar{h}L_{Ri}} \\ \tilde{\Phi}_{Rig} &= \sum_j (\delta_{jg}/d_{ij})^{\theta} \\ \bar{L}_g &= \sum_i L_{Rig} \\ v &= \left(\sum_j (\xi_j/A_j)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \end{aligned}$$

$$
\begin{aligned}
X &= \beta \sum_{ig} L_{Rig} \bar{y}_{ig} \\
N_j &= \xi_j^{-\sigma} A_j^{\sigma-1} v^{\sigma-1} X \\
\frac{\tilde{L}_{Fjg}}{N_j} &= \left(\frac{\delta_{jg}}{\alpha_g \xi_j} \right)^{-\sigma_F} \\
L_{ijg} &= \frac{(\delta_{jg}/d_{ij})^\theta}{\sum_s (\delta_{sg}/d_{is})^\theta} L_{Rig} \\
\tilde{L}_{Fjg} &= \sum_i \gamma \pi_{j|ig}^{-1/\theta} \frac{L_{ijg}}{d_{ij}}
\end{aligned}
$$

This is exactly the same as the system characterizing the competitive equilibrium with equal home ownership when $(\delta_{jg}, \xi_j, \kappa_i, v, -\mu_g) = (w_{jg}, W_j, r_{Ri}, P, y^e)$ and y^e is the equal share of expenditure on housing.[107]

Part 2: Welfare Elasticity

Using the envelope theorem, the change in welfare to a change in commute costs between i and j is

$$
\frac{\partial \bar{U}_g}{\partial d_{ij}} = -\frac{L_{ijg}}{d_{ij}} \left(\theta \mu_g + \frac{\tau_{ig} \Phi_{Rig}}{L_{Rig}} \right) \quad \Leftrightarrow \frac{\partial \ln \bar{U}_g}{\partial \ln d_{ij}} = -\frac{1}{\bar{U}_g} L_{ijg} w_{ijg}
$$

where $w_{ijg} \equiv \gamma \tilde{\Phi}_{Rig}^{1/\theta}$ is average labor income for those in cell (i, j, g). The FOC for $\bar{U}_g$ implies $\bar{U}_g = \sum_{ig} \bar{y}_{ig} L_{Rig}$. Since all expenditure goes towards goods and housing, we know that

$$
\begin{aligned}
&\sum_{ijg} L_{ijg} e_{ig} = \sum_{ijg} L_{ijg} w_{ijg} + \sum_{ijg} L_{ijg} \left(r_i \bar{h} + (1-\beta) \left(e_{ig} - r_i \bar{h} \right) \right) \\
\Leftrightarrow &\sum_{ijg} L_{ijg} \bar{y}_{ig} = \frac{1}{\beta} \sum_{ijg} L_{ijg} w_{ijg}
\end{aligned}
$$

since $\bar{y}_{ig} = e_{ijg} - r_i \bar{h}$. Therefore we have $\frac{\partial \ln \bar{U}_g}{\partial \ln d_{ij}} = \beta \frac{L_{ijg} w_{ig}}{\sum_i L_{ijg} w_{ig}}$. Using that $d \ln d_{ij} = \kappa dt_{ij}$, we can therefore write the welfare elasticity to a system of shocks $\{dt_{ij}\}$ as

$$
d \ln \bar{U}_g = \beta \kappa \sum_{ij} \frac{L_{ijg} w_{ig}}{\sum_i L_{ijg} w_{ig}} dt_{ij}
$$

as required. ■

Proof of Proposition A.1

Part 1: Commuter Market Access

Note that $L_{ij} = \gamma_i \delta_j \kappa_{ij}$ implies that $L_{Ri} = \gamma_i \sum_j \delta_j \kappa_{ij} = \gamma_i \Phi_{Ri}$ and $L_{Fj} = \delta_j \sum_i \gamma_i \kappa_{ij} = \delta_j \Phi_{Fj}$. Substituting these into each other yields the system of equations

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

[107] That is, each member of group g owns an equal share in a portfolio that owns all housing resided in by group g.

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

Substituting the first into the second we get

$$\Phi_{Fj} = \sum_j K_{ij}^F \frac{1}{\sum_s K_{is}^R \Phi_{Fs}^{-1}} \equiv T_j(\Phi_F)$$

where $K_{ij}^F \equiv L_{Rj}\kappa_{ji}$ and $K_{ij}^R \equiv L_{Fj}\kappa_{ij}$ are observed given the data. By inspection we see T is strictly increasing and homogenous of degree one, so by the results in Fujimoto and Krause (1985) there exists a unique solution to the system $\Phi_F = T(\Phi_F)$.

Part 2: General Gravity Model

I first show existence and establish sufficient conditions for uniqueness of the model. Equilibrium in the labor market requires that demand $\tilde{L}_{Fj} = A_j \delta_j^\alpha$ equals supply $\tilde{L}_F = \delta_j^\delta \tilde{\Phi}_{Fj}$ which implies

$$\delta_j^{\alpha-\delta} = \sum_i K_{ij}^\delta \gamma_i^\epsilon \Phi_{Ri}^\zeta$$

where $K_{ij}^\delta \equiv \kappa_{ij}^\delta / A_j$. Similarly, equating demand $L_{Ri} = B_i \gamma_i^\beta \Phi_{Ri}^\gamma$ for residents with supply yields

$$\gamma_i^{\beta-1} \Phi_{Ri}^\gamma = \sum_j K_{ij}^\gamma \delta_j$$

where $K_{ij}^\gamma \equiv \kappa_{ij}\kappa_{ij}^y B_i$. Thus, equilibrium can be written as

$$\begin{aligned}
\delta_j^{\alpha-\delta} &= \sum_i K_{ij}^\delta \gamma_i^\epsilon \Phi_{Ri}^\zeta \\
\gamma_i^{\beta-1} \Phi_{Ri}^\gamma &= \sum_j K_{ij}^\gamma \delta_j \\
\Phi_{Ri} &= \sum_j K_{ij}^\Phi \delta_j
\end{aligned}$$

which in the form of equation (1) in Allen et. al. (2015) with coefficient matrices

$$\Gamma = \begin{pmatrix} \alpha-\delta & 0 & 0 \\ 0 & \beta-1 & \gamma \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & \epsilon & \zeta \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

According to proposition 1 of their paper, it remains to characterize when the spectral radius of the matrix

$$|B\Gamma^{-1}| = \begin{pmatrix} 0 & |\frac{\epsilon}{\beta-1}| & |\zeta - \frac{\epsilon\gamma}{\beta-1}| \\ |\frac{1}{\alpha-\delta}| & 0 & 0 \\ |\frac{1}{\alpha-\delta}| & 0 & 0 \end{pmatrix}$$

is less than one. It suffices to find as $x > 0$ such that $|B\Gamma^{-1}|x \leq x$. Solving the system of inequalities yields the parameter restriction in the proposition. Existence also follows from proposition 1 of Allen et. al. (2015).

The final part of the proposition comes from substituting out the shifters in terms of employment, residence and market access in the expression for equilibrium in the market for residence ($L_{Ri} = B_i \left(\frac{L_{Ri}}{\Phi_{Ri}}\right)^{\beta} \Phi_{Ri}^{\gamma}$) and employment ($\tilde{L}_{Fj} = A_j \left(\frac{\tilde{L}_{Fj}}{\tilde{\Phi}_{Fj}}\right)^{\frac{\alpha}{\delta}}$) and rearranging.

■

Comment The second part of the proposition shows that when additional structure on the demand for residents and workers across the city, population and employment can be written as log-linear functions of commuter market access. In addition to the log-linear functional forms, this structure requires knowing values for the parameters $\alpha, \beta, \gamma, \delta, \epsilon$ and ζ.[108] The result in part (i) implies that knowledge of these parameters as well as data on residence, employment and commute costs allows one to solve for the endogenous objects $\{\delta_i, \gamma_i\}$ and location characteristics $\{A_i, B_i\}$ that rationalize the observed data. While supply is upward sloping in the shifters $\{\delta_i, \gamma_i\}$, multiple equilibria may occur when demand is upward sloping (determined by the constants α, β). This will be the case in the presence of strong spillovers as seen above. As seen in the paper and in section C.4, the framework can accommodate additional factors so long as equilibrium in factor markets collapses into log-linear demand for labor and residents

[108] Most models impose additional restrictions between these parameters, which reduces the number of parameters one needs to know (see the above for examples).