THE UNIVERSITY OF CHICAGO

THE AGGREGATE AND DISTRIBUTIONAL EFFECTS OF URBAN TRANSIT INFRASTRUCTURE: EVIDENCE FROM BOGOTA'S TRANSMILENIO

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE UNIVERSITY OF CHICAGO

BOOTH SCHOOL OF BUSINESS

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

JOHN NICHOLAS TSIVANIDIS

CHICAGO, ILLINOIS

JUNE 2018

To my father, who taught me the value of education.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my advisor Chang-Tai Hsieh for his continuous guidance and support throughout my time as a graduate student. My most cherished memories from these years are the hours spent working through ideas on his office whiteboard. He always challenged me to push for the big picture, and took the time to provide insight on the details. His continuous support - from sounding out preliminary ideas to supporting my many trips to Bogotá - is something I will always appreciate and never forget.

I am also indebted to my committee members Marianne Bertrand, Erik Hurst and Robert E. Lucas Jr. I learned so much from working with Marianne on our contract labor project, and am grateful for her detailed feedback on various rounds of my paper and presentation. I will always appreciate her generosity through financial support, letters of recommendation and numerous introductions for other projects. From taking his second year PhD class to navigating the job market, I benefitted immensely from Erik who taught me how to craft and communicate economic questions and ideas. Lastly, it was an honor to work with Bob who was always generous with his time to attend workshops and read and critique numerous drafts of my dissertation. His work on economic growth was an inspiration to me long before my time at Chicago. Once I started thinking about questions of growth and development I, too, found it hard to think about anything else.

I have benefitted greatly from discussions with the numerous faculty at the University of Chicago. Jonathan Dingel is one of the smartest people I have met, and his generosity with his time and comments developed my research a great deal. I also thank Rodrigo Adao, Fernando Alvarez, Rick Hornbeck, Ralph Ossa, Felix Tintelnot, Nancy Stokey, Owen Zidar, as well as participants across many University of Chicago workshops for their comments. 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. Lastly, I relied on a great deal of generosity from people in helping me access the data for this project. I would like to express a particular thanks to Andres Villaveces who was immensely helpful in providing data, as well as to Marco Antonio Llinas Vargas, Helmut Menjura and those at DANE who granted access to various microdata. From translating my initial emails into Spanish to inviting me into his family's home during my time in Bogotá, I thank Santiago Caicedo for his support throughout our time in the program and the friendship that came out of it.

This thesis would not have been possible without the support of my friends and family. I'm grateful to the friends I made in Chicago for discussing preliminary research ideas and hanging out in our spare time, whenever we could find some. My friends in London gave constant support and perspective on the many trips home for Christmas. My siblings and my Irish family, especially my Auntie Jackie, made those times all the more special. My dog, Molly, was always generous with late night cuddles.

None of this would have been possible without the constant love and support of my wife, Bettina, who believed in me so much she was willing to quit her career in England and start from scratch in the US for me. She is the best thing to have happened to me. To my newborn son, Finn, whom I hope I can provide all the same opportunities I was given in life.

Finally, to my parents who sacrificed so much for me. My mother, who worked so hard in and out of the house to support me and my siblings, is my role model for strong women and Irish wit. My father came from Greece to England as a teenager to pursue higher education and a better life. He taught me the value of education, and without him this dissertation would never have been written. I dedicate it to him, and hope I can be as good a father as he has been to me.

Abstract

How large are the benefits to improving transit in cities, and how are the gains shared between low- and high-skilled workers? This paper uses detailed tract-level data to analyze the construction of the world's largest Bus Rapid Transit (BRT) system–TransMilenio–in Bogotá, Colombia. First, I build a quantitative general equilibrium model of a city where low- and high-skill workers sort over where to live, where to work, and whether or not to own a car. Second, I develop a new reduced form methodology derived from general equilibrium theory to evaluate the effects of transit infrastructure based on "commuter market access", and use it to empirically assess TransMilenio's impact on city structure. Third, I structurally estimate the model and quantify the effects of the system. I find that while the system caused increases in welfare and output larger than its cost, the gains accrued slightly more to high-skilled workers. The incidence of public transit across skill-groups is determined not only by who uses it most, but also by how easily individuals substitute between commutes, whether the system connects workers with employment opportunities, and equilibrium adjustment of housing and labor markets. Finally, adjusting zoning regulations to allow increased building densities in affected locations would have led to higher welfare gains. This underscores the benefits to cities from pursuing a unified transit and land use policy.

1 Introduction

How large are the economic gains to improving public transit systems within cities and how are they distributed between low- and high-skilled workers? With 2.5 billion people predicted to move into cities by 2050, mostly in developing countries, governments will spend vast sums on mass transit systems 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. Yet measuring the benefits of these systems is challenging: while individuals save time on any particular commute, their decisions of where to live and work will change as new alternatives become attractive and land and labor markets 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 I construct 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 aggregate and distributional effects of urban transit systems. First, I build a quantitative general equilibrium model of a city where low- and high-skill workers sort over where to live, where to work, and whether or not to own a car. Second, I develop a new reduced form methodology derived from general equilibrium theory to evaluate the effects of changes in commuting networks in cities. I show that a wide class of models (nesting a special case of my own) contain a log-linear relationship between outcomes and the transit network through "commuter market access" (CMA). For individuals this reflects access to jobs while for firms it reflects access to workers; both are easily computed using data on employment and residence across the city. I use the implied regression framework to empirically evaluate the effect of TransMilenio on outcomes such as population, employment and house prices. Third, I instrument for changes in market access

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

to identify the model's structural elasticities, and use the estimated model to quantify the effects of the system and counterfactual policies.

I have three main results. First, changes in CMA perform better than traditional distance-based approaches in explaining the heterogeneous adjustment of population, employment and housing markets to TransMilenio. This suggests the framework can be applied elsewhere to improve predictions about the effects of transit on the spatial organization of cities. Second, I find the system provided large aggregate gains for the city, increasing average welfare by 3.5% and output by 2.73% (net of construction and operating costs) at my most conservative estimates. However, these gains would have been around one fourth larger had the government implemented a complementary change in zoning policy to allow housing supply to respond where it was most needed. Third, I find that high-skilled workers benefitted slightly *more*, which is surprising given the reliance of the low-skilled on public transit.

To build intuition, I find certain key channels explain the incidence of improved public transit across worker groups. The first is mode choice: the group that relies on public transit benefits more. This operates in favor of the low-skilled who are poorer and less willing to pay for cars. The second is the elasticity of commuting decisions to commute costs, which determines how willing individuals are to bear high commute costs to work in a particular destination. In the model, this is determined by the heterogeneity of workers' match-productivities with firms in different locations. 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. In the data, I find high-skilled workers' commuting choices are less sensitive to relative differences in commute costs - suggesting a greater dispersion of match productivity, consistent with other evidence - so they bear a greater incidence along this channel.[2] Lastly, the geography of the city and transit network matter: where house

[2]See Lee (2015) for an estimates of this dispersion across educations groups using wage data in the US and developing countries. One might expect rich, high-skilled workers to be more sensitive to commute costs

prices appreciate and wages adjust, whether the system connects locations of dense residence with well-paid jobs, and how these characteristics differ where each group lives and works. In Bogotá these favor the high-skilled.

I compare my results with those using the standard approach in transportation economics to evaluate the gains from new infrastructure based on the value of time savings often used by institutions such as the World Bank.[3] In this framework, welfare gains are equal to the time saved on each commute times the marginal value of time. I find that who benefits from new infrastructure is driven primarily by mode choice: the low-skilled gain more than the high-skilled, with zero gains accruing to landlords.[4] Accounting for the additional channels suggested by the theory and supported by the data, my methodology sheds new light on the distributional effects of commuting infrastructure. My findings suggest investments in public transit are a less precise way to target welfare improvements for the poor than is implied by this alternative partial equilibrium approach.

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 single-use lanes with express and local services, passengers pay at station entrances using smart cards, and buses are boarded at stations rather than at roadside. BRT provides an attractive alternative to subways in rapidly growing developing country cities since they are able to deliver similar reductions in commuting times at a fraction of the cost, and are much faster to build.[5] I collect new sources of data covering 2,800

since their value of time (VoT) is higher. Indeed, in the model VoT is proportional to wages and it is precisely because of this that the high-skilled are more willing to pay for faster transit (i.e. cars). However, when individuals decide where to work, they choose based on *relative* differences in wages net of commute costs across locations. When the dispersion of wages is high (e.g. due to the heterogeneity of match-productivity across employers), these choices are less sensitive to commute costs because differences in net wages are driven mostly by wages at destination rather than by commute costs. Individuals are therefore more willing to bear a high commute cost to work at a particular location.

[3]See Small and Verhoef (2007) for details on the methodology, and Mackie et. al. (2005) for its use.

[4]For comparability with my model, I compute percentage welfare gains under this approach. This differences out changes in the level of gains across groups driven by alternative values of time (such as that due to average wages).

[5]For example, the per mile cost of construction of the subway in Colombia's second largest city, Medellín, was ten times that of TransMilenio, all the while maintaining similar system speeds. Moreover, TransMilenio took less than eighteen months to construct and open, compared to the twelve years taken by Metro Medellín.

census tracts on residence, employment, commuting patterns, and land markets spanning the system's construction.

Prior to TransMilenio's opening, low-skilled workers commuted using a network of informal buses which were on average 30% slower than cars. To understand the implications of improving public transit on worker welfare, I develop a quantitative general equilibrium model of a city where workers choose where to live, where to work, and how to commute. Non-homothetic preferences mean that the high-skilled live in high amenity neighborhoods and are more likely to own cars. Individuals work in different locations due in part to differential demand for skills from firms across the city. For example, retail and manufacturing establishments demand more low-skilled workers while real estate and financial service businesses rely on the high-skilled. 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.

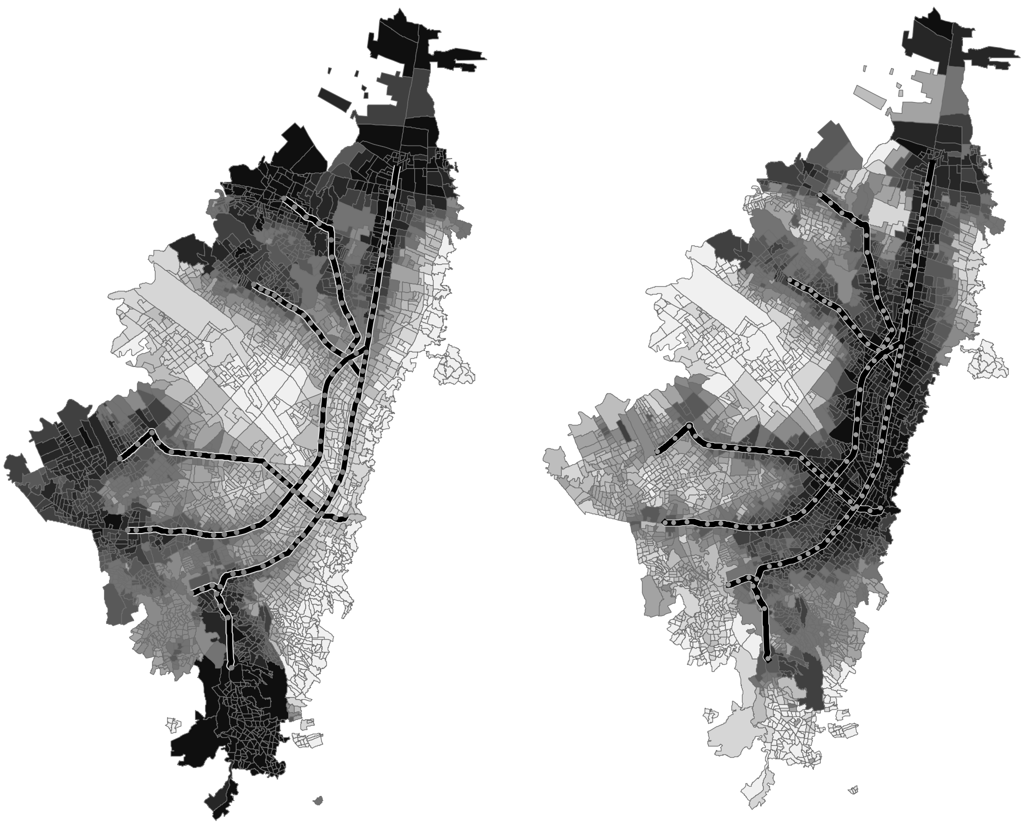
A large literature estimates average treatment effects of transit based on proximity to stations. In contrast, I show that for a wide class of models featuring a gravity equation for commute flows the full direct and indirect effects of the entire transit network on firms and workers can be summarized by a single variable: CMA. Importantly, these terms are easily computed using data on residence and employment in the city (available from censuses, or alternative sources such as cellphone metadata records), as well as a measure of commute costs. Figure 1 plots the change in CMA as a result of TransMilenio. For residents, this captures access to high paying jobs through the commuting network. Tracts towards the edge of the city far from the high density of jobs in the center experienced a much larger improvement in market access. For firms, it reflects access to workers. Central locations benefit most from increased access to workers supplied along all spokes of the network.[6]

[6] Note that firm commuter access also increases away from the center-North of the city. This is due to the high density of (low-skill) workers in the South, as discussed in the next section.

Changes in market access capture a wide heterogeneity in treatment effects from TransMilenio (separately for workers and firms) that would be missed by looking at distance to the system alone.

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 darker shades indicating a larger increase in CMA. Black line shows the TransMilenio routes as of 2006. See Section 7 for full discussion.

In a special case of my model, the equilibrium has a reduced form in which outcomes such as population, employment and house prices can be written as log-linear functions of CMA. Moreover, I show any model with log-linear demand for residents and workers across the city has a similar representation. The framework is therefore isomorphic to a number of alternative assumptions over production technologies, housing supply and worker preferences. I use the implied regression specifications to empirically evaluate the impact of TransMilenio through improvements in market access.

To address non-random route placement, I predict TransMilenio's location in two ways. First, I use a historical tram system built by 1921. Second, using engineering estimates for the cost of building BRT on different types of land use, I solve for least-cost construction paths connecting terminals at the end of the system with the central business district (CBD) as was the intent of the government. I then construct instruments for the change in CMA had TransMilenio been built along these predicted routes.[7]

My identification assumption is that these instruments have only an indirect effect on outcomes through the probability of TransMilenio being built. Relative to distance-based analyses, a key advantage of my approach is that I can control for the distance to these instruments to capture potential direct effects and rely only on residual variation in predicted CMA growth for identification. I run falsification tests exploiting the timing of station openings as well as using residual variation in market access conditional on distance to stations to provide additional evidence that the effects are causal.

I find that changes in CMA perform well in predicting the heterogeneous response of population, employment and land markets in response to TransMilenio. Improvements in resident CMA also led to growth in commute distances and wages, supporting the intuition that it measures access to jobs. Interestingly, the system caused a re-sorting of workers by skill group. The high-skilled moved into high-amenity, expensive neighborhoods in the North while the low-skilled moved into poorer neighborhoods in the South. This suggests that transit has the potential to increase residential segregation between skill groups in cities.[8]

In the final part of the paper, I structurally estimate the full (non-linear) model. 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 in the model is a log-linear function of market access, my instruments provide exogenous variation

[7] I exclude the targeted neighborhoods surrounding portals and the CBD from the analysis.

[8] Heilmann (2016) notes a similar finding following the opening of the light rail system in Dallas.

in the number of individuals living and working across the city. This allows me to identify these key elasticities through a Generalized Method of Moments (GMM) procedure.

I estimate an agglomeration elasticity roughly three times the size of median estimates in the US but close to other studies using experimental approaches. I provide one of the first estimates identified using variation within a less developed country city, suggesting that these forces can be particularly strong in poorer countries. I find a substantial elasticity of amenities to the college share of residents, reflecting the endogeneity of neighborhood characteristics like crime.

The model performs well in matching a number of non-targeted moments such as income, employment and commute flows by skill group. The amenities and productivities recovered from the model correlate well with observable proxies like local homicide rates and the slope of land. I check the robustness of my results to alternative parameter values and incorporate home ownership, alternative timing assumptions and the employment of domestic servants in model extensions. Lastly, I use the extreme assumptions of either zero or infinite mobility costs between Bogotá and the rest of Colombia to bound the impact of TransMilenio on welfare, population, land rents and output.

The system led to large aggregate gains in worker welfare and output. Productive locations were able to "import" more workers through the commuting network. This suggests better transit can improve the spatial allocation of labor within cities. The increase in output greatly exceeded construction and operating costs, supporting the notion that BRT can be a profitable investment. Population decentralized, as improved access to jobs made distant neighborhoods more attractive. Land use became more specialized: the share of floorspace used for commercial (residential) purposes increased in central (outlying) locations where firm (resident) commuter access increased the most.

High-skill workers benefit slightly more from the system. The incidence of public transit across skill-groups is determined not only by who uses it most, but also by how easily individuals substitute between commutes, whether the system connects workers with employment opportunities, and equilibrium adjustment of housing and labor markets. Landlords benefit from house price appreciation where transit access improves.

The effect of different parts of the network is heterogeneous: lines serving poor neighborhoods in the South of the city, as well as a cable car connecting hillside slums with a TransMilenio slated to open in 2018, disproportionately benefit the low-skilled. The conclusion that the low-skilled benefit less than what is implied by mode choice alone remains generalizable, though, since existing evidence suggests that the key elasticities that vary across groups have similar relative magnitudes in other countries and Bogotá is by no means unusual in its spatial configuration. Moreover, the methodology developed in this paper can be applied to the specific geography and transit systems of any city where data on residence and employment are available to predict the effects of new infrastructure.

The estimated model allows to me assess the impact of counterfactual policies. In the first exercise, I simulate the removal of the feeder bus network that transports individuals in the outskirts of the city to terminals at the end of lines using existing road infrastructure at no additional fare. This part of TransMilenio increases welfare more than any other single line of the network. This underlines the potential for 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.

In the second exercise, I evaluate the welfare impacts of a "Land Value Capture" (LVC) scheme under which development rights to increase building densities near stations are sold by the government to developers, and the extent to which the revenues could have financed the system's construction. Similar schemes have seen great success in Asian cities such as Hong Kong and Tokyo.[9] In contrast, 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 where it was needed. I compare the effects of two alternative policies. The first increases permitted densities within a certain distance of stations,

[9]See Hong et. al. (2015) for a review.

while the second allocates the same number of permits based on predicted growth in CMA. The CMA-based policy increases welfare gains from TransMilenio by around 23%, while government revenues cover at least 18% of the construction costs. Under the distance-based scheme welfare and government revenues only increase by around half that amount. These policies disproportionately benefit low-skilled workers by dampening house price appreciation towards the edge of the city where they live. My findings suggest large returns to the pursuit of an integrated transit and land use policy, and highlight the applicability of CMA as an instrument to guide government policy.

The rest of the paper proceeds as follows. Section 2 discusses the paper's contribution to the literature. Second 3 presents the context of Bogotá and TransMilenio. Section 4 develops the model and Section 5 outlines the reduced form framework it delivers. Section 6 describes the data. Section 7 presents the reduced form estimation results while Section 8 structurally estimates the non-linear model which Section 9 uses to quantify the effects of TransMilenio. Section 10 simulates the effects of counterfactual policies. Section 11 concludes.

2 Relation to Previous Literature

This paper contributes to the literatures on urban economics and economic geography.

Within a large body of work that documents the association between transit and urban structure, a smaller strand exploits the opening of new systems to establish a causal relationship. These papers typically measure changes in population and property prices as a function of distance to the CBD (Baum-Snow 2007; Baum-Snow et. al. 2017; Gonzalez-Navarro and Turner 2016) or distance to stations (Gibbons and Machin 2005; Glaeser et. al. 2008; Billings 2011). However, when spatial units are interlinked (as is likely in cities where locations are connected through commuting), spillovers across treatment and control locations confound causal inference from such comparisons.[10] If these linkages lead to heterogeneous responses,

[10]Identification of causal impacts of transit infrastructure from relative comparisons requires that the treatment effect on one location is independent of whether another is treated. This is known as the stable

then average treatment effects estimated in one context will no longer be externally valid in another. My approach confronts these challenges by explicitly measuring 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.[11]

A long literature has examined the link between access to goods markets and economic development across regions within countries (see Redding 2010 for a review). Recent work has combined data at the regional level with natural experiments that impact the cost of trading goods across space to examine the effects on local outcomes such as factor prices and population through access to goods markets (Redding and Sturm 2008; Donaldson forthcoming; Donaldson and Hornbeck 2015).[12] Yet these models contain no notion of commuting within cities, and therefore are silent on the effects of infrastructure that reduces the cost of moving people rather than goods across space. I consider a different class of urban commuting models where individuals can live and work in separate locations, and show reduced form relationships between outcomes and measures of access to workers and jobs apply in these settings. These measures can be recovered from data on residential population and employment using less model structure than is typically relied on in the economic geography literature; a full discussion is provided in Section 5. The structural part of this paper uses the

unit treatment value assumption (SUTVA), and can be violated if locations interact. While this can be overcome if spatial units are sufficiently aggregated so that interaction across locations is minimal, this is unlikely to hold for highly local geographies within cities. See Donaldson (2015) for a full discussion.

[11]Within the urban planning literature, the notion of "accessibility" as a determinant of land use and prices within cities dates back as far as Hansen (1959). However, the idea has received little attention within urban economics which instead has focused on models where distance from the CBD is a summary statistic for the spatial configuration of the city (Alonso 1964; Mills 1967; Muth 1969; Lucas and Rossi-Hansberg 2002). There is a literature that attempts to measure accessibility in polycentric cities based on cross-sectional comparisons (see Anas et. al. 1998 and Handy and Niemeier 1997 for reviews). A small handful of recent papers measure the correlation between residential land prices and an accessibility index given by the distance-weighted sum of employment (Osland and Thorsen 2008; Ahlfeldt 2011; McArthur et. al. 2012). In contrast, I show similar indices reflecting market access can be derived from general equilibrium theory (separately for firms and workers), and use a change in the transit network to estimate its causal impact on a range of urban outcomes. My focus on accessibility links to recent work on the relationship between urban form and consumption-related trips (Couture 2016, Duranton and Turner 2017).

[12]Bartelme (2015) generalizes the market access approach within a gravity framework in Allen et. al. (2017) in a non-experimental setting. There is a long literature using an empirical market access approach in international trade literature e.g. Redding and Venables (2004).

reduced form moments to guide estimation of a non-linear model, allowing me to quantify the distributional effects of transit across worker skill groups.

This paper contributes to the growing body of quantitative work featuring gravity equations for commute flows (Ahlfeldt et. al. 2015; Allen et. al. 2015; Monte et. al. 2017; Owens et. al. 2017). I show that these models share a common measure that summarizes the effect of transit on the supply of residents and workers across locations. These measures are easily computed using data and population and employment, and can be used to guide reduced form analysis of changes in the network. This paper is also the first to structurally estimate a gravity model by combining a large-scale construction of commuting infrastructure with moments derived directly from log-linear relationships between resident and firm CMA and model outcomes.[13]

I extend the approach of these papers to incorporate multiple types of workers, firms and transit modes, which is necessary to assess the distributional effects of urban policies.[14] Importantly, I show that by incorporating multiple types of firms with different demand for worker groups, one can invert the model to solve for unobserved group-specific wages that rationalize the data. Allowing for differences in wages between skill groups across the city turns out to be quantitatively important for assessing the distributional effects of transit.[15]

A large literature has studied the relationship between population density and outcomes such as wages and productivity (see Rosenthal and Strange 2004 for a review). A smaller strand of work uses potentially exogenous sources of variation in the density of economic activity to estimate these spillovers (Greenstone et. al. 2010; Kline and Moretti 2014; Ahlfeldt et. al. 2015). Other papers examine how amenities depend on the composition of local

[13]Severen (2016) estimates a model similar to Ahlfeldt et. al. (2015) using the expansion of the Los Angeles subway, but does not use the market access approach developed in this paper in his methodology.

[14]To my knowledge, Redding and Sturm (2016) is the only other paper in the recent quantitative urban literature to incorporate multiple types of workers. However, they test only qualitative predictions of their model. I also incorporate non-homothetic demand for amenities rather than generating sorting through differences in preferences alone.

[15]Without this additional structure, in order to solve for wages one would either need to assume wages are identical for each worker group in every location (i.e. skill groups are perfect substitutes in production) or observe both residence and employment by skill group (which is extremely rare to obtain at small spatial scale).

residents (Bayer, Ferreira and McMillan 2007; Guerrieri, Hartley and Hurst 2013; Diamond 2016). To my knowledge, this paper provides the first intra-city estimates of productivity and amenity spillovers within a developing country city, using changes in the transit network as labor and resident supply shocks that provide sources of identifying variation.[16]

3 Background

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, and there is a stark divide between rich and poor.[17] In this section, I provide background on the city and its transit system.

3.1 Structure of Bogotá

Residence and Employment Bogotá is characterized by a high degree of residential segregation between the rich and poor. Defining high-skill or college 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.[18] The high-skilled are much more likely to live in the North, with low-skilled workers located primarily in the city's South and periphery. Panel (b) shows that these poorer neighborhoods have a much higher population density, reflecting the concentration of smaller housing units that are crowded in.

High- and low-skilled residents work in different kinds of jobs and neighborhoods. Table 1 shows the share of workers employed in each one-digit industry with post-secondary edu-

[16] Using the relationship between earnings and population density across cities, Chauvin et. al. (2016) find that the connection between density and incomes are about similar, 40% higher and 400% higher in Brazil, India and China respectively when compared to the US.

[17] Colombia is the eleventh most unequal country in the world according to the ranking of Gini coefficients from the World Bank for the most recently available year. The income distribution in Bogotá has a slightly higher Gini than the country as a whole (author's calculation using GEIH data from DANE in 2014). Other figures from DANE.

[18] All datasets are described in detail in Section 6. In this section, population data comes from the 1993 census, employment location data comes from the 1990 economic census, other employment data is from DANE's GEIH and ECH labor surveys and mobility data is from DANE's mobility surveys.

cation. 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.

Taken together, this shows substantial differences in where the high- and low-skilled live and work.[19]

Commuting Prior to TransMilenio In 1995 the average trip to work in Bogotá took 55 minutes, more than double the average commute in US cities. The vast majority of these commutes was taken by bus (73%), followed by car (17%) and walking (9%).[20] 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. However, 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).[21] Disregard of bus stops promoted boarding and alighting along curbs, further reducing traffic flows.

[19]Given the non-uniform distribution of jobs by skill intensity, it may be that differential residential patterns of skill groups are driven by differences in access to high- and low-skill jobs (defined by employment share) rather than forces towards residential sorting. In additional results available upon request, I show that in the cross-section access to jobs is able to explain at most 26% of the variation in the college share across census tracts even when accounting for the skill intensity of jobs. This suggests the majority of residential sorting is driven by neighborhood attributes other than access to jobs.

[20]The average commute in US cities has increased from 21 minutes in 1980 to 26 minutes in 2015. For an analysis using US census data, see Ingraham, C. (2016) "The Astonishing Human Potential Wasted on Commutes." Washington Post, Feb. 26. Data from Bogotá comes from the 1995 Mobility Survey. Bicycles and motos account for the remaining 1% of commutes.

[21]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.

The result was that while the crowding of Bogotá's streets slowed traffic overall, buses were much slower than cars. Table 2 uses commuting microdata from the Mobility Survey to compare 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, the burden of slow public transit fell disproportionately on the city's low-skill population. Column (3) of the shows that low-skill Bogotanos were about 29% more likely to use buses as opposed to cars, which is also robust to controlling for differences in the trip composition in column (4). This dependence on slow public transportation meant that the low-skilled faced a different distribution of commute times than the high-skilled.[22]

3.2 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.[23] Phases 2 and 3 added an additional 70km in 2006 and 2011, creating a network spanning the majority of the city (Figure 4). 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).[24] 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).

[22]The results are the same when examining the relationship between income and bus use. Individuals in the bottom and middle terciles of the income distribution are 32% and 24% more likely to commute by bus respectively.

[23]In many cases anticipation of a system may predate its inauguration. However, TransMilenio went from a "general idea" to implementation in only 35 months (Hidalgo and Graftieux 2005).

[24]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.

The system involves exclusive dual bus lanes running along the median of arterial roads in the city separated from other traffic.[25] 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.[26]

There are two main reasons why BRT provides an attractive alternative to subways in rapidly growing cities. First, it delivers similar reductions in commuting times at a fraction of the cost: the average per kilometer construction cost is one-tenth of rail (Menckhoff 2005). Second, BRT is much faster to construct. An illustrative comparison is that of TransMilenio and Metro de Medellín, the subway system in Colombia's second largest city. While both achieve similar system speeds, Medellín's metro cost eleven times as much as TransMilenio and took twelve years from announcement to opening.[27] BRT has the additional advantage that it can be put down quickly and cheaply to address congestion in rapidly growing cities, but in the years after can be easily repurposed into roadways if subways are built or can remain as part of a multimodal public transit network.[28]

[25] As shown in the appendix, concrete barriers separate TransMilenio from other lanes and have helped the city to achieve essentially complete compliance.

[26] 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.

[27] The difference in times from planning to opening between BRT and subways is not specific to Bogotá. While a comprehensive source on construction times is hard to find, stories of such instances are not. In India planning for subways in Delhi and Bangalore started eighteen and eight years before inauguration respectively, while the BRT in Ahmedabad took only 4 years. New York's second avenue subway line opened on January 1st 2017 having been originally proposed in 1919. In Bogotá, there were a total of ten attempts to introduce heavy rail between 1947 and 1997, thwarted by high capital costs and vested interests of the public transportation sector (Lleras 2003).

[28] Many cities currently operate both subways and BRT systems, for example Mexico City, Medellín and Guangzhou.

While BRT is not without drawbacks, 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). Since the cost component was an important determinant of route selection, final lines were placed along wide arterial roads that were cheaper to convert. The number of car lanes was left unchanged either because existing busways were converted or due to road widening.[29]

Two features of the choice process merit emphasis. 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 far cheaper to construct along the widest arteries of the city, whose availability was limited and determined in large part by the city's historical evolution. I leverage both in constructing instruments for the system's layout.

A notable feature of TransMilenio was that it was rolled out so quickly, primarily to complete a portion of the system within mayor Peñalosa's term that ran between 1998 and

[29]See Hidalgo and Graftieux (2005) for a discussion of existing busways on phase one corridors, and Wright and Hook (2007) who report road widening during phase two. Inspection of satellite images confirms that the number of road lanes for other traffic was unchanged (see appendix for examples). That certain routes already contained median busways did not mean that there was efficient bus transit available along them (e.g. Avenida Caracas). While these lanes shared many similar features to TransMilenio, including dual bus lanes and bus stops, within a few years of the opening in 1990 the "the scheme 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).

2001.[30] The unanticipated nature of the system's construction, combined with the staggered opening of lines across three phases, provide additional sources of time series variation I use in my 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, I show in the appendix 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. Cities such as Hong Kong and Tokyo have had great success in implementing LVC schemes which increase permitted densities around new stations but charge developers for the right to build there. These policies achieve the dual aim of increasing housing supply and raising revenue to finance the construction of the system.[31] In counterfactuals, I quantitatively assess the welfare gains from TransMilenio had Bogotá pursued a similar policy.

Trip Characteristics and Effects on Congestion In the appendix, I provide additional details on the way in which TransMilenio is used and its effects on other modes which I briefly summarize here. First, TransMilenio is a quantitatively important mode of transit that is more likely to be for longer trips compared to other modes.[32] Second, TransMilenio 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. Third, TransMilenio use appears to have come primarily from substitution away from buses. Fourth, conditional on car ownership the rich and poor are equally likely to use TransMilenio, consistent with the

[30] Peñalosa's upheaval of the status quo faced entrenched opposition both from the incumbent bus industry and car owners, ultimately leading him to be voted out of office in 2001.

[31] By increasing the response of housing quantities rather than prices, these policies also shift some of the incidence of the infrastructure from land owners to city residents. For a comprehensive review of LVC schemes, see Hong et. al. (2015).

[32] This is commensurate with the fixed time costs of entering and exiting stations, and time spent walking between stations and trip origins and destinations.

similar fares charged compared to traditional buses.

In general, BRT is likely to have complex and ambiguous effects on the speeds of other modes as commuters substitute between modes and equilibrium speeds respond to the changing volumes of vehicles. Both data limitations and the challenge of incorporating these forces within a general equilibrium model put a full analysis of these forces beyond the scope of this paper. However, in the appendix I use commuting microdata to show that there were no significant changes in car and bus speeds on routes where TransMilenio was built compared to other control trips. This suggests my abstraction from the effects of TransMilenio on other mode speeds appears to be a reasonable approximation to reality, and is consistent with recent evidence.[33]

4 A Quantitative Model of a City with Heterogeneous Skills

This section presents a general equilibrium model of a city. High- and low-skill workers decide where to live, where to work, and how to commute between a large number of discrete locations. Individuals are attracted to neighborhoods with nice amenities, good access to jobs and low house prices. Public transit is available to everyone to commute between home and work, 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. Some locations are more productive than others. Each industry differs in its demand for skills: for example, hotels and restaurants demand more low-skilled workers while financial services require more high-skilled individuals. Since industries may be located in

[33] Akbar and Duranton (2017) estimate congestion in Bogotá and find that during times primarily used for commuting, the elasticity of speed with respect to the number of travelers is a mere 0.06. Similarly, Akbar et. al. (2017) find that only 15% of differences in driving speeds in Indian cities are due to congestion (which is broader than the number of vehicles traveling). The vast majority of the variance in speeds is due to uncongested travel speeds. Lastly, Duranton and Turner (2012) find that for the US vehicle-kilometers travelled (VKT) increase one for one with roadway lane kilometers, and as a further implication, find no evidence that the provision of public transportation affects VKT.

different places, wages for low- and high-skill workers will differ across the city. Each location has a fixed amount of floorspace supply which landowners allocate to either 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.

The setup differs from recent quantitative urban models (e.g. Ahlfeldt et. al. 2015) along two key dimensions. First, I add in multiple skill groups of workers, commute modes and industries. This allows me to assess the distributional effects of public transit systems. Second, I incorporate non-homothetic demand for cars and residential amenities to match the sorting patterns documented in the data.

Despite the interactions between labor and land markets across thousands of locations that occur through the city's commuting network, there will be a single measure that summarizes the effect of the entire network on outcomes in any location given by its CMA. This will be integral to my empirical analysis and structural estimation in Sections 7 and 8 respectively.

4.1 Model Setup

The city is comprised of a discrete set of locations $i \in I$. Locations differ by their total amount of floorspace (which can be used for either residential or commercial purposes), productivities, amenities as well as their access to the transit network which determines the time it takes to reach any other location in the city.[34]

The city is populated by different worker skill groups indexed by $g \in G = \{L, H\}$, each of which has a fixed population $\bar{L}_g$.[35] Each worker has an idiosyncratic preference for each combination of where to live and whether or not to own a car, as well as a match-productivity with firms in each location, and chooses the combination that maximizes their

[34] 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. In Section 10, I explore the impact of allowing floorspace supply to respond to the system.

[35] This is the "closed city" assumption. In quantitative exercises, I also consider the alternative extreme of perfect mobility between the city and the rest of the country ("open city" assumption).

utility. I assume timing is such that workers first choose where to live and whether or not to own a car, and then choose where to work.[36] Firms in different industries $s \in S$ produce using labor and commercial floorspace under perfect competition. Absentee landlords own floorspace which they allocate to residential and commercial use to maximize profits. In equilibrium, wages and the price and use of floorspace adjust to clear land and labor markets.

4.2 Workers

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 common components of how members of that group enjoy living 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 where they work ($\epsilon_j(\omega)$) and their preference for each residence-car ownership pair ($\nu_{ia}(\omega)$).

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. Cars also provide an amenity benefit capturing the potential for improved leisure benefits, but come at a fixed cost of ownership $p_a > 0$.[37]

Assuming that 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

$$\max_{C_i(\omega), H_{Ri}(\omega)} u_{iag} C_i(\omega)^{\beta} (H_{Ri}(\omega) - \bar{h})^{1-\beta} \nu_{ia}(\omega)$$

[36] I consider alternative timing assumptions in Section 9.

[37] In the appendix, I outline a third stage mode choice problem in which individuals decide how to commute between home and work conditional on their decision on car ownership. Car owners can choose between cars and public modes (walk, bus, TransMilenio) while non-car owners may only use public transportation. The result is that car owners face different average commute times for each trip; these are what I report in this section.

$$\text{subject to} \;\; C_i(\omega) + r_{Ri} H_{Ri}(\omega) + p_a a = \frac{w_{jg}\epsilon_j(\omega)}{d_{ija}}$$

Solving for the optimal demand for housing and consumption good yields the following expression for indirect utility

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

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.[38]

In contrast to models with homothetic preferences (e.g. Ahlfeldt et. al. 2015), the fixed nature of expenditures on cars and housing allows me to match the Engel curves I document for car ownership and housing expenditure,[39] and drives sorting of workers over car ownership and residential neighborhoods by income. When cars are quicker than public modes of 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 neighborhoods with cheaper housing. Since housing is expensive in high amenity locations in equilibrium, the poor (rich) sort into low (high) amenity neighborhoods.

Workers first choose where to live and whether or not to own a car, and then choose where to work. I now solve their problem by backward induction.

---

[38]The consensus within the literature is that a semi-log gravity equation best fits the commuting data within cities, which will come from this specification of commute costs d_{ija} (e.g. Fortheringham and O'Kelly 1989). I assume the commute cost affects productivity at work (i.e. by reducing effective labor supply) rather than overall utility, since this simplifies the gravity equation for commute flows. In Monte Carlo exercises I show that in a simulated city, the effect of improving commuting infrastructure is quantitatively similar if commute costs reduce utility directly.

[39]See the appendix for both figures and explanations of their construction.

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 in locations across the city.[40] I assume this is drawn from a multivariate Frechet distribution

$$F_g(\epsilon_1,\ldots,\epsilon_J) = \exp\left(-\left[\sum_j \tilde{T}_g \epsilon_j^{-\frac{\tilde{\theta}_g}{1-\rho_g}}\right]^{1-\rho_g}\right).$$

The parameter $\tilde{\theta}_g$ measures the dispersion of productivities for type-g workers (comparative advantage), with a higher $\tilde{\theta}_g$ corresponding to a smaller dispersion, while the parameter ρ_g determines the correlation of an individual's talent across locations (absolute advantage). If $\rho_g = 1$ then draws are perfectly correlated within individuals while if $\rho_g = 0$ then they are perfectly uncorrelated. The scalar $\tilde{T}_g$ controls the overall level of productivities for workers in a particular group.

With these draws in hand, linearity of (1) means that workers simply 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

$$\pi_{j|iag} = \frac{(w_{jg}/d_{ija})^{\theta_g}}{\sum_s (w_{sg}/d_{isa})^{\theta_g}} \equiv \frac{(w_{jg}/d_{ija})^{\theta_g}}{\Phi_{Riag}} \tag{2}$$

where $\theta_g \equiv \tilde{\theta}_g/(1-\rho_g)$ reflects the relative strength of comparative advantage.

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),

[40] In additional results available upon request, I show this 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.

then commuting decisions 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.

Resident Commuter Market Access Expected income prior to drawing the vector of match productivities is directly related to the denominator in (2) through

$$\bar{y}_{iag} = T_g \Phi_{Riag}^{1/\theta_g}, \tag{3}$$

where T_g is a transformation of the location parameter of the Frechet distribution.[41]

I define the term Φ_{Riag} as *Resident Commuter Market Access* (RCMA). This summarizes the effect of the entire commuting network on the supply of residents to a location: it rises when a location is close (in terms of commute costs) to well-paid jobs. I return to the content and measurement of CMA in the next section.

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 in order 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

$$L_{Riag} = \lambda_U \left(u_{iag} \left(\bar{y}_{iag} - p_a a - r_{Ri} \bar{h} \right) r_{Ri}^{\beta - 1} \right)^{\eta_g} \tag{4}$$

where λ_U is an equilibrium constant.[42]

[41]In particular, $T_g \equiv \gamma_{\theta,g} \tilde{T}_g^{1/\theta_g}$ and $\gamma_{\theta,g} = \Gamma\left(1 - \frac{1}{\theta_g(1-\rho_g)}\right)$ where $\Gamma(\cdot)$ is the gamma function.

[42]In particular, $\lambda_U = \bar{L}_g(\gamma_{\eta,g}/\bar{U}_g)^{\eta_g}$ where $\gamma_{\eta,g} = \Gamma\left(1 - \frac{1}{\eta_g}\right)$ and $\bar{U}_g$ is the overall level of utility for group-g individuals. Expected utility prior to learning match productivities is given by $U_{iag\omega} = u_{iag}\left(\bar{y}_{iag} - p_a a - r_{Ri}\bar{h}\right) r_{Ri}^{\beta-1} \nu_{ia\omega}$.

Intuitively, workers are more attracted to locations with high amenities, expected incomes and low house prices, with an elasticity determined by the dispersion of their idiosyncratic preferences η_g. The entire transit network only matters for individuals' residential choices in so far as it affects RCMA, which determines workers expected incomes through (3).[43]

4.2.3 Aggregation

Firm Commuter Market Access and Labor Supply Using the commuting probabilities (2), 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

$$L_{Fjg} = w_{jg}^{\theta_g} \Phi_{Fjg} \qquad (5)$$
$$\text{where } \Phi_{Fjg} = \sum_{i,a} d_{ija}^{-\theta_g} \frac{L_{Riag}}{\Phi_{Riag}}$$

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. I define the term Φ_{Fjg} as a location's *Firm Commuter Market Access* (FCMA). It summarizes the effect of the entire commuting network for firms in a location through its effect on labor supply. 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.[44]

Worker Welfare I equate the overall welfare of group-g residents with the expected utility

[43] Locations will be populated by members of group g only if they are desirable ($\bar{u}_{iag} > 0$) and affordable ($\bar{y}_{iag} - p_a a - r_{Ri}\bar{h} > 0$). Thus, the expression for residential populations in (4) applies only for active locations $\mathcal{A}_{Rg} = \left\{(i,a) : \bar{u}_{iag} > 0,\ r_{Ri} < (\Phi_{Riag}^{1/\theta} - p_a a)/\bar{h}\right\}$ that are both desirable and affordable for members of group-g, and is zero otherwise. For clarity, I omit this additional notation in the text.

[44] In particular, $\bar{\epsilon}_{jg} = T_g \sum_{i,a} \frac{\pi_{j|iag}^{-1/\theta_g}}{d_{ija}} \frac{\pi_{j|iag} L_{Riag}}{\sum_{r,o} \pi_{j|rog} L_{Rrog}}$. Under the Frechet distribution, the average productivity of workers who commute to j from (i,a) is inversely related to the share who choose to do so through $T_g \pi_{j|iag}^{-1/\theta_g}$ reflecting the selection of less productive workers as the share increases.

prior to drawing their idiosyncratic preferences in the first stage given by[45]

$$\bar{U}_g = \gamma_{\eta,g} \left[\sum_{i,a} \left(u_{iag} \left(\bar{y}_{iag} - p_a a - r_{Ri} \bar{h} \right) r_{Ri}^{\beta - 1} \right)^{\eta_g} \right]^{1/\eta_g} \tag{6}$$

4.3 Firms

Technology There are $s \in \{1, \ldots, S\}$ industries which produce varieties differentiated by location in the city under perfect competition. Output is freely traded, and consumers have CES preferences over each variety with elasticity of substitution $\sigma_D > 1$.[46] Firms produce using a Cobb-Douglas technology over labor and commercial floorspace

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

where the labor input is a CES aggregate over the effective labor across skill groups 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

[45]Here $\gamma_{\eta,g} \equiv \Gamma\left(1 - \frac{1}{\eta_g}\right)$ is a constant.

[46]This is the numeraire good introduced in the consumer's problem. While the assumption of representative firms with a fixed location seems restrictive, in the next section I show this production technology is isomorphic to more realistic setups.

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, the demand for labor and commercial floorspace is

$$\tilde{L}_{Fjgs} = \left(\frac{w_{jg}}{\alpha_{sg} W_{js}} \right)^{-\sigma} N_{js} \tag{7}$$

$$H_{Fjs} = (1 - \alpha_s) \frac{X_{js}}{r_{Fj}} \tag{8}$$

where X_{js} is firm sales.[47]

4.4 Floorspace

Market Clearing There is a fixed amount of floorspace H_i in any location, a fraction ϑ_i of which is allocated to residential use and $1 - \vartheta_i$ to commercial use. For any allocation, market clearing for residential floorspace requires that the supply of residential floorspace $H_{Ri} = \vartheta_i H_i$ equals demand:

$$r_{Ri} = (1 - \beta) \frac{E_i}{H_{Ri} - \beta \bar{h} L_{Ri}} \tag{9}$$

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:

$$r_{Fj} = \frac{\sum_s (1 - \alpha_s) \left(W_{js}^{\alpha_s} r_{Fj}^{1-\alpha_s} / A_{js} \right)^{1-\varsigma} X}{H_{Fj}}. \tag{10}$$

[47] From CES demand sales are given by $p_{js}^{1-\sigma_D} X$ where $X = \sum_i \beta (E_i - \bar{h} r_{Ri} L_{Ri})$ is total spending on goods in the city and $E_i = \sum_{g,a} (\bar{y}_{iag} - p_a a) L_{Riag}$ is total spending on goods and housing from residents in i.

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

$$\begin{aligned} \vartheta_i = 1 \;\; &\text{if } r_{Ri} > (1-\tau_i)r_{Fi} \\ (1-\tau_i)r_{Fi} = r_{Ri} \;\; &\forall\{i : \vartheta_i \in (0,1)\} \\ \vartheta_i = 0 \;\; &\text{if } (1-\tau_i)r_{Fi} > r_{Ri} \end{aligned} \tag{11}$$

4.5 Externalities

Productivities A long literature points to the importance of productivity spillovers in cities.[48] I allow a location's productivity to depend on an exogenous component that reflects features independent of the density of economic activity (e.g. access to roads, slope of land) as well as a production externality that depends on the density of employment in that location

$$A_{js} = \bar{A}_{js}\left(\frac{\tilde{L}_{Fj}}{T_j}\right)^{\mu_A}, \tag{12}$$

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.[49]

Amenities Similarly, I allow amenities in a neighborhood to depend on an exogenous component which also varies by car ownership (e.g. leafy streets, close to getaways surrounding

[48] This idea dates back at least to Adam Smith (1776), and was articulated more fully in Marshal (1890). Two prominent examples establishing this relationship are Ciccone and Hall (1996) using regional data and Ahlfeldt et. al. (2015) using intra-city data. See Rosenthal and Strange (2004) for a review.

[49] Unlike Ahlfeldt et. al. (2015), I do not allow for spillovers across locations given spatial units in my analysis are census tracts. The authors find very local spillovers across space which go to zero within 15 minutes of walk time. Rossi-Hansberg et. al. (2010) who find spillovers from revitalized houses fall approximately one half every 1,000 feet. However, in the appendix I show that the regression approach can be extended to include such spillovers.

the city) and a residential externality that depends on the college share of residents

$$u_{iag} = \bar{u}_{iag} \left(\frac{L_{RiH}}{L_{Ri}} \right)^{\mu_{U,g}}. \tag{13}$$

In contrast to existing urban models (e.g. Ahlfeldt et. al. 2015), endogenous amenities depend on the composition of residents 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.[50] This externality provides an 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 the amenities even more. While this sorting force could be driven by the subsistence housing requirement alone, I allow the strength of residential externalities $\mu_{U,g}$ to potentially differ across groups so that some groups may prefer to live near the high-skilled all else equal. I let the data speak to the relative strength of these forces towards residential segregation in estimation.

4.6 Equilibrium

I now define general equilibrium in the city.

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 $\{\bar{h}, \beta, \alpha, p_a, \kappa, \theta_g, \rho_g, T_g, \eta_g, \alpha_{sg}, \sigma_D, \sigma, \mu_A, \mu_U\}$, an equilibrium is defined as a vector of endogenous objects $\{L_{Riag}, L_{Fjg}, w_{jg}, r_{Ri}, r_{Fi}, \vartheta_i, \bar{U}_g\}$ such that

1. **Labor Market Clearing** The supply of labor by individuals (5) is consistent with demand for labor by firms (7),

[50] Evidence for the US discussed in Section 2 also suggests amenities in cities depend on the composition of residents.

2. **Floorspace Market Clearing** The market for residential floorspace clears (9) and its price is consistent with residential populations (4), the market for commercial floorspace clears (10) and floorspace shares are consistent with land owner optimality (11),

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

With this definition in hand, I now characterize existence and uniqueness of equilibria in this economy.

Proposition 1. *An equilibrium exists in this city. Moreover, in a special case of the model with one group of workers, firms and commute modes and no non-homotheticities ($\bar{h} = p_a = 0$) and a fixed allocation of floorspace, a sufficient condition for the equilibrium to be unique is that*

$$\begin{aligned}
\mu_A &\leq 1-\alpha+\frac{\sigma+\theta-1}{(\sigma-1)(\theta-1)} \\
\mu_U &\leq \frac{1+\eta(1-\beta)}{\eta}-\frac{\beta}{\theta-1} \\
\beta(\sigma-1)\mu_A+\sigma\mu_U &\leq \frac{\sigma}{\eta}+\sigma(1-\beta)+\beta(1+(\sigma-1)(1-\alpha))
\end{aligned}$$

Relative to existing papers (Ahlfeldt et. al. 2015, Allen et. al. 2016), the model adds multiple groups of workers, industries and transit modes, along with two fixed expenditures that potentially influence the set of locations inhabited by each group. Despite these additional features, the first part of the proposition ensures an equilibrium still exists in the city. The second part of the proposition shows that in a special case of the model, the equilibrium is unique only if spillovers are sufficiently weak. In the presence of strong spillovers, fundamental productivities and amenities become less important and different urban configurations can be supported as equilibria. While multiplicity does not pose a problem for my estimation strategy (which only requires that two equilibria be observed), I address it through my equilibrium selection rule when solving for counterfactual equilibria.

4.7 Intuition for Welfare Effects

To build intuition for the channels through which changes in the transit network affect welfare, I totally differentiate the expression for average utility (6) and assume that $\bar{h} = p_a = 0$.[51] The change in utility in response to a small change in commute costs is given by

$$d\ln \bar{U}_g = \sum_{i,a} \lambda_{iag} \left(\underbrace{-\sum_j \pi_{j|iag} d\ln d_{ija}}_{\text{Partial Equilibrium}} + \underbrace{d\ln u_{iag} + \sum_j \pi_{j|iag} d\ln w_{jg} - (1-\beta) d\ln r_{Ri}}_{\text{General Equilibrium}} \right)$$

where $\lambda_{iag} = L_{Riag}/\bar{L}_g$ is the share of type-g individuals living in i under car ownership a and $\pi_{j|iag}$ is the conditional commuting probability.

There are both partial and general equilibrium effects of reductions in commute costs on worker welfare. The partial equilibrium effect is greater if costs are reduced between locations where many people live and work (reflected through λ_{iag} and $\pi_{j|iag}$ respectively). General equilibrium effects depend on the response of wages and amenities (which raise welfare) and residential floorspace prices (which reduce welfare). Standard approaches to measuring the benefits of improvements in commuting infrastructure capture only the partial effects through the value of time savings (e.g. Small and Verhoef 2007). I assess the importance of accounting for general equilibrium forces in quantitative exercises.

The effects of improvements in public transit differ across skill groups in a way that is ex ante ambiguous. First, within any residential location the low-skilled rely more on public transport and so put more weight on the improvement. Second, worker groups can differ in the elasticity of commute flows to commute costs (controlled by θ_g). The group with less sensitive commute decisions is more willing to tolerate high commute costs in the initial equilibrium, and thus put more weight on costly commutes (reflected through $\pi_{j|iag}$). If the percentage drop in commute costs as a result of the improvement is greater for longer

[51] With non-homotheticities, the expression is very similar. Notably, there is an additional coefficient on the change in house prices that increases with income: this reflects that poorer individuals spend a greater share of income on housing and are thus hurt more by price increases.

commutes, then welfare gains will be greater for this group. Third, low- and high-skill individuals live and work in different locations (reflected both through λ_{iag} and $\pi_{j|iag}$). This creates different exposure to house price appreciation resulting from the system. Moreover, individuals benefit most when commute costs fall between locations where they live and work (reflected through the product $\lambda_{iag} \times \pi_{j|iag}$).

The direction of these forces depends on parameter estimates as well as the geography of the city and its transit improvements. While this decomposition helps develop intuition, the estimated model allows me to quantify the net effect on worker welfare.

5 Using The Model To Guide Empirical Work

In this section, I show that in a special case the model's equilibrium has a reduced form representation in which outcomes such as population, employment and floorspace prices can be written as log-linear functions of CMA. In fact, I show this framework applies to a wide class of models featuring a gravity equation for commute flows and thus is robust to a number of alternative modeling assumptions.

I use the model-implied regression framework to empirically evaluate the effect of TransMilenio, and evaluate the performance of gravity models in predicting the changes observed in the data. However, since this simplified framework is unable to assess the distributional effects of the system across worker groups, I turn to structurally estimating the full model in Section 8.[52]

Commuter Market Access: Measurement and Intuition Consider a simplification of the model with one group of workers, firms and transit modes and a fixed allocation of residential and commercial floorspace. Using the labor supply curve (5) to substitute for

[52]One concern might be how "valid" this framework is as a model validation exercise, since it abstracts away from the multiple layers of heterogeneity I claim are in the data generating process. In the appendix, I conduct a Monte Carlo exercise in which I run the same regression specifications on data simulated from my full model. The two models are consistent in the sense that using the data simulated from the full model, the non-parametric relationships between outcomes and CMA in the simplified regression model are log-linear.

wages in the expression for RCMA in (2), the CMA terms can be expressed as the solution to the following system of equations

$$\Phi_{Ri} = \sum_j d_{ij}^{-\theta} \frac{L_{Fj}}{\Phi_{Fj}} \tag{14}$$

$$\Phi_{Fj} = \sum_i d_{ij}^{-\theta} \frac{L_{Ri}}{\Phi_{Ri}} \tag{15}$$

RCMA reflects access to well-paid jobs. It is greater when a location is close (in terms of commute costs) to other locations with high employment, particularly so when these other locations have low 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 have low access to jobs (lowering the wage individuals are willing to work there for).[53] I show below that the solution to this system of equations exists and is unique, 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.

Regression Framework In this simplified model, the equilibrium reduces to the following system

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

[53] These expressions are closely related to commute-distance weighted sums of employment and residence respectively, reminiscent of the discussion of accessibility in Hansen (1959). In the appendix, I reproduce my main results using these alternative measures to show robustness to measuring CMA using less model-dependent measures.

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

where $\tilde{\Phi}_{Fi} = \sum_i d_{ij}^{-\theta} \frac{L_{Ri}}{\Phi_{Ri}} \Phi_{Ri}^{\frac{1}{\theta}}$ is adjusted firm commuter access capturing access to effective units of labor.[54]

The first line determines the supply of residents given residential floorspace prices. The second line is a market clearing condition for residential floorspace, which provides an inverse demand equation for residents. Together, these supply and demand curves determine equilibrium in the market for residents. The third and fourth lines are labor supply and demand schedules that determine equilibrium in the labor market. The fifth line is a market clearing condition that determines equilibrium in the market for commercial floorspace.[55]

Taking logs, stacking the equations, substituting out for wages and considering long-differences between two time periods, the change in endogenous variables can be written as the following system

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

where $\Delta \ln Y = \begin{bmatrix} \Delta \ln L_{Ri} & \Delta \ln r_{Ri} & \Delta \ln r_{Fi} & \Delta \ln \tilde{L}_{Fi} \end{bmatrix}'$ is a vector of log changes in endogenous variables, A is a matrix of coefficients reflecting the interdependence between endogenous variables, B_R and B_F are vectors of coefficients controlling the direct effects of changes in market access on outcomes, and e is a vector of structural residuals containing changes in fundamentals $\bar{u}_i, \bar{A}_i, H_{Ri}, H_{Fi}$.[56] Premultiplying by the coefficient matrix A yields

[54] To economize on notation, amenities and productivity depend on residence and employment rather than their density. This doesn't affect the results since the units of land in the denominator are absorbed into the structural error.

[55] This can be substituted into the expression for labor demand to eliminate commercial floorspace prices from the system. I retain this since I explore the response of these prices to firm CMA in the empirics. The gravity generalization reduces this and isomorphic models into equilibrium systems in only population, employment, and market access.

[56] The explicit form of A and the reduced form residual $A^{-1}e_i$ is given in the appendix. Residuals for residential outcomes contain changes in exogenous amenities and residential floorspace, while those for commercial outcomes contain changes in exogenous productivities and commercial floorspace.

the reduced form

$$\Delta \ln Y_i = A^{-1}B_R \Delta \ln \Phi_{Ri} + A^{-1}B_F \Delta \ln \tilde{\Phi}_{Fi} + A^{-1}e_i \tag{16}$$

where the reduced form coefficients are given by

$$A^{-1}B_R = \begin{bmatrix} \frac{\beta\eta}{\theta(1+\eta(1-\beta-\mu_U))} \\ \frac{1+\eta(1-\mu_U)}{\theta(1+\eta(1-\beta-\mu_U))} \\ 0 \\ 0 \end{bmatrix}, \ A^{-1}B_F = \begin{bmatrix} 0 \\ 0 \\ \frac{\alpha+(\alpha(\sigma-1)+1)\mu_A}{\theta\sigma/(\sigma-1)+(\theta-1)((1-\alpha)(\sigma-1)\mu_A-\alpha)} \\ \frac{\sigma/(\sigma-1)}{\theta\sigma/(\sigma-1)+(\theta-1)((1-\alpha)(\sigma-1)\mu_A-\alpha)} \end{bmatrix}$$

The regression specification (16) reflects the total change of outcomes in response to changes in market access. This reflects both the direct effect (in the B_R and B_F coefficient vectors) and the indirect effect (in A^{-1}) as the response to improved CMA filters through labor and land markets. The block structure of reduced form coefficients means that residential (commercial) outcomes depend only on changes in residential (commercial) CMA, so that the specification for each outcome has a simple univariate specification.

The following proposition shows that this reduced form representation and the ability to retrieve measures of market access using only the gravity equation for commuting of commuters across the city is shared by a wide class of gravity models. For brevity, a complete formal statement of the proposition and its proof are provided in the appendix.

Proposition 2. ***(i)* Measuring CMA** *In a gravity commuting model with commute flows $L_{ij} = \gamma_i \delta_j \kappa_{ij}$ where $\gamma_i, \delta_j > 0$ are endogenous, the supply of residents and workers is log-linear and given by $L_{Ri} = \gamma_i \Phi_{Ri}$ and $L_{Fi} = \delta_i \Phi_{Fi}$, where Φ_{Ri}, Φ_{Fi} are uniquely determined by data $\{L_{Ri}, L_{Fi}\}$ and parameters $\{\kappa_{ij}\}$.*

***(ii)* Isomorphisms** *In a gravity commuting model with log-linear demand for residents and labor $\tilde{L}_{Fj} = A_j \delta_j^{\alpha}$ 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 $\tilde{\Phi}_{Fj} = \sum_i \frac{L_{Ri}}{\Phi_{Ri}}\kappa_{ij}\Phi_{Ri}^{\epsilon}$, in equilibrium residence and effective employment can be expressed as log-linear functions of $\Phi_{Ri}, \tilde{\Phi}_{Fi}$ and constants A_i, B_i. The equilibrium always exists and is unique when $|\epsilon(\beta-1)-\gamma| \leq |\beta-1||\alpha-1|$.

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.[57] The first part of 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. In the appendix, I show this framework accommodates iso-elastic housing supply, alternative production technologies (e.g. Eaton and Kortum 2002, and individual entrepreneurs who choose where to produce across the city) and worker preferences (such as utility over leisure). Thus, the regression framework I take to the data is robust to a host of alternative modeling assumptions.[58]

Relation to Market Access Literature In the economic geography literature, individuals live and work in the same location and goods trade is subject to trade costs. A class of models contain measures of market access reflecting the ease for consumers (firms) to buy (sell) goods from (to) other locations. This is well-suited to study the effects of barriers to goods trade between regions, but is silent on the effect of commuting infrastructure within cities. I consider a different class of urban models where goods trade is costless but indi-

[57]See McDonald and McMillen (2010) for a review of the evidence in support of gravity in commute flows. All the quantitative models in Section 2 feature gravity equations for commute flows. Finally, note that my full model only exhibits gravity in commute flows conditional on location of residence. I show in the next section that there exist unique measures of CMA in my model given the observed data. In fact, part (i) of proposition 2 can be easily extended to include models with gravity in commute flows conditional on location of residence.

[58]In the appendix, I also show the framework can be used when workers have preference rather than productivity shocks over employment locations, so that there is no difference between effective labor and the number of workers. In addition, most models impose additional restrictions between $\alpha, \beta, \gamma, \delta$ and ϵ which reduces the number of parameters one needs to know (see the appendix for examples).

viduals can live and work in separate locations. The previous exposition shows how these models contain measures of accessibility of residents (firms) to jobs (workers). The similarity is natural given that both rely on a gravity equation to model the flows of goods or factors across space.

The other key difference is that CMA can be calculated from observable data using less model structure than is typically required in economic geography settings. In my framework, the only condition I need is the gravity equation governing the supply of commuters across the city. In economic geography settings, additional structure such as symmetric trade costs, balanced trade and goods market clearing is typically required 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 deliver separate notions of resident and firm CMA in this paper. This distinction is important given that changes in firm and resident CMA capture very different sources of variation.[59]

6 Data

In this section I provide an overview of the primary datasets used in the analysis. Additional details are provided in the data appendix.

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.[60] These are contained within larger spatial units including 19

[59]In the economic geography literature, balanced trade and symmetric trade costs imply that firm and consumer market access collapse to a single measure. Imposing balanced trade in commuters in my setting would require the number of workers in a location to equal the number of residents, which is clearly 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. This is available between countries, but rarely between small regions within countries. While Bartelme (2015) only requires symmetric trade costs and balanced trade, Donaldson and Hornbeck (2015) also impose goods market clearing. The reason is that they only observe population rather than expenditure in each location. In contrast, I use only the gravity equation determining the supply of commuters to recover the CMA measures directly from data residential population and employment, which are typically available in cities from censuses and new sources such as cellphone metadata records.

[60]Number reported is for tracts with positive population. Almost all tracts (2,768) have positive population

localities and 113 planning zones (UPZs).

My 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. I define college-educated individuals as those with some post-secondary education defined by their last complete year of study. In 2015, DANE provides population totals at the UPZ. I combine this with the share of college-educated workers in each UPZ in the GEIH survey in that year (described below) to construct population by skill group. This allows me to compute separate growth rates of college and non-college residents between 2005 and 2015 within each UPZ. I then calculate 2015 census tract population by skill group by inflating the 2005 totals by these growth rates. Details are provided in the appendix.

I use two sources of data on employment. 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 establishment size is not provided. While I tend to use the census and CCB datasets separately, a concern is that the spatial distribution of registered employment may be different from that of total employment. In the appendix, I show 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 different types of neighborhoods, suggesting both that the CCB data is representative of overall employment. Since I rely on establishment counts to proxy for employment in the CCB data, I also show that establishment count and employment densities are highly correlated in years where both are available.

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

in 2005. For comparison, tracts in Bogotá are about 60% smaller than those in New York City which had an average of 4,067 residents in the 2010 census.

properties (Ruiz and Vallejo 2015).[61] 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.[62] 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. As described in the appendix, 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).[63] As a result, I show 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.

[61] I confirmed this high coverage by overlaying the shapefile for available properties over satellite images of the city.

[62] For example, in 2008 property taxes accounted for 19.8% of Bogotá's tax revenues (Uribe Sanchez 2015).

[63] 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.

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.[64] To construct the time to traverse each edge of the network, I assign speeds in order to match both reported values in the literature as well as the distribution of commute times observed in the Mobility Surveys.[65]

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).

7 Empirical Analysis

In this section, I use the log-linear relationships between endogenous outcomes and CMA derived in Section 5 to empirically assess the effect of TransMilenio on land and labor markets.

[64]For example, the TransMilenio network is the union of pedestrian paths, trunk lines and feeder routes; the latter two can only be entered at stations. As described in detail in the appendix, one mode may have different speeds depending on the part of the network. For example, cars have different speeds on primary, secondary and tertiary roads.

[65]While I provide evidence speeds were not changing on routes affected by TransMilenio relative to other locations, the appendix shows that aggregate speeds fell between 1995 and 2005 (a period of city expansion) but remained relatively constant thereafter. I assign speeds to separately match the commute data for each period, and use the average computed time in the main analysis. In robustness exercises, I run specifications with alternate commute times to ensure my results are not sensitive to this choice.

7.1 Approach and Identification

Taking logs of the expression for residential outcomes in the first two entries of the reduced form system (16) delivers my baseline specification

$$\Delta \ln Y_{Rit} = \beta \Delta \ln \Phi_{Rit} + \alpha_\ell + \gamma' X_{it} + \epsilon_{Rit} \tag{17}$$

That is, I regress changes in (log) residential outcome Y_{Rit} in census tract i in year t on changes in (log) RCMA Φ_{Rit}, as well as a set of controls that contain census tract characteristics X_{it} as well as locality fixed effects α_ℓ. An equivalent specification holds for commercial outcomes, which instead depend on FCMA. These CMA terms are defined by the system of equations in population, employment and commute costs $d_{ijt}^{-\theta}$ in (14) and (15). The elasticity β is 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.[66] Typically this regression will be estimated in long-differences over a pre- and post-period.

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.[67] 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.

[66] Of course, the reduced form elasticities are outcome-specific but I omit this additional notation here. Note also that for firm outcomes, I use the unadjusted firm commuter access instead of the adjusted term that reflects units of effective labor supplied for clarity. The results are qualitatively unchanged if I use the adjusted term (the measures have a 0.99 correlation).

[67] In order to compute the market access terms, I require values for $d_{ij}^{-\theta} = \exp(-\theta\kappa t_{ij})$. The estimation of θ_g, κ is outlined in the next section; I measure θ and t_{ij} by averaging over skill group and car ownership values respectively weighting by population shares of each category. 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.

Challenges to Identification There are two key challenges to estimating the specification (17).

First, changes in CMA contain population and employment in both periods. Local productivity and amenity shocks (contained in the error term) that drive movements in residence and employment will therefore be mechanically correlated with changes in CMA. I address this by instrumenting for changes in CMA holding population and employment fixed at their initial values in the system (14) and (15).[68] This isolates the variation in CMA due to the change in commute costs.

Second, growth in CMA 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 those that were lagging. I therefore construct two instruments for TransMilenio routes as described below (further details are in the appendix). These in turn allow me to instrument for the change in CMA as follows. In ArcMap, I compute the commute times had the system been built along each instrument. Plugging these into (14) and (15) 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 defined as the difference between this predicted CMA under TransMilenio and its value in the initial period without the system.

The first instrument takes as given the government's overall strategy of connecting portals at the edge of the city with the CBD as given, excludes those areas from the analysis, and constructs the routes that would have been built if the sole aim had been to minimize costs. I construct these routes by first digitizing 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

[68]I exclude the location itself when calculating its predicted change in CMA to address the possible correlation between initial residence and employment and unobserved shocks to changes in residential and employment outcomes. In robustness checks, I extend this to exclude all tracts within 1.5km of a location.

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 in ArcMap. This will be a valid instrument under the reasonable assumption that these routes should be uncorrelated with trends in amenities and productivities (conditional on controls).

The second instrument exploits the location of a tram system opened in 1884, which was last extended in 1921 and stopped operating in 1951. The tram was built along wide arterial roads in the city, which should predict the location of TransMilenio since these are cheaper to convert to BRT than narrow ones. Moreover 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.[69]

My 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, might have a direct effect on outcomes. A key feature of my approach is that I can control for distance to the instrumented routes (distance to the tram, distance to main roads) and use only residual variation in predicted CMA growth. This helps controls for direct effects of the instruments, and is not possible in a distance-based approach since these control variables are themselves the instruments. I include additional historical variables to control for other direct effects of the historical instrument.[70] 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.

[69] Similar approaches have been used in the trade and economic geography literature to predict the location of present day infrastructure based on historical instruments (Baum Snow 2007; Duranton and Turner 2012) or construction costs (Faber 2015, Alder 2017).

[70] These include 1933 population and distance to main roads in 1933. To further reduce concerns over direct effects of the tram on outcomes, I extend the tram lines to the edge of the city (which also greatly improves its predictive power over TransMilenio placement).

7.2 Results: Main Outcomes

Main Outcomes Table 3 presents the main results.[71] 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. Columns (1) and (2) report the results from the OLS regressions where the change in CMA is measured using both the change in commute costs as well as the change in population and employment. In most cases, the point estimates are slightly lower in column (2) (my preferred specification with full controls) due to the positive correlation of changes in market access with initial land market and demographic characteristics that caused treated locations to grow faster over the period.

Columns (3) and (4) run the baseline IV specification, which instrument for the total change in market access holding employment and population fixed at their initial levels. The point estimates tend to fall slightly, reflecting the positive mechanical correlation previously discussed.[72]

Columns (5) and (6) instrument for the change in market access both by 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

[71] 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 will 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. For example, population regressions using differences between 1993 and 2005 measure changes in market access induced by phase one (opened between 2000 and 2003). 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). I explore whether future station openings predict prior growth in outcomes in falsification regressions below. My main 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.

[72] F-Stats are not reported for clarity; they are extremely high in columns (3) and (4).

the 1990s and may therefore 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 remain constant as additional controls are added provides additional evidence in support of the exclusion restriction holding.

Finally, columns (7) and (8) use both the tram and LCP instruments. The coefficients remain stable compared to using the LCP instrument alone, and in all but one case I fail to reject validity of the overidentification restrictions.

Heterogeneous Effects of Transit Figure 5 plots the non-parametric relationship between (residual) growth in outcomes and (residual) changes in CMA. The relationship appears approximately log-linear for each outcome, supporting the functional form 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 In the appendix, I report a number of additional results which I summarize here.

First, I use less model-dependent measures of resident and firm CMA. 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).[73] The results are robust to using this alternative measure, suggesting my findings are not sensitive to using the structure implied by gravity models. That the coefficient from this measure differs statistically from the coefficient on CMA suggests that the adjustments in the gravity-based definition capture meaningful variation. Second, I run falsification tests to check that changes in CMA induced by particular lines are not associated with growth

[73] 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.

in outcomes before they open. Third, I condition on distance to stations to show that the effects are driven by changes in market access rather than characteristics of stations (such as changes in foot traffic, pollution and complementary infrastructure). Fourth, I assess the response of variables to changes in market access to distant locations more than 1.5km away. Both of these empirical approaches are not possible with a distance-based empirical approach. Fifth, I use alternative speeds to compute the commute times for each mode. Sixth, I vary the commute elasticity θ to 1.5 and 0.5 times its estimated value. Seventh, I include all census tracts in the analysis, rather than those within 3km of a station. Eighth, I run unweighted regressions for employment and population regressions which are weighted in the main specifications. Finally, I use Conley (1999) HAC standard errors (compared to the baseline estimates which cluster by census tract) to allow for arbitrary spatial correlation of errors across tracts within 500m of each other. That my results are robust to these alternative specifications provides additional evidence in support of the causal effect of TransMilenio on urban outcomes through improvements in CMA.

Comparison with Distance Band-Based Predictions In the appendix, I compare the predictions for residential house price growth in the CMA model with those from a distance-based regression of the change in floorspace prices on two dummies for being closer than 750m from a station and between 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).[74]

7.3 Results: Additional Outcomes

Commute Distance Table 4 examines whether TransMilenio led to changes in commuting distances. Column (1) shows that changes in market access caused by TransMilenio

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

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 (17) 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 CMA led to increases in commute distances, suggesting the system made employment in more 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 likely reflects both their greater reliance on public transit as well as a greater sensitivity of commute flows to commute costs as shown in the next section.

College Share A key question surrounding the debate on the effects of public transit is whether it leads to a re-sorting of worker (skill) 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). However, 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 heterogeneous responses across types of census tracts. For example, the model predicts that the high-skilled are more willing to pay for improved access to jobs in neighborhoods with high amenities.[75] In columns (2) to (4), I test whether the response differed by tracts according to the college share of the surrounding neighborhood.[76]

[75] Assuming one mode of transit for the moment and the same wage for skill groups (i.e. $\sigma_L \to \infty$), log-linearizing the expression for residential populations (4), the change in residential population can be written as

$$\Delta \ln L_{Rig} \approx \eta_g \frac{\mu_{ig}^Y}{\theta_g} \Delta \ln \Phi_{Ri} - \eta_g (1 - \beta + \mu_{ig}^R) \Delta \ln r_{Ri} + \eta_g \Delta \ln u_{ig}$$

where $\mu_{ig}^Y = \frac{T_g \Phi_{Ri}^{1/\theta_g}}{T_g \Phi_{Ri}^{1/\theta_g} - r_{Ri}\bar{h}}$ and $\mu_{ig}^R = \frac{r_{Ri}\bar{h}}{T_g \Phi_{Ri}^{1/\theta_g} - r_{Ri}\bar{h}}$. Note μ_{ig}^R is greater in expensive neighborhoods and when individuals are poor. Thus poor, low-skilled workers are more sensitive to house price appreciation in expensive neighborhoods and are less willing-to-pay for improved CMA than the high-skilled.

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

The results show that the college share of a tract's residence 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 would not trade off these benefits for the lower amenities in poorer locations in the South. In contrast, the low-skilled were more likely to move into neighborhoods with a lower initial college share. Overall, this shows that TransMilenio increased residential segregation between the low- and high-skilled.

Wages Lastly, Table 6 examines the impact of market access on wages reported by individuals across UPZs. I run a difference-in-difference specification similar to (17) to examine how the effect of residential market access on log average hourly wage over the past month reported by full-time workers between 18 and 55 who usually work at least 40 hours a week. 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 than for the low-skilled. [77]

Unfortunately the labor surveys are only cross-sectional, so I am unable to distinguish changes in wages within individuals from those due to re-sorting. These results should therefore be interpreted with caution. With this caveat, 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-skill workers suggests they may benefit more

around each 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, but this is subject to a mechanical bias from including the lagged value of the dependent variable. The results are also robust to proxying for college share using a dummy for localities in the North of the city, since this is where most of Bogotá's college educated live (Figure 2).

[77] Effects on other labor market outcomes are available upon request. In summary, 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.

(in terms of increased income) from improved transit, a topic I return to in the quantitative section.

8 Structural Estimation

Having empirically established the causal effect of TransMilenio on land and labor market outcomes through improved CMA, in this section I structurally estimate the full model from Section 4. Since this model contains multiple groups, it allows a quantitative assessment of the distributional effects of TransMilenio.

The section proceeds as follows. I first describe 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. I then outline the procedure to estimate the model's parameters. Finally, I present the estimation results and model diagnostics.

8.1 Model Inversion

The model contains location characteristics, such as productivities, amenities and land use wedges, that are unobserved but needed to solve for counterfactual equilibria. While the presence of agglomeration forces allows for the possibility of multiple equilibria, a key advantage of my approach is that I am able to recover unique values of composite productivities and amenities that rationalize the observed data as an equilibrium.

There is a key difference in the process to solve for unobservables between this paper and recent quantitative urban models (e.g. Ahlfeldt et. al. 2015). In those models, there is one group of workers. Given data on where individuals live and work, it is straightforward to solve for the vector of wages that rationalize the distribution of residence and employment given observed commute costs. With wages in hand, recovering the remaining unobservables from the model's equilibrium conditions is straightforward.

In a model with different skill groups of workers, one would need data both on where skill groups live and work for this procedure to work. While data on residence by skill group are typically available from population censuses, I am unaware of similar datasets that provide employment by skill group within small spatial units within cities. This is where the model's multiple industries becomes useful. In the data I observe employment by industry. Intuitively, given the different skill-use intensities across industries, the relative employment by industries in a location should be informative about the relative employment across skill groups. The following proposition formalizes this intuition, and shows that a unique vector of group-specific wages can be recovered using data on residence by skill and employment by industry. Obtaining the remaining unobservables is straightforward.

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

(ii) Productivities and Amenities *Given the 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 composite amenities and productivities $\{u_{iag}, A_{js}\}$ which rationalizes the observed data as an equilibrium of the model.*

8.2 Parameter Estimation

My procedure to estimate the parameters of the model can be summarized as follows:

Step 1. Calibrate and estimate a subset of parameters without solving full model.

Step 2. Solve for wages using parameters from step 1.

Step 3. Estimate remaining elasticities via GMM using moments similar to reduced form analysis.

Step 4. With all parameters in hand, invert the model to recover unobservables.

I now describe my procedure to structurally estimate the parameters of the model.

8.2.1 Parameters Calibrated to Exogenous Values

I calibrate $\{\sigma, \sigma_D, \alpha_s\}$ to existing values from the literature. I set the elasticity of substitution between labor skill groups to $\sigma = 1.3$ 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. I set this to be equal across industries. I set the elasticity of substitution of demand 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 how I estimate the parameters $\{\beta, \alpha_{sg}, \kappa, \theta_g, \rho_g, T_g\}$ using relationships from the model.

8.2.2 Parameters Estimated without Solving the Model

Share Parameters I estimate $1 - \beta = 0.24$ to match the long-run housing expenditure share in Bogotá.[78] I estimate the labor shares α_{sg} 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. Since I assume that firms outside Bogotá aggregate labor using Cobb-Douglas technology, these labor cost shares identify α_{sg}.

Commute Costs The appendix 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 chosen where to live and work individuals choose which mode of transport to commute with given their idiosyncratic preference for each mode. These preferences are drawn from a Generalized Extreme Value distribution that allows for a nested preference structure across public and private nests. The commute time

[78] See the appendix for the Engel curve for housing and details on its estimation.

between each pair of census tracts for each mode is computed in ArcMap. This allows me to estimate κ by estimating the mode choice model to the commute microdata by Maximum Likelihood. This is identified from the sensitivity of individuals' mode choice decisions to differences in their commute times for a particular commute. I do so using the 2015 Mobility Survey.

Table 7 reports the results. I obtain an estimate of $\kappa = 0.012$, very close to the estimate of 0.01 in Ahlfeldt et. al. (2015). The last entry reports the estimate for the parameter λ governing the correlation of preference shocks within the public nest. When $\lambda = 1$ there is complete independence between draws in the public nest, while $\lambda = 0$ implies perfect correlation. The estimate of $\lambda = 0.14$ suggests idiosyncratic preferences are highly correlated within public nests i.e. commuters on public transport tend to take the quickest mode. Given this order of magnitude difference in sensitivity to commute times, my baseline specifications assume that users take the quickest public mode of transportation available but imperfectly substitute across cars and public transit.[79] The remaining parameters reflect average preferences for each mode relative to walking (conditional on commute time). Intuitively, cars are most attractive followed by buses and TransMilenio. That TransMilenio is least desirable likely reflects the high crowds using the system as well as the inconvenience of having to walk between stations and final origins and destinations.

Skill Distribution The gravity equation for commute flows in (2) combined with the specification of commute costs $d_{ija} = \exp(\kappa t_{ija})$ implies a semi-log gravity equation for (conditional) commute flows

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

where γ_{iag} and δ_{jg} are fixed effects and ε_{ijag} is an unobserved component of commute times.

[79]In the appendix, I explore the sensitivity of my results to alternative aggregation methods. Note that, as can be seen from the mode choice model in the appendix, the semi-elasticity of mode choices to commute times is κ for public vs private nests and κ/λ for modes within the public nest.

In order to leave sufficient residual variation conditional on fixed effects, I aggregate to the locality level and use commuting data from the 2015 Mobility Survey. θ_g is identified from the sensitivity of commuting decisions to commute costs, conditional on trip origin, trip destination and car ownership. Given the presence of zeros in the data, I estimate the model using Poisson Pseudo-Maximum Likelihood (PPML) as suggested by the trade literature on gravity equations (e.g. Santos Silva and Tenrayo 2006).

While the fixed effects absorb any unobserved factors varying by origin and destination, one concern is whether there are origin-destination specific unobserved components of commute costs correlated with trip time. I address this in two ways. First, I include direct measures of factors other than time that may determine the attractiveness of a commute and examine how this changes the results. I measure the average number of crimes along a route, the average house price, as well as the share of the trip that occurs along a main road.[80] Second, I instrument for commute times using the LCP and Tram instruments.

The results are reported in Table 8. Column (1) shows that high-skill workers are less sensitive to commute costs than low-skill workers with a semi-elasticity of -0.0242 compared to -0.0336.[81] In column (2) I control for other observable factors that may affect the cost of commuting. These characteristics are not significant determinants of commute flows, and the point estimates are unchanged. In columns (3) and (4) I instrument for commute times and find that the estimates are remarkably stable. Taken together, these findings suggest the following. First, having controlled for unobservables that vary by origin and destination, commuting decisions are primarily driven by commute times rather than other observed non-time factors. Second, the stability of OLS and IV point estimates suggests unobserved

[80] Using car trips for car owners and bus trips for non-car owners, I compute the least-cost routes between each origin-destination pair in 2015. I then intersect these routes with a 50m buffer around primary roads to compute the share along a primary road, as well as census tracts. Since the latter provides the share of each route lying within each census tract, I merge this with the log average house price in 2012 and the average number of crimes between 2007 and 2012 to create averages for each route.

[81] This does not imply one should observe the high-skilled taking longer commutes. It means that from any location of residence, the low-skilled will be less willing to commute to locations with high commute costs ceteris paribus. Average commute times and distances are greater for the low-skilled in Bogotá (as in many other cities) since they live further from high employment densities in central areas.

determinants of commute costs are not correlated with commute times (conditional on fixed effects). In other words, the endogenous placement of TransMilenio was driven by origin and destination unobservables rather than origin-destination specific unobservables.[82]

Given the estimate of κ, the point estimates from column (4) correspond to $\theta_H = 2.054$ and $\theta_L = 2.840$. Both the overall magnitude and the 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).[83]

8.2.3 Parameters Estimated Solving the Full Model

It remains to estimate the parameters $\{\bar{h}, p_a, T_g, \eta_g, \mu_A, \mu_{U,g}\}$.

In the appendix, I show that given the parameter estimates in the previous section, there is a unique vector of parameters $\{\bar{h}, p_a, T_g\}$ that matches the average expenditure share on housing, the average expenditure on cars, and the college wage premium respectively. I solve for them in the process of recovering the model's unobservables, and allow them to vary in each period to exactly match each wave of data.

The final step is to solve for the residential supply elasticity η_g and spillover parameters

[82]My cross-sectional results suggest the observed driving times are uncorrelated with other factors that might affect commuting behavior conditional on origin-skill-car ownership and destination-skill fixed effects. However, in the appendix I also estimate the relationship using changes in commuting patterns between 1995 and 2015 to difference out any time-invariant original-destination-skill-car ownership unobservables. There are two reasons why I do not use this as the baseline specification. First, there is a pronounced city-wide reduction in car and bus speeds between 1995 and 2005. While I show this is uncorrelated with TransMilenio routes, a concern is that changes in computed times driven by changes in driving speeds will be greater for longer trips which introduces an endogeneity problem of its own. Second, work has shown that 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) which is the case in my cross-sectional specification with fixed effects at the origin-car ownership-group and destination-group level. However, I am not aware of any result for the case with three fixed effects that applies in the time-differenced specification (this includes the two fixed effects from the cross-sectional regression interacted with time dummies, plus an origin-car ownership-destination-group fixed effect). IV-PPML also failed to converge in the two-period model. Regardless, I report the results from the PPML as well as a least squares model (where both OLS and IV estimators converge) in the appendix. The PPML point estimates are very similar to those from the cross-section, and I show in robustness exercises that the quantitative results are qualitatively unchanged when using these alternative θ estimates.

[83]I acknowledge the estimation procedure does not account for noise introduced by prior estimates. In ongoing work I am estimating all parameters jointly. Finally, while the only parameter that matters for the computation of equilibria is the ratio $\theta_g = \tilde{\theta}_g/(1-\rho_g)$, I am also able to use wage data to separate $\tilde{\theta}_g$ and ρ_g in my setting. I report the results and procedure in the appendix.

$\mu_A, \mu_{U,g}$. While the parameters estimated so far were identified using cross-sectional data, these require exogenous variation in the density of residence of skill groups and employment across the city. I therefore exploit 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 (4) delivers the following expression for residential population growth across skill groups

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

where $\Delta \ln V_{iag} \equiv \Delta \ln \left(T_g \Phi_{Riag}^{1/\theta_g} - r_{Ri} \bar{h} - p_a a \right) - (1-\beta) \Delta \ln r_{Ri}$ is the change in indirect utility from living in (i, a) net of changes in amenities, γ_ℓ and Cont_j are locality fixed effects and tract characteristics (to partially control for changing fundamentals) and $\Delta \ln \epsilon_{Riag}$ is a residual that reflects unexplained growth in productivity (i.e. residual variation in $\Delta \ln \bar{u}_{iag}$). To identify η_g, I require a source of exogenous variation in the common component of utility from living in a location $\Delta \ln V_{iag}$. To identify the strength of spillovers $\mu_{U,g}$, I require a separate source of exogenous variation in the college share of residents $\Delta \ln L_{RiH}/L_{Ri}$.

I instrument for the change in indirect utility using the instruments for the change in RCMA. Two additional instruments provide separate variation in the share of college residents. First, tracts which 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 $Z^k_{Diff,i} = \Delta \ln \bar{\Phi}^k_{RiH} - \Delta \ln \bar{\Phi}^k_{RiL}$ for $k \in \{LCP, Tram\}$.[84] Second, I augment this by interacting the change in CMA for high-skilled residents with the house price in the initial period $Z^k_{Rents,i} = \Delta \ln \bar{\Phi}_{RiH} \times \ln r^{2000}_{Ri}$. That this should differentially predict entry of high-skilled residents is a direct consequence of log-linearizing the expression for residential populations (4), which implies that poorer, low-skilled residents are less likely

[84] I define $\bar{\Phi}^k_{Rig} \equiv \sum_a \Phi^k_{Riag}$ to be the sum of RCMA across car ownership within a location-skill group.

to move into expensive neighborhoods due to their increased expenditure on housing.[85] The moment condition I use to identify η_g and $\mu_{U,g}$ is therefore[86]

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

Productivity Moment Recall that firm sales are given by $X_{js} \propto \left(W_{js}^{\alpha_s} r_{Fj}^{1-\alpha_s}\right)^{1-\sigma_D} A_{js}^{\sigma_D - 1}$. Commercial floorspace prices are observed. Wages are recovered from model inversion in proposition 3 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 $A_{js} \propto W_{js}^{\alpha_s} r_{Fj}^{1-\alpha_s} X_{js}^{1/(\sigma_D - 1)}$ 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. Recall from (12) that composite productivity depends on employment as well as location fundamentals $A_{js} = \bar{A}_{js}(\tilde{L}_{Fj}/T_j)^{\mu_A}$. Taking logs of this expression and including a set of control variables to (partially) capture changing fundamentals yields

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

where γ_ℓ and Cont_j are locality fixed effects and tract characteristics and $\Delta \ln \epsilon_{Fjs}$ is a

[85] See footnote 75 for exposition.

[86] I also include orthogonality conditions with each control variable, and demean each variable by locality prior to estimation to purge out fixed effects. My 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). I explore robustness to using both phases 1 and 2 in robustness checks.

residual that reflects unexplained growth in productivity (i.e. residual variation in $\Delta \ln \bar{A}_{js}$).

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. Thus, TransMilenio 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 I use to identify μ_A is therefore[87]

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

8.2.4 GMM Results

Main Results Table 9 presents the main results.[88] Three comments are in order. First, the estimate of the productivity externality of 0.237 is large. Ahlfeldt et. al. (2015) obtain an estimate of 0.07 using a similar framework in Berlin, while the estimates in the literature have tended to lie within the 0.03-0.08 range reviewed in the survey by Rosenthal and Strange (2004). However, other experimental approaches in the US have obtained estimates as high as 0.12 (Greenstone, Hornbeck, and Moretti 2010) and 0.2 (Kline and Moretti 2014). The majority of existing evidence is within the US and other developed countries. The returns to agglomeration may be higher in developing countries due to factors like a lack of road

[87]I define $\bar{\Phi}^k_{Fig} \equiv \sum_a \Phi^k_{Fiag}$ as before. I include orthogonality conditions with each control variable, and demean each variable by locality prior to estimation to purge out fixed effects. Moreover, since the unit of observation differs across firm outcomes (tract-industry) and residential outcomes (tract-skill group-car ownership), and there are no interdependencies between parameters across moment conditions (i.e. μ_A only affects the productivity moment condition while $\eta_g, \mu_{U,g}$ only affect the amenity moment conditions), I estimate the equations separately via GMM rather than in one joint system, since it avoids the need to make arbitrary aggregation up to consistent units without any loss of information.

[88]I provide a set of 10 different starting values for the solver drawn from a uniform grid between 0 and 10 for each parameter, and find the solver converges to the same minimum each time. This suggests the objective function is well-behaved and my estimates do not result from a local minimum.

infrastructure or high crime, both of which are certainly at play in Bogotá. To my knowledge, this is the first intra-city estimate of agglomeration in less developed countries using quasi-experimental variation. However, in counterfactuals I turn spillovers off completely as well as set them to a smaller values similar to Ahlfeldt et. al. (2015) to ensure my quantitative results are not driven by this estimate alone.

Second, the residential population elasticity is greater for low-skilled than high-skilled. I interpret this as reflecting other factors such as home ownership that make the residential locations of the high-skilled more sticky. These elasticities are larger than the commute elasticities, underscoring the benefits from estimating residential and employment location decisions as a two stage problem.

Third, the spillover parameters for residential amenities are around twice as large as those in Ahlfeldt et. al. (2015), and larger for high-skilled. It appears the share of college-educated residents in a tract increases the amenities from living there, the high-skilled value living around each other more than the low-skilled, and these endogenous forces appear stronger in Bogotá than existing evidence for developed countries.

Robustness In the appendix, I check the robustness of these estimates. First, I control for log distance to the closest TransMilenio station (instrumented using the log distance to the instruments). I find TransMilenio stations decrease both productivities and amenities, likely reflecting increased foot traffic and pollution near stations. Second, I show the estimates are qualitatively similar when measuring the TransMilenio network as of 2006 (rather than 2003). Since the raw tract-level population data is from 2005, my preferred specification uses changes due to the first phase of the system. Finally, I vary the elasticity of substitution of demand (from 4 to 9) and elasticity of substitution between skill groups (from 1.3 to 2.5).[89] The point estimates are largely robust to the elasticity of substitution of labor, but the agglomeration point estimate is mechanically related to the demand elasticity

[89]The value of 2.5 is estimated by Card (2009) for skill groups using regional data in the US.

as evident in the moment condition above. While my preferred estimate lies in the middle of the observed range, this underscores the need to check the robustness of my quantitative results to alternate values of this parameter.

8.3 Non-targeted Moments: Model vs Data

In this section, I evaluate the performance of the model by comparing the model's predictions for moments not targeted in estimation.

Wages Figure 6 compares the average wage for each skill group earned by residents of each locality with that observed in 2014 in the GEIH data. The latter was not used in the procedure to estimate parameters determining wages. We see the two variables are highly correlated with values of 0.528 for non-college and 0.592 for college workers. However, while most observations lie along the 45-degree line for low-skilled workers, there is noticeable deviation for the richest localities amongst high-skill workers. 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 In the model, amenities and productivities represent characteristics that make locations more or less desirable to individuals and firms who might choose to locate there. Panel A of Table 10 shows that neighborhoods with less crime are associated with higher amenities. Panel B shows that productivities are higher in tracts with less crime, a flatter slope and a higher density of roads.[90] Overall, the model performs well at capturing features that affect the desirability of locations in the city.

Commute Flows The model solves for commute flows by first recovering wages that

[90]Road density could of course affect productivity directly rather than through affecting the supply of labor as emphasized through commuting in the model. For example, better roads might make it easier to ship goods from or order supplies to offices. Slope might affect productivity through delivery accessibility in the same way.

rationalize the observed distribution of residential population by skill and employment by industry, and then predicts the commute flows between origin, destination and car ownership pairs according to the gravity equation (2). I test the performance of the model's assumptions by comparing these implied commute flows with those observed within each cell in the 2015 Mobility Survey (again aggregating to the locality level). Other than the share of car owners in each UPZ, this data was not used in estimation or in solving for the model's unobservables. We see that the model performs very well matching the commute flows observed in the data, even when looking within car ownership groups (which will fit well by construction). Most importantly, the fit is even across college and non-college workers, suggesting the method used to back out wages by skill group using the location of employment by industries performs well in predicting commute flows.

Employment By Skill Group To provide more evidence that the model performs well in fitting the distribution of employment by skill groups in Bogotá, I compare 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 baseline model in which labor skill groups are perfect substitutes and share the same commute elasticity (set to the average value). 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 workers all receive the same wage across the city and have the same sensitively 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 Northern UPZs close to where they live, and vice versa for the poor who live in the South. This pattern is clearly counterfactual to the distribution in the data shown in panel (c). By contrast, the baseline model performs much 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 is 0.406 compared to 0.256 in the simplified model.

9 Quantifying the Effect of TransMilenio

In this section, I use the estimated model to quantify the impact of TransMilenio by simulating the effect of its removal from the 2012 equilibrium.[91]

9.1 Removing the System

Main Results Table 11 presents the effect of TransMilenio on GDP, total rents and welfare. Each entry reports the negative of the percentage change in each variable from removing the first two phases of the system. 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 3.12%-3.92%, total city rents by 3.29%-3.72% and worker welfare by around 3.5-3.9%, the higher number referring to the case with spillovers.

In the open city, welfare is fixed to the reservation level in the wider economy. Instead, gains to workers can be read off of changes in population. The effects of TransMilenio are large, increasing the population of the low-skilled by 8.56%-10.74% and the high-skilled by 9.54%-12.30%. Given the increase in factor supply, it is no surprise that the increase in GDP between 10.34%-15.59% is much larger than in the closed city. The population influx fuels greater house price appreciation: gains shift from workers to land owners who see total rents rise between 13.15-16.28% due to TransMilenio.[92]

[91] I refer to the "2012 equilibrium" as the post-TransMielnio 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. See the appendix for additional information on the algorithm used to solve for counterfactual equilibria. As highlighted by Proposition 1, there may be multiple equilibria in the presence of spillovers. The selection rule I use 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.

[92] The relative magnitude of the effects on outcomes such as output and floorspace values across open and closed city models is similar to those in Ahlfeldt et. al. (2016) in the context of changing the subway network in Berlin.

TransMilenio improved the spatial allocation of employment and residence. Figure 9 plots the change 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 with high employment to grow the most, the system's efficiency gains are driven by an improvement in the spatial allocation of labor. Panel (b) shows similar patterns hold for residence, but the effects are more muted.

The model captures the increase in residential segregation documented in the reduced form analysis. In panel (c) of Figure 9, I plot the interaction terms from a regression of the change in college share on a full interaction between a dummy for a tract's initial college share quantile and the change in log RCMA. This is the analogous to the regression in Table 5, only this time using the model to produce counterfactual data for the city's structure without TransMilenio. Tracts in the top three quantiles of the initial college share distribution increase their share of high-skill residents in response to improved transit. In contrast, those in the bottom two quantiles experience a net inflow of low-skill workers. The total effect on residential segregation is mild, though: the index of dissimilarity for residential locations increases by 0.28% due to TransMilenio.

Somewhat surprisingly, we see that in the closed city model the welfare of high-skill workers rises more than for the low-skilled, leading to a 0.37% rise in welfare inequality (defined as $\bar{U}_H/\bar{U}_L$). I now turn to understanding the channels through which these differential welfare gains accrue.

Decomposing the Channels Table 12 decomposes the channels through TransMilenio affects welfare. For each model, I begin by setting spillovers to zero and simulate the counterfactual equilibrium without TransMilenio using the unobservables retrieved from the 2012 equilibrium. In rows (1) to (6) of the table, I report the change in welfare from adding back the TransMilenio network allowing for different margins of adjustment. In rows (7) and (8), I report the change in welfare in full general equilibrium with productivity and amenity spillovers respectively.[93]

In Model 1, worker skill groups have the same elasticities θ and η (set to their average values) and enter as perfect substitutes in production. Skill groups face the same wage in each location, and labor supply from any place of residence is identical. Differences in relative employment by skill are driven solely by variation in residential locations.

Row (1) adds TransMilenio holding all choices and prices fixed. The low-skilled benefit far more due both to their dependence on public transit and tendency to live in the outskirts where access to jobs improves the most. In rows (2) to (4), different choices are allowed to adjust in partial equilibrium (i.e. wages, rents and land use decisions are held constant). The ability to change where individuals work has the largest impact on welfare, followed by where to live and finally by car ownership choice. Altogether, this reorganization increases the average welfare gain from the system from 3.77% to 5.35%. In row (5), housing markets adjust. House price appreciation substantially reduces welfare gains to workers. Moreover, this attenuates the relative gains of the low-skilled precisely because house prices appreciate most where they live further from the CBD: the reduction in welfare inequality halves from 1.404% to 0.774%. Row (6) allows for full general equilibrium adjustment, so that wages also change in response to the shift in labor supply. Wages fall where more workers commute, so this lowers welfare. Finally, rows (7) and (8) set productivity and amenity spillovers to their estimated values and the results are qualitatively similar.

Taken together, the low-skilled benefit in this model due to their greater reliance on public transit, but are hurt more by house price appreciation in the outskirts. On net, they benefit the most from the system with inequality falling 0.885%.

In model 2 skill groups differ in their labor supply elasticities θ_g. In row (1), we see that this has an immediate impact on attenuating the relative welfare gain of the low-skilled.

[93] For each model, both the baseline and counterfactual equilibria are re-computed. This involves solving for updated values for $T_H, \bar{h}, p_a$ and w_{jg}. The results from the decomposition differ slightly when spillovers are set to their estimated values, since the counterfactual equilibrium without TransMilenio is different. While the decomposition in Table 12 is easiest to isolate each channel, in the appendix I report the analogous table with spillovers. I also provide additional details on the decomposition procedure, including alternative formulas to calculate the welfare change under each mobility assumption.

This highlights the crucial role played by θ_g in determining the welfare gains from transit: a lower elasticity implies the high-skilled are less willing to substitute between employment locations, and the incidence of commute costs falls more broadly on their shoulders.[94] The qualitative impact of the remaining channels remains similar. Differences in θ_g are enough to reduce the drop in inequality from 0.885% to 0.185%.

Model 3 allows groups to also differ in their residential elasticities η_g. This pushes the gains slightly more in favor of the high-skilled, who are less able to substitute between residential locations, but the overall results remain qualitatively similar.

Lastly, model 4 considers the full model where skill groups are imperfect substitutes in production. Wages for low- and high-skill workers now differ across the city due to differences in the location of high- and low-skill intensive industries. Relative to the previous columns, this accounts for whether the system connects workers with the "right" jobs. High- (low-) skill workers benefit most if they are connected with well-paid jobs in high-skill (low-skill) intensive industries. This skews the gains even more in favor of the high-skilled, for whom TransMilenio connects the high concentration college-educated neighborhoods in the North with the dense high-skilled intensive jobs in the financial district and CBD. In contrast, low-skill intensive industries are located in more dispersed locations throughout the South close to where the low-skill already live. Additionally, since skill groups are imperfect substitutes in production the reduction in wages due to the labor supply shift more acutely affects low-skilled workers who use public transit the most. Taken together, these features have a large effect on inequality. In combination with the differences in commute elasticities high-skilled residents benefit more from the system.

To summarize, there are three key channels determining the incidence of public transit across worker groups. The first is mode choice: the group that relies on public transit

[94]Note that θ_g impacts both the substitution patterns as well as the calibrated wages. In results available upon request, I compute the welfare gains in an intermediate scenario in the specification in row 1 by using the wages from Model 1 but the θ_g from Model 2 to compute the welfare gains. The majority of the observed drop in inequality observed in Table 12 is due to θ_g's role on substitutability between employment destinations.

benefits more, and this operates in favor of the low-skilled. The second is the elasticity of commuting decisions to commute costs, which determines how willing individuals are to bear high commute costs to work in a particular destination. The third is geographic factors such as where house prices appreciate and whether the system connects locations of dense residence with well-paid jobs. These last two channels operate in favor of the high-skilled. The net effect is that welfare inequality between the low- and high-skilled increased by 0.369% as a result of TransMilenio.

Costs vs Benefits How did the output gains from TransMilenio compare with the costs of the system? Panel A of Table 13 provides a breakdown of the costs and benefits of the system (see appendix for details on cost calculations). Even using the most conservative estimate in column (1), I find that the net present value of the net increase on GDP was about $50bn, or a net increase of 2.73% in the steady-state level of GDP. This suggests the system was a highly profitable investment for the city.

9.2 Robustness and Model Extensions

In the appendix, I explore the robustness of my quantitative results to alternative parameter values. The effects on output, rents and welfare are qualitatively unchanged.[95]

In Table 14, I explore the sensitivity of the quantitative results to a number of model extensions.

First, I assume that the shocks by workplace location affect preferences rather than productivity.[96] In this model, TransMilenio no longer acts as a positive supply shock to each location (holding employment decisions constant). As a result, wages do not fall via

[95]These include: increasing the values of θ and η by 50%, using alternative values of θ estimated via PPML in two periods, setting spillovers to one third of their estimated values (to match the magnitude of productivity spillovers in Ahlfeldt et. al. 2015), using a larger elasticity of substitution across labor skill groups $\sigma_L = 2.5$, measuring the distribution of employment using the 2005 census rather than the 2015 CCB (to address whether missing informal establishments impacts the results), and using alternative values of the elasticity of demand $\sigma = 3, 9$.

[96]In the preference shock and joint decision model, I assume that there are no fixed costs i.e. $p_a = \bar{h} = 0$. Rows 3 and 4 should therefore be compared with row 2.

this channel and welfare gains from the system rise. Since labor supplied by each worker is unchanged by commute costs, the effect on output falls by more than two thirds. However, this difference is eliminated by the increase in labor supply from population growth in the open city model.

Second, I allow workers to make a joint decision over where to live and work.[97] One worry is that when workers choose first where to live and then where to work, they may face ex-post regret over their residential choice. The results are qualitatively unchanged from the baseline model, suggesting the timing assumption has little quantitative impact.

Third, I confront the fact that neither census nor CCB employment data cover 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 previous model may under-estimate the welfare gains to the low-skilled by ignoring the fact that TransMilenio likely improved access to domestic services jobs in the homes of the college educated in the North. On the other hand, high-skilled workers also benefit from this increased labor supply which lowers the cost of consuming domestic services. In the appendix, I extend the baseline model to include employment in domestic services and outline its calibration. The fifth row of Table 14 shows that allowing for employment in domestic services has little effect on the distributional effects: the benefits for low- and high-skilled workers roughly balance out.

Fourth, I incorporate home ownership. One concern is whether, by assuming all individuals are renters, I understate welfare gains to the low-skilled who may disproportionately benefit from appreciation in the city's outskirts if they own their home. However, acting against this is the fact that the high-skilled are more likely to own their home with an ownership rate of 0.603 compared to 0.457 for the low-skilled in 2015. In the appendix, I extend the model as follows. I assume local land rents are collected into a fund, a fraction of which

[97] I assume that workers draw a joint preference for each residence-employment pair drawn from a Frechet with shape θ_g. I estimate θ_g from the implied gravity equation for unconditional commute flows, finding that $\hat{\theta}_L = 3.058$ and $\hat{\theta}_H = 1.772$.

is re-distributed to low- and high-skilled workers based on the home ownership rates. The remainder (i.e. rents from unowned residential floorspace and all commercial floorspace) is paid into an aggregate portfolio which is owned in equal shares by all residents.[98] The final row of Table 14 shows that home ownership does indeed close the gap in welfare gains between high- and low-skilled workers, but not enough to reverse it.[99]

Comparison with VTTS Approach The typical approach to evaluate the gains from commuting infrastructure is based on the Value of Travel Time Savings (VTTS) approach (e.g. Small and Verhoef 2007). In this framework, the benefits from new infrastructure are given by the marginal value of time times the amount of time saved. In the appendix, I provide more details on how I use the commuting microdata to evaluate the gains using this method.

Table 15 presents the results. Welfare gains under VTTS are driven solely by mode choice: the low-skilled gain more than the high-skilled in row (1). This approach is silent on changes in output and house prices. In row (2), I present the results in partial equilibrium from simple model where worker skill groups have the same elasticities θ and η and enter as perfect substitutes in production. As discussed above, the low-skilled benefit more in this model and the percentage point difference in welfare gains is almost exactly as under VTTS (0.846 vs. 0.807). Row (3) presents the results from my full model in partial equilibrium. The low-skilled benefit more but the relative gains are attenuated, because of differences in elasticities and relative wages across the city. Output increases by 2.85% as individuals have more time to work at their workplace. Lastly, row (3) shows the effects of the system under full general equilibrium adjustment. Gains shift from workers to land owners as floorspace

[98]This provides an upper bound of the effects of home ownership on the relative welfare of low-skill workers, since it is likely that a greater share of high-skilled workers own multiple properties and therefore would receive a greater share of the aggregate portfolio.

[99]The effect of TransMilenio on output is greater in this model since house price appreciation increases expenditure by residents who own the housing stock, instead of being spent outside of the city by absentee landlords. The effect on welfare is attenuated since income now depends on the sum of labor income and income from home ownership, and the direct effect of TransMilenio is greater on the former in proportionate terms.

prices adjust, flipping the relative welfare gains in favor of the high-skilled. Output rises by 3.92%, so that the general equilibrium channels account for approximately 20% of the gains in output.

Welfare gains under VTTS are driven solely by mode choice: the low-skilled gain more than the high-skilled, with zero gains accruing to landlords. In contrast, my model accounts for differences in commuting elasticities across groups, general equilibrium adjustment of the housing market, and a more localized geography of where each group lives and works. Capturing these additional channels suggested by the theory and supported by the data, my framework concludes that high-skilled workers in fact benefit the most with substantial gains also accruing to landowners.[100]

10 Policy Counterfactuals

10.1 Impact of Different Lines and Planned Cable Car

Table 16 evaluates the effects of different portions of the system.

Row (1) evaluates the impact of adding a Cable Car to the slums in the hills of Ciudad Bolivar in the South. This system is planned to open in 2019. The aggregate effects of the line are small due to its modest size, but it benefits the low-skilled workers who are more likely to live in targeted areas. Row (2) simulates the effect of removing line H connecting the Southern most portions of the city with the CBD. Since this area has a much greater density of low-skilled residents, the welfare effects are greater for the poor. Row (3) examines the impact of removing line A connecting the Northern parts of the city to the CBD. The

[100] A couple of comments are in order. First, VTTS accounts for differences in residence and employment locations across groups revealed in commute data. However, these are typically only representative between larger units (e.g. localities in my setting) rather than the census tracts considered in my approach. Second, while the percentage point difference in welfare gains in row (2) of Table 15 is almost identical to that in the VTTS approach, the overall level of gains is higher. This is due to the exponential functional form that links commute times to commute costs in my model, compared to the linear form under VTTS. Third, the partial equilibrium entries do not match those in Table 12. This is because I construct the partial equilibrium benefits in Table 15 by setting spillovers to their estimated values. These figures therefore match the alternative table for the decomposition reported in the appendix.

effects are slightly larger since a high density of businesses lie along this line. This line benefits the high-skilled who live in the North of the city. Thus, the effect of different parts of the network is heterogeneous. Row (4) simulates the effect of removing the feeder system connecting outlying areas with portals using buses that run on existing roadways. This 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.

Lastly, I isolate the contribution of route placement to TransMilenio's welfare gains. To do so, I compute the counterfactual from applying a uniform reduction in times across all commutes and modes equal to the average percentage change from the system. Row (5) presents the results. The increase in welfare is only three quarters of that from TransMilenio. This gap reflects how the system disproportionately reduced commute times between locations where many individuals live and work as discussed in Section 4.7. The relative gain for the high-skilled falls in this counterfactual since TransMilenio connected high-density residence and employment locations more for these individuals. Output also increases by less, since the system increased labor supply to central, productive locations where FCMA improved the most. Taken together, this highlights how placing lines along heavily commuted routes increases the gains from infrastructure spending.

10.2 Land Value Capture

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 where it was needed. I show in the appendix 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. These policies achieve the dual aim of increasing housing supply and raising revenue to finance the construction of the system.

I evaluate the impact of TransMilenio if housing supply had responded to the opening of the system. In the most extreme case, I assume that housing supplies freely adjust to reflect long-run adjustment. This provides a useful upper bound on welfare loses from restrictive zoning. I then simulate the effect of two potential LVC schemes implemented by the government. 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.[101] 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.[102] Details on these model extensions are provided in the appendix. I compare the two equilibria from first removing TransMilenio and then adding it back under each housing supply model.

Table 17 presents the results. In the closed city model, welfare increases by 24.4% more under free adjustment. Under the LVC schemes, welfare improves by 23% and 6.9% under the CMA and distance-band policies respectively. Taking the CMA scheme as an example, welfare rises by 24.9% for low-skilled individuals versus 18.3% for the high-skilled. More accomodative zoning benefits the low-skilled by dampening house price appreciation towards the edge of the city where they live.

Panel B of Table 13 converts the government revenue from these policies into fractions

[101]These schemes have a number of benefits over property taxes. They are likely to incur less opposition from stakeholders, are less distortive, 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 vaires across alternative candidate policies.

[102]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.

of overall capital costs of constructing the system. In the more conservative closed city case, the distance-band based permit measure recoups only 9.9% of the capital cost of the system, compared to the 17.9% earned using the commuter market access-based permit. In the open city, these increase to 27% and 50% respectively. These figures fall slightly when spillovers are shut down, but the policy implications are qualitatively unchanged.

My results suggest the potential for large welfare gains to governments pursuing a unified transit and land use policy. These policies can also be used to finance the construction of public transit. Additionally, the comparison with the distance-based policy underscores how measures of CMA can be a used as parsimonious tools for governments to guide the allocation of rezoning. By targetting central and outlying areas where accessibility for residents and firms increase the most, the CMA-based policy concentrates new construction where it is most needed.

11 Conclusion

This paper assesses the aggregate and distributional consequences of improving public transit infrastructure in cities. I leverage the the construction of the world's largest BRT system in Bogotá, Colombia to make three distinct contributions to the urban economics literature. First, I built a novel quantitative general equilibrium model of a city where low- and high-skill workers sort over where to live, where to work, and whether or not to own a car. Second, I develop a new reduced form methodology derived from general equilibrium theory to evaluate the effects of changes in commuting networks in cities. Third, I estimate the structural model and use it to quantify the effects of the system and counterfactual policies.

I find that the new reduced form methodology explains the heterogeneous adjustment of resident- and firm-related outcomes across the city. The BRT system led to large increases in welfare and output (net of construction and operating costs), but these would have been around one fourth larger had the government implemented a more accommodative zoning policy. This underscores the benefits to cities from pursuing a unified transit and land use policy. Accounting for the full general equilibrium impact of the system, I find that high-skilled workers benefit slightly more. This suggests improving public transit is a less precise policy tool to target welfare improvements for the poor than implied by the traditional VTTS approach that focuses on mode choice alone.

References

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.

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.

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.

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.

Chauvin, J., Glaeser, E., Ma, Y., and K. Tobio (2016), "What is Different About Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States", Working Paper.

Ciccone, A., and R. Hall (1996), "Productivity and the Density of Economic Activity," *American*

Economic Review, 86(1), 54–70.

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.

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.

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.

Galle, S., Rodriguez-Clare, A., and M. Yi (2017), "Slicing the Pie: Quantifying the Aggregate and Distributional Consequences of Trade", 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.

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.

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 Bogotn", 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.

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.

Osland, L., and I. Thorsen (2008), "Effects on housing prices of urban attraction and labor-market accessibility", *Environment and Planning*, 40: 2490-2509.

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, Whitepaper.

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: 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 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)	0.470***	0.355***	0.208**	0.333***	0.463***	0.565***	0.435***	0.557***
	(0.091)	(0.074)	(0.097)	(0.079)	(0.152)	(0.118)	(0.158)	(0.122)
N	1,943	1,943	1,943	1,943	1,943	1,943	1,943	1,943
F-Stat					851.31	852.24	410.63	418.15
Over-ID p-value							0.52	0.16
ln(Residential Pop)	0.271**	0.283**	0.203	0.209	0.325*	0.348*	0.303*	0.327*
	(0.136)	(0.138)	(0.138)	(0.140)	(0.178)	(0.185)	(0.181)	(0.188)
N	1,997	1,997	1,997	1,997	1,997	1,997	1,997	1,997
F-Stat					1,747.80	1,706.02	859.94	844.56
Over-ID p-value							0.83	0.95
Panel B: Firms								
ln(Comm Floorspace Price)	0.223*	0.228*	0.259**	0.277**	0.261	0.278*	0.338**	0.338**
	(0.122)	(0.128)	(0.124)	(0.128)	(0.162)	(0.168)	(0.165)	(0.171)
N	1,884	1,884	1,884	1,884	1,884	1,884	1,884	1,884
F-Stat					1,265.27	1,114.61	812.06	733.28
Over-ID p-value							0.09	0.19
Comm Floorspace Share	0.164***	0.162***	0.163***	0.166***	0.124**	0.122*	0.131**	0.130**
	(0.043)	(0.045)	(0.044)	(0.046)	(0.059)	(0.064)	(0.061)	(0.065)
N	1,981	1,981	1,981	1,981	1,981	1,981	1,981	1,981
F-Stat					1,361.75	1,183.55	876.75	772.27
Over-ID p-value							0.52	0.57
ln(Establishments)	1.423***	0.880**	1.185***	0.780**	1.573***	1.488**	1.090*	0.974*
	(0.365)	(0.354)	(0.374)	(0.361)	(0.598)	(0.604)	(0.592)	(0.591)
N	1,724	1,724	1,724	1,724	1,724	1,724	1,724	1,724
F-Stat					224.50	264.90	333.56	293.00
Over-ID p-value							0.12	0.01
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 CMA in first differences. Each column corresponds to a specification. Only tracts further than 500m from a portal and the CBD (and less than 3km from a station) are included. Controls are as described in previous table, other than distance to tram which 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 very high (>10,000) 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 instrument and the change induced by the tram instrument. Robust standard errors 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 All lnDist	(5) IV All lnDist
lnRCMA	0.957***	0.541**	0.383	0.951**	0.310
	(0.212)	(0.252)	(0.300)	(0.404)	(0.466)
lnRCMA X High Skill					-0.147**
					(0.058)
N	9,088	22,119	22,119	17,212	19,920
R^2	0.07	0.10			
F-Stat				72.88	18.26
Over-ID p-value				0.52	0.55
UPZ FE	X	X	X	X	X
Locality FE X Post FE	X	X	X	X	X
Log Dist CBD X Region FE X Post FE	X	X	X	X	X
Trip Controls X Post FE	X	X	X	X	X
Tract Controls X Post FE	X	X	X	X	X
Historical Controls X Post FE	X	X	X	X	X
Educ X Post FE					X

Note: Observation is a trip, only trips to work are included. Column (1) reports coefficients from a regression of the probability 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 the 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 All
Δ lnRCMA	-0.011	-0.046	-0.040	-0.062
	(0.030)	(0.029)	(0.029)	(0.047)
Δ lnRCMA X HighColl		0.051*	0.063*	0.111**
		(0.027)	(0.033)	(0.047)
N	1,886	1,886	1,886	1,886
R^2	0.27	0.27		
F-Stat				123.61
Over-ID p-value				0.55
Locality FE	X	X	X	X
HighColl FE		X	X	X
Log Dist CBD X Region FE	X	X	X	X
Tract Controls	X	X	X	X
Historical Controls	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 (excluding the tract itself) and then setting the 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 and changing only commute costs, excluding the census tract itself from the variable construction. Column (4) instruments using the change in CMA using the LCP and Tram instruments. Only tracts further than 500m from a portal and the CBD (and less than 3km from a station) are included. Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6: Wages

Outcome: $\ln$Wage	(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 workers 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, while time of day controls include dummies for the hour of trip departure. Each have choice-varying coefficients. Only trips during rush hour 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)
High-Skill X ln Commute Cost	-0.0242***	-0.0240***	-0.0239***	-0.0234***
	(0.0036)	(0.0036)	(0.0035)	(0.0036)
Low-Skill X ln Commute Cost	-0.0336***	-0.0333***	-0.0333***	-0.0329***
	(0.0043)	(0.0043)	(0.0042)	(0.0042)
Crime		-0.005		-0.002
		(0.024)		(0.019)
House Price		-0.318		-0.352
		(0.463)		(0.353)
Primary Road		0.942		0.828
		(0.624)		(0.580)
N	1,444	1,444	1,444	1,444
Origin-Skill-Car Ownership Fixed Effects	X	X	X	X
Destination-Skill Fixed Effects	X	X	X	X
Method	PPML	PPML	IV-PPML	IV-PPML

Note: Outcome is the conditional commuting shares between localities in 2015. Observation is an origin-destination-skill-car ownership 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 and 2 use PPML estimated under a GLM routine. Columns 3 and 4 implement IV-PPML with a 2-step GMM routine, 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. Crime, house price and primary road include the average number of crime per year from 2007-2014, the average log house price in 2012, and the share of the trip that takes place along a primary road along the least-cost routes between origin and destination. In columns 1 and 2 standard errors are clustered by origin-destination locality; in columns 3 and 4 heteroscedasticity robust errors are recovered from the GMM variance matrix. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 9: GMM Results

Parameter	Estimate
Panel A: Firms	
μ_A	0.237***
	(0.089)
Panel B: Workers	
η_L	3.595***
	(0.861)
η_H	3.261***
	(0.697)
μ_U^L	0.250***
	(0.031)
μ_U^H	0.342***
	(0.048)

Note: Estimates are from two-step GMM procedure separately for firms at the tract-industry level with 6137 observations and for workers at the tract-group-car ownership level with 7036 observations. Controls include log distance to CBD interacted with region fixed effects, commercial floorspace share in 2000, and log population density and college share in 1993 for employment moment conditions. Spillover parameter estimates obtained via delta method: original parameter clusters $\eta_L \mu_U^L$ and $\eta_H \mu_U^H$ are 0.898 (0.222) and 1.114 (0.176) respectively. Only tracts within 3km of the network and those more than 500m from portals and the CBD are included. Standard errors clustered at the tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 10: Amenities and Productivities: Model vs Data

Panel A: Amenities	(1) ln Crime Density	(2) ln Crime Density	
Amenity Elasticity	-0.115***	-0.238***	
	(0.016)	(0.012)	
Skill	College	Non-College	
R^2	0.12	0.40	
N	551	548	
Panel B: Productivities	(1) ln Crime Density	(2) ln Slope	(3) ln Roads
Productivity Elasticity	-0.043*	-0.155***	0.087***
	(0.025)	(0.010)	(0.012)
R^2	0.01	0.24	0.08
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 either 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 11: Effect of Removing Phases 1 and 2 of TransMilenio

	No Spillovers	Spillovers
Panel A: Closed City		
GDP	3.119	3.918
Rents	3.285	3.721
Welfare Low	3.444	3.814
Welfare High	3.651	4.169
Inequality	0.215	0.369
Panel B: Open City		
GDP	10.347	15.596
Rents	13.145	16.275
Population Low	8.562	10.744
Population High	9.543	12.303
Relative Population	1.072	1.747

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 1: Same η, θ			**Model 2:** Diff θ			**Model 3:** Diff η, θ			**Model 4:** Full Model		
	Low	High	Ineq.	Low	High	Ineq.	Low	High	Ineq.	Low	High	Ineq.
Choices Fixed	4.163	2.870	-1.349	3.947	3.810	-0.143	3.937	3.827	-0.115	4.560	4.434	-0.132
Emp Adjust	5.488	4.230	-1.331	5.389	4.702	-0.726	5.380	4.725	-0.692	5.371	4.817	-0.585
Emp, Car Adjust	5.536	4.286	-1.323	5.435	4.774	-0.699	5.428	4.790	-0.675	5.419	4.885	-0.565
All Choices Adjust	5.749	4.426	-1.404	5.651	4.936	-0.758	5.652	4.938	-0.758	5.647	5.036	-0.648
All Choices, Rents Adjust	3.959	3.216	-0.774	3.792	3.677	-0.119	3.788	3.681	-0.112	3.807	3.758	-0.050
GE, No Spillovers	3.710	2.893	-0.849	3.512	3.432	-0.083	3.509	3.435	-0.077	3.444	3.651	0.215
GE, Prod Spillovers Only	4.008	3.211	-0.831	3.807	3.682	-0.131	3.804	3.684	-0.124	3.691	3.943	0.261
GE, Spillovers	4.408	3.562	-0.885	4.219	4.042	-0.185	4.163	4.010	-0.160	3.814	4.169	0.369

Note: Table shows the welfare gain from phases 1 and 2 of TransMilenio under varies adjustment scenarios. For each model, rows (1)-(6) are computed by first simulating the counterfactual equilibrium without TransMilenio using unobservables from the 2012 equilibrium, with spillovers set to zero. Each entry then reports the welfare gain from the equilibrium with TransMilenio under the adjustment permitted in each row. Analagous to previous tables, the welfare gain is defined as the (negative of the) percentage change in welfare from going from the equilibrium with TM to that without it. Welfare inequality is defined as before. In row (1), all worker choices, rents, wages and land use decisions are held fixed. In row (2), workers employment decisions are allowed to adjust. In row (3), employment and car ownership choices change. In row (4), employment, residential and car ownership decisions adjust. In row (5), all individual choices can change while rents and land use decisions adjust to equilibriate housing markets. In row (6), there is full general equilibrium adjustment. Rows (7) and (8) then report the welfare gain from removing the system under each spillover condition.In model 1, both worker groups are assigned the same (average) η and θ parameters and are assumed to be perfect substitutes in production (i.e. $\sigma_L \rightarrow \infty$). In model 2, worker groups differ by their estimated θ parameters. In model 3, worker groups differ both by their estimated θ and η parameters. In model 4, workers differ both by their estimates θ and η parameters and are imperfectly substitutable within firms in the way described in the text.

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)	57,359	72,052	190,282	286,812
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)	50,258	64,952	183,181	279,711
% Net Increase GDP	2.73	3.53	9.96	15.21
Panel B: Land Value Capture				
LVC Band Revenue (mm)	93	113	240	315
As share of capital costs	8.18	9.91	21.08	27.72
LVC CMA Revenue (mm)	170	203	464	571
As share of capital costs	14.96	17.86	40.82	50.21

Note: All numbers in millions of 2016 USD. NPV calculate over a 50 year time horizon with a 5% discount rate. Each column describes 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
	Low Skill	**High Skill**	**Output**	**Output**
Baseline	3.814	4.169	3.918	15.596
No Fixed Costs	3.864	4.107	3.922	20.484
Preference Shocks	4.620	4.788	1.184	20.796
Joint Decision	4.106	4.178	3.937	20.725
Domestic Services	3.746	4.066	3.861	15.231
Home Ownership	3.758	3.851	4.224	17.337

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. For each model, columns 1 and 2 report the percentage change in low- and high-skill worker welfare, column 3 reports the percentage change in output in the closed city model which column 4 reports the change in output in the open city model. Row 1 reports results from the baseline model. Row 2 presents those from the model without fixed costs (i.e. $\bar{h} = p_a = 0$). Row 3 reports the model with preference rather than productivity shocks by location of employment. Row 4 presents the model where there is a joint decision for residence and employment locations. Row 5 reports results from the model with employment as domestic servants. Row 6 shows results for the model with home ownership.

Table 15: Comparison with Value of Time Savings Calculation

	Welfare Low	**Welfare High**	**Output**	**Rents**
VTTS	4.203	3.396	0.000	0.000
Partial Eqbm, Simple Model	4.902	4.056	2.908	0.000
Partial Eqbm	4.461	4.357	2.848	0.000
General Eqbm	3.814	4.169	3.918	3.721

Note: The first row reports the percentage change in each variable due to TransMilenio from a Value of Time Savings approach as described in the appendix. The second row reports the values from my model with the same θ, η and perfect substitutes in production, in partial equilibrium where commute times change but all prices and location decisions are fixed. The third row does the same in the baseline model. The fourth row reports the values from my model with full general equilibrium adjustment. Partial equilibrium results computed in the same way as in Table 13, but with spillovers set to their estimated values.

Table 16: Effect of Different System Components

	Welfare Low	**Welfare High**	**Output**
Add Cable Car	0.064	0.057	0.033
Remove Line South	1.609	1.560	1.045
Remove Line North	1.540	1.666	1.698
Remove Feeders	1.864	1.933	1.585
Uniform Reduction	2.993	3.146	3.716

Note: Table shows the (negative of the) value of the percentage change in welfare from removing a piece of the TransMilenio (existing, future or hypothetical) network. These counterfactuals are adding the Cable Car system, removing line H in the south, removing line A in the north, removing the feeder system, and a uniform 5.8% reduction in all commute times.

Table 17: Effect of Adjusting Housing Supply, and Land Value Capture Scheme

Panel A: Closed City

	Output	Housing	Rents	Welfare Low	Welfare High	Gvt Revenue
Fixed Supply	3.918		3.721	3.814	4.169	
Free Adjustment	4.264	2.169	1.174	4.817	5.013	
LVC, Bands	4.171	1.007	2.731	4.073	4.462	0.115
LVC, CMA	4.218	1.977	1.396	4.765	4.954	0.207

Panel B: Open City

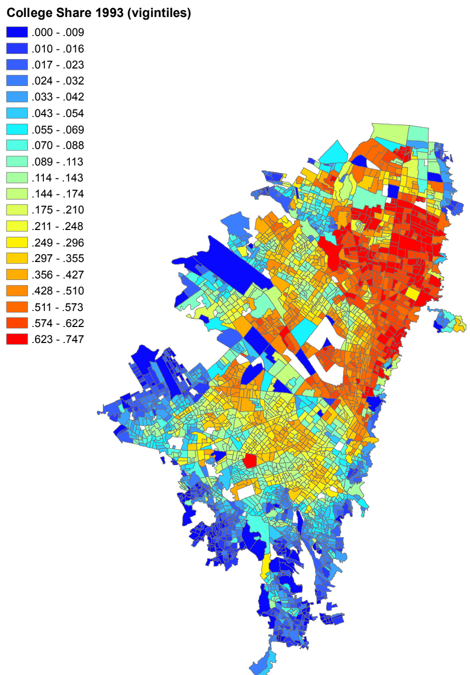
	Output	Housing	Rents	Pop. Low	Pop. High	Gvt Revenue
Fixed Supply	15.596		16.275	10.744	12.303	
Free Adjustment	28.900	12.431	16.936	22.433	23.665	
LVC, Bands	18.279	2.792	16.233	12.654	14.545	0.357
LVC, CMA	21.585	5.128	17.038	16.118	17.412	0.646

Note: Table shows the percentage change in each outcome going from the equilibrium with TransMilenio to that without TransMilenio under the housing supply conditions indicated in each row, with spillovers set to their estimated values. As in previous exercises, 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. Row (1) is the case with fixed housing supply. Row (2) is the case of freely adjusting housing supply. Row (3) 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. Government revenue from the scheme is given in column (6) as a percentage of pre-TM GDP. Row (4) shows the results of the scheme based on predicted changes in commuter market access as described in the text.

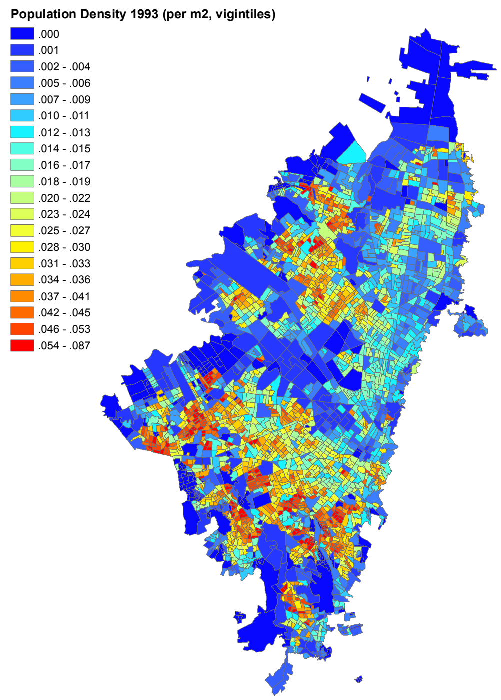
Figures

Figure 2: Population Density and Demographic Composition in 1993

(a) College Share



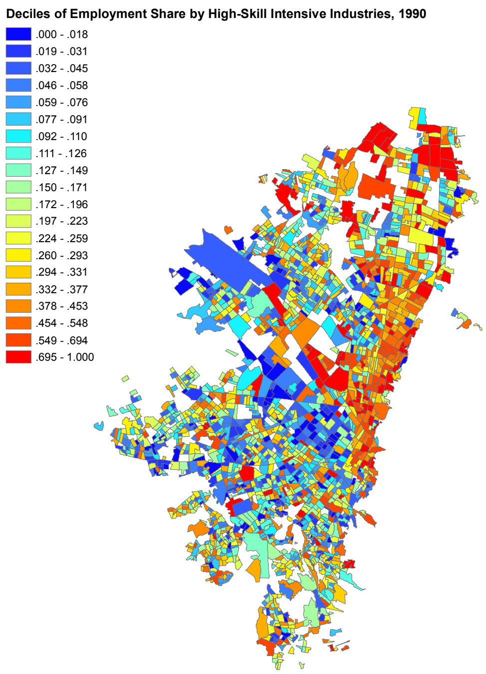
(b) Population Density



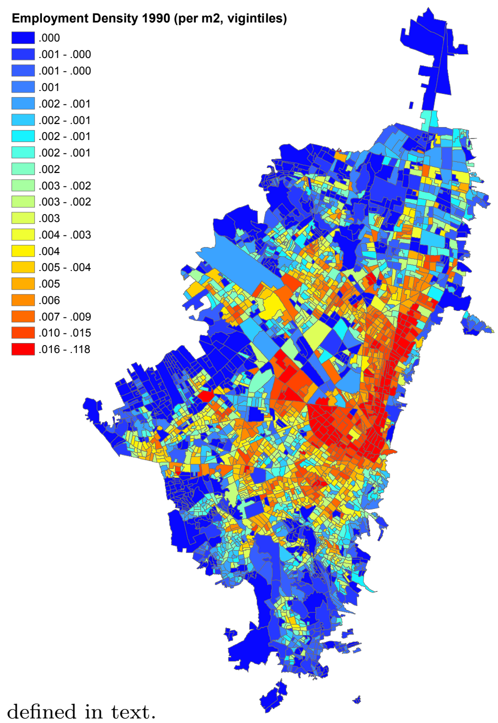
Note: Data is from 1993 Census.

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: TransMilenio Routes

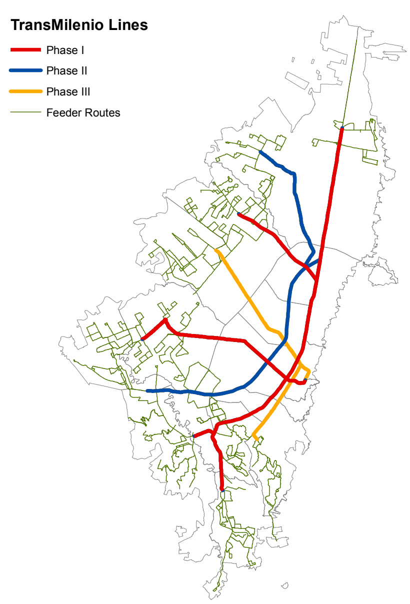
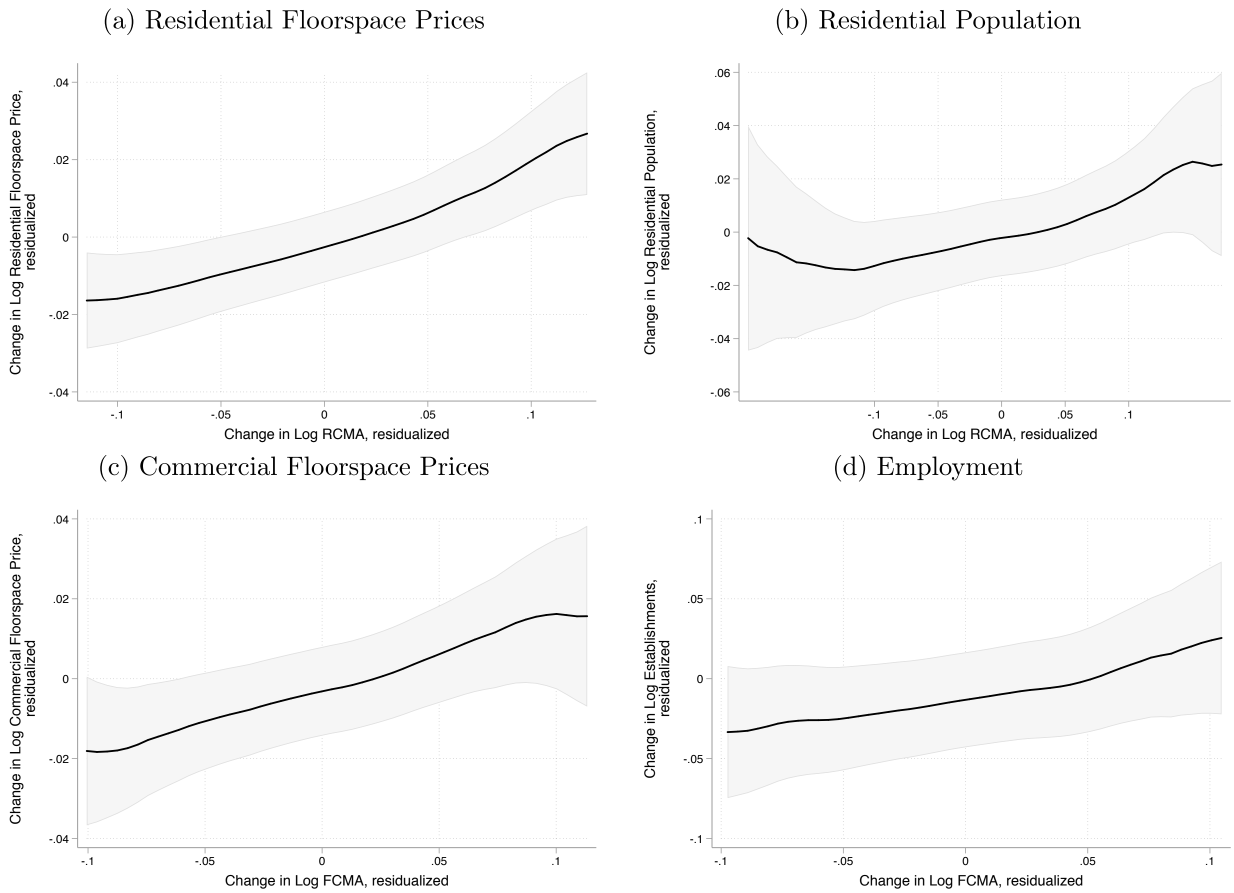
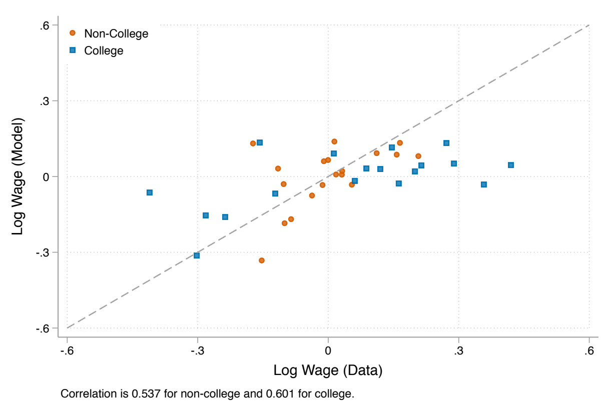


Figure 5: Non-Parametric Relationship Between Outcomes and Commuter Market Access



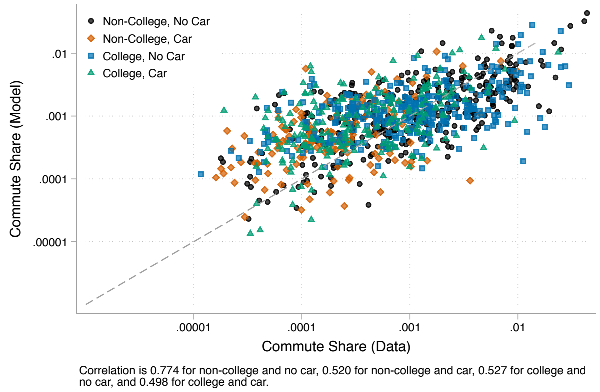
Note: Plot shows the non-parametric relationship between outcomes and CMA. Specifications correspond to the reduced form from column (4) of main table in which CMA is measured holding population and employment fixed at their initial levels, with the full set of baseline controls included, and is regressed directly on outcomes.

Figure 6: Wages: Model vs. Data



Note: Plot 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).

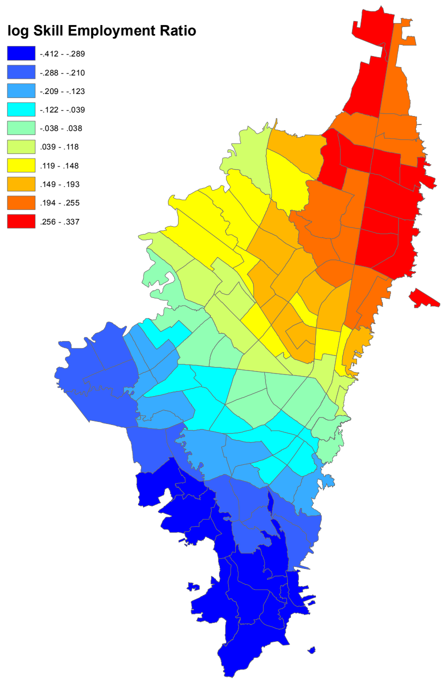
Figure 7: Commute Flows: Model vs. Data



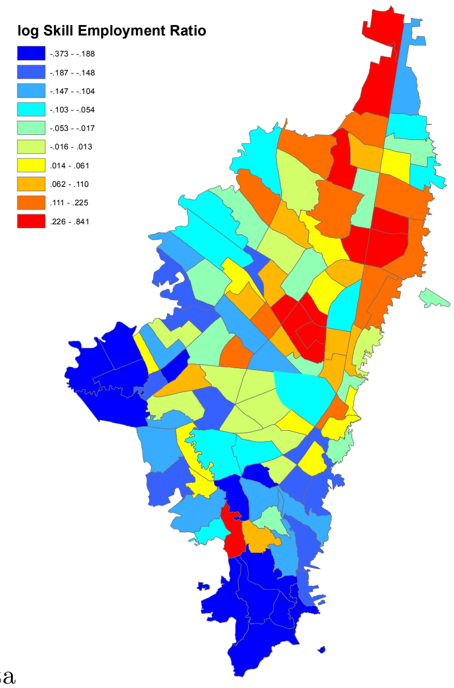
Note: Observation is a locality origin-destination pair, skill group and car ownership combination. Plot shows relationship between share of commuters choosing each (i, j, a) pair in the model vs those doing so in the 2015 Mobility Survey.

Figure 8: Relative Employment by Skill by UPZ: Model vs Data

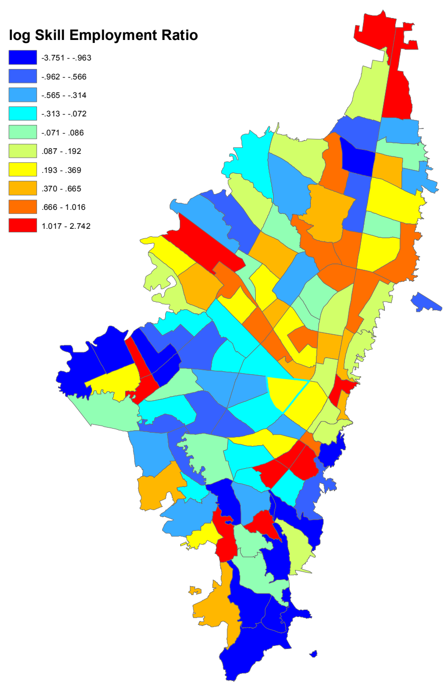
(a) Model: Perfect Subsitutes & Same θ



(b) Model: Baseline Estimates



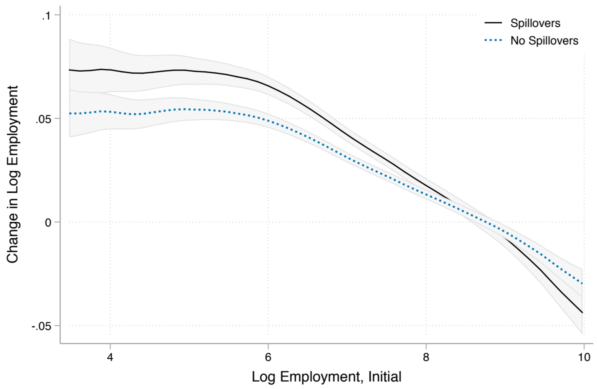
(c) Data



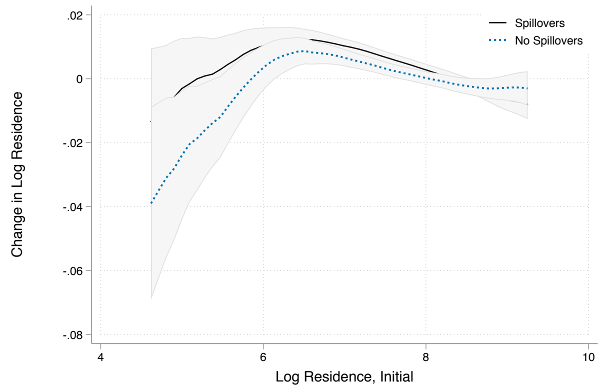
Note: Panel (a) shows the deciles of the distribution of the log skill employment ratio $\ln L_{FjH}/L_{FjL}$ by UPZ in the model when skill groups are perfect substitutes in production and have the same value of θ (equal to the average value in the population. Panel (b) shows the distribution for the baseline model. Panel (c) shows the distribution in the 2015 Mobility Survey. Correlation between data in panel (a) and (c) is 0.256, while that between panel (b) and (c) is 0.406.

Figure 9: Simulated Changes in Outcomes

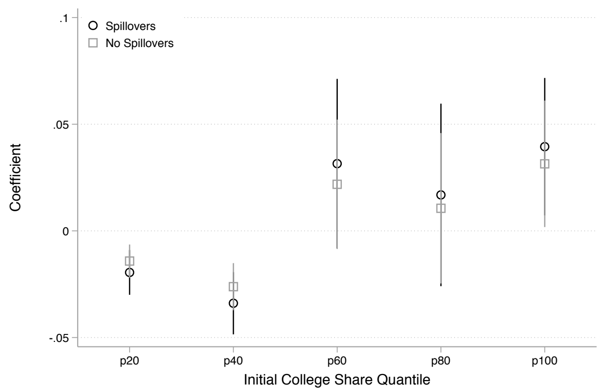
(a) Employment



(b) Residential Population



(c) Change in College Share vs RCMA



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. Panel (c) plots the interaction terms from a regression of the change in college share on a full interaction between a dummy for a tract's initial college share quantile and the change in log RCMA.

A Additional Tables

Table A1: Additional OLS Results

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Residents								
ln(Res Floorspace Price)		0.406***	0.146*	0.182**	0.294***	0.349***	0.284***	0.352***
		(0.085)	(0.088)	(0.088)	(0.073)	(0.090)	(0.065)	(0.085)
	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)		0.294***	0.151	0.180*	0.206*	0.268**	0.175*	0.306*
		(0.100)	(0.110)	(0.109)	(0.108)	(0.127)	(0.100)	(0.159)
	N	2,028	2,028	2,028	2,028	2,028	2,028	1,757
	R^2	0.11	0.11	0.14	0.16	0.16	0.14	0.15
Panel B: Firms								
ln(Comm Floorspace Price)		0.185*	0.232**	0.232**	0.260**	0.377***	0.207*	0.221**
		(0.102)	(0.109)	(0.112)	(0.116)	(0.144)	(0.106)	(0.109)
	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		0.199***	0.173***	0.175***	0.160***	0.156***	0.147***	0.111***
		(0.037)	(0.039)	(0.040)	(0.041)	(0.049)	(0.037)	(0.038)
	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)		0.373	0.989***	0.848**	0.703**	0.831**	0.696**	0.854***
		(0.331)	(0.336)	(0.340)	(0.326)	(0.409)	(0.296)	(0.308)
	N	1,753	1,753	1,753	1,753	1,753	1,753	1,751
	R^2	0.19	0.21	0.25	0.30	0.30	0.30	0.30
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). 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), tracts with zero change excluded. Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A2: OLS Results: Robustness

		(1) Baseline	(2) Alt Times	(3) Alt Times	(4) Alt Times	(5) Slow	(6) Fast	(7) High θ	(8) Low θ	(9) No Cutoff	(10) Unweighted	(11) Conley SE
Panel A: Residents												
ln(Res Floorspace Price)		0.294***	0.513***	0.319***	0.316***	0.285***	0.300***	0.201***	0.547***	0.222***		0.294***
		(0.073)	(0.128)	(0.079)	(0.078)	(0.073)	(0.071)	(0.057)	(0.120)	(0.069)		(0.104)
	N	1,975	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	0.55	0.58		0.55
ln(Res Population)		0.206*	0.335*	0.222*	0.236**	0.189*	0.231**	0.140*	0.371*	0.269***	0.380**	0.206
		(0.108)	(0.186)	(0.118)	(0.109)	(0.106)	(0.107)	(0.080)	(0.189)	(0.102)	(0.158)	(0.172)
	N	2,028	2,028	2,028	2,028	2,028	2,028	2,028	2,028	2,480	2,372	2,028
	R^2	0.16	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)		0.260**	0.486**	0.281**	0.272**	0.283**	0.232**	0.223**	0.352*	0.251**		0.260**
		(0.116)	(0.204)	(0.126)	(0.122)	(0.121)	(0.109)	(0.093)	(0.181)	(0.114)		(0.136)
	N	1,914	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		0.15
Comm Floorspace Share		0.175***	0.308***	0.190***	0.182***	0.185***	0.162***	0.142***	0.250***	0.164***		0.175***
		(0.040)	(0.069)	(0.043)	(0.041)	(0.042)	(0.037)	(0.033)	(0.061)	(0.040)		(0.048)
	N	2,071	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	0.07	0.07	0.06	0.06		0.07
ln(Establishments)		0.703**	1.012*	0.771**	0.579*	0.660*	0.697**	0.446*	1.342***	0.758**	0.339	0.703**
		(0.326)	(0.571)	(0.354)	(0.346)	(0.340)	(0.309)	(0.264)	(0.507)	(0.321)	(0.293)	(0.389)
	N	1,753	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	0.30	0.29	0.27	0.30
Time Specification			Pref. Wgt	Min. All	Unweighted	Slow	Fast					
Locality Fixed Effects		X	X	X	X	X	X	X	X	X	X	X
CBD X Region Controls		X	X	X	X	X	X	X	X	X	X	X
Tract Controls		X	X	X	X	X	X	X	X	X	X	X
Historical Controls		X	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. 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, with robust standard errors reported in parentheses. 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 an unweighted average commute time from the logit model. Column (5) uses slower commute times that best match average speeds in the post-period. Column (6) uses faster ones that do the same for the pre-period. Column (7) uses a larger value of θ equal to 1.5 times its baseline estimate, while column (8) uses a smaller value. Column (9) relaxes the 3km distance cutoff and includes all census tracts, while column (10) provides unweighted estimates for outcomes where weighted regressions are presented in the main tables. Column (11) provides spatial HAC standard errors (Conley (1999)), using a spatial weight of 1 for tracts within 500m of each other and zero otherwise. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A3: 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)		0.308***	0.288***	0.422***	0.414***	0.289***	0.273***	0.467***	0.461***
		(0.067)	(0.073)	(0.113)	(0.116)	(0.062)	(0.067)	(0.101)	(0.105)
	N	1,823	1,823	1,823	1,823	1,975	1,975	1,975	1,975
	F-Stat			645.99	320.07			845.50	414.61
	Over-ID p-value				0.57				0.06
ln(Res Population)		0.342***	0.264**	0.337*	0.317*	0.305***	0.227*	0.358**	0.342**
		(0.125)	(0.127)	(0.182)	(0.184)	(0.117)	(0.118)	(0.156)	(0.159)
	N	1,883	1,883	1,883	1,883	2,028	2,028	2,028	2,028
	F-Stat			1,313.43	652.27			1,842.18	908.10
	Over-ID p-value				0.84				0.93
Panel B: Firms									
ln(Comm Floorspace Price)		0.230**	0.268**	0.241	0.275*	0.211*	0.257**	0.238	0.288*
		(0.117)	(0.118)	(0.156)	(0.155)	(0.112)	(0.114)	(0.154)	(0.154)
	N	1,766	1,766	1,766	1,766	1,914	1,914	1,914	1,914
	F-Stat			838.56	610.24			895.54	653.97
	Over-ID p-value				0.54				0.31
Comm Floorspace Share		0.136***	0.141***	0.095	0.106*	0.171***	0.173***	0.137**	0.146**
		(0.041)	(0.042)	(0.060)	(0.060)	(0.040)	(0.041)	(0.058)	(0.059)
	N	1,861	1,861	1,861	1,861	2,013	2,013	2,013	2,013
	F-Stat			879.55	634.07			947.73	683.02
	Over-ID p-value				0.50				0.50
ln(Establishments)		0.661**	0.628*	1.390**	0.891*	0.758**	0.697**	1.208**	0.744
		(0.325)	(0.331)	(0.541)	(0.533)	(0.312)	(0.320)	(0.580)	(0.553)
	N	1,616	1,616	1,616	1,616	1,753	1,753	1,753	1,753
	F-Stat			215.59	234.55			167.05	222.58
	Over-ID p-value				0.01				0.03
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. Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A4: Falsification Tests

	(1) ResPr	(2) ResPr	(3) ResPop	(4) ResPop	(5)	(6)
Panel A: Residents						
ln(RCMA)	0.180***	0.184***	0.214**	0.15 3		
	(0.061)	(0.061)	(0.109)	(0.107)		
ln(RCMA) Later Phase		0.087		0.037		
		(0.128)		(0.114)		
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)	0.233**	0.217**	0.101***	0.102***	0.634*	0.621*
	(0.110)	(0.110)	(0.030)	(0.030)	(0.331)	(0.329)
ln(FCMA) Later Phase		0.468		0.066		1.692
		(0.387)		(0.097)		(1.367)
N	1,910	2,011	2,013	2,119	1,753	1,845
R^2	0.06	0.06	0.10	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. In Panel A, log residential floorspace price per m2 and log residential populations correspond to columns (1)-(2) and (3)-(4) respectively. In Panel B, log commercial floorspace price per m2, commercial floorspace share and log number of establishments correspond to columns (1)-(2), (3)-(4) and (5)-(6) respectively. For each outcome, the first column reports the baseline specification (reduced form) where CMA growth is measured using the change in commute costs induced by TransMilenio holding residence and employment fixed at their initial levels. The second column adds an additional variable containing the growth in this CMA measure induced by opening later phases of the system. For residential population (where growth is measured between 1993 and 2005) this includes the growth in CMA going from phase 1 of the system to phases 2 and 3 (which opened in 2005 and 2011, respectively). For establishment counts (where growth is measured between 2000 and 2015) and land markets (where growth is measured between 2000 and 2007), this includes the growth in CMA going from phase 2 to 3. All specifications include the full set of controls described in the baseline OLS specification. Tracts closer than 3km from a station of the phases in question are included.Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A5: Theta Estimation Robustness: Estimation in Changes

	PPML	OLS	IV
HighSkill X ln Commute Cost	-0.026***	-0.031***	-0.030*
	(0.009)	(0.009)	(0.017)
LowSkill X ln Commute Cost	-0.029***	-0.033***	-0.033*
	(0.009)	(0.010)	(0.018)
N	2,610	1,562	1,562
F-Stat			20.59
Over-ID p-value			0.46
Origin-Destination-Skill-Car Ownership Fixed Effects	X	X	X
Post-Origin-Skill-Car Ownership Fixed Effects	X	X	X
Post-Destination-Skill Fixed Effects	X	X	X

Note: Outcome is the conditional commuting shares between localities. 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 run a difference-in-difference specification using variation in times within each origin-destination-car ownership cell between 1995 and 2015 for identification. Post is a dummy for 2015. Column 1 runs PPML estimation, while 2 and 3 run OLS and IV-2SLS estimates respectively (dropping all zero observations). Standard errors are clustered by origin-destination locality. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A6: GMM Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Firms												
μ_A	0.237***	0.263***	0.237***	0.331***	0.141	0.179*	0.321***	0.190**	0.286***	0.367***	0.241***	0.237***
	(0.089)	(0.082)	(0.089)	(0.063)	(0.115)	(0.103)	(0.096)	(0.085)	(0.058)	(0.067)	(0.054)	(0.089)
log(Dist TM)				0.037***		0.011						
				(0.014)		(0.021)						
Panel B: Workers												
η_L	3.600***	3.595***	3.595***	4.269***	3.406***	4.809***	3.595***	3.595***	3.649***	3.649***	3.649***	5.958***
	(0.858)	(0.861)	(0.861)	(0.902)	(0.872)	(1.299)	(0.861)	(0.861)	(0.838)	(0.838)	(0.838)	(1.113)
η_H	3.268***	3.261***	3.261***	4.007***	1.109	2.300***	3.261***	3.261***	3.518***	3.518***	3.518***	4.610***
	(0.695)	(0.697)	(0.697)	(0.748)	(0.766)	(0.885)	(0.697)	(0.697)	(0.707)	(0.707)	(0.707)	(0.791)
log(Dist TM)				0.041		0.053						
				(0.028)		(0.040)						
μ_U^L	0.250***	0.250***	0.250***	0.242***	0.290***	0.283***	0.250***	0.250***	0.247***	0.247***	0.247***	0.166***
	(0.031)	(0.031)	(0.031)	(0.028)	(0.030)	(0.027)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)	(0.020)
μ_U^H	0.342***	0.342***	0.342***	0.306***	0.875*	0.529***	0.342***	0.342***	0.323***	0.323***	0.323***	0.268***
	(0.048)	(0.048)	(0.048)	(0.038)	(0.471)	(0.133)	(0.048)	(0.048)	(0.043)	(0.043)	(0.043)	(0.031)
Full Controls			X	X	X	X	X	X	X	X	X	X
Basic Controls	X	X										
Comm Floorshare Controls		X										
Network Year	2003	2003	2003	2003	2006	2006	2003	2003	2003	2003	2003	2003
Labor Demand Elasticity	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	2.5	2.5	2.5	1.3
Demand Elasticity	6	6	6	6	6	6	4	9	6	4	9	6
$\bar{h} = 0$												X

Note: Estimates are from 2-step GMM procedure separately for firms at the tract-industry level with 6137 observations and for workers at the tract-group-car ownership level with 7036 observations. Column (1) reports estimates using only basic controls (fixed effects, historical controls, distance to tram dummy and log distance to main road) from main specs. Column (2) adds in the initial share of floorspace allocated to commercial use. Column (3) adds full controls from baseline specification. Column (4) adds a control for log distance to the nearest TransMilenio station, instrumented with the distance to the LCP instrument. Column (5) measures the network as of 2006, while column (6) controls for distance to nearest station using the network in this year. Columns (7) and (8) use alternate values for the demand elasticity, while columns (9)-(11) use alternate values for the labor demand elasticity and demand elasticity. Column (12) 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 at the tract reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A7: Effect of TransMilenio: Robustness

	Output	**Rents**	**Welfare Low**	**Welfare High**
Baseline	3.918	3.721	3.814	4.169
Larger θ	2.964	3.068	2.887	3.150
Larger η	3.599	3.379	3.843	4.310
Alt θ	4.052	3.867	4.012	4.009
Smaller Spillovers	3.387	3.406	3.811	3.811
$\sigma_L = 2.5$	3.894	3.629	3.886	4.202
Census Employment	4.401	3.515	4.361	5.201
$\sigma = 3$	3.875	3.934	3.981	4.258
$\sigma = 9$	4.004	3.222	3.477	3.953

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 1.5 times their estimated values respectively. Row (4) uses alternative values of theta that are more similar across groups using the PPML estimates from the gravity regression estimated in first differences. 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 A8: Welfare Decomposition with Spillovers

	Model 1: Same η, θ			**Model 2:** Diff θ			**Model 3:** Diff η, θ			**Model 4:** Full Model		
	Low	High	Ineq.	Low	High	Ineq.	Low	High	Ineq.	Low	High	Ineq.
Choices Fixed	4.902	4.056	-0.889	4.722	4.537	-0.194	4.706	4.550	-0.164	4.461	4.357	-0.109
Emp. Adjust	5.351	4.135	-1.285	5.226	4.606	-0.654	5.225	4.638	-0.620	5.261	4.731	-0.560
Emp + Car Adjust	5.396	4.189	-1.276	5.268	4.675	-0.626	5.270	4.701	-0.600	5.307	4.796	-0.539
All Choices Adjust	5.601	4.322	-1.354	5.474	4.830	-0.681	5.484	4.843	-0.678	5.530	4.942	-0.622
Rents, All Choices Adjust	4.380	3.496	-0.925	4.191	3.980	-0.221	4.135	3.949	-0.194	4.120	3.995	-0.131
GE	4.408	3.562	-0.885	4.219	4.042	-0.185	4.163	4.010	-0.160	3.814	4.169	0.369

Note: Table shows the percentage welfare gain from phases 1 and 2 of TransMilenio under alternative adjustment scenarios. For each model, I first simulate the counterfactual equilibrium without TransMilenio using unobservables from the 2012 equilibrium. Each entry then reports the welfare gain from the equilibrium with TransMilenio under the adjustment permitted in each row. Analagous to previous tables, the welfare gain is defined as the (negative of the) percentage change in welfare from going from the equilibrium with TM to that without it. Welfare inequality is defined as before. In row (1), all worker choices, rents, wages and land use decisions are held fixed. In row (2), workers employment decisions are allowed to adjust. In row (3), employment and car ownership choices change. In row (4), employment, residential and car ownership decisions adjust. In row (5), all individual choices can change while rents and land use decisions adjust to equilibriate housing markets. In row (6), there is full general equilibrium adjustment. In model 1, both worker groups are assigned the same (average) η and θ parameters and are assumed to be perfect substitutes in production (i.e. $\sigma_L \to \infty$). In model 2, worker groups differ by their estimated θ parameters. In model 3, worker groups differ both by their estimated θ and η parameters. In model 4, workers differ both by their estimates θ and η parameters and are imperfectly substitutable within firms in the way described in the text.

Table A9: 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 A10: 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 A11: 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 A12: 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 included. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A13: 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 applied 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 A14: 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 included 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 A15: Separating $\tilde{\theta}_g$ from ρ_g

	High Skill	Low Skill
$\tilde{\theta}$	1.410	1.704
	(0.008)	(0.006)
Implied ρ	0.314	0.400

Note: Table reports MLE estimates of $\tilde{\theta}$ from ECH/GEIH. Implied ρ is backed out using values for $\theta\kappa$. Standard errors reported in parentheses.

Table A16: 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 A17: Relationship between Predicted and Observed Times Over Time

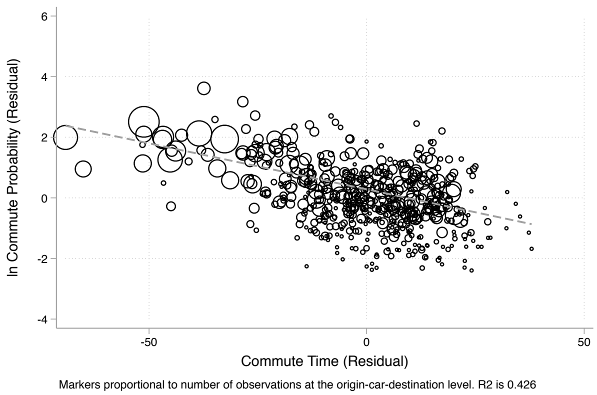
	(1)	(2)	(3)	(4)
ln(Predicted Time)	0.705***	0.511***	0.655***	0.697***
	(0.034)	(0.020)	(0.032)	(0.023)
Post	0.317*	-0.662***	0.151	
	(0.190)	(0.126)	(0.216)	
ln(Predicted Time) X Post	0.018	0.187***	0.046	
	(0.051)	(0.030)	(0.052)	
Car				-0.037
				(0.167)
TM				0.020
				(0.193)
ln(Predicted Time) 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 A1: Fit of Gravity Commuting Model

(a) Overall



(b) By Group

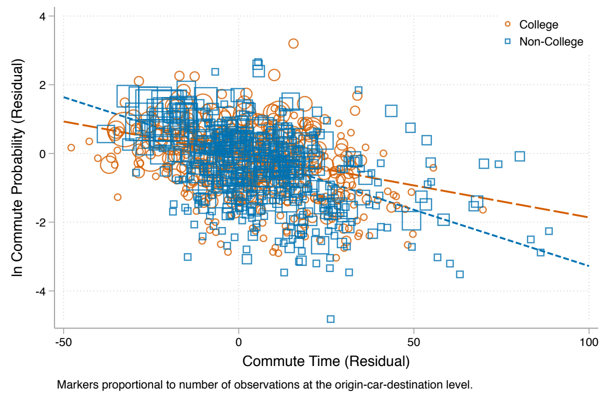


Figure A2: TransMilenio

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

(b) TransMilenio Station, Avenida Caracas (Norte)

Figure A3: Examples of Road Conversion

(a) Portal del Sur, 2002

(b) Portal del Sur, 2010

Figure A4: Engel Curves for Car Ownership and Housing

(a) Car Ownership

(b) Housing Expenditure

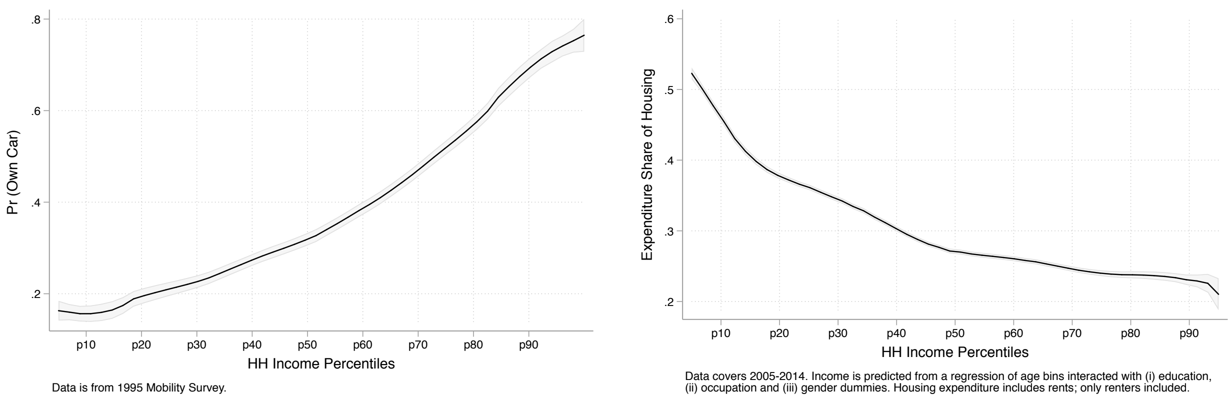
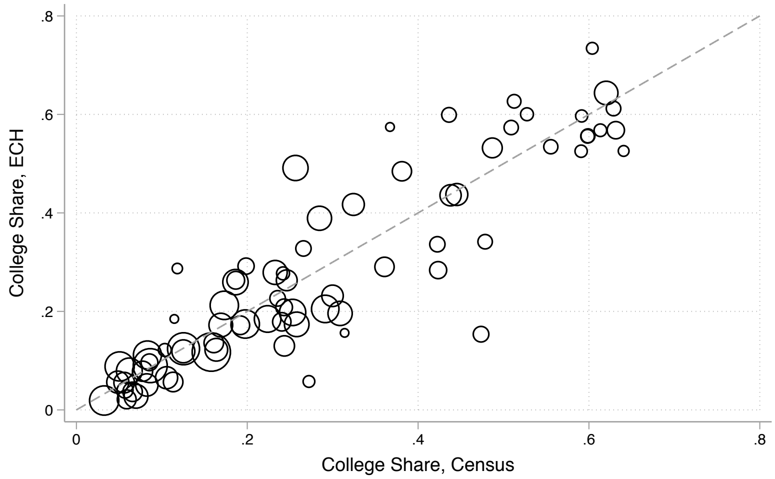
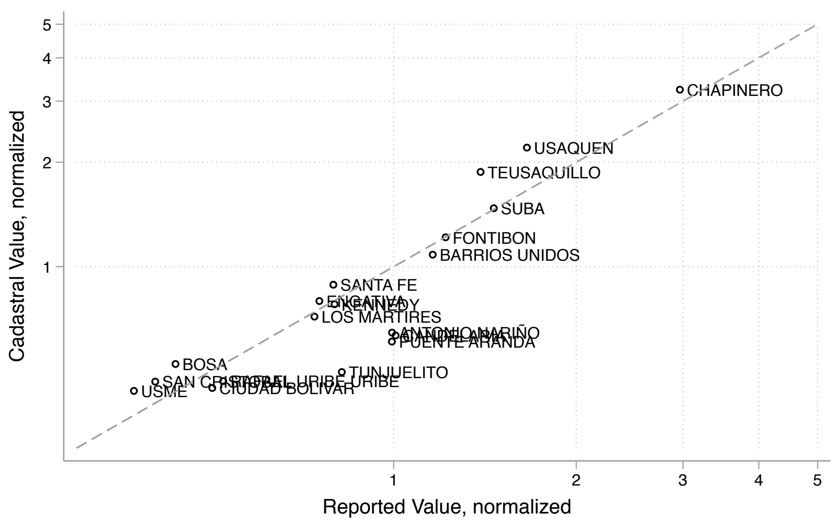


Figure A5: College Share in Census vs ECH, 2005



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

Figure A6: Cadastral vs Reported Property Values

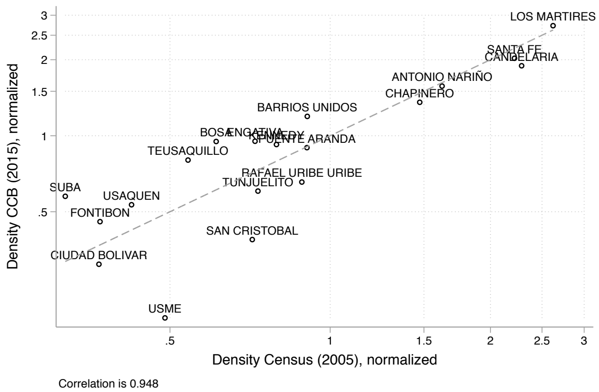


Correlation is 0.948.

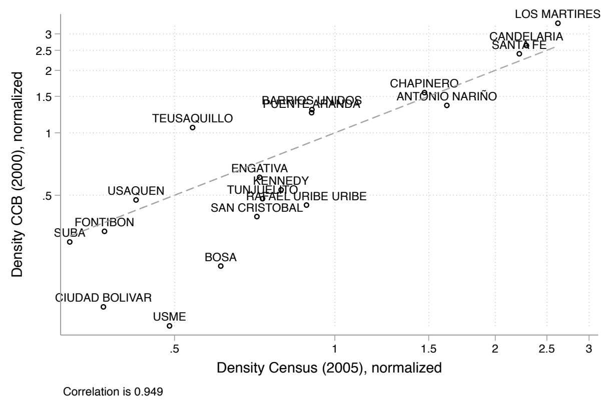
Note: Reported value is the reported purchase price per room as observed in the Multipurpose survey in 2014, for properties bought after 2005 (both the purchase price and year are reported. The cadastral value is the average residential property value per m2 in the locality in that year. Prices are averaged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

Figure A7: Employment in Chamber of Commerce vs Census

(a) 2015 Establishment Comparison by Locality



(b) 2000 Establishment Comparison by Locality



(c) Establishment Comparison by Sector

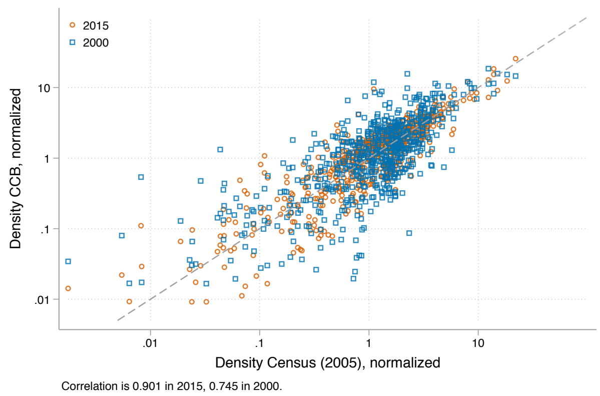
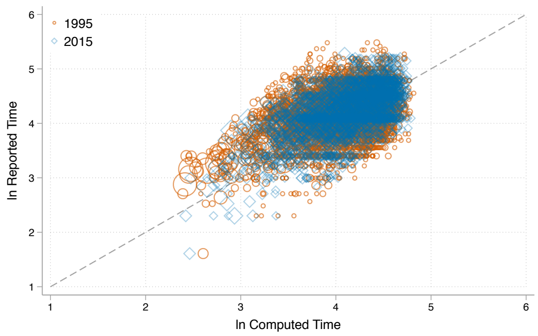


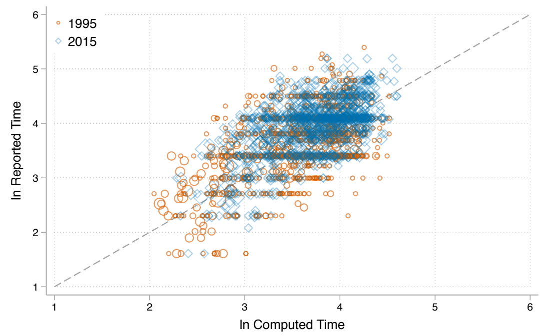
Figure A8: Computed vs Observed Commute Times

(a) Buses



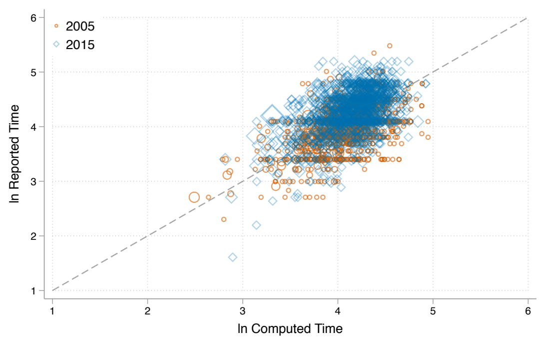
Regression slope is 0.522 in 1995 with an R2 of 0.307, and 0.715 in 2015 with an R2 of 0.406.

(b) Cars



Regression slope is 0.723 in 1995 with an R2 of 0.373, and 0.746 in 2015 with an R2 of 0.357.

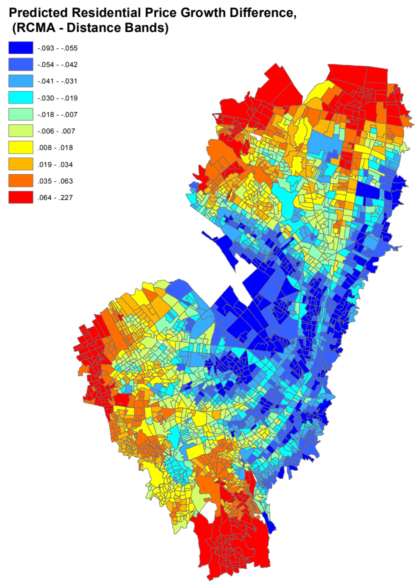
(c) TransMilenio



Regression slope is 0.657 in 2005 with an R2 of 0.308, and 0.713 in 2015 with an R2 of 0.261.

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 (regressions with reported coefficients weighted by this number).

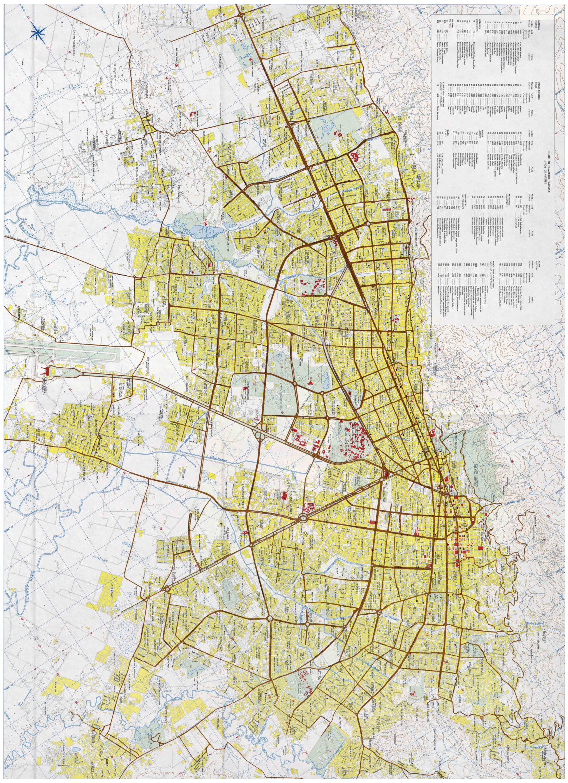
Figure A9: CMA vs Distance Band Predictions For Floorspace Values



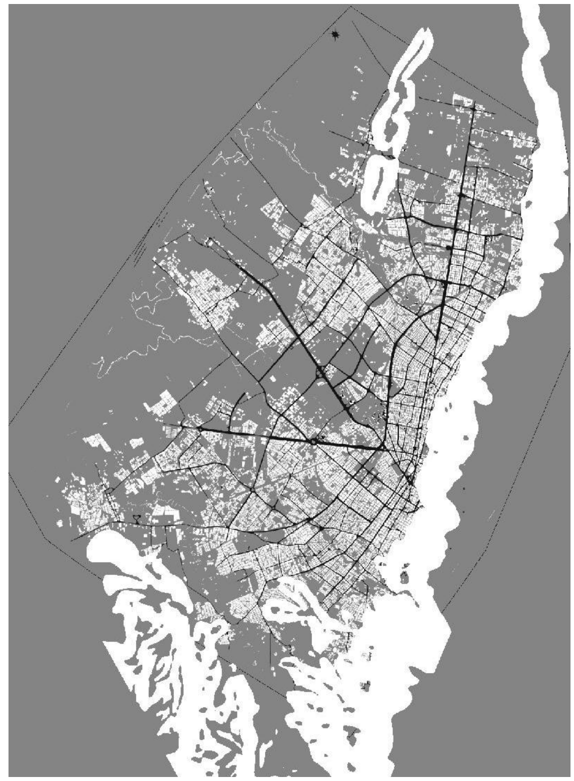
Note: Plot shows the log difference in predicted residential floorspace price growth between the commuter market access specification and the distance band based model. Only tracts within 3km of a TransMilenio station are plotted. The dissimilarity index between the predicted changes, which varies between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

Figure A10: Instruments

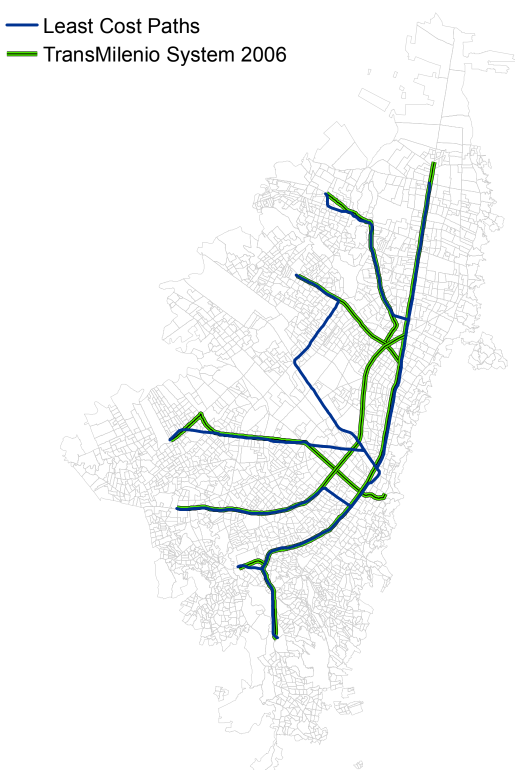
(a) Raw Land Use Map 1980



(b) Cost Raster



(c) LCP Instrument



(d) Tram Instrument

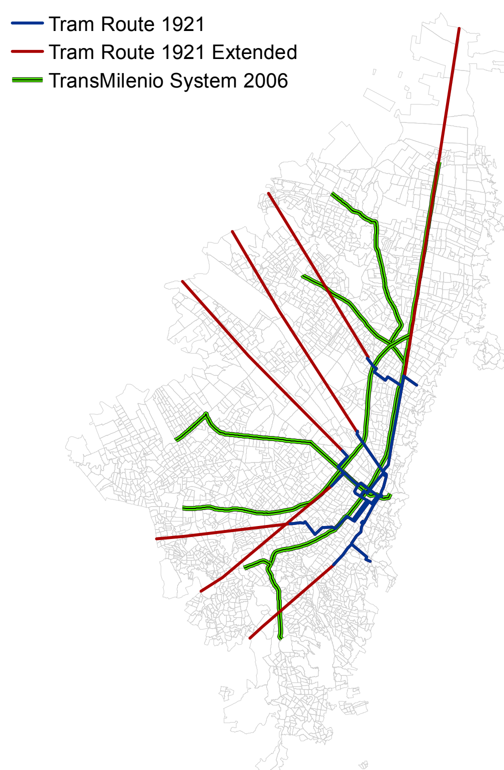
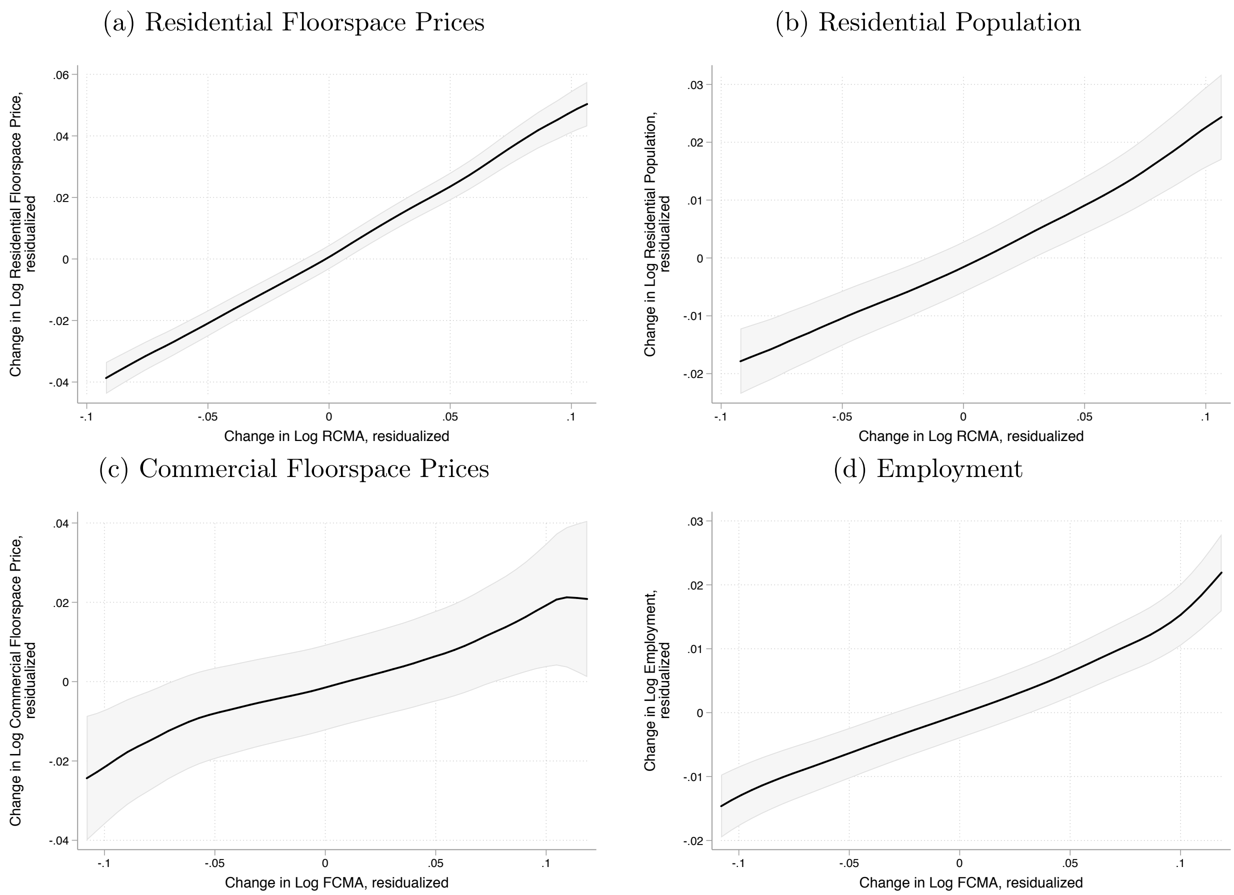


Figure A11: Monte Carlo: Non-Parametric Relationship Between Outcomes and Commuter Market Access



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 F.8.

C Data Appendix

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

C.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. While the education groups are slightly more detailed in 2005, 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.[103]

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.[104] 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 A5 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).

[103] 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.

[104] 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.

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.[105] I include only trips that originate or end in municipal Bogotá in the analysis.[106] 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.[107] For each trip, the data report the departure time, arrival time, purpose of the trip, mode, as well as origin and destination UPZ.[108] Since all trips are reported, these include commutes (trips to work) as well as for other purposes (e.g. shopping, seeing friends). Reported modes are often quite detailed (e.g. 25 options in 2011); I often aggregate into car, bus, TransMilenio, and others (walking, bicycle, motorbike). Trips on TransMilenio trunk and feeder buses are reported separately, so I consider TransMilenio trips to be those involving at least one stage on a trunk bus (multiple modes can be reported in a single trip).

Housing

As described in the main text, the mission of the cadastre is to keep the city's geographical information up to date and thus 98.6% of the city's more than 2 million properties

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

[106] 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)

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

[108] 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.

are included.[109] 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).[110] 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

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

[110] 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).

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).[111] 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,[112] 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 A6 plots the average cadastral price against the reported purchase price, normalizing each variable to have unit mean. The measures have a high correlation coefficient of 0.947, with the majority of observations lying along the 45-degree line. Importantly, there appears to be no deviation of the relationship for expensive neighborhoods, which we would expect if cadastral values were systematically over- or under-valuing these properties.[113] 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

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

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

[113] 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.

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 A16 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.[114] 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

[114] 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.

actual changes (for example, if informal employment is more likely to be located in certain areas than others).[115] 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 A7 plots the density of establishments in each locality in the CCB dataset in each year on the y-axis against the density of establishments in the 2005 census on the x-axis, normalizing both variables to have unit geometric mean. Both figures show a reassuringly tight relationship, with correlations of 0.948 and 0.949 respectively. Importantly, the majority of localities lie along the 45-degree line regardless of whether they are poor (Ciudad Bolivar, Kennedy, Bosa, Tunjuelito) or rich (Chapinero, Usaquen), implying that the coverage is fairly uniform across different types of neighborhoods. Panel (c) confirms that the uniform coverage holds across smaller spatial units, by comparing establishment counts across 631 sectors.

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.[116] 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.[117] 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

[115] 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.

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

[117] 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.

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 which were available to me. 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.[118] 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

The city provides a geodatabase for use in ArcMap containing spatial datasets on the features of Bogotá. From the road layer I extract shapefiles for primary, secondary and tertiary roads. Walk routes consist of the union of the road network in addition to some smaller pedestrian-only paths. The routes of the bus official bus system (which was integrated towards the end of 2012) are also provided. Given that the aim of the government was to bring the provision of existing routes under one integrated system, I use these current routes to measure the location of the bus network throughout the period.[119] Since buses tended to ignore posted

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

[119] 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.

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.[120] 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 its non-monocentric layout.

Finally, 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).

C.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 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

[120]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.

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.[121] 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 E.3 provided evidence that speeds were not changing on routes affected by TransMilenio relative to other locations, Table A10 shows that aggregate speeds were not quite constant over the period. There was a significant reduction in speeds between 1995 and 2005 (a period of city expansion), which remained relatively constant thereafter. I therefore seek to assign two sets of speeds to match the distribution of observed commute times in the "pre" and "post" periods. In the main results, I use the average of both but provide evidence in 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

[121]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.

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.[122]

Figure A8 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 A17, 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

[122]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.

are insignificant for cars and TransMilenio, but is positive for the case of buses. However, inspection of Figure A8 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.[123] 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.

C.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.[124] 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.[125] I then digitize a land use map of the city in 1980 produced by the United States Defense Mapping Agency (Figure A10, 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 A10 shows the results, with lighter shades representing higher cost.

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

[124] 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).

[125] 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.

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 A10 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.[126] 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 A10 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.

D Additional Information on TransMilenio

Trip Characteristics Table A9 presents some descriptives of trips taken in Bogotá in 2015. Three points are worth emphasizing. First, TransMilenio is an important mode of transit constituting 16% of all trips, exceeding the 13.7% taken by cars but less than the roughly 34% taken by bus and walking. Second, the average TransMilenio trip is 10.5km

[126]The chief of the Liberal Party was assassinated during an international conference in Bogota in 1948, afterwhich 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.

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 A10 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 documents that while overall speeds in the city have fallen as Bogotá's population increased from 5.6 to 7.8 million between 1995 and 2015, TransMilenio is on average 26.7% faster than buses and roughly the same speed as trips taken by cars.[127][128] 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 A11 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.

[127]Note that these are observed speeds rather than system speeds: TransMilenio buses are reported to operate faster than the results in Table A10 suggest, but queueing at stations and time taken to walk between stations and final destinations decrease average observed speeds.

[128]Of course, these average speeds are likely conflated by the different nature of trips taken across modes (such as TransMilenio being used for longer trips, which are typically faster than short trips). Section E.1 in the appendix presents a regression-based comparison of speeds across these modes that controls 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.

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.[129] 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.

E Supplementary Empirical Results

E.1 Regressions of Relative Speed over Time

Table A12 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 A10. 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

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

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.

E.2 Effect of TransMilenio on Growth in Floorspace

In Table A13, 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.[130] 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

[130] 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).

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.

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.

E.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,[131] 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

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

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 A14 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.[132]

E.4 Engel Curve for Housing

In Figure A4, 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

[132]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.

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 A4 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.

F Supplementary Quantitative Results

F.1 Estimating $\tilde{\theta}_g$ and ρ_g

In the paper, I estimate the parameter cluster $\theta_g = \tilde{\theta}_g/(1-\rho_g)$ using the gravity equation for commute flows. It turns out that this cluster is all that matters for computing equilibria in the model. However, in this section I show how comparative and absolute advantage can be separated using the estimate for θ_g and wage data via maximum likelihood.

An appealing feature of the Frechet distribution is that the distribution of the maximum wage of type-g individuals within a location i under car ownership a is also Frechet and given by $F_{iag}(w) = \exp\left(-(T_g\Phi_{Riag})^{1-\rho_g} z^{-\tilde{\theta}_g}\right)$. Taking logs of the associated density and summing over all residential location and car ownership combinations yields the following log-likelihood for observed wages

$$\ell_n\left(\tilde{\theta}_g, \tilde{T}_g\right) = \ln\left(\tilde{\theta}_g \tilde{T}_g\right) - (\tilde{\theta}_g + 1)\frac{1}{N}\sum_i \ln w_i - \tilde{T}_g \frac{1}{N}\sum_i w_i^{-\tilde{\theta}_g}$$

where I have defined $\tilde{\theta}_g \equiv \theta_g(1-\rho_g)$ and $\tilde{T}_g \equiv T_g^{1-\rho_g}\sum_{i,a}\Phi_{Riag}^{1-\rho_g}$. I therefore estimate $\tilde{\theta}_g$ via MLE and report the results in Table A15. In addition, the third row reports the value of the correlation parameter ρ_g implied by these estimates for $\tilde{\theta}_g$ and θ_g. Low-skill workers have a lower degree of comparative advantage (i.e. a higher $\tilde{\theta}_g$, indicating less dispersion) and greater within-individual correlation of match-productivity (i.e. a higher ρ_g).

F.2 Calibrating $T_H, \bar{h}, p_a$

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

$$\widehat{WP} = \frac{T_H \sum_{ia} \Phi_{RiaH}^{1/\theta_H} \lambda_{iaH}}{\sum_{ia} \Phi_{RiaL}^{1/\theta_L} \lambda_{iaL}}$$

$$w_g = F_g(w_g; L_{Fs}, L_{Rg}, T_H)$$

where $\widehat{WP}$ 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 H.

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,[133] $\hat{\omega}_H = 0.3075$ 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 $T_H = 0.81$, $\bar{h} = 1.045$ and $p_a = 101.697$.

F.3 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

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

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

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

   (c) Compute L^t_R from

$$L^t_{Riag} = \bar{L}_g \frac{\left(u_{iag}(T_g \Phi^{1/\theta}_{Riag} - \bar{h}r^t_{Ri} - p^t_a a)r^{\beta-1}_{Ri}\right)^{\eta_g}}{\sum_{r,o}\left(u_{rog}(T_g \Phi^{1/\theta}_{Rrog} - \bar{h}r_{Rr} - p^t_o o)r^{\beta-1}_{Rr}\right)^{\eta_g}} \quad \forall i \in \mathcal{A}^t_g$$

   where $\mathcal{A}^t_g = \left\{(i,a) : T_g\Phi^{1/\theta}_{Riag} - \bar{h}r^t_{Ri} - p^t_a a > 0\right\}$, $p^t_a = p_a P^t$ and $L^t_{Riag} = 0$ otherwise.

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

   (e) Update the main variables

$$\tilde{w}_{jg} = \left[\frac{(P^t)^{\sigma-1}X^t\sum_s B_{isg}A^{\sigma-1}_{is}(W^t_{is})^{\sigma_L-(1+\alpha_s(\sigma-1))}(r^t_{Fi})^{-(1-\alpha_s)(\sigma-1)}}{\Psi^t_{jg}}\right]^{\frac{1}{\theta_g+\sigma_L-1}}$$

$$\tilde{r}_i = \frac{E^t_i + (1-\alpha)Y^t_i}{H_i}$$

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

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

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

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

$\beta) \sum_{a,g}(T_g(\Phi^t_{Riag})^{1/\theta_g} - p^t_a a - r^t_i \bar{h})L^t_{Riag}$ is expenditure on housing.

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

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

F.4 Welfare Decomposition

In this section, I outline the procedure to perform the welfare decomposition reported in the paper. First, I simulate the effect of removing the TransMilenio network from the 2012 equilibrium. I then calculate the effect of adding back TransMilenio allowing for different margins of adjustment. Labor and land market adjustment simply requires imposing the appropriate equilibrium condition from the previous section. However, the expression to compute the change in welfare depends on the assumption made over which choices are allowed to adjust. Since the table reports these welfare changes under partial equilibrium where wages, floorspace prices and land use are fixed (at their levels without the network in the initial, counterfactual equilibrium) I impose this restriction here too to economize on notation.

Consider the first case where all choices are fixed. Using the result from the Frechet distribution that the expected income of workers who have chosen (i,a) is equal across destinations, expected income of workers of type-g commuting to j from (i,a) is given by $\tilde{y}_{iag} = \bar{y}_{iag} - r_i\bar{h} - p_a a$. Following the change in commute costs $\hat{d}_{ija} \equiv d'_{ija}/d_{ija}$, the new gross income for commuters to j is $\bar{y}_{iag} \frac{1}{\hat{d}_{ija}}$. Thus, the new gross income for residents of i is

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

is the average chnage in commute costs holding all choices constant. Thus, the change in net income can be written as

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

The change in average utility is then

$$\hat{\bar{U}} = \sum_{i,a} \pi_{ia|g} \hat{\tilde{y}}_{iag}.$$

Next, consider the case where employment decisions can adjust. In this model, expected income is defined as in the text in both equilibria so the change can be written as

$$\hat{\bar{y}}_{iag} = \left[\sum_j \pi_{j|iag} \hat{d}_{ij}^{-\theta_g} \right]^{1/\theta_g}$$

In the previous model, the change in expected income was determined by the commute weighted average of the change in commute costs since choices were fixed. In this model, the change instead reflects individuals' changing employment location decisions. In this case, the change in average net income is

$$\hat{\tilde{y}}_{iag} = \frac{\bar{y}_{iag} \hat{\bar{y}}_{iag} - r_i \bar{h} - p_a a}{\tilde{y}_{iag}}$$

and the expression for the average change in utility $\hat{\bar{U}} = \sum_{i,a} \pi_{ia|g} \hat{\tilde{y}}_{iag}$ is unchanged from before.

Now suppose that both employment and car ownership are allowed to change. The change in expected income is the same as in the previous model, but now average utility for residents of i is given by $\bar{U}_{ig} = \left[\sum_a \left(u_{iag}\tilde{\bar{y}}_{iag} r_i^{\beta-1}\right)^{\eta_g}\right]^{\frac{1}{\eta_g}}$. Thus, in partial equilibrium changes we have

$$\hat{\bar{U}}_{ig} = \left[\sum_a \pi_{a|ig} \hat{\tilde{\bar{y}}}_{iag}^{\eta_g}\right]^{\frac{1}{\eta_g}}$$

where $\pi_{a|ig}$ is the share of car owning residents of i in the initial equilibrium. The change in overall utility is

$$\hat{\bar{U}}_g = \sum_i \pi_{i|g} \hat{\bar{U}}_{ig}$$

where $\pi_{i|g}$ is the share of type-g workers living in i in the initial equilibrium.

Finally, where all decisions are allowed to adjust, the change in welfare is determined by the expression in the paper

$$\hat{\bar{U}}_g = \left[\sum_{i,a} \pi_{ia|g} \hat{\tilde{\bar{y}}}_{iag}^{\eta_g}\right]^{\frac{1}{\eta_g}}$$

where $\pi_{ia|g}$ are the residential choice probabilities from the paper in the initial equilibrium.

F.5 Model with Employment in Domestic Services

I first note 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.

As discussed above, 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.

Model Extension Given these observations, I now extend the baseline model to incorporate domestic services in the following way. First, I assume that only high skilled households consume domestic services while only low skill workers are used in its production. Second, I assume that 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.

To solve for wages, I extend the system of equations to incorporate the additional domestic servant sector. In particular, the system becomes

$$D_{ig}(w) = w_{ig}^{\theta_g}\left[\sum_s \frac{L_{Rsg}}{\sum_k w_{kg}^{\theta_g} d_{sk}^{-\theta_g}} d_{si}^{-\theta_g}\right] - \left[\sum_s \frac{(w_{ig}/\alpha_{sg})^{-\sigma}}{\sum_h (w_{ih}/\alpha_{sh})^{-\sigma}} \frac{\bar{\epsilon}_{is}}{\bar{\epsilon}_{ig}} L_{Fis} + L_{FiD}\mathbb{I}_{gL}\right]$$

where $\mathbb{I}_{gL}$ is a dummy for whether g is L, and L_{FiD} is employment in domestic services as described above.

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

$$L_{Riag} = \bar{L}_g \frac{\left(u_{iag}(T_g \Phi_{Riag}^{1/\theta} - \bar{h} r_{Ri} - p_a a) r_{Ri}^{\beta-1} w_{Li}^{\beta_g^D}\right)^{\eta_g}}{\sum_{r,o} \left(u_{rog}(T_g \Phi_{Rrog}^{1/\theta} - \bar{h} r_{Rr} - p_o o) r_{Rr}^{\beta-1} w_{Lr}^{\beta_g^D}\right)^{\eta_g}} \text{ , } g = H.$$

Amending these two equilibrium conditions, the procedure to solve for unobservables and compute counterfactual equilibria is unchanged.

F.6 Model with Variable Housing Supply

In this section, I provide details on the model with variable housing supply elasticity described in the paper. I assume 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, I introduce a residual $\zeta_i = H_i/T_i(\eta r_i)^{\frac{\eta}{1-\eta}}$ 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, I make a conservative choice for the housing elasticity $\eta/(1-\eta) = 0.7$ 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).

F.7 Model with Home Ownership

In this section, I extend the model to allow individuals to own their own homes. In the data, I observe home ownership rates of 0.603 and 0.457 for college and non-college individuals respectively in 2015. I therefore assume that land rents in a location are split between three sources. First, a share λ_{Li} goes to low-skill workers who live there. This is determined by the share of low-skilled workers who own their home and total expenditures of the low-skilled residents in i on housing. Second, a share λ_{Hi} goes to high-skill workers which is defined analagously. Third, the remainder (i.e. rents from the remaining residential floorspace and all commercial floorspace) goes into an aggregate portfolio which is owned in equal shares by all residents.[134]

Total land rents in location i are

$$\Pi_i = r_{Ri}H_{Ri} + r_{Fi}H_{Fi}$$

Letting o_L and o_H be the shares of home owners in the data, this implies that

$$\begin{aligned}
\lambda_{iL} &= \frac{o_L \times r_{Ri}H_{RiL}}{\Pi_i} \\
\lambda_{iH} &= \frac{o_H \times r_{Ri}H_{RiH}}{\Pi_i} \\
1 - \lambda_{iL} - \lambda_{iH} &= \frac{(1-o_L)r_{Ri}H_{RiL} + (1-o_H)r_{Ri}H_{RiH} + r_{Fi}H_{Fi}}{\Pi_i}
\end{aligned}$$

are the shares of local land rents paid to low-skill workers, high-skill workers and the aggregate portfolio respectively.

Payments to the aggregate portfolio are $\tilde{\Pi} = \sum_i (1 - \lambda_{iL} - \lambda_{iH})\Pi_i$ while payments to the each group from the local fund are $\lambda_{ig}\Pi_i$. This implies that each individual receives $\tilde{\pi}$ and π_{ig} from each respectively where

$$\tilde{\pi} = \frac{\tilde{\Pi}}{\bar{L}} \quad \text{and} \quad \pi_{ig} = \frac{\lambda_{ig}\Pi_i}{L_{Rig}}.$$

Income from living in location i is equal to the sum of labor income and income from home ownership given by

$$\bar{y}_{iag} = T_g \Phi_{Riag}^{1/\theta_g} + \tilde{\pi} + \pi_{ig}.$$

The remaining equilibrium equations are unchanged.

[134] I could also assume that individuals own group-specific shares of the aggregate portfolio. It is likely that high-skill workers would own greater shares of this aggregate portfolio than the low-skilled, since a greater share of them are likely to own multiple properties. By giving workers an equal share of aggregate land rents, my setup therefore provides an upper bound of the effects of home ownership on the relative welfare of low-skill workers.

Calibration and Counterfactuals I calibrate the shares λ_{ig} to exactly match the data in the initial equilibrium. Since unobservables now need to be solved for jointly, I amend the procedure outlined in the proof of proposition 3 as follows.[135]

1. The procedure to obtain $\Phi_{Riag}, W_{is}, A_{js}, \xi_{Fi}$ is unchanged.

2. Pick a vector of residential housing quantity unobservables ξ^0_{Ri}. Given ξ^t_{Ri}

   (a) Compute $\Pi^t_i = r_{Ri}\tilde{H}_{Ri}\xi^t_{Ri} + r_{Fi}\tilde{H}_{Fi}\xi_{Fi}$

   (b) Pick λ^0_g. Given λ^τ_g

      i. Compute payment from aggregate and local stock $\tilde{\pi}^\tau = \frac{\sum_i (1-\lambda^\tau_{iL}-\lambda^\tau_{iH})\Pi_i}{\bar{L}}$ and $\pi^\tau_{ig} = \frac{\lambda^\tau_{ig}\Pi_i}{\bar{L}_{Rig}}$.

      ii. Solve for $p^\tau_a, \bar{h}^\tau$ from

$$p^\tau_a = \frac{\hat{\omega}_C}{\sum_{i,g}\lambda^C_{ig}\frac{P}{T_g\Phi^{1/\theta_g}_{Riag}+\tilde{\pi}^\tau+\pi^\tau_{ig}}}$$

$$\bar{h}^\tau = \frac{\hat{\omega}_H-(1-\beta)}{\sum_{i,a,g}\frac{r_{Ri}L_{Riag}}{T_g\Phi^{1/\theta_g}_{Riag}-p_a a+\tilde{\pi}^\tau+\pi^\tau_{ig}}\lambda_{iag}}$$

      iii. Compute expenditure on housing

$$E^\tau_{Hig} = \sum_a \bar{h}^\tau r_{Ri}L_{Riag} + (1-\beta)\left(T_g\Phi^{1/\theta}_{Riag} - \bar{h}^\tau r_{Ri} - p^\tau_a a + \tilde{\pi}^\tau + \pi^\tau_{ig}\right)L_{Riag}$$

      iv. Compute λ' from

$$\lambda'_{ig} = \frac{o_g \times E^\tau_{Hig}}{\Pi^t_i}$$

      and update $\lambda^{\tau+1} = \zeta\lambda^\tau + (1-\zeta)\lambda'$. Continue until convergence.

   (c) Compute $u^t_{iag} = \frac{(L_{Riag}/\lambda^t_{Lg})^{1/\eta_g} r^{1-\beta}_{Ri}}{(T_g\Phi^{1/\theta}_{Riag}-\bar{h}^t r_{Ri}-p^t_a a+\tilde{\pi}^t+\pi^t_{ig})}$.

[135] For example, the λ_{ig}'s depend on the housing quantity unobservables ξ_{Ri}, ξ_{Fi} which in turn depend on the λ_{ig}'s through expenditure on housing.

(d) Compute $\xi'_{Ri} = \frac{\sum_g E^t_{Hig}}{\tilde{H}_{Ri} r_{Ri}}$, update $\xi^{t+1}_{Ri} = \zeta \xi'_{Ri} + (1-\zeta)\xi^t_{Ri}$ and continue until convergence.

With λ_{ig} in hand, counterfactuals are computed in the same way as before only this time using the amended definition for expected income in each location. As with other unobservables, I assume that the shares λ_{ig} are constant across equilibria.

F.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.[136] 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 A11 plots the results. Each relationship displays a tight, log-linear relationship.

[136] 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 suggests that the reduced form regressions in the paper are consistent as a model validation exercise for the full model.

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

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

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

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

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

$$\begin{aligned}
\pi_{m|ija} &= \pi_{k|ija} \times \pi_{m|ijka} \\
&= \frac{\left(\sum_{n\in\mathcal{B}_k} \exp\left(b_n - \frac{\kappa}{\lambda_k} t_{ijn}\right)\right)^{\lambda_k}}{\sum_{k'}\left(\sum_{n\in\mathcal{B}_{k'}} \exp\left(b_n - \frac{\kappa}{\lambda_{k'}} t_{ijn}\right)\right)^{\lambda_{k'}}} \times \frac{\exp\left(b_m - \frac{\kappa}{\lambda_k} t_{ijm}\right)}{\sum_{n\in\mathcal{B}_k} \exp\left(b_n - \frac{\kappa}{\lambda_k} t_{ijn}\right)}
\end{aligned} \tag{18}$$

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

$$\begin{aligned}
\bar{d}_{ija} &= \exp\left(\kappa \bar{t}_{ija}\right) \\
\text{where } \bar{t}_{ij0} &= -\frac{\lambda}{\kappa} \ln \sum_{m\in\mathcal{B}_{\text{Public}}} \exp\left(b_m - \frac{\kappa}{\lambda} t_{ijm}\right) \\
\bar{t}_{ij1} &= -\frac{1}{\kappa} \ln\left(\exp(b_{car} - \kappa t_{ij\text{Car}}) + \exp\left(\kappa \bar{t}_{ij0}\right)\right)
\end{aligned} \tag{19}$$

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

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

which is the expression in the paper.[137]

G.2 Full Definition of CMA Regression Coefficient Matrices

Each term in the system reported in the text is given by

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

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

Computing $A^{-1}B_R$ and $A^{-1}B_F$ gives the expressions in the paper. The reduced form error $A^{-1}e_i$ is a linear combination of changes in the exogenous components of amenities and productivities, as well as changes in exogenous residential and commercial floorspace given by

$$A^{-1}e_i = \begin{bmatrix} \frac{\eta}{1+\eta(1-\beta-\mu_U)} \left(\Delta \ln \bar{u}_i + (1-\beta)\Delta \ln H_{Ri}\right) \\ \frac{1}{1+\eta(1-\beta-\mu_U)} \left(\eta \Delta \ln \bar{u}_i - (1-\eta\mu_U)\Delta \ln H_{Ri}\right) \\ \frac{1}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha-\mu_A(1-\alpha)(\sigma-1))} \big((\theta+\mu_A(\sigma-1)(\theta-1))(\sigma-1)\Delta \ln \bar{A}_i - (\alpha(\sigma-1)+\theta)\Delta \ln H_{Fi}\big) \\ \frac{(\theta-1)(\sigma-1)}{\theta\sigma-(\theta-1)(\sigma-1)(\alpha-\mu_A(1-\alpha)(\sigma-1))} \big(\Delta \ln \bar{A}_i + (1-\alpha)\Delta \ln H_{Fi}\big) \end{bmatrix}$$

[137] 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.3 Mapping Models to Gravity Framework

G.3.1 My Model

Consider the model characterized by the system of equations in section 5.1 of the paper. Using that $L_{ij} = \pi_{j|i} L_{Ri}$ we have that

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

so that commute flows are of the gravity form. Note that from this expression, we can compute the supply of residents $L_{Ri} = \gamma_i \sum_j \delta_j \kappa_{ij} = \gamma_i \Phi_{Ri}$ and $L_{Fj} = \delta_j \sum_i \gamma_i \kappa_{ij} = \delta_j \Phi_{Fj}$ by straightforward accounting.

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

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

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

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

So we have $\beta = -\frac{1}{\eta(1-\beta-\mu_U)}$ and $\gamma = \frac{\eta\beta-\theta}{\theta\eta(1-\beta-\mu_U)}$.

G.3.2 Ahlfeldt et. al. (2015)

The model in Ahlfeldt et. al. (2015) with fixed supplies of residential and commercial floorspace, no spillovers across locations and productivity rather than preference shocks is the same as that above, but with with joint employment and residential decisions ($\eta = \theta$) and perfect competition ($\sigma \to \infty$). Thus, this falls within this class. This also highlights formally the isomorphism between different timing assumptions in the model with no non-homotheticities.

G.3.3 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 Frechey 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

$$L_{Fj} = A_j w_j^{-\theta_F - 1} P^{\theta_F} E$$

which is of the form in proposition 2.

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 that depends on the elasticity of demand 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 profits

$$
w_j \ell_j = (\sigma - 1)\pi_j
$$

Since total employment is simply $L_{Fj} = N_j \ell_j$, we find that

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

where λ_F is an equilibrium constant.

In this example, 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.

G.3.4 Alternate Housing

Elastic Supply In deriving the regression equations used in the paper, housing supply was assumed to be perfectly inelastic. In this section, I show the same specifications can be extended to incorporate a 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

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

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

No Housing The model can also exclude housing by setting $\alpha = \beta = 1$ (as in Allen and Arkolakis 2014), since idiosyncratic preferences provide a dispersion force that ensures activity does not converge to a single location.

G.3.5 Leisure

Suppose individuals have Cobb-Douglas preferences over goods, housing and leisure and we allow for labor-leisure decision and time affects total endowment in a model with a joint location decision

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

Indirect utility is then

$$
U_{ij\omega} = \frac{u_i w_j^{1-\gamma} r_{Ri}^{-\beta}}{d_{ij}} \epsilon_{ij} \Rightarrow L_{ij} = \left(u_i w_j^{1-\gamma} r_{Ri}^{-\beta} \right)^{\theta} d_{ij}^{-\theta}
$$

where $d_{ij} \equiv \frac{1}{1-t_{ij}}$ so that commute flows are still of the gravity form. Since the rest of the model is unchanged, this setup fits the requirements of proposition 3.

G.4 Gravity Framework for Models with Preference Shocks

In this section I show how a similar regression framework applies to models with preference rather than productivity shocks. There is an analogous statement of proposition 3 in this setup that generalizes this example, available upon request.

Consider the simplified model with one group of workers, firms and commute modes when the $\epsilon_{j\omega}$ are preference rather than productivity shocks. The equilibrium equations are given by

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

where $\bar{y}_i = \sum_j \pi_{j|i} w_j$ is average income. Note that this can computed having solved for residential and firm commuter market access, since $w_j = (L_{Fj}/\Phi_{Fj})^{1/\theta}$. Using that $\bar{y}_i = \frac{\tilde{\Phi}_{Ri}}{\Phi_{Ri}}$ where $\tilde{\Phi}_{Ri} = \sum_j w_j^{1+\theta}/d_{ij}^{\theta}$, we can write this in changes as the linear system

$$\begin{bmatrix} 1-\eta\mu_U & \eta(1-\beta) & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & (\sigma-1)(1-\alpha) & \frac{(\sigma-1)(\alpha-\mu_A\theta)+\theta+1}{\theta} \\ 0 & 0 & 1+(\sigma-1)(1-\alpha) & \frac{(\sigma-1)(1-\mu_A\theta)}{\theta} \end{bmatrix} \begin{bmatrix} \Delta\ln L_{Ri} \\ \Delta\ln r_{Ri} \\ \Delta\ln r_{Fi} \\ \Delta\ln L_{Fi} \end{bmatrix}$$
$$= \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \Delta\ln\tilde{\Phi}_{Ri} + \begin{bmatrix} \frac{\eta}{\theta} \\ -1 \\ 0 \\ 0 \end{bmatrix} \Delta\ln\Phi_{Ri} + \begin{bmatrix} 0 \\ 0 \\ \frac{1+\alpha(\sigma-1)}{\theta} \\ \frac{\sigma-1}{\theta} \end{bmatrix} \Delta\ln\Phi_{Fi} + \begin{bmatrix} \eta\Delta\ln\bar{u}_i \\ -\Delta\ln H_{Ri} \\ (\sigma-1)\Delta\ln\bar{A}_i \\ (\sigma-1)\Delta\ln\bar{A}_i - \Delta\ln H_{Fi} \end{bmatrix}$$

Log-linearizing the term $\Delta\ln\tilde{\Phi}_{Ri}$ around the initial equilibrium we find that

$$\Delta\ln\tilde{\Phi}_{Ri} \approx (1+\theta)\sum_j \frac{w_j^{1+\theta}d_{ij}^{-\theta}}{\sum_s w_s^{1+\theta}d_{is}^{-\theta}}\Delta\ln(w_j/d_{ij}) - \theta\sum_j \frac{w_j^{\theta}d_{ij}^{-\theta}}{\sum_s w_s^{\theta}d_{is}^{-\theta}}\Delta\ln(w_j/d_{ij})$$

when the weights are approximately equal $\frac{w_j^{1+\theta}d_{ij}^{-\theta}}{\sum_s w_s^{1+\theta}d_{is}^{-\theta}} \approx \frac{w_j^{\theta}d_{ij}^{-\theta}}{\sum_s w_s^{\theta}d_{is}^{-\theta}}$, we have that $\Delta\ln\tilde{\Phi}_{Ri} \approx \Delta\ln\Phi_{Ri}$ and thus the system can be approximated as

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

where $B_R = \begin{bmatrix} \frac{\eta}{\theta} & 1 & 0 & 0 \end{bmatrix}$ and the rest of the coefficients are as given above.

G.5 Allowing for Spatial Spillovers in Productivity and Amenities

In my setting, I use data at the census tract and so, given the evidence of rapidly decaying spatial spillovers, model the productivity and amenities in a location depending only on own-location activity. However, other frameworks such as Ahlfeldt et. al. (2015) allow for such spillovers. In this section, I show the market access regression specifications apply in these models.

Consider the expression for employment which, using the result of the previous section, is

$$L_{Fi} = \tilde{\kappa}_1 H_{Fi}^{\frac{\theta(1-\alpha)}{1+\theta(1-\alpha)}} A_i^{\frac{\theta}{1+\theta(1-\alpha)}} \Phi_{Fi}^{\frac{1}{1+\theta((1-\alpha)}}$$
$$A_i = \bar{A}_j \left[\sum_k d_{jk}^{-\delta_A} L_{Fk}\right]^{\mu_A}$$

The only difference is now productivity depends on employment in all surrounding locations, with the spatial decay $\delta_A > 0$ determining how fast these spillovers decline over space. Thus, my model is a special case of this setup with $\delta_A \to \infty$.

Log-linearizing changes in productivity around the initial equilibrium yields

$$\Delta \ln A_j = \Delta \ln \bar{A}_j + \mu_A \sum_k \lambda_{jk}^A \Delta \ln L_{Fk}$$

where $\lambda_{jk}^A = \frac{d_{jkt-1}^{-\delta_A} L_{Fkt-1}}{\sum_{k'} d_{jk't-1}^{-\delta_A} L_{Fk't-1}}$. Productivity growth depends on the change in fundamental productivity, and a geometric average of the change in employment in all other locations in the city, with weights reflecting the importance of each location in determining a given location's productivity in the initial equilibrium. Distant locations (with small $d_{jkt-1}^{-\delta_A}$) will carry little weight, while nearby locations matter a lot.

Replacing this into the expression for employment growth yields

$$\Delta \ln L_{Fj} = \frac{\theta}{1+\theta(1-\alpha)} \Delta \ln \bar{A}_j + \frac{\mu_A \theta}{1+\theta(1-\alpha)} \sum_k \lambda_{jk}^A \Delta \ln L_{Fk}$$
$$+ \frac{\theta(1-\alpha)}{1+\theta(1-\alpha)} \Delta \ln H_{Fj} + \frac{1}{1+\theta(1-\alpha)} \Delta \ln \Phi_{Fj}$$

Writing this in matrix form we find

$$\Delta \ln L_F = (I - \mu_A \tilde{\theta} \Lambda_A)^{-1} \left(\tilde{\theta} \Delta \ln \bar{A} + \tilde{\theta}(1-\alpha) \Delta \ln H_F + \frac{1}{1+\theta(1-\alpha)} \Delta \ln \Phi_F \right)$$

where $\Lambda_A = \left[\lambda_{ij}^A\right]_{ij}$. Changes in market access have two effects on employment in a location. First, there is a direct effect through which an increase in market access increases labor supply to that block. This direct effect enters with elasticity $\frac{1}{1+\theta(1-\alpha)}$. Second, there is a spillover effect through which an increase in employment in one location increases the productivity of surrounding locations, which in turn increases the productivity of their surrounding locations, and so on. The total effect is captured by the matrix

$$(I - \mu_A \tilde{\theta} \Lambda_A)^{-1} = \underbrace{I}_{\text{direct effect}} + \underbrace{\sum_{k=1}^{\infty} \left(\mu_A \tilde{\theta} \Lambda_A \right)^k}_{\text{indirect effect}}$$

which resembles a spatial Leontief matrix. In this case, the regression equation becomes

$$\Delta \ln L_{Fj} = \frac{\gamma_{jj}^A}{1+\theta(1-\alpha)} \Delta \ln \Phi_{Fj} + \frac{1}{1+\theta(1-\alpha)} \sum_{k \neq j} \gamma_{jk}^A \Delta \ln \Phi_{Fk} + \nu_{Fj}$$

So in this case, the same specification applies but this time the effect of changes in CMA is heterogeneous, and will be greater for locations with a greater share of proximate employment (high γ_{jj}^A). Note that as $\delta_A \to \infty$, we get $\gamma_{jk}^A \to 0 \ \forall k \neq j$ and $\gamma_{jj}^A \to 1$ so that the second effect drops out and the treatment effect becomes constant once more. An equivalent approximation applies to specifications for residential outcomes when amenities decay over space.

G.6 Value of Travel Time Savings Approach

The typical approach to evaluate the gains from commuting infrastructure is based on the Value of Travel Time Savings (VTTS) approach (e.g. Small and Verhoef 2007). In this framework, the benefits from new infrastructure are given by the marginal value of time times the amount of time saved

$$VTTS_{ij} = \frac{\partial U}{\partial t} \Delta t_{ij}$$

where $\Delta t_{ij} \equiv t'_{ij} - t_{ij}$ is the amount of time saved across the equilibria with and without the new infrastructure. While the marginal value of time can vary with trip and person characteristics depending on the model, typical approaches estimate this to be a constant equal to approximately 50% of prevailing wages (e.g. Victoria Transport Policy Institute 2016).

Since $\frac{\partial U}{\partial t}$ is constant within worker groups, we have that

$$VTTS_{ij} = \frac{\partial U}{\partial t}(t'_{ij} - t_{ij}) \Rightarrow \widetilde{VTSS}_{ij} = \frac{\Delta t_{ij}}{t_{ij}}$$

where $\widetilde{VTSS}_{ij}$ is $VTTS_{ij}$ expressed in percentage terms. The aggregate VTTS by group is computed by aggregating over all commutes

$$\widetilde{VTSS}_g = \sum_{i,j} \frac{L^{pub}_{ijg}}{\bar{L}_g} \widetilde{VTSS}_{ijg}$$

I compute this measure using the commuting data by locality in the 2015 Mobility Survey, where t and t' correspond to the times with and without TransMilenio respectively.

G.7 Sufficient Statistic for Welfare

In this section, I show how the special case of the model considered in Section 5 of the paper admits a sufficient statistic for the welfare gains from commuting infrastructure. For simplicity, consider the model without spillovers. From the expression for residential populations $L_{Ri} = (\gamma/\bar{U})^{\eta}(u_i r_{Ri}^{\beta-1} \Phi_{Ri}^{1/\theta})^{\eta}$, welfare can be written as

$$\bar{U} \propto L_{Ri}^{-\frac{1}{\eta}} u_i r_{Ri}^{\beta-1} \Phi_{Ri}^{1/\theta}.$$

Population mobility ensures this holds for each location. Considering a change in commute costs from d to d' and letting $\hat{x} = x'/x$ denote the relative change in variable x across equilibria, market clearing for residential floorspace implies $\hat{r}_{Ri} = \hat{\Phi}_{Ri}^{1/\theta}\hat{L}_{Ri}$. Substituting this into the expression for welfare yields

$$\hat{\bar{U}} = \hat{L}_{Ri}^{-\left(\frac{1}{\eta}+1-\beta\right)}\hat{\Phi}_{Ri}^{\frac{\beta}{\theta}}$$

Thus, the change in RCMA, the change in residential population, and the elasticities β, η, θ are sufficient statistics for the welfare gains from the change in commuting infrastructure. Note that in a model with fixed residential populations $\hat{L}_{Ri} = 1$ so the change in RCMA and elasticities β, θ are sufficient statistics for welfare.[138]

H Proofs

Proof of Proposition 1

Existence (Open City)

Let $\mathcal{D}_{Fs} = \left\{i : \bar{A}_{is} > 0\right\}$ with $\mathcal{D}_F = \cup_s \mathcal{D}_{Fs}$, and let $\mathcal{D}_{Rg} = \{i : \exists a \text{ s.t. } \bar{u}_{iag} > 0\}$ with $\mathcal{D}_R = \cup_g \mathcal{D}_{Rg}$ be the sets of desirable locations for firms and workers respectively.[139] It is trivial to show that every $i \in \mathcal{D}_F$ will have positive employment (and therefore $\vartheta_i < 1$)

[138] In a class of trade models (e.g. Arkolakis, Costinot and Rodriguez-Clare 2012), the welfare gains from trade relative to a counterfactual equilibria under autarky can be written in terms of parameters and the share of expenditure on domestic goods in the observed equilibrium. This is particularly attractive since, if model parameters are known, this latter statistic is commonly observed in the data. Unfortunately, no similar reduction exists for my model with commuting. For example, letting C and A denote the values of variables with commuting and in autarky (i.e. where the population of each location is fixed and there is no commuting) then

$$\frac{\bar{U}^C}{\bar{U}^A} = \left(\frac{L_{Ri}^C}{L_{Ri}^A}\right)^{-\left(\frac{1}{\eta}+1-\beta\right)}\left(\frac{w_i^C}{w_i^A}\right)^{\beta}\lambda_i^{-\frac{\beta}{\theta}}$$

where λ_i is the share of residents in i working in i in the equilibrium with commuting. Thus, one needs to observe values of variables in the counterfactual autarkic equilibrium to know the welfare gains from commuting as a whole.

[139] I consider each location in the city to be "relevant" in the sense that $i \in \mathcal{D}_R \cup \mathcal{D}_F \ \forall i \in I$.

because of an Inada condition-type property of firms' demand for floorspace that ensures the market clearing price of commercial floorspace r_{Fi} will go to infinity as the share allocated goes to zero (see expression below). However, as discussed in the text the set of active locations for residents are those which are both desirable and affordable:

$$\mathcal{A}_{Rg} = \left\{(i,a) : \bar{u}_{iag} > 0,\ r_{Ri} < (T_g \Phi_{Riag}^{1/\theta} - p_a a)/\bar{h}\right\}.$$

Therefore, I define active locations for firms (i.e. locations with positive equilibrium employment) as $\mathcal{A}_{Fs} = \left\{i : \bar{A}_{is} > 0\right\}$ and let $\mathcal{A}_F = \cup_s \mathcal{A}_{Fs}$ be the set of locations with positive employment (i.e. where there exists one industry with positive productivity). This set is exogenous and determined solely by fundamental productivities. I define active locations for residents as above, and likewise let $\mathcal{A}_R = \cup_g \mathcal{A}_{Rg}$ be the set of locations with positive residence (i.e. where there exists one skill group with positive residence). This set is endogenous and determined both by fundamental amenities and residential floorspace prices.

The system of equations characterizing the model equilibrium $\{L_{Riag}, \tilde{L}_{Fig}, w_{jg}, r_{Ri}, r_{Fi}, \vartheta_i, \bar{U}_g\}$ is

$$L_{Riag} = \bar{L}_g \frac{\left(u_{iag}(\Phi_{Riag}^{1/\theta_g} - \bar{h} r_{Ri} - p_a a) r_{Ri}^{\beta-1}\right)^{\eta_g}}{\sum_{r,o \in \mathcal{A}_{Rg}} \left(u_{rog}(\Phi_{Rrog}^{1/\theta_g} - \bar{h} r_{Rr} - p_o o) r_{Rr}^{\beta-1}\right)^{\eta_g}} \quad \forall (i,a) \in \mathcal{A}_{Rg}, \quad 0 \text{ otherwise}$$

$$r_{Ri} = \frac{(1-\beta)\sum_{g;a:(i,a)\in\mathcal{A}_{Rg}} \left(\Phi_{Rig}^{1/\theta_g} - p_a a\right) L_{Riag}}{\vartheta_i H_i - \bar{h}\beta \sum_{g,a} L_{Riag}} \quad \forall i \in \mathcal{A}_R, \quad 0 \text{ otherwise}$$

$$\tilde{L}_{Fig} = T_g \sum_{r,o \in \mathcal{A}_{Rg}} \pi_{j|rog}^{\frac{\theta_g - 1}{\theta_g}} \frac{1}{d_{rjo}} L_{Rrog} \quad \forall i \in \mathcal{A}_F$$

$$\tilde{L}_{Fig} = w_{ig}^{-\sigma_L} P^{\sigma-1} E \sum_s B_{isg} A_{is}^{\sigma-1} W_{is}^{\sigma_L - (1+\alpha_s(\sigma-1))} r_{Fi}^{-(1-\alpha_s)(\sigma-1)} \quad \forall i \in \mathcal{A}_F, \quad 0 \text{ otherwise}$$

$$r_{Fi} = \frac{\sum_s A_{is}^{\sigma-1} (W_{is}^{\alpha_s} r_{Fi}^{1-\alpha_s})^{1-\sigma} P^{\sigma-1} E}{(1-\alpha)(1-\vartheta_i) H_i} \quad \forall i \in \mathcal{A}_F, \quad 0 \text{ otherwise}$$

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

$$
\vartheta_i = \begin{cases} 0 & \text{if } r_{Ri} < (1-\tau_i) r_{Fi} \\ \in (0,1) & \text{if } r_{Ri} = (1-\tau_i) r_{Fi} \\ 1 & \text{if } r_{Ri} > (1-\tau_i) r_{Fi} \end{cases}
$$

where $p_a = \tilde{p}_a P$ is the price of a car denominated in units of the final good, $B_{isg} \equiv \alpha_s \alpha_{sg}^{\sigma_L}$ is a constant and the auxiliary variables $\Phi_{Riag}, E, W_{is}, A_{is}, u_{iag}$ are defined in the text.

Equating supply and demand for labor, and repeating the floorspace market clearing conditions allows us to write the equilibrium (up to scale) as

$$
\begin{aligned}
w_{ig} &= F_{ig}^{w}(w, r_R, r_F) \\
&\equiv \begin{cases} \left[\frac{\sum_{r,o \in \mathcal{A}_{Rg}} \pi_{j|rog}^{\frac{\theta_g - 1}{\theta_g}} \frac{1}{d_{rj}} L_{Rrog}}{\sum_s B_{isg} A_{is}^{\sigma-1} \left(\sum_h \alpha_{sh}^{\sigma_L} w_{ih}^{1-\sigma_L} \right)^{\frac{\sigma_L - (1+\alpha_s(\sigma-1))}{1-\sigma_L}} r_{Fi}^{-(1-\alpha_s)(\sigma-1)}} \right]^{-1/\sigma_L} (P^{\sigma-1} E)^{\frac{1}{\sigma_L}} & \text{for } i \in \mathcal{A}_F \\ 0 & \text{otherwise} \end{cases} \\
r_{Fi} &= F_i^{r_F}(w, r_R, r_F) \\
&\equiv \begin{cases} \frac{\sum_s A_{is}^{\sigma-1} (W_{is}^{\alpha_s} r_{Fi}^{1-\alpha_s})^{1-\sigma} P^{\sigma-1} E}{(1-\alpha)(1-\vartheta_i) H_i} & \text{for } i \in \mathcal{A}_F \\ 0 & \text{otherwise} \end{cases} \\
r_{Ri} &= F_i^{r_R}(w, r_R, r_F) \\
&\equiv \begin{cases} \frac{(1-\beta) \sum_{g;a:(i,a) \in \mathcal{A}_{Rg}} \left(\Phi_{Rig}^{1/\theta_g} - p_a a \right) L_{Riag}}{\vartheta_i H_i - \bar{h}\beta \sum_{g,a:(i,a) \in \mathcal{A}_{Rg}} L_{Riag}} & \text{for } i \in \mathcal{A}_R \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Note that $r_{Ri} \geq 0 \ \forall i$ since the numerator is always weakly positive (i.e. if $p_a a > \Phi_{Rig}^{1/\theta}$ then $L_{Riag} = 0$) and we know that while $L_{Ri} > 0$ we also have that $\vartheta_i H_i > \bar{h}\beta L_{Ri}$ since otherwise $r_{Ri} \to \infty$ and $L_{Ri} \to 0$, a contradiction. Define

$$x = \begin{bmatrix} \begin{bmatrix} w_{ig} \end{bmatrix}_{i=1\ldots,N,g=1,\ldots,G} \\ \begin{bmatrix} r_{Ri} \end{bmatrix}_{i=1\ldots,N} \\ \begin{bmatrix} r_{Fi} \end{bmatrix}_{i=1\ldots,N} \end{bmatrix} \in \mathbb{R}^{N(G+2)}, \quad F(x) = \begin{bmatrix} \left[F_{ig}^{w}(x)\right]_{i=1,\ldots,N,g=1,\ldots,G} \\ \left[F_{i}^{r_F}(x)\right]_{i=1,\ldots,N} \\ \left[F_{i}^{r_R}(x)\right]_{i=1,\ldots,N} \end{bmatrix} \in \mathbb{R}^{N(G+2)}$$

Then an equilibrium is a solution to the system $x = F(x)$. I proceed by establishing that the function F satisfies the conditions of Brouwer's fixed point theorem.

To start, I establish that F is homogenous of degree one in (w, r_R, r_F). To begin, by inspection note that $L_{Riag}, \tilde{L}_{Fig}, \pi_{j|iag}, A_{is}, u_{iag}, \vartheta_i$ are all homogenous of degree zero in (w, r_R, r_F). Moreover, $\Phi_{Riag}^{1/\theta}$ and P are homogenous of degree one, which implies that so is $p_a = \tilde{p}_a P$ and E. This provides homogeneity properties of each of the terms in the definition of F. Some straightforward algebra then implies that F^w, F^{r_F}, F^{r_R} are all homogenous of degree one and thus so is F.

I therefore consider the rescaled problem

$$x_i = \tilde{F}_i(x) = \frac{F_i(x)}{\sum_k F_k(x)}$$

Since F is homogenous of degree one, $\tilde{F}(x)$ is homogenous of degree zero and thus we can restrict our search to the unit simplex, a compact convex subset of $\mathbb{R}^{N(G+2)}$. It remains to show that F is continuous. First, notice that as locations enter and exit the set of active locations $\mathcal{A}_{Rg}$ we have that $L_{Riag} = 0$. Thus, the numerator in F^w is continuous even as changes in x induce changes in set $\mathcal{A}_{Rg}$ (and, for the same reason E is also continuous at such points). By continuity of all other terms, F^w is continuous. F^{r_F} is continuous by inspection. Finally, notice that $\lim_{L_{Ri}\to 0} r_{Ri} = 0$ so that F^{r_R} remains continuous even as resident groups enter and exit the location along the extensive margin. By continuity of all other terms, F^{r_R} is continuous. So F is a continuous function mapping a compact convex set to itself (since $\tilde{F}_i(x) \in [0,1]$) and thus by Brouwer's fixed point theorem there exists a fixed point $x = \tilde{F}(x)$.

Finally, since $\tilde{F}(x)$ is homogenous of degree zero we know that at any solution $x^* = F(x^*)$ we also have $x^* = F(\lambda x^*)$ for any $\lambda > 0$. Therefore, pick $\lambda = \frac{1}{\sum_k F_k(x^*)}$ and then λx^* is a solution to the original system $x = F(x)$. The levels of population of each group are solved from the condition that the welfare of each worker group equals its reservation utility $\bar{U}_g$.

Existence (Closed City)

In the open city model, despite the fact that the set of active locations is an equilibrium outcome, there will always be a positive population of each worker skill group in an equilibrium (i.e. the sets of active locations $\mathcal{A}_{RL}, \mathcal{A}_{RH}$ will both be non-empty). Suppose there was positive population of only one group, then CES production by firms would mean that their wage goes to zero at every location meaning their population everywhere must be zero, a contradiction. Instead, if floorspace prices grow so large that the population of the low-skill group tends to zero, CES demand for workers implies their market clearing wage will tend to infinity ensuring they can afford to live in at least one location.[140]

In the closed city model, population does not adjust to ensure an interior equilibrium. Instead, either all the city's residents can afford to live there or none of them can since the definition of L_{Rig} implies that $\sum_{(i,a)\in\mathcal{A}_{Rg}} L_{Riag} = \bar{L}_g$. In this case, land owner optimality ensures positive population of both groups. We need to check that at an equilibrium

$$\sum_{(i,a)\in\mathcal{A}_{Rg}} \vartheta_i H_i \lambda^H_{iag} > \bar{h}\bar{L}_g$$

where λ^H_{iag} is the share of i's residential housing consumed by type (g,a). Consider the

[140] I ignore the trivial degenerate equilibrium of the city shutting down completely, i.e. when $\bar{L}_L = \bar{L}_H = 0$. Starting at this point, wages, residential and commercial floorspace prices are zero and markets would be in equilibrium. Note that the potential for the city to shut down does not impact the existence proof above. This requires us to show, for example, that given a vector (w, r_F) the function F is continuous in r_R which was established. Clearly (conditional on (w, r_F)) for r_R high enough the city would shut down, but at such a value the equilibrium values of (w, r_F) would respond too.

problem of a landowner i at an equilibrium i who would shut down the mass of residents with his choice of ϑ_i. That is, there is some $\vartheta_i^* \in (0,1)$ such that the inequality above holds with equality. If she sets $\vartheta_i \leq \vartheta_i^*$ then $L_{Rkag} = 0\ \forall k$ so the population of type-g residents shuts down, which implies $\tilde{L}_{Fkg} = 0\ \forall k$ and thus wages and commercial rents will be zero, none of the other group will live in the city either and rents to the landowner will be zero. If she sets $\vartheta_i > \vartheta_i^*$, then there is a positive population of both groups, positive wages and commercial rents at active commercial locations, and positive rents for her since location i is desirable and thus attracts a positive residential population. Therefore, at an equilibrium it must be that all residents can fit in the city. The same existence proof as above applies since the function F is unaffected by the closed or open city assumption.

Uniqueness

Consider the special case of the model with one group of workers, firms and commute modes and no non-homotheticities ($\bar{h} = p_a = 0$), with a fixed allocation between residential and commercial floorspace. The equilibrium is given by

$$
\begin{aligned}
L_{Ri} &= \lambda_L \left(\bar{u}_i \Phi_{Ri}^{1/\theta} r_{Ri}^{\beta-1} \right)^{\frac{\eta}{1-\eta\mu_U}} \\
r_{Ri} &= (1-\beta) \frac{\Phi_{Ri}^{1/\theta} L_{Ri}}{H_{Ri}} \\
\tilde{L}_{Fi} &= w_i^{\theta-1} \sum_s \frac{L_{Rr}}{\Phi_{Rr}} d_{ri}^{-\theta} \Phi_{Rr}^{\frac{1}{\theta}} \\
r_{Fi} &= \left(\frac{A_i^{\sigma-1} w_i^{-\alpha(\sigma-1)} P^{\sigma-1} E}{(1-\alpha) H_{Fi}} \right)^{\frac{1}{1+(\sigma-1)(1-\alpha)}} \\
\tilde{L}_{Fi} &= \frac{1}{\alpha} w_i^{\alpha(1-\sigma)-1} A_i^{\sigma-1} r_{Fi}^{(1-\sigma)(1-\alpha)} P^{\sigma-1} E \\
\Phi_{Ri} &= \sum_j (w_j/d_{ij})^{\theta}
\end{aligned}
$$

where $\lambda_L = \bar{L}(\gamma_\eta/\bar{U})^{\frac{\eta}{1-\eta\mu_U}}$.

Substituting the commercial floorspace condition into the expression for labor demand, we find that

$$\tilde{L}_{Fi} = \lambda_F B_i w_i^{-\frac{\sigma}{1+(\sigma-1)(1-\alpha-\mu_A)}}$$

where $\lambda_F = (P^{\sigma-1}E)^{\frac{1}{1+(\sigma-1)(1-\alpha-\mu_A)}}$ is an equilibrium constant and B_i is defined as $B_i \equiv \left(\bar{A}_i^{\frac{1}{1-\alpha}}(1-\alpha)H_{Fi}\alpha^{-\frac{1+(\sigma-1)(1-\alpha)}{(\sigma-1)(1-\alpha)}}\right)^{\frac{(\sigma-1)(1-\alpha)}{1+(\sigma-1)(1-\alpha-\mu_A)}}$.

Substituting the residential floorspace condition into the expressions for resident supply, we find that

$$L_{Ri} = \tilde{\lambda}_L C_i \Phi_{Ri}^{\frac{\eta\beta}{\theta(1+\eta(1-\beta-\mu_U))}}$$

where $\tilde{\lambda}_L = \lambda_L^{\frac{1-\eta\mu_U}{1+\eta(1-\beta-\mu_U)}}$ and $C_i = \left(\frac{\bar{u}_i^{1/(1-\beta)}H_{Ri}}{1-\beta}\right)^{\frac{\eta(1-\beta)}{1+\eta(1-\beta-\mu_U)}}$.

These two equations provide allocations of residence and employment consistent with floorspace market clearing as a function of wages. It remains to determine wages from labor market clearing. Equating labor supply and demand and substituting in these results yields

$$x_i = \lambda \sum_j D_{ij} \left(\sum_s d_{js}^{-\theta} x_s^{-\frac{\theta}{\tilde{\sigma}}}\right)^{\tilde{\theta}} \equiv T_i(x)$$

where $\tilde{\sigma} \equiv \left(\frac{\sigma}{1+(\sigma-1)(1-\alpha-\mu_A)} + \theta - 1\right)$, $\tilde{\theta} \equiv \frac{1}{\theta}\left[\frac{\eta\beta}{1+\eta(1-\beta-\mu_U)} - (\theta-1)\right]$, $\lambda = \frac{\tilde{\lambda}_L}{\lambda_F}$, $D_{jr} = \frac{d_{ri}^{-\theta}C_r}{B_i}$ and $x_j = w_j^{-\tilde{\sigma}}$. Now, since the operator on the right-hand since is homogenous and continuous, we can show there exists a solution using the same logic as in the existence proof above.

To show uniqueness, I use the method of Allen et. al (2015) result by writing the system as

$$x_i = \lambda \sum_j D_{ij} \Phi_{Rj}^{\tilde{\theta}}$$

$$\Phi_{Ri} = \sum_j d_{ij}^{-\theta} x_j^{-\frac{\theta}{\tilde{\sigma}}}$$

This satisfies the system of equations (1) in their paper with coefficient matrices $\Gamma =$

I and $B = \begin{pmatrix} 0 & \tilde{\theta} \\ -\frac{\theta}{\tilde{\sigma}} & 0 \end{pmatrix}$, where I is the identity matrix. Then, when $\tilde{\sigma} > 0$ and $\tilde{\theta} < 0$ we have that

$$|B\Gamma^{-1}| = \begin{pmatrix} 0 & -\tilde{\theta} \\ \frac{\theta}{\tilde{\sigma}} & 0 \end{pmatrix}$$

where $|\cdot|$ denotes the element-wise absolute value. The system $x = T(x)$ has a unique solution when the spectral matrix of $|B\Gamma^{-1}|$ is less than one. This is equivalent to finding its largest eigenvalue, itself equivalent to finding conditions under which there exists an $x > 0$ such that $|B\Gamma^{-1}|x \leq x$. Solving this inequality yields

$$\beta(\sigma - 1)\mu_A + \sigma\mu_U \leq \frac{\sigma}{\eta} + \sigma(1-\beta) + \beta(1 + (\sigma - 1)(1-\alpha)).$$

Finally, notice that without spillovers while $\tilde{\sigma} > 0$, for $\tilde{\theta} < 0$ we require that

$$\theta > \frac{1+\eta}{1+\eta(1-\beta)}$$

and since $\beta \in [0,1]$, the RHS varies between 1 and $1+\eta$ so that a sufficient condition is that $\theta > 1+\eta$. With spillovers, a sufficient condition for uniqueness is

$$\begin{aligned} \mu_A &\leq 1 - \alpha + \frac{\sigma + \theta - 1}{(\sigma - 1)(\theta - 1)} \\ \mu_U &\leq \frac{1 + \eta(1-\beta)}{\eta} - \frac{\beta}{\theta - 1} \\ \beta(\sigma - 1)\mu_A + \sigma\mu_U &\leq \frac{\sigma}{\eta} + \sigma(1-\beta) + \beta(1 + (\sigma - 1)(1-\alpha)) \end{aligned}$$

This completes the proof.[141] ■

[141] What's the content of the restriction $\theta > \frac{1+\eta}{1+\eta(1-\beta)}$? Consider the labor supply equation

$$\tilde{L}_{Fj} = \sum_i \frac{L_{Ri}}{\Phi_{Ri}} d_{ij}^{-\theta} \Phi_{Ri}^{\frac{1}{\theta}}$$

Proof of Proposition 2

I first provide a formal statement of proposition 2. Section G.3 provides the mapping from particular models to the general framework.

Proposition. *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*

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

(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 $\tilde{\Phi}_{Fj} = \sum_i \frac{L_{Ri}}{\Phi_{Ri}} \kappa_{ij} \Phi_{Ri}^{\epsilon}$ then*

1. *An equilibrium always exists and is unique whenever $|\epsilon(\beta - 1) - \gamma| \leq |\beta - 1||\alpha - 1|$*

2. *The economy has a reduced form representation where residence and employment can*

$$= \quad w_j^{\theta-1} \sum_i \lambda_U C_i \lambda_U^{-\frac{\eta(1-\beta)}{1+\eta(1-\beta-\mu_U)}} \left[\frac{(1-\beta)C_i}{H_{Ri}}\right]^{-\frac{\eta(1-\beta)}{1+\eta(1-\beta-\mu_U)}} d_{ij}^{-\theta} \Phi_{Ri}^{\frac{1}{\theta}\left[\frac{\eta\beta}{1+\eta(1-\beta-\mu_U)} - (\theta-1)\right]}$$

When Φ_{Ri} increases in a location, two things happen. First the relative wage from j has fallen so that share of people in i commuting to j falls. Second, more people move into i so the total number of residents increases. That $\tilde{\theta} < 0$ is a restriction that the first effect dominates the second, which resembles a condition that firms face upward sloping labor supply.

be written as

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

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

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

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

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

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

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

where $K_{ij}^{\delta} \equiv \kappa_{ij} A_j^{1-\alpha}$. 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

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

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

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

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

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

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.

Finally, note that existence of a solution up-to-scale can be shown be collapsing the 3 equations into a single system in δ_j

$$\delta_j = \left[\sum_i K_{ij}^{\delta} \left[\sum_s K_{is}^{\gamma} \delta_s \right]^{\frac{1}{\beta-1}} \Phi_{Ri}^{\epsilon - \frac{\gamma}{\beta-1}} \right]^{\frac{1}{\alpha-\delta}} = F_j(\delta)$$

where Φ_{Ri} is a function of $\{\delta_j\}$ as defined above.Since F is homogenous, the function $\tilde{F}_i(\gamma) = \frac{F_i(x)}{\sum_r F_r(x)}$ is homogenous of degree zero. We can therefore restrict the search to the unit simplex, and so $\tilde{F}$ maps a compact convex set to itself and by Brouwer's fixed point theorem there exists a solution. We then scale the solution by the same method as in the proof of proposition 1.

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 $\alpha, \beta, \gamma, \delta$ and ϵ.[142] 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 G.3, the framework can accommodate additional factors so long as equilibrium in factor markets collapses into log-linear demand for labor and residents

[142]Most models impose additional restrictions between these 5 parameters, which reduces the number of parameters one needs to know (see the above for examples).

Proof of Proposition 3

Part 1: Wages

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

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

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

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

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

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

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

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

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

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

Next, break down the probability as

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

So

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

Next, note that

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

Putting these results together, we have that

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

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

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

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

- **Monotonicity**. Transform the system into log-space. From Euler's theorem since F_1

is homogenous of degree zero we know for any vector $d \ln w$ we have that

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

Thus there is no change in response to a vector of proportionate changes. 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}.

Amenities are retrieved from the resident supply condition

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

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.[143] ■

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