A Model of Spatial Job Referral Networks

Carlos A. Mesa-Guerra
Cornell University

July 12, 2023

Abstract

The aim of this paper is to provide new insights on the contribution of referrals to the overall employment of workers, their commuting, residence, and employment locations by linking the urban and social space. Using the well-known fact that workers rely on their social contacts to find jobs, I develop a model in which workers face a tradeoff between searching for employment outside their neighborhood of residence driven by the strength of their social network and the commuting cost. The model incorporates three key facts from the data: dual search methods (direct vs. indirect), network congestion, and spatial mismatch. The model is extended to capture first-order features of cities and to simulate the effect of shutting down job referrals. Results suggest that networks may suffer from congestion effects, reducing workers employability, particularly for immigrants, but allow workers to find better matches. Overall, in the absence of job referral networks, unemployment, wages, welfare, and output fall. However, in the case of immigrants, removing search through the network increases welfare. The results suggest that referrals allow workers to find jobs closer to where they live, avoiding the wage cost of commuting. Shutting down referrals increases the frequency of commuting, especially for ethnic groups.

Keywords: social networks; job search; urban structure; immigration.

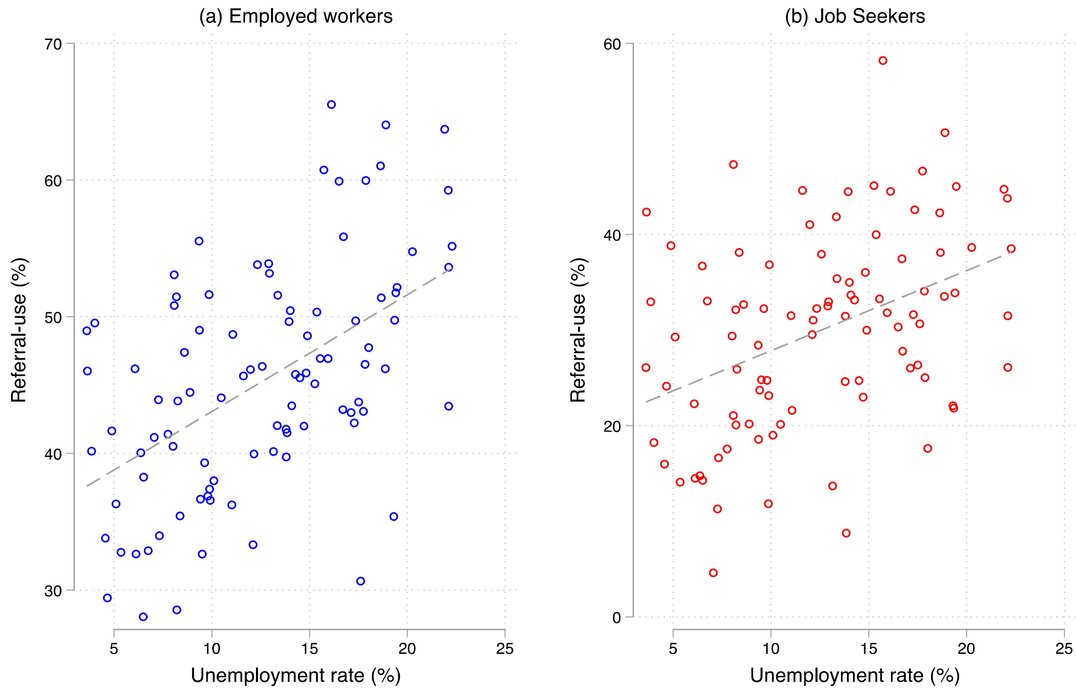
1. Introduction

Both workers and firms report extensive use of referral networks in the job search, with over half of jobs acquired using social contacts. However, locations in which workers report using personal contacts to find jobs more intensively are also those with higher unemployment rates (see Figure 1). This fact seems at odds with standard theories of how referrals work in the labor market which have usually abstracted from the spatial structure of job information networks. For instance, the leading descriptions of the mechanism through which referrals operate in the labor market, either through a reduction of search frictions by having workers share information about job opportunities or as a screening device for the firm by transmitting information about the quality of workers, generally predict positive effects of referrals in terms of increasing the probability of being hired.

The aim of this paper is twofold. First, I model the interdependency between the urban and the social space as part of the job search process, attempting to build bridges between the search and matching literature and the urban spatial equilibrium literature. The model incorporates three key ingredients from observed facts in the literature:[1] *(i)* workers can find jobs through two distinct channels: by directly learning about the job opening or through information received from a social contact (indirect channel); *(ii)* search through the network is characterized by a non-monotonic relationship between the job matching rate and the network size, making clear the tradeoff between information diffusion and congestion forces that is seen in larger networks; *(iii)* distance to jobs has a negative impact on workers' search outcomes, thus, linking the urban structure (i.e., the housing market) and the labor market. Second, I develop a quantitative implementation of the stylized model and provide new insights on the contribution of referrals to the overall employment of workers, their commuting, residence, and employment locations.

[1] These facts are presented in an accompanying paper (Mesa-Guerra, 2023).

Figure 1: Referrals and Unemployment in Space



Notes: The Figure shows the spatial relationship between the use of referrals to find jobs and unemployment across neighborhoods (Zonal Planning Units–UPZ) in Bogotá. Referral-use measures the share of employed (or unemployed) workers who report having found (or are searching for) their current job through friends, family, or acquaintances. *Source:* 2021 EMB.

There have been some attempts linking the urban and social space to understand the effect of social interactions in workers' labor market outcomes. The first to my knowledge to model the link between labor and housing markets is Selod & Zenou (2006). Although there is no explicit description of the mechanisms behind the social network, they define the rate at which workers gather information about jobs as function of local social connections and embed this in a linear city model to explain urban segregation based on racial preferences. More recently, Zenou (2015) provides an explanation for the mismatch between workers residential location and labor outcomes using a monocentric city model—in the spirit of the Alonso-Muth-Mills model.

Following these early contributions, in the first part of the paper, I embed social interactions in a tractable general equilibrium framework of an urban labor market. The basic structure of the model introduces an urban structure to a Markov process of job search. In the model, both the social and geographical spaces are key determinants of workers' labor market outcomes. The main takeaway from the stylized model is that commuting may be efficient for some workers because of search frictions and spatial mismatch (i.e., increasing distance from suitable employment opportunities). Workers face a tradeoff between searching outside their neighborhood of residence given by the strength of their social network and the commuting cost. Thus, workers will only search outside their neighborhood if their job finding probability compensate the commuting cost.

The second part of the paper presents a quantitative version of the stylized model. The model is extended to capture first-order features of the data while retaining tractability. First, it allows for numerous discrete locations that differ in residential amenities, resulting in more complex spatial interactions. Second, workers choose optimal housing consumption and have idiosyncratic preferences for residential locations which determines the distribution of the labor force. Third, it allows workers to search for jobs anywhere around the city. As in standard urban models, congestion forces are driven by an inelastic supply of land (or housing space) and commuting costs that are increasing in travel time.

Using the tractability of model, I derive closed-form expressions for several key moments that are calibrated to exactly replicate the observed data. Among these are the fraction of workers living in each location and the fraction of currently employed workers commuting to other locations within the city. To the best of my knowledge, some of these expressions are new to the literature. In particular, the probability of commuting between residential and workplace locations matches the ratio of job finding probabilities for a given residential location. The quantitative model is then used to estimate the extent to which referrals drive workers' employment and urban mobility patterns. I simulate the effect of shutting down job referrals and increasing search efficiency through formal channels. This attempts to isolate the effects of geographic proximity from information diffusion that is valuable for forming good matches in the labor market. Results are presented for the total population and for a sample of immigrant workers.

I find that referrals networks may suffer from congestion effects driven by spatial mismatch, reducing workers employability, particularly for immigrants, but allow workers to find better matches. For instance, when the network is absent, the unemployment rate drops 0.1 percentage points for the complete sample and 0.9 percentage points when only looking at immigrants. These effects are consistent with workers having to expand their job search range within the city and having their outside option drop, which captures the value of unemployment. Unpacking the aggregate effects, neighborhoods where referrals are used more intensively gain the most when workers are not allowed to search for jobs using their social contacts.

In addition, simulation results indicate that around 0.2% of the observed wage of all employed and 2.6% of wages of immigrant workers can be attributed to the use of referrals, holding constant search efficiency through other channels. These results are lower to recent findings in the literature do not seem to be explained by a lack of job offers. Also, shutting down referrals increases the frequency of commuting, especially for immigrants. Simulating an increase of a standard deviation in the measure of direct access to job information increases commuting substantially for both groups.

Overall, referrals account for 1.5% of workers' welfare, 1.4% of the expected utility of unemployment, and 3.3% of output in the city. However, in the case of immigrants, once we remove search through the network, welfare increases by 5.5% with both the expected utility of search and output decreasing by 1.2% and 0.6%, respectively.

Related literature. This paper contributes to the large literature studying the effect of job referrals on labor market outcomes.[2] One central challenge that has proven difficult is distinguishing endogenous social interactions from sorting into neighborhoods or social groups.[3] Agents may choose who to be friends with, or where

[2] Comprehensive reviews of the literature can be found in Ioannides & Loury (2004), Beaman (2016), and Topa (2019).
[3] The influential work of Manski (1993) discusses the problem of separately identifying correlated and contextual effects (sharing the same sources of information or individual characteristics) from endogenous social effects in empirical analysis.

to live, based on their expectation of the social interaction effect that we are trying to estimate. Since both workers and referrals are not randomly assigned, estimates on the effect of referrals are likely to suffer from bias resulting from individual sorting on unobservable attributes. To address this concern some studies have relied on exogenous variation in the use of referrals from experimental settings (Beaman & Magruder, 2012; Pallais & Sands, 2016), random (or quasi-random) assignment of workers to neighborhoods (e.g., Damm, 2014), or have used novel empirical strategies that focus on very local interactions, e.g., at the block-level (Bayer et al., 2008; Hellerstein et al., 2011; Schmutte, 2015). My contribution is to estimate the effect of referrals through the lens of a model. By modelling the interdependency between urban and social space, I will account for the sorting of immigrants within neighborhoods.

The paper also complements the theoretical literature that has modeled referrals as an information exchange mechanism in which potential employees inform each other about job opportunities, reducing search frictions (Holzer, 1988; Topa, 2001; Calvó-Armengol, 2004; Calvó-Armengol & Zenou, 2005).[4] In particular, referral networks may play a role in increasing the arrival rate of job offers, improving the type of offers received, or the rate at which employers and jobseekers are matched. I contribute to this literature by focusing on the geographical or spatial dimension of networks and understanding how they are influenced by aggregate market conditions.[5] Now, this paper is closely related to recent work by Lester et al. (2021) who studies the effect of referrals on labor market outcomes through the lens of a model, combining observed information on workers' job search methods. Their model adopts a flexible approach to modelling referrals that lets the data dictate the relationship between job search methods and rate at which a worker meets a firm. In

[4] An alternative way of how information can be exchanged within networks is by using referrals as a screening device for the firm—*i.e.*, transmit information about the quality of workers, as in Montgomery (1991), Simon & Warner (1992), Galenianos (2013), and Dustmann *et al.* (2016).

[5] Calvó-Armengol (2004) and Galeotti & Merlino (2014) are two examples in the literature that analyze the influence of labor market conditions in the use and effectiveness of social networks in the job search process.

contrast to his paper, I derive a specific micro-founded expression for the matching function following Calvó-Armengol & Zenou (2005) and embed this in an urban structure. This is needed to account for complex spatial interactions that mediate through commuting patterns.

The paper also adds to the broader literature on the role played by social networks in supporting new labor market entrants in their new locations, particularly as it concerns to immigrants or rural-urban migration. For a broad overview of this literature, I refer the reader to the surveys by Munshi (2014, 2020). While large empirical evidence supports the fact that social networks affect the spatial mobility of workers and their employment prospects, little is known about their contribution to aggregate outcomes. Two recent studies have structurally estimated the role of networks in determining migration decisions (Munshi & Rosenzweig, 2016) and the growth of businesses (Dai et al., 2019). The contribution of the paper to this literature is to estimate the aggregate contribution of networks to labor market outcomes and urban patterns, and understand the differences observed in the data in the use of referrals across groups.

Lastly, this paper contributes to the rapidly growing literature in urban economics that closely maps models of internal city structure to the data. These models are part of a larger literature on quantitative trade and spatial models (Eaton & Kortum, 2002; Redding & Rossi-Hansberg, 2017). My starting point is the seminal paper by Ahlfeldt et al. (2015). In contrast to the earlier theoretical contributions by Fujita & Ogawa (1982) and Lucas & Rossi-Hansberg (2002) using linear and symmetrical circular city models, respectively, Ahlfeldt et al. allow for numerous discrete locations that can differ in their productivity, amenities, transportation infrastructure, and supply of residential and commercial land.

This paper builds on the recent work introducing frictional labor markets into a quantitative spatial framework (Kline and Moretti, 2013; Fournier, 2021; Bilal, 2023). Perhaps most closely related is the work by Fournier (2021), which embeds a job search framework à la Diamond (1981) and Mortensen & Pissarides (1994) in an urban quantitative model to study the spatial mismatch hypothesis. However, the focus of that paper is more on the impact of spatial mismatch and place-based policies, while my focus is more on the distinct effects of referrals. This paper also departs from the literature by deriving an aggregate matching function that fails to exhibit constant returns to scale, consistent with a large body of empirical work.

The remainder of the paper is structured as follows. Section 2 describes a simple version of the model, using only two locations. Section 3 lays out the quantitative implementation of the model and describes de structural estimation. Section 4 computes counterfactuals. Section 5 concludes.

2. The Model

The simple model presented here introduces an urban structure to a Markov process of job search. While parsimonious, this framework intends to capture some important factors that explain a worker's residential and workplace location within the city such as the presence of referral networks, commuting costs, consumption and productive amenities, and individual preferences. This will allow us to understand the conditions under which workers search for employment (and work) across locations.

For ease of exposition, the model is presented in different building blocks. In Section 2.1, I start by assuming workers' residential location is fixed. In Section 2.2, I introduce job referrals. Finally, Section 2.3 allows workers to choose their residential location.

2.1. Fixed residential location and no referrals

The first part of the model bears similarity with Coulson et al. (2001). For the moment, the model abstracts from worker's preferences over tradable goods, optimal housing demand and supply, and consumption amenities.

2.1.1. Setup

Consider a city comprised of two neighborhoods, 1 and 2. The city is populated by a measure $\bar{N}$ of workers, each of whom inelastically supplies a unit of labor. Workers are either employed or unemployed and only search for jobs when unemployed. Assume workers residential neighborhood is predetermined and cannot move (e.g., workers may be randomly assigned to neighborhoods) but workers are free to search for jobs and work in either of the two neighborhoods. In addition to their initial residential assignment, workers differ according to their relative "ability" to find a job outside their neighborhood of residence, indexed by $s \in S \equiv [0,1]$. For instance, this could reflect differences in the number of direct contacts that provide information about employment opportunities. All agents are infinitely lived and discount the future at the common rate $\rho > 0$.

Before proceeding, let me introduce some notation. Let L_{jk} $\left(U_{jk}\right)$ denote the number of workers living in j who are employed (search for work) in neighborhood k. Let $N_{Rj} = N_{j1} + N_{j2}$ denote the residential population at neighborhood j and $N_{Ek} = N_{1k} + N_{2k}$ denote the labor supply at workplace k, where $N_{jk} = L_{jk} + U_{jk}$. The number of employed and unemployed workers living in j is $L_{Rj} = L_{j1} + L_{j2}$ and $U_{Rj} = U_{j1} + U_{j2}$, respectively. The number of employed and unemployed workers working or searching for a job in k is $L_{Ek} = L_{1k} + L_{2k}$ and $U_{Ek} = U_{1k} + U_{2k}$, respectively.

2.1.2. Timing

Time is discrete and continues forever. At the beginning of each period, there is a number of employed and unemployed workers and firms post V_k vacancies. I refer to $v_k = V_k / N_{Ek}$ as the job arrival or vacancy rate and assume $V_k < N_{Ek}$. In each period some workers find jobs, and some do not. The individual hiring probability of an unemployed worker with search ability s is denoted by $\lambda_{jk}(s, U_{Ek}, V_k) \equiv \lambda_{jk}(s)$. If an unemployed worker finds a job, he starts working at the beginning of the following period.[6] At the end of the period, jobs dissolve with probability δ, which determines the employment level at the start of the next period. Job separation is taken only to depend on the general state of the economy and hence is treated as exogenous to each location.

I will often omit the time subscript and focus on the steady state of this dynamic model.

2.1.3. Workers

The indirect utility of worker i living in neighborhood j and commuting to neighborhood k (either to work or search for a job) is:

$$\mathrm{v}_{ijk} = w_{jk} - d_{jk}, \tag{1}$$

where w_{jk} is the wage rate and d_{jk} is the cost of commuting to work or search for a job between residence j and workplace k. Let $w_{\emptyset} = b > 0$ denote the income equivalent of the utility flow that a worker obtains when unemployed. This is assumed to be derived from home production and is common across locations. Commuting costs are assumed to be symmetric and non-commuters do not incur in commuting costs such that $d_{12} = d_{21}$ and $d_{11} = d_{22} = 0$.

2.1.4. Firms

Consider a firm as a unit of production that can be filled by a worker. For our purposes it does not make a difference if one interprets each firm as a single "job" which hires one worker or if each firm can hire as many workers as it likes. Firms may enter freely into each location but to find a worker a firm needs to post a vacancy which involves a fixed cost g. Coordination failures or partial information about jobs

[6] This has the practical implication of separating search outcomes in the current period with the payoff of being employed.

results in a "frictional" labor market, i.e., firms and workers do not meet instantaneously.

2.1.5. Equilibrium

DEFINITION 1. *A spatial equilibrium consists of a wage (w_{jk}) and vacancies posted (V_k) in each location, the number of workers commuting to each workplace (N_{jk}), the number of unemployed workers living in a given neighborhood who search for work in each workplace (U_{jk}), and a critical search value ($\bar{s}$), such that the labor market and commuting equilibrium is solved for.*

The labor market equilibrium requires solving for: *(i)* the wage and labor market tightness in each location and *(ii)* the steady-state labor market equilibrium. I first start by solving the labor market equilibrium and then proceed to solve for the commuting equilibrium in order to derive the distribution of workers in each neighborhood.

2.1.6. Wage determination and labor market tightness

Denote respectively by J_{jk}^F and J_k^V the intertemporal profit of a filled job and an open vacancy at the beginning of period t, and before vacancies are posted. Note that J_{jk}^F is the value of a vacancy posted in location k and filled by a worker from location j. Define the expected or average probability of finding a job for an unemployed worker searching in workplace k as $\bar{\lambda}_k \equiv \sum_j \int_s \lambda_{jk}(s)\, ds$. Then $U_{Ek}\bar{\lambda}_k$ determines the number of matches per unit of time in workplace k, and we can define $f(\theta_k^{-1}, s) = \frac{U_{Ek}}{V_k}\bar{\lambda}_k$ as the probability for a firm of filling a vacancy in workplace k.[7] Note that U_{Ek}/V_k is the inverse of the traditional definition for labor market tightness (θ_k).

[7] Note that since a vacancy is filled by exactly one worker, in equilibrium, the number of vacancies filled must be equal to the number of unemployed workers who find a job.

The Bellman equations are given by:

$$J^F_{jk,t} = y_k - w_{jk} + \frac{1}{1+\rho}\left[(1-\delta)J^F_{jk,t+1} + \delta J^V_{k,t+1}\right] \tag{2a}$$

$$J^V_{k,t} = -g + \frac{f_t(\theta_k^{-1}, s)}{1+\rho} J^F_{jk,t+1} + \frac{1 - f_t(\theta_k^{-1}, s)}{1+\rho} J^V_{k,t+1}. \tag{2b}$$

A job already filled at the beginning of period t and before vacancies are posted has a payoff $(y_k - w_{jk})$ but the worker-firm match can be kept with probability $(1-\delta)$ at the end of the period or can be destroyed with probability δ, which implies Eq. (2a). On the other hand, a firm opening a vacancy at the beginning of period t incurs in cost g which may be filled with probability $f(\theta_k^{-1}, s)$ or remain unfilled with complementary probability, which implies Eq. (2b).

At steady state, $J^F_{jk,t} = J^F_{jk,t+1} = J^F_{jk}$, $J^V_{k,t} = J^V_{k,t+1} = J^V_k$. Following the literature, I assume that firms post vacancies up to a point where they yield zero profit, $J^V_k = 0$ (free-entry condition). From Eqs. (2a), (2b), and the free-entry condition, we can get an expression for the *labor demand curve* at workplace location k (in the U, V plane):

$$\frac{y_k - \mathbb{E}_{|j}\left[w_{jk}\right]}{\rho + \delta} = \frac{g}{f(\theta_k^{-1}, s)} \qquad k = 1, 2. \tag{3}$$

Note that in contrast to the standard expression in the search and matching literature (Pissarides, 2000), in this expression the job-filling rate depends on the relative "ability" of workers to find a job outside their neighborhood.

I now turn to the dynamic problem of workers. Denote by J^E_{jk} and J^U_{jk} the intertemporal utilities of an employed and unemployed worker, respectively. The Bellman equations are given by:

$$J_{jk,t}^{E} = w_{jk} - d_{jk} + \frac{1}{1+\rho}\left[(1-\delta)J_{jk,t+1}^{E} + \delta J_{jk,t+1}^{U}\right] \tag{4a}$$

$$J_{jk,t}^{U} = b - d_{jk} + \frac{1}{1+\rho}\left[\lambda_{jk,t}(s)\, J_{jk,t+1}^{E} + \left(1-\lambda_{jk,t}(s)\right)J_{jk,t+1}^{U}\right]. \tag{4b}$$

The first two terms on the right-hand-side of Eq. (4a) and Eq. (4b) represents workers' flow value when employed or unemployed, respectively. The last term in Eq. (4a) reflects the expected lifetime value if an employed worker were to lose his job with probability δ. Similarly, in Eq. (4b), an unemployed worker living in neighborhood j and searching for work may find a job in neighborhood k with probability $\lambda_{jk}(s)$. If he finds a job, he obtains a positive income stream equal to the difference between the expected value of becoming employed in neighborhood k and his current expected lifetime utility.

At steady state, $J_{jk,t}^{E} = J_{jk,t+1}^{E} = J_{jk}^{E}$, $J_{jk,t}^{U} = J_{jk,t+1}^{U} = J_{jk}^{U}$, and $\lambda_{jk,t} = \lambda_{jk}$. The worker surplus is then given by:

$$J_{jk}^{E} - J_{jk}^{U} = \frac{1+\rho}{\rho+\delta+\lambda_{jk}(s)}\left(w_{jk} - b\right) \qquad k = 1,2. \tag{5}$$

The worker accepts any wage such that $J_{jk}^{E} \geq J_{jk}^{U}$ (i.e., when $w_{jk} \geq b$). The firm accepts any wage such that $J_{jk}^{F} \geq J_{k}^{V}$.

ASSUMPTION 1. Firms and workers bargain over the wage using a generalized Nash-bargaining process, where workers have bargaining power $\beta \in [0,1]$:

$$w_{jk} = \operatorname*{Arg\,max}_{w_{jk}} \left(J_{jk}^{E} - J_{jk}^{U}\right)^{\beta} \left(J_{jk}^{F} - J_{k}^{V}\right)^{1-\beta} \qquad k = 1,2. \tag{6}$$

Using Eqs. (2a) and (4a), the first-order condition satisfies:

$$(1-\beta)\left(J_{jk}^{E}-J_{jk}^{U}\right)=\beta\left(J_{jk}^{F}-J_{k}^{V}\right) \qquad k=1,2. \tag{7}$$

Using the free-entry condition, Eqs. (2a), (2b) and (5), and observing the definition for $f(\theta_k^{-1}, s)$, we have the wage-setting condition:

$$w_{jk}=\beta\left(y_k+g\frac{\lambda_{jk}(s)}{\bar{\lambda}_k}\frac{V_k}{U_{Ek}}\right)+(1-\beta)b. \tag{8}$$

In search-matching models, Eq. (8) is taken to be the *labor supply curve* which is upward-sloping. Since in equilibrium V_k and U_{Ek} depend on the search ability of workers, as does λ, the wage itself depends on s.

2.1.7. Steady-state labor market equilibrium

In steady state, flows in and out employment must be equal such that:

$$\bar{\lambda}_{jk}U_{jk}=\delta\left(N_{jk}-U_{jk}\right), \tag{9}$$

where $\bar{\lambda}_{jk}=\int_s \lambda_{jk}(s)\,ds$. This can be rewritten as:

$$U_{jk}=\frac{\delta}{\delta+\bar{\lambda}_{jk}}N_{jk}, \tag{10}$$

and job seekers in each workplace location U_{Ek} (residential location U_{Rj}) is given by summing Eq. (10) across residential locations j (workplace locations k). A steady-state labor market equilibrium $\left(U_{Ek}^{*}(s), V_k^{*}(s), w_{jk}^{*}(s)\right)$ is given by the intersection of the labor demand and supply curves (Eqs. (3) and (8)) and satisfying Eq. (10).

2.1.8. Commuting equilibrium

ASSUMPTION 2. The individual job finding probability is $\lambda_{jk}(s)$, where $\lambda_{jj}(s) = (1-s)\Lambda(U_{Ek}, V_k)$ and $\lambda_{jk}(s) = s\Lambda(U_{Ek}, V_k)$ for $s \in S$, $j = 1, 2$, and $j \neq k$, where $\Lambda(U_{Ek}, V_k) \geq 0$, $\Lambda'(U_{Ek}) > 0$, $\Lambda'(V_k) > 0$, $\lim_{U_{Ek}\to\infty} \Lambda(U_{Ek}, V_k) = V_k$, and $\lim_{V_k\to\infty} \Lambda(U_{Ek}, V_k) = U_{Ek}$.

Without loss of generality and for ease of exposition, set $b = 0$. From Eq. (4a) and Eq. (4b), the net present value accruing to a worker living in j who searches for work in location k is:

$$J_{jk}^{U} = \frac{1+\rho}{\rho}\left[\frac{\lambda_{jk}(s)}{\rho+\delta+\lambda_{jk}(s)} w_{jk} - d_{jk}\right]. \tag{11}$$

It follows from Eq. (11) that J_{jk}^{U} is increasing in s for $j \neq k$. This indicates that if commuting flows occurs in equilibrium, then it is done by those with the highest relative search "ability" (i.e., those with the greatest value of s) as they have to bear the cost of commuting. The following proposition follows from Eq. (11):

PROPOSITION 1. *Given U_{Ek} and V_k there is a unique critical value, $\bar{s} \geq 0$, such that every worker with search ability $s \in [0, \bar{s}]$ optimally searches for a job at $j = k$ (within his neighborhood), whereas every worker with search ability $s \in (\bar{s}, 1]$ searches for a job at $j \neq k$ (outside his neighborhood).*

The critical values $\bar{s}$ solves:

$$\begin{aligned} J_{j1}^{U} &= \left[\frac{\lambda_{j1}(\bar{s})}{\rho+\delta+\lambda_{j1}(\bar{s})}\right] w_{j1} - d_{j1} \\ &= \left[\frac{\lambda_{j2}(\bar{s})}{\rho+\delta+\lambda_{j2}(\bar{s})}\right] w_{j2} - d_{j2} = J_{j2}^{U}, \qquad \forall j = 1, 2. \end{aligned} \tag{12}$$

This is easily seen from Figure 2, which plots the value functions for unemployed workers living in neighborhood 1, J_{11}^U and J_{12}^U, against search ability $s \in S$. Since $\lambda_{11}(s)$ is a decreasing function of s, J_{11}^U is monotonically decreasing in s. On the contrary, J_{12}^U is increasing in s. A unique interior equilibrium solution for $\bar{s}^*$ requires that the wage net of commuting costs is large enough. Particularly, it requires:

$$\frac{w_{jk}}{d_{jk}} > 1 + \frac{(\rho + \delta)}{s\Lambda_k} \quad \text{if} \quad j \neq k. \tag{13}$$

Note that when Eq. (13) holds with equality, then we have a corner solution, and all workers search only in their own neighborhood. Assuming search ability s is distributed across workers according to a uniform distribution, the probability density function is $f(s) = 1$ for all $s \in S$ and the fraction of workers living in j who commute to work or search for a job to k is:

$$\frac{N_{jk}}{N_{Rj}} = \begin{cases} \int_0^{\bar{s}} f(s)ds = \bar{s} & \text{if} \quad j = k \\ \int_{\bar{s}}^1 f(s)ds = 1 - \bar{s} & \text{if} \quad j \neq k \end{cases}, \tag{14}$$

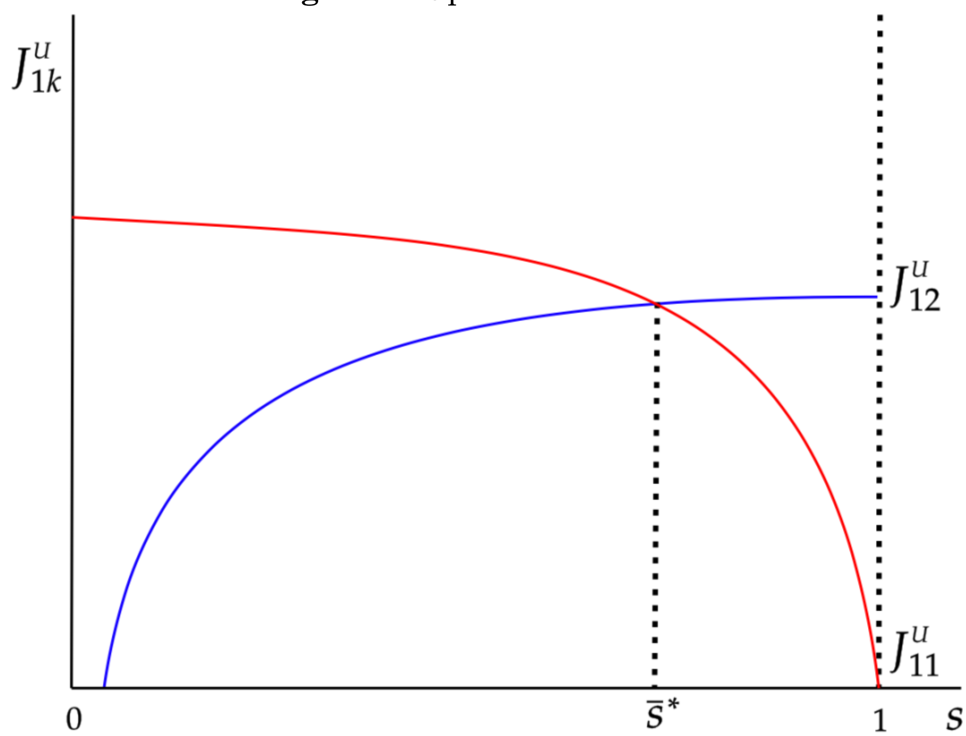
and the local labor supply at each neighborhood k is:

$$N_{Ek} = N_{1k} + N_{2k}, \quad k = 1, 2. \tag{15}$$

2.1.9. Summary

A city is comprised of two neighborhoods. We have an initial allocation of workers in the two neighborhoods and firms free entering each neighborhood. All the workers do is just choose where to search for a job, which depends on the ability to find a job outside their community. The wage is going to be determined by Nash bargaining. The number of vacancies is going to be determined by free entry. And the workers searching is just going to be determined by the initial allocation of workers across the neighborhoods, which is exogenous. Then a fraction of them will decide to search outside their neighborhood, which depends on the distribution of search ability. In equilibrium, there's going to be four wages, two numbers of firms in each community which determines the number of vacancies, a distribution of workers searching in each location, and a critical value for search ability.

Figure 2: Optimal Job Search



Why don't we get an equilibrium where the number of firms entering just clears the local labor market for residents in each neighborhood if the cost of entering either one is the same? In other words, why don't just firms enter in the two neighborhoods such that no worker has to look for a job in the other neighborhood? Because firms have different productivities in both locations and workers face commuting costs. Thus, there is a mismatch, and some workers commute to the more productive neighborhood or where more jobs (vacancies) are available. Therefore, commuting may be efficient for some workers. At the end, workers face a tradeoff between the probability of finding a job and the commuting cost. So, workers will move if their job finding probability compensate the commuting cost. If not, then it's not efficient for workers to search outside of their residential location.

2.2. Job search with social interactions

I now extend the basic framework to allow workers to use personal contacts when searching for jobs. This follows Calvó-Armengol & Zenou (2005).

The labor market is modelled as one in which workers share job information with social contacts. Unemployed individuals may find jobs through indirect (i.e. referrals) or direct methods. The benefit of a worker of having social contacts is that it increases the probability of finding a job through an increase in information flows about employment opportunities.

NETWORK STRUCTURE. Workers only interact with other individuals from the same residential neighborhood and the network is assumed to be symmetric (i.e., the relationship between two connected members is reciprocal).

NETWORK SIZE. In each period, a worker randomly meets a group of s workers among all his neighbors, with $s \leq N_{Rj}$.

Note that I have changed the definition of s from Section 2.1.

2.2.1. Job search

At the beginning of each period, an exogenous number of vacancies V_k are posted in each workplace location. Firms only advertise vacancies using help-wanted signs on their windows, which bears no cost, so job information is not equally available to everyone. Assuming an unemployed worker is not rereferred, he can only know of an open vacancy if he walks past the firm.

INFORMATION DIFFUSION. The information transmission protocol is the following:

i. A worker (employed or unemployed) can directly learn about a vacancy at location k with probability v_k.
ii. If the worker is unemployed, he takes the job (formal method).
iii. If the worker is employed at k, he passes on the information to one of his unemployed network members, with uniform probability.
iv. If all members within the network are employed, then the job offer is lost.
v. Once the information arrives to an unemployed worker the vacancy is immediately filled.

Let u_j be the unemployment rate in neighborhood j and $\pi_{jk|j}$ the fraction of workers from neighborhood j that work in neighborhood k. Because workers meet randomly each period, on average, each worker living in neighborhood j meets su_j unemployed workers and $(1-u_j)s$ employed workers. The probability that an unemployed worker living in neighborhood j is referred by a neighbor to a job in workplace k is given by:

$$q_{jk}(s, u_j, \pi_{jk}, v_k) = 1 - \underbrace{\left[1 - v_k(1-u_j)\pi_{jk|j}\frac{1-(1-u_j)^s}{su_j}\right]^s}_{\substack{\text{Prob. that none of his } s \text{ contacts refers} \\ \text{unemployed worker } i}}, \tag{16}$$

where $v_k(1-u_j)\pi_{jk|j}$ is the probability that an employed worker knows of a job opening, $1-(1-u_j)^s$ is the probability that someone is unemployed, and $1/su_j$ is the probability that a particular worker gets the referral. It is trivial to see that q_{jk} is increasing and concave in s, with a global maximum at s_{max}, decreasing and convex in u_j, increasing and strictly concave in v_k and π_{jk}.[8] In an accompanying paper (Mesa-Guerra, 2023), I provide evidence that the relationship between network size and the job finding probability is non-monotonic.

Since unemployed workers find jobs using direct and indirect methods, the probability that an unemployed worker living in j finds a job in k is now given by:

$$\lambda_{jk}\left(s, u_j, \pi_{jk}, v_k\right) = v_k + (1 - v_k) \cdot q_{jk}\left(s, u_j, \pi_{jk}, v_k\right). \tag{17}$$

Because this probability is independent across individuals, the number of matches per unit of time (i.e., the matching function) in location k is given by $\sum_j \lambda_{jk} U_{jk}$. This contrast with the previous definition in Section 2.1.6.

2.2.2. *Equilibrium with referral networks*

Note that the properties of net present value for an unemployed worker given by Eq. (11) are immediately deduced from the new definition of the individual job finding probability given by Eq. (17). J_{jk}^U is increasing between 0 and $\bar{s}$ and decreasing between $\bar{s}$ and N_{Rj}.

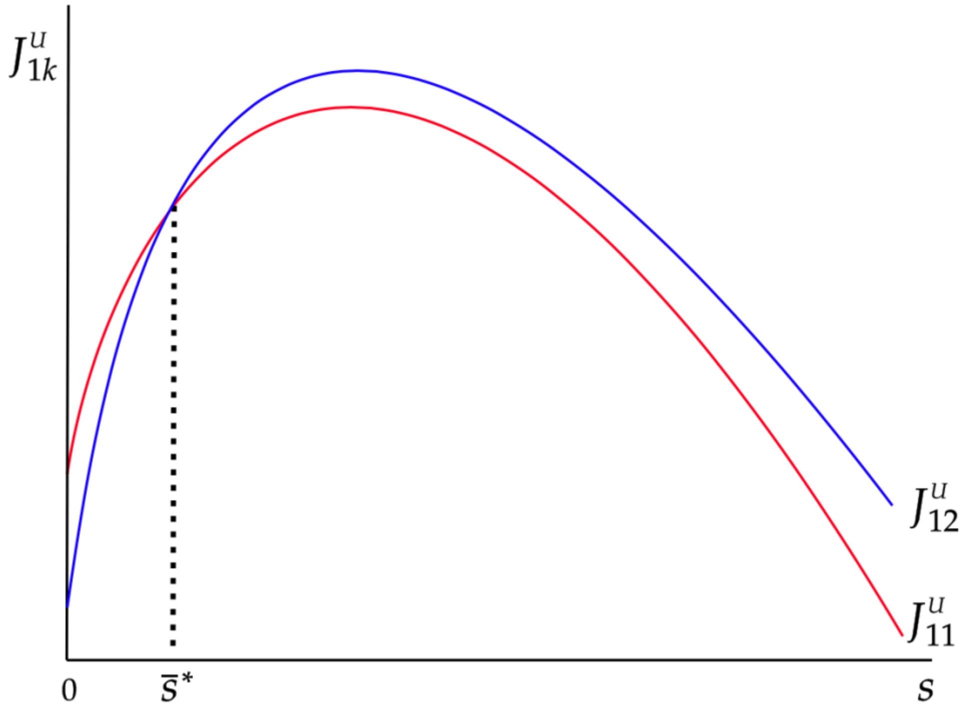
Assume the probability of finding a job through direct methods is the same in both neighborhoods (i.e., $v_k = v$).

PROPOSITION 2. *Fix u_j and π_{jk} and assume $y_1 \leq y_2$ holds,*[9] *then it is possible to show that for a set of values for $\{\rho, \delta, v, w_{jk}, d_{jk}\}$ there exists a unique interior critical value, $\bar{s}$, such that every worker with a network size $s \leq \bar{s}$ optimally searches for a job at $j = k$ (within his neighborhood), whereas every worker with network size $s > \bar{s}$ searches for a job at $j \neq k$ (outside his neighborhood). And we can show that $\bar{s} < s_{max}$.*

[8] The proofs that give rise to these properties can be found in Calvó-Armengol & Zenou (2005).
[9] In an accompanying paper (Mesa-Guerra, 2023), I provide evidence that there is a wage premium for commuting, i.e., the further workers commute, the higher their wages.

As illustrated in Figure 3, both J_{11}^U and J_{12}^U are strictly concave function of $s \in S$. A unique interior equilibrium solution for $\bar{s}^*$ requires that commuting costs are not too large or too small.[10]

Figure 3: Optimal Job Search with Referrals



2.3. Allowing workers to choose their residential location

Up to this point workers' neighborhood of residence was predetermined and permanent but workers were able to commute between locations. I now relax the mobility assumption and allow workers to also choose their residential location. I adopt a variant of the Rosen-Roback-type model as in Kline & Moretti (2014) to allow

[10] Particularly, it requires:

$$\left(\frac{v}{\rho+\delta+v}\right)(w_{12}-w_{11}) < d_{12}$$
$$< \frac{(\rho v+\delta v+v^2+\xi_1^2\pi_{11}\pi_{12})(w_{12}-w_{11})+\xi_1(\rho+\delta+v)(\pi_{12}w_{12}-\pi_{11}w_{11})+\xi_1 v(\pi_{11}w_{12}-\pi_{12}w_{11})}{(\rho+\delta+v+\xi_1\pi_{11})(\rho+\delta+v+\xi_1\pi_{12})},$$

where $\xi_1 = v - vu_1 - v^2 + v^2u_1$.

workers to have a "taste" for locations but distinguishing between place of residence and place of work (or job search).[11] The decisions of workers are modeled using discrete choice.

2.3.1. Workers

Workers first choose a neighborhood to live in and then where to search for jobs. Each worker inelastically supplies a unit of labor and rents a single unit of housing. The indirect utility of worker i living in neighborhood j and commuting to neighborhood k (either to work or search for a job) is now:

$$\mathrm{v}_{ijk} = \underbrace{w_{jk} - d_{jk} - r_j + A_j}_{\mathrm{v}_{jk}} + e_{ij}, \tag{18}$$

where r_j is the local rent level and A_j is the mean consumption value of local amenities. Without loss of generality and for ease of exposition, set $w_{\emptyset} = b = 0$. The term v_{jk} measures the average worker utility of each residence-workplace pair (common component across individuals) and e_{ij} represents individual-specific preferences for living in neighborhood j (idiosyncratic component). Thus, workers have idiosyncratic preferences for residential locations but not for workplace locations. Workers are imperfectly mobile within the city due to individual location preferences and commuting costs. For instance, personal networks can create high relocation costs.

The idiosyncratic component of utility (e_{ij}) is assumed to be i.i.d. according to a type I extreme value distribution with scale parameter μ and mean zero. From the properties of the extreme value distribution, the difference in idiosyncratic preferences for neighborhood 1 and 2 is distributed across workers according to a logistic distribution:

[11] Kline & Moretti (2014) assume that locations are distinct labor markets, such that workers cannot live in one area and work in another.

$$(e_{i1} - e_{i2})/\mu \sim \text{logistic}(0, 1), \tag{19}$$

where μ characterizes the importance of idiosyncratic preferences for locations. A higher value of μ implies a lower response to differences in the common component of utility. Workers will require large differences in amenities, rents, or in the job finding probability in order to move. To simplify the analysis, assume only unemployed workers are able to relocate, meaning that an employed worker loses his job if he decides to move.[12]

2.3.2. Housing Market

All housing units are owned by absentee landlords. Workers and landowners are assumed to be separate agents.[13] Assume that the number of housing units in neighborhood j is equal to the number of workers choosing to live in the neighborhood N_{Rj}. I allow for upward sloping housing supply curves in each neighborhood as in Kline & Moretti (2014). Assume the supply of housing is given by:

$$r_j = z_j N_{Rj}^{\eta_j}, \tag{20}$$

where η_j controls the elasticity of housing supply which is assumed to be exogenously determined by geography and zoning regulations (Glaeser & Ward, 2009; Saiz, 2010).

2.3.3. Equilibrium

[12] A similar approach is used by Fournier (2021). This assumption is also consistent with the fact that a worker arriving to the city (e.g., an immigrant) usually first chooses where to live and then chooses where to work.

[13] Immigrant workers, especially those who arrived recently, usually do not own their residence.

The new market equilibrium requires solving: *(i)* the residential location and commuting flows that determine the rent outcome (*urban location equilibrium*) and *(ii)* a search and matching equilibrium that determines the flows in and out of employment (*labor market equilibrium*).

The analysis proceeds through a process similar to backward induction. I start by taking the optimal search results presented in Sections 2.1 and 2.2. Given where workers search for jobs, I determine the fraction of workers living in each neighborhood. I then derive the two key steady-state relationships: the equilibrium rents and the steady-state level of unemployment in each location.

2.3.4. Urban location equilibrium

Given where workers optimally search for jobs (k^*), an unemployed worker chooses where to live to maximize his lifetime utility, taking as given local amenities, rents, wages, commuting costs, and the location decisions of other workers. An unemployed worker then maximizes (follows from Eq. (11)):

$$\max_{j\in\{1,2\}} \frac{\rho}{1+\rho} J^{U}_{ijk^*}. \tag{21}$$

A worker chooses to live in neighborhood 1 if and only if $e_{i1} - e_{i2} > \tilde{v}_{2k^*} - \tilde{v}_{1k^*}$, where $\tilde{v}_{jk^*} = \varphi(\lambda_{jk^*})w_{jk^*} + A_j - r_j - d_{jk^*}$ and $\varphi(\lambda_{jk^*}) \equiv \lambda_{jk^*}/(\rho + \delta + \lambda_{jk^*})$. Given Eq. (19), the probability that an unemployed worker lives in neighborhood 1 is:

$$P(e_{i2} - e_{i1} \leq \tilde{v}_{1k^*} - \tilde{v}_{2k^*}) = \Phi\left(\frac{\tilde{v}_{1k^*} - \tilde{v}_{2k^*}}{\mu}\right), \tag{22}$$

where $\Phi(\cdot) \equiv \exp(\cdot)/[1 + \exp(\cdot)]$. Note that this expression is equivalent to the fraction of unemployed workers who choose to locate in neighborhood 1, $\mathbf{u}_{R1} \equiv U_{R1}/U$. Other things equal, the fraction of unemployed workers in neighborhood 1 is increasing in in the difference between the expected payoff if the worker finds a job in k^* net of rents and commuting costs $(\varphi(\lambda_{1k^*})w_{1k^*} - r_1 - d_{1k^*}) - (\varphi(\lambda_{2k^*})w_{2k^*} - r_2 - d_{2k^*})$ and in the difference in the mean value of amenities $A_1 - A_2$.

In equilibrium, the marginal unemployed worker must be indifferent between neighborhoods; thus, his relative preference for neighborhood 2 over neighborhood 1 $(e_{i2} - e_{i1})$ must equal the difference in net present values $(\tilde{v}_{1k^*} - \tilde{v}_{2k^*})$:

$$\begin{aligned}\mu\Phi^{-1}(\mathbf{u}_{R1}) = \varphi(\lambda_{1k^*})w_{1k^*} - \varphi(\lambda_{2k^*})w_{2k^*} - (r_1 - r_2) - (d_{1k^*} - d_{2k^*}) \\ + (A_1 - A_2),\end{aligned} \tag{23}$$

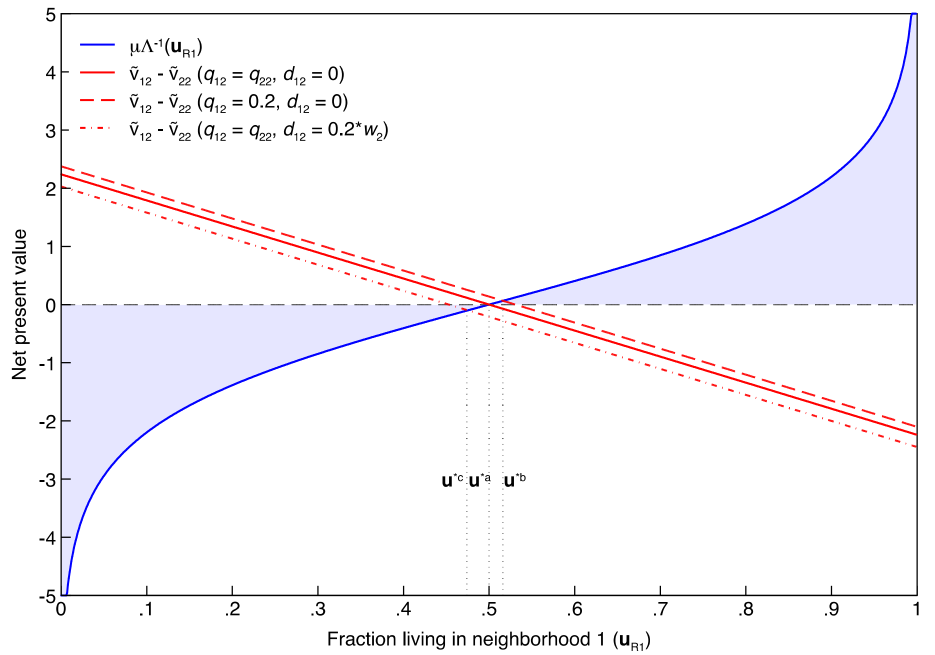
where $\Phi^{-1}(\cdot)$ is the logit function. Given μ, positive changes in net expected payoffs or amenities in neighborhood 1 creates migration from neighborhood 2 to neighborhood 1.

To see the role played by job-referral networks and location fundamentals in determining the population distribution across neighborhood, we can use Eqs. (8), (17) and (19) to rewrite the equilibrium condition:

$$\begin{aligned}\mu\Phi^{-1}(\mathbf{u}_{R1}) = (\varphi_{1k^*} - \varphi_{2k^*})\beta y_{k^*} + (\varphi_{1k^*}\lambda_{1k^*} - \varphi_{2k^*}\lambda_{2k^*})(g_k\theta_{k^*}/(\lambda_{1k^*} + \lambda_{2k^*})) \\ - [z_1(L_{R1} + \mathbf{u}_{R1}U)^{\eta_1} - z_2(\bar{N} - L_{R1} - \mathbf{u}_{R1}U)^{\eta_2}] - (d_{1k^*} - d_{2k^*}) \\ + (A_1 - A_2).\end{aligned} \tag{24}$$

Figure 4 illustrates the case in which the two neighborhoods are initially identical in terms of amenities, housing supply, and the job finding rate (with respect to workplace 2) so that half the workers live in neighborhood 1 and the other half in neighborhood 2. Considering that employed workers cannot move (L_{R1} is fixed at the beginning of each period), one can think of the upward sloping (blue line) as defining the supply curve to neighborhood 1 while the downward sloping (red line) as the relative demand to live in neighborhood 1 versus neighborhood 2. At point $\mathbf{u}^{*a}$ the marginal worker is indifferent between both neighborhoods. The shaded area measures all other workers' expected payoff (except for the marginal worker) associated with choosing the neighborhood they strictly prefer.

Figure 4: Population Distribution in Equilibrium



This figure was constructed by setting $k^* = 2$, $\mu = 1$, $\rho = 0.04$, $\delta = 0.02$, $\beta = 0.1$, $g = 1$, $z_1 = z_2 = 1$, $\eta_1 = \eta_2 = 0.5$, $\upsilon_2 = 0.02$, $q_{22} = 0.1$, $\theta_2 = 1$, $y_2 = 10$, $A_1 = A_2 = 0$, $\bar{N} = 1{,}000$, $U = 100$, $L_{R1} = (\bar{N} - U)/2$.

Figure 4 can be used to assess graphically the effect of differences in referral-networks and commuting costs on the distribution of workers between neighborhood 1 and 2. For instance, a larger difference (e.g., 10 percentage points) in the probability of being referred to a job in neighborhood 2 for workers living in neighborhood 1 versus those living in neighborhood 2 increases the fraction of workers who choose to live in neighborhood 1 ($\mathbf{u}^{*b}$). However, the fraction of workers choosing to live in neighborhood 1 is decreasing in the value of the commuting cost ($\mathbf{u}^{*c}$); in Fig. 3 the cost of commuting from neighborhood 1 to neighborhood 2 is modeled as being 20 percent of the nominal wage.

The equilibrium rent is given by re-arranging Eq. (23):

$$r_1^* = \varphi(\lambda_{1k^*})w_{1k^*} - \varphi(\lambda_{2k^*})w_{2k^*} + r_2 - (d_{1k^*} - d_{2k^*}) + (A_1 - A_2) - \mu\Phi^{-1}(\mathbf{u}_{R1}^*). \quad (25)$$

Equilibrium in the housing market is obtained by equating Eqs. (20) and (25).

2.3.5. Labor market equilibrium

In equilibrium, employment in each neighborhood has to be equal to the total labor inputs used by firms. Because workers from different neighborhoods are assumed to be perfect (and homogeneous) substitutes in production:

$$L_{Ek} = \sum_j L_{jk} \,. \quad (26)$$

The commuting clearing condition requires equating the measure of workers employed in neighborhood k with the measure of workers living in j choosing to commute to neighborhood k:

$$L_{Ek} = \sum_j \pi_{jk|j} L_{Rj} \,. \quad (27)$$

Now, the conditional commuting probabilities $\left(\pi_{jk|j}\right)$ follows from Proposition 2. Finally, in steady state, the local unemployment levels are determined by flows in and out of employment as there is no relocation of workers within the city:

$$\lambda_{jk} U_{jk} = \delta L_{jk} \quad (28)$$

Therefore, in steady state, the probability that a worker living in neighborhood j is employed at workplace k $\left(l_{jk}\right)$ must also be consistent with the ratio of employment to local labor supply for each (j,k) pair:

$$l_{jk} = \frac{L_{jk}}{N_{jk}} = \frac{\lambda_{jk}(v_k, q_{jk})}{\delta + \lambda_{jk}(v_k, q_{jk})}. \tag{29}$$

The literature has extensively shown that the decentralized market equilibrium is not socially efficient because of the presence of both search and network externalities. Either there is too much unemployment, creating congestion for unemployed workers to find jobs, or there is too little unemployment, creating congestion for firms to fill vacancies.

3. Quantitative Implementation

In this section, I show how to extend the stylized model laid out in Section 2 to match observed data across locations (e.g., population, employment, rents, and wages). This framework allows for more realistic features of cities and labor markets. It incorporates an urban structure consisting of many locations which results in more complex spatial interactions. This more general structure extends the standard urban quantitative models by including labor market frictions with social interactions.[14]

Each location is unique in terms of amenities, match productivities, which are extended to vary also by residential location, and geography, and has a supply of firms (jobs). With spatial frictions, workers care more about some locations, especially those locations that are closer. Forward-looking agents predict the implications of their decisions in the future, not only in their location but everywhere in the city. Workers choose optimal housing consumption and have idiosyncratic preferences for living in different locations.

Another difference with the stylize model is that unemployed workers can search for jobs in any location and do not incur in commuting cost. As workers can search for jobs anywhere around the city, introducing commuting costs would add an extra layer of complexity without adding further insights. However, since workers

[14] See Ahlfeldt et al. (2015) for a standard presentation of an urban spatial equilibrium model that captures first-order features of the data.

are endowed with a unit of time, the model introduces a measure of direct information access about vacant jobs across locations. I allow the match output of a filled vacancy to be specific to a residential-workplace pair. Finally, I move away from the discrete-time model and introduce its continuous-time counterpart in order to simplify some of the derivations and be consistent with the literature; the solutions are virtually identical.

In Section 3.1, I give an overview of the model. All derivations are shown in Section A.1 in the Online Appendix. In Section 3.2, I describe the structural estimation procedure.

3.1. Model

Consider a closed city populated by a measure $\bar{N}$ of workers, composed of both employed L and unemployed U workers. The city is composed of multiple distinct locations. Residential locations are indexed by j, while k indexes work locations. Locations differ in terms of their residential and productive amenities, supply of land, social networks, and commuting times between any two locations within the city.

Jobs are filled probabilistically via a matching function. With probability $\upsilon_{jk} = \Theta_{jk}\upsilon_k$ an unemployed worker living in j hears about a vacant job in k directly, where $\upsilon_k = V_k\,/(L_{Ek} + U)$ is the job arrival or vacancy rate in workplace k and Θ_{jk} is a measure of direct access to job information about available jobs.[15] However, workers may also hear about vacancies through indirect methods (referrals). As described in Section 2, the matching function is characterized by a non-monotonic relationship between the job finding rate, λ_{jk}, and the network size, s.[16] In very large networks, on average, unemployed workers hear more about available jobs through their social network but, at the same time, it is more likely that multiple job opportunities reach

[15] The measure of direct access to job information can be interpreted as information loss (for values lower than 1) and may be a function of distance between locations or search efficiency. I calibrate this directly using the structure of the model.
[16] Although the matching functions fails to be homogeneous of degree one it is increasing and strictly concave in both the number of unemployed workers and job vacancies.

the same unemployed worker, slowing down information diffusion and the matching process. Workers are assumed to only search when unemployed.

Whether searching or employed, workers maximize utility over consumption of a single final good, chosen as the numeraire, and consumption of housing, for which they pay rent, r_j. In contrast to the stylized model presented in Section 2, each location is endowed with a fixed supply of housing space owned by absentee landlords, H_j, and workers spend a fraction $(1-\alpha)$ of income on housing. In addition, workers have idiosyncratic preferences shocks over residential locations which are drawn from an independent Gumbel distribution. As before, workers are imperfectly mobile. Unemployed workers are free to locate anywhere but, after finding a job, workers cannot relocate to reduce commuting costs.[17] After choosing a residential location, workers enjoy the mean value of amenities A_j, which is assumed to enter multiplicatively in the value function.

When employed, workers receive a wage, w_{jk}, and pay a commuting cost which is modeled as a reduction in effective units of labor, $d_{jk} = \exp(\kappa\tau_{jk})$. When unemployed, workers receive unemployment benefit b and do not incur in commuting cost. In steady state, workers do not move, and the unemployment rate $u_j = U_j/N_{Rj}$ in each location will be determined by equating flow rates in and out of employment.

Firms may post vacancies which entail a fixed cost g. When a vacancy is filled, it creates a match output y_{jk}. Wages are set via Nash bargaining over the match surplus and worker's share of the match surplus is β. Matches may be split at an exogenous rate δ. Free entry of firms drives the value of an unfilled vacancy in each workplace location to zero.

Finally, workers and firms are risk neutral, infinitely lived, and discount future payoffs in continuous time at rate ρ.

[17] This assumption implies that employed workers do not sort into neighborhoods or networks based on information received by the worker at the work location. In other words, workers do not choose to live close to their new colleagues after finding a job but may end up working together with others from the same residential location.

DEFINITION 2. *Given the model's parameters* $\{\alpha, \kappa, \rho, \delta, \beta, s_j\}$, *unemployment benefit b, exogenous location-specific characteristics* $\{\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{y}, \mathcal{g}, \boldsymbol{\tau}\}$, *and total city population* $\overline{N}$, *the steady-state general equilibrium of the model is characterized by the vector* $\{\boldsymbol{r}, \boldsymbol{V}, \boldsymbol{w}, \boldsymbol{\lambda}, \boldsymbol{N_R}, \boldsymbol{\pi_R}, \boldsymbol{\pi^E}\}$.

The following seven relationships characterize the equilibrium vector:

(i) Housing market clearing:

$$r_j = (1-\alpha)\frac{N_{Rj}}{H_j}\left[(1-u_j)\left(\sum_k \pi^E_{jk|j}(w_{jk}/d_{jk})\right) + u_j b\right]. \tag{30}$$

(ii) Equilibrium entry condition:

$$\frac{\mathbb{E}_{|j}[y_{jk} - w_{jk}]}{\rho + \delta} = \mathcal{g}\frac{V_k}{\sum_j U_j \lambda_{jk}}. \tag{31}$$

(iii) Wage-setting condition:

$$w_{jk} = \frac{\beta y_{jk} + (1-\beta)\dfrac{(r_j)^{\alpha-1}}{(\rho+\delta+\sum_k \lambda_{jk})}\left((\rho+\delta)b + \sum_k \dfrac{\lambda_{jk} w_{jk}}{d_{jk}}\right)}{\beta + (1-\beta)(r_j)^{\alpha-1} d_{jk}^{-1}}. \tag{32}$$

(iv) Job-finding probability:

$$\lambda_{jk} = v_{jk} + (1-v_{jk})\left[1 - \left(1 - v_k(1-u_j)\pi^E_{jk|j}\frac{1-(1-u_j)^{s_j}}{s_j u_j}\right)^{s_j}\right]. \tag{33}$$

(v) Labor market clearing:

$$N_{Rj} = \frac{\delta + \sum_k \lambda_{jk}}{\delta} U_j. \tag{34}$$

(vi) Residential choice probability:

$$\pi_{Rj} = \frac{\exp\left[A_j\left(r_j\right)^{\alpha-1}\left(\rho + \delta + \bar{\lambda}_j\right)^{-1}\left((\rho+\delta)b + \sum_{k=1}^{\mathcal{L}} \lambda_{jk} w_{jk} d_{jk}^{-1}\right)\right]}{\sum_{\ell=1}^{\mathcal{L}} \exp\left[A_\ell(r_\ell)^{\alpha-1}\left(\rho + \delta + \bar{\lambda}_\ell\right)^{-1}\left((\rho+\delta)b + \sum_{k=1}^{\mathcal{L}} \lambda_{\ell k} w_{\ell k} d_{\ell k}^{-1}\right)\right]}. \tag{35}$$

(vii) Commuting probability:

$$\pi_{jk|j}^{E} = \frac{L_{jk}}{L_{Rj}} = \frac{\lambda_{jk}}{\sum_k \lambda_{jk}} \equiv \frac{\lambda_{jk}}{\bar{\lambda}_j}. \tag{36}$$

And populations add up to the city total, i.e., $\bar{N} = \sum_j N_{Rj}$.

As shown by Ahlfeldt et al. (2015), in the absence of endogenous agglomeration forces, commuting costs and the inelastic supply of housing act as congestion forces ensuring the existence of a unique equilibrium. Under strictly positive, finite, and exogenous location characteristics there is strictly positive and finite values of both wages and rents in equilibrium such that some workers commute to and some workers live in a particular location. Complete specialization (i.e. zero residents or commuters) would imply $A_j = 0$ or $y_k = 0$.

In this case of endogenous agglomeration forces, however, there is potential for multiple equilibria. A worker's location choice depends, among other things, on the job finding probability which in turn depend on the location and search decisions of all workers. Thus, agglomeration forces are captured by network externalities. The structure and size of the network of social contacts affects workers' residential choices. This reflects the fact that, for example, immigrants may prefer to live close to other immigrants in the hope of getting more information about available jobs. These externalities impose structure on how amenities (in a broad sense) in a given location are affected by the characteristics of other locations and create spillovers to neighboring locations through commuting. As described in Section 3.2, by conditioning on a set of parameters and a combination of observed endogenous variables and fundamentals, the equilibrium conditions of the model determine unique values of the unobserved location-specific characteristics.

3.2. Structural Estimation

This section describes how the equilibrium conditions presented in Section 3.1 are used to obtain unobserved amenities, housing supply, productivities, and entry costs that rationalize the observed data as an equilibrium of the model. The calibration and estimation of a subset of parameters without solving the full model is also discussed.

3.2.1. Data description

The quantitative implementation of the model requires data on employment and unemployment by place of residence, commuting flows and commuting times between locations, wages across locations, and rents. I use data from the 2021 household survey for Bogotá (Colombia), *Encuesta Multipropósito de Bogotá*—EMB. The EMB provides information for more than 100 thousand households, including detailed demographic and employment information for all working age individuals. The employment information includes direct information on referrals and job search methods, the residential location of all workers, and the place of work (up to the census block). Since the EMB contains geocoded information on the place of work for only 35% of the sample of employed workers,[18] I use the observed commuting probabilities between pairs of locations and total employed population in each location to input the remaining commuting flows.

[18] An additional 29% of the sample has imputed zeros as coordinates for the workplace location. In the case of immigrant workers, 40% of the sample has precise information on the workplace location with an additional 38% reporting zeros as coordinates.

Data for Bogotá is available at different levels of spatial disaggregation. The finest available disaggregation are blocks, followed by sectors which are similar to a U.S. census tract, planning zones (UPZs), and localities. Because the backbone of the analysis is the geographical information for each residential-workplace pair, which requires a large enough sample of workers across pairs, the primary geographic unit used in the analysis are UPZs. These planning zones consist of large neighborhoods that share common socio-economic characteristics and are often used to coordinate urban development policies. UPZs have an average size of 366,262 square meters and an average population of 69,684 in 2021.

I will present results using information on all working-age workers and immigrants, for which I will consider only information for Venezuelan-born immigrants. Between 2015 and 2021, Colombia received over 1.7 million Venezuelan-born immigrants, representing about 4% of the Colombian population. About 20% of the total stock of immigrants established in Bogotá.

Finally, additional data to calibrate some of the parameters of the model comes from the Colombian labor force survey (*Gran Encuesta Integrada de Hogares*, GEIH). I use data from 2014 to 2021.

3.2.2. Parametrization and estimation of structural parameters

I calibrate the model at a monthly frequency. Parameters $\{\alpha, \kappa, \rho, \delta, \beta, s_j\}$ and unemployment benefit b are either taken directly from existing values from the literature or estimated directly from the data.

The discount rate ρ is set to 0.004 to match an annual discount factor equal to 0.95 or equivalently a 5% annual interest rate. I follow the methodology proposed by Shimer (2005a) to estimate the separation rate, δ, using the GEIH data. The monthly average value over the sample period is 0.033 (or 3.3%).

I estimate the share of housing expenditure in workers' income, $1-\alpha$, using data for Bogotá from the GEIH.[19] Estimates show that housing consumption represented 30% of total household income on average over the sample period. This value is close to those found in the literature. Based on this, I set $\alpha = 0.7$ to match the long-run expenditure share. Common values for the bargaining power, β, either calibrated or estimated, generally range between 0.05 and 0.15 in the literature (e.g., Hagedorn & Manovskii, 2008; Card et al., 2013, 2018; Jäger et al., 2020). I present estimates using a value of 0.1.

Using trip-level data for Bogotá from the 2015 Mobility Survey, Tsivanidis (2019) estimate the rate of spatial decay or disutility from commuting, κ, using a mode choice model. His estimate of $\kappa = 0.012$ is very similar to estimates in the literature, particularly the value of 0.01 reported in Ahlfeldt et. al. (2015).

The network size, s_j, is taken to maximize the probability of finding a job through the network using a standard optimization algorithm that searches over possible alternative values for the parameter vector. Finally, the flow value of unemployment, b, is set to zero.[20]

3.2.3. Model inversion

The model can be used to recover unobserved location characteristics, such as amenities or productivities. Regardless of whether the model has a single equilibrium or multiple equilibria, with values on a set of parameters and observed data, the equilibrium conditions of the model can be used to determine unique values of the unobserved location-specific characteristics.

[19] I use all reported income earned by all members of a household during the month of the survey. Housing expenditure is defined as monthly rent, so only households renting a unit are included.
[20] Colombia does not have universal unemployment insurance. Using match observed data on the average value of government transfers in the city does not affect the results.

PROPOSITION 3. *(i) Given data on total population* $\{\boldsymbol{N_R}\}$ *and unemployment* $\{\boldsymbol{U_R}\}$ *by residence, commuting flows* $\{\boldsymbol{L}\}$*, rents* $\{\boldsymbol{r}\}$*, wages*$\{\boldsymbol{w}\}$*, and travel times between locations* $\{\boldsymbol{\tau}\}$*, in addition to a known value for the separation rate* $\{\delta\}$*, there exist a unique vector of job finding probabilities* $\{\boldsymbol{\lambda}\}$ *that is consistent with the data being an equilibrium of the model.*

(ii) Given data $\{\boldsymbol{N_R}, \boldsymbol{U_R}, \boldsymbol{L}, \boldsymbol{r}, \boldsymbol{w}, \boldsymbol{\tau}\}$*, known values for the parameters* $\{\alpha, \kappa, \rho, \delta, \beta\}$*, and a vector of job finding probabilities* $\{\boldsymbol{\lambda}\}$*, there exist unique vectors of unobserved location-specific characteristics* $\{\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{y}, \boldsymbol{g}, \boldsymbol{\Theta}\}$ *that are consistent with the data being an equilibrium of the model.*

Since we have the same number of observed endogenous variables $\{N_{Rj}, U_j, L_{jk}, r_j, w_{jk}\}$ as unobserved location characteristics $\{A_j, H_j, y_k, g_k, \Theta_{jk}\}$, Proposition 3 implies that the model is exactly identified (i.e., there are zero degrees of freedom). Therefore, we cannot evaluate the performance of the model using the observed data because unobserved location characteristics are calibrated so as to guarantee that the model exactly rationalizes the data. Nevertheless, since the model rationalizes the observed data, it can be used to compute counterfactuals.

4. Counterfactual Analysis

In this section, I start by using the calibrated model to explore the contribution of networks on employment, wages, and the patterns of urban mobility. To do so, I start by simulating the implications of shutting down the network $(s_j \rightarrow 0)$ and evaluate the labor market outcomes and resulting residential location. I assume that shutting down the referral channel has no effect on the arrival rate of job offers through direct channels. I then simulate the effect of shutting down the network but increase the measure of information access between residence-workplace pairs by one standard deviation, $\Theta_{jk} + \sigma_\Theta$. This captures the fact that while referrals are no longer a channel to get information about job offers, workers might search more efficiently using the direct channel.

As discussed above, endogenous agglomeration forces create the potential for multiple equilibria. When solving for counterfactual equilibria, I follow the literature and assume the selection rule of solving for the closest counterfactual equilibrium to the observed equilibrium. This is done by using the observed equilibrium values as the initial guess to start the algorithm described in Section A.3 in the Online Appendix.

Table 1 reports the changes (relative to the observed equilibrium) in the aggregate unemployment rate, the average monthly wage (conditional on being employed), and the commuting frequency of workers. Several interesting facts emerge from the simulations. First, and perhaps most striking, referrals seem to constraint workers' ability to find jobs. Shutting down referrals slightly reduces the unemployment rate for the total population but decreases it by roughly 1 percentage point when only looking at immigrants. One possible explanation is that workers living in locations that rely more on referrals to find jobs and have higher unemployment (as seen in Figure 1) now have an incentive to expand their search range or relocate within the city, increasing the probability of finding a job. Another possible explanation is that because workers cannot find jobs using their personal contacts, workers outside option declines, increasing the acceptance rate of offers. Assuming that unemployed workers search more efficiently while the network is absent (counterfactual *(b)*) reduces even further the unemployment rate.

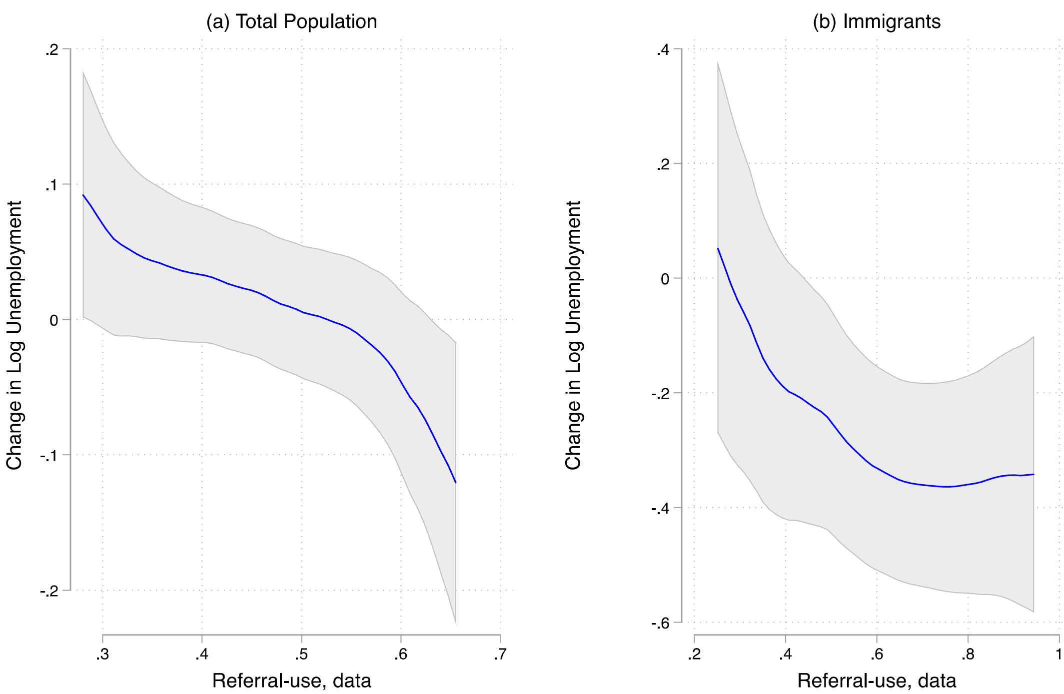
Figure 5 plots the changes in log unemployment in each neighborhood (UPZ) from shutting down the network against the fraction of workers who report in the data having used referrals to find their current job. Panel (a) shows that neighborhoods where referrals are used more intensively gain the most when workers are not allowed to search for jobs using their social contacts. Panel (b) shows larger and more persistent patterns for immigrants.

Table 1: Change in Unemployment, Wages and Commuting

Exercise	Outcome		
	Unemployment rate	Average monthly wage (COP)	Commuting frequency
Panel A. Total population			
Benchmark	13.2%	\$2,718,214	81.6%
(a) Shut down referrals $(s_j \to 0)$	–0.1 pp.	–0.2%	–2.9 pp.
(b) Shut down referrals $(s_j \to 0)$ and $\Theta_{jk} + \sigma_{\Theta}$	–1.5 pp.	+2.4%	+14.7 pp.
Panel B. Immigrant population			
Benchmark	13.2%	\$1,602,207	65.2%
(a) Shut down referrals $(s_j \to 0)$	–0.9 pp.	–2.6%	+4.3 pp.
(b) Shut down referrals $(s_j \to 0)$ and $\Theta_{jk} + \sigma_{\Theta}$	–2.0 pp.	+2.1%	+22.4 pp.

Notes: The Table reports estimates from counterfactual exercises based on the calibrated model for neighborhoods in the city. Panel A reports results using the total population (employed and unemployed) in the Bogotá. Panel B reports results using only information about immigrants. Results are reported for three outcomes: the aggregate unemployment rate, the average monthly wage, and the fraction of workers who commute to work to a different neighborhood. Average wage is weighted by the employed population in each neighborhood. Results are presented as percentage point changes (pp.) or percent changes relative to the observed equilibrium. In each panel, the top row reports the observed equilibrium results (benchmark).

Figure 5: Change in Unemployment from Shutting the Network



Notes: The Figure shows the change in the log unemployment level between the observed equilibrium and the first counterfactual, which removes the network, against the fraction of workers who report having found their current job through personal contacts. I plot the local polynomial smoothing along with the 95% confidence interval. Panel (a) presents results for the total population, while panel (b) uses only information on immigrant workers.

Second, results suggest that referrals do not have a large effect on wages when using the complete sample, with only 0.2% of the average wage being explained by the use of referrals. However, using only the sample of immigrants suggest that approximately 2.6% of a workers' wage can be attributed to the use of referrals, holding constant search efficiency through other channels. Both results are lower to the overall estimates presented by Lester et al. (2021) in a sample of U.S. workers, but the results for immigrants presented here lie somewhere in between their estimates of the effect of business referrals and referrals from friends and family. Considering that in the absence of referrals workers are less likely to be unemployed, a drop in wages does not seem to be explained by a lack of job offers. Results seem to suggest that referrals allow for matches closer to a workers' residential location in the case of immigrants. Now, simulating an increase of a standard deviation in the measure of direct access to job information increases commuting substantially.

Consistent with the latter, results in the third column imply that referrals allow immigrant workers to find jobs closer to where they live, avoiding the wage cost of commuting. Shutting down referrals increases the frequency of commuting by about 4 percentage points for immigrant workers but reduces commuting by 3 percentage points for the total population. These effects are consistent with empirical evidence on the commuting patterns of immigrant workers. Since immigrants are more likely to work closer to home (35% of immigrant do not commute to work outside their neighborhood of residence relative to just 18% using the whole sample), in the absence of suitable employment opportunities closer to a workers' place of residence—likely coming from referrals, unemployed workers increase their search outside their own neighborhood.

Finally, Table 2 reports changes in workers' welfare, the expected utility of search (defined by Eq. A.17), and output.[21] Overall, referrals increase welfare by 1.5%, the expected utility that unemployed workers get from searching by 1.4%, and output in the city by 3.3%. These effects, however, are dissipated once we allow for

[21] The welfare function maximizes the utility of employed and unemployed workers discounted over time.

an increase in the rate at which workers find jobs through formal methods. Strikingly, once we remove search through the network, welfare increases by 5.5% in the case of immigrants, but both the expected utility of search and output decrease. These results suggest that immigrant networks may suffer from congestion effects, reducing workers employability, but allow workers to find better matches.

Table 2: Change in Welfare, Utility of Search and Output

	Outcome		
Exercise	Welfare	Utility of search	Output
Panel A. Total population			
(a) Shut down referrals $(s_j \to 0)$	–1.5%	–1.4%	–3.3%
(b) Shut down referrals $(s_j \to 0)$ and $\Theta_{jk} + \sigma_\Theta$	+22.2%	+16.6%	+27.4%
Panel B. Immigrant population			
(a) Shut down referrals $(s_j \to 0)$	+5.5%	–1.2%	–0.6%
(b) Shut down referrals $(s_j \to 0)$ and $\Theta_{jk} + \sigma_\Theta$	+7.0%	+5.0%	+14.9%

Notes: The Table reports estimates from counterfactual exercises based on the calibrated model for neighborhoods in the city. Panel A reports results using the total population (employed and unemployed) in the Bogotá. Panel B reports results using only information about immigrants. Results are reported for three outcomes: workers' welfare, the expected utility of search, and output. Results are presented as percent changes relative to the observed equilibrium.

5. Conclusion

The model presented provides a mechanism explaining why locations in which workers rely more on personal contacts to find jobs are also those with higher unemployment rates. As networks suffer from congestion effects driven by spatial mismatch, particularly for ethnic minorities, workers are faced with fewer suitable employment opportunities. Because the use of personal contacts dominates the search process, it is costly for unemployed workers (in terms of time and access to information) to commute around the city to search for jobs directly. If workers mainly rely on their personal contacts and these are increasingly unemployed or working closer to where they live, then job opportunities are limited for the large pool of unemployed workers, creating congestion in the diffusion of job information.

By modelling the interdependency between the urban and the social space and evaluating the implications of spatial frictions in the location and movement of workers and jobs, we can uncover new mechanisms that determine the access to good jobs. The findings presented in this paper suggest that labor market interventions targeting high unemployment and dense areas by increasing the access to information about suitable job opportunities may be welfare-improving.

Finally, future research could assess and empirically measure the optimal network size (or population density) across neighborhoods, particularly in the presence of ethnic enclaves.

References

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

Bayer, P., Ross, S., & Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy, 116*(6), 1150–1196.

Beaman, L. A. (2016). Social Networks and the Labor Market. In Y. Bramoullé, A. Galeotti, & B. Rogers (Eds.), *The Oxford Handbook of the Economics of Networks.* Oxford Universtity Press.

Beaman, L. A., & Magruder, J. (2012). Who Gets the Job Referral? Evidence from a Social Networks Experiment. *American Economic Review, 102*(7), 3574–3593.

Bilal, A. (2023). The Geography of Unemploymen. *The Quarterly Journal of Economics*, qjad010, https://doi.org/10.1093/qje/qjad010.

Calvó-Armengol, A. (2004). Job Contact Networks. *Journal of Economic Theory, 115*(1), 191–206.

Calvó-Armengol, A., & Zenou, Y. (2005). Job Matching, Social Network and Word-of-Mouth Communication. *Journal of Urban Economics, 57*, 500–522.

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics, 36*(S1), S13–S70.

Card, D., Devicienti, F., & Maida, A. (2013). Rent-sharing, Holdup, and Wages: Evidence from Matched Panel Data. *Review of Economic Studies, 81*(1), 84–111.

Coulson, N. E., Laing, D., & Wang, P. (2001). Spatial Mismatch in Search Equilibrium. *Journal of Labor Economics, 19*(4), 949–972.

Dai, R., Mookherjee, D., Munshi, K., & Zhang, X. (2019). The Community Origins of Private Enterprise in China. Working Paper, Peking University, Beijing.

Damm, A. P. (2014). Neighborhood Quality and Labor Market Outcomes: Evidence from Quasi-random Neighborhood Assignment of Immigrants. *Journal of Urban Economics, 79*, 139–166.

Diamond, P. A. (1981). Mobility Costs, Frictional Unemployment, and Efficiency. *Journal of Political Economy, 89*(4), 798–812.

Dustmann, C., Glitz, A., & Schönberg, U. (2016). Referral-based Job Search Networks. *Review of Economic Studies, 83*, 514–546.

Eaton, J., & Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica, 70*(5), 1741–1779.

Fournier, J. (2021). Spatial Mismatch. Mimeo, Massachusetts Institute of Technology.

Friedberg, R. M. (2000). You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics, 18*(2), 221–251.

Fujita, M., & Ogawa, H. (1982). Multiple Equilibria and Structural Transformation of Non-Monocentric Urban Configurations. *Regional Science and Urban Economics, 12*(2), 161–196.

Galenianos, M. (2013). Learning About Match Quality and the Use of Referrals. *Review of Economic Dynamics, 16*, 668–690.

Galeotti, A., & Merlino, L. P. (2014). Endogenous Job Contact Networks. *International Economic Review, 55*(4), 1201–1226.

Glaeser, E. L., & Ward, B. A. (2009). The Causes and Consequences of Land Use Regulation. *Journal of Urban Economics, 65*(3), 265–278.

Hagedorn, M., & Manovskii, I. (2008). The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited. *American Economic Review, 98*(4), 1692–1706.

Hellerstein, J. K., McInerney, M., & Neumark, D. (2011). Neighbors and Coworkers: The Importance of Residential Labor Market Networks. *Journal of Labor Economics, 29*(4), 659–695.

Holzer, H. J. (1988). Search Method Use by Unemployed Youth. *Journal of Labor Economics, 6*(1), 1–20.

Ioannides, Y. M., & Loury, L. D. (2004). Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature, 53*, 1056–1093.

Jäger, S., Schoefer, B., Young, S., & Zweimüller, J. (2020). Wages and the Value of Nonemployment. *The Quarterly Journal of Economics, 134*(4), 1905–1963.

Kline, P., & Moretti, E. (2013). Place Based Policies with Unemployment. *American Economic Review: Papers and Proceedings, 103*(3), 238–243.

Kline, P., & Moretti, E. (2014). People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs. *Annual Review of Economics, 6*(1), 629–662.

Lester, B., Rivers, D. A., & Topa, G. (2021). Job Referrals and Labor Market Outcomes: It's Not Who You Know, It's How You Know Them. Unpublished manuscript.

Lucas, R. E., & Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica, 70*(4), 1445–1476.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies, 60*(3), 531–542.

Mesa-Guerra, C. A. (2023). Immigrant Networks in the Labor Market. *Unpublished manuscript*.

Montgomery, J. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *American Economic Review, 81*(5), 1408–1418.

Mortensen, D. T., & Pissarides, C. A. (1994). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies, 61*(3), 397–415.

Munshi, K. (2014). Community Networks and the Process of Development. *Journal of Economic Perspectives, 28*(4), 49–76.

Munshi, K. (2020). Social Networks and Migration. *Annual Review of Economics, 12*, 503–524.

Munshi, K., & Rosenzweig, M. (2016). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review, 106*(1), 46–98.

Pallais, A., & Sands, E. G. (2016). Why the Referential Treatment? Evidence from Field Experiments on Referrals. *Journal of Political Economy, 124*(6), 1793–1828.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory, 2nd Edition.* Cambridge, MA: MIT Press.

Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative Spatial Economics. *Annual Review of Economics, 9*, 21–58.

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

Schmutte, I. M. (2015). Job Referral Networks and the Determination of Earnings in Local Labor Markets. *Journal of Labor Economics, 33*(1), 1–32.

Selod, H., & Zenou, Y. (2006). City Structure, Job Search and Labour Discrimination: Theory and Policy Implications. *The Economic Journal, 115*(514), 1057–1087.

Simon, C., & Warner, J. (1992). Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure. *Journal of Labor Economics, 10*(3), 306–330.

Topa, G. (2001). Social Interactions, Local Spillovers and Unemployment. *Review of Economic Studies, 68*, 261–295.

Topa, G. (2019). Social and Spatial Networks in Labour Markets. *Oxford Review of Economic Policy, 35*(4), 722–745.

Tsivanidis, N. (2019). Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio. Mimeo, University of California, Berkeley.
Zenou, Y. (2015). Spatial versus social mismatch. *Journal of Urban Economics, 74*, 113–132.

Online Appendix

A Model of Spatial Job Referral Networks

A Quantitative Implementation Appendix

A.1 Model

Geography. Consider a closed city populated by a measure $\bar{N}$ of workers. The city is composed of $\mathcal{L}$ distinct locations, indexed by $j, k \in \{1, \ldots, \mathcal{L}\}$. Residential locations are indexed by j, while k indexes work locations. Each location can be thought of as a neighborhood or, to a smaller scale, a block within the city. Locations differ in terms of their residential and productive amenities, supply of land, social networks, and commuting times between any two locations within the city.

Preferences. Workers derive utility from the consumption of the single final good (c_j), chosen as the numeraire, and consumption of housing (h_j), for which they pay rent (r_j). Utility is assumed to take the take the following form: $\left(\frac{c_j}{\alpha}\right)^{\alpha}\left(\frac{h_j}{1-\alpha}\right)^{1-\alpha}$, $0 < \alpha < 1$. At the same time, workers have idiosyncratic preferences shocks (e_{ij}) over residential locations which are independently and identically distributed across individuals, locations, and time.

The budget constraint faced by a worker depends on his employment status. When employed, a worker's income depends on the wage rate (w_{jk}) net of the cost of commuting to work from j to k, $d_{jk} = \exp(\kappa\tau_{jk}) \in [1, \infty)$, which is modeled as a reduction in effective units of labor and is increasing with travel time (τ_{jk}).[1] If the worker is unemployed, he receives an unemployment benefit or value of home production (b) and does not incur in commuting cost.

The indirect utility of a worker, conditional on their employment status, can be expressed as:

$$\mathrm{v}_{ijk}^{E} = w_{jk}\left(r_j\right)^{\alpha-1} d_{jk}^{-1}\,, \tag{A.1a}$$

$$\mathrm{v}_{ij}^{U} = b\left(r_j\right)^{\alpha-1}\,. \tag{A.1b}$$

As workers alternate between employment and unemployment their intertemporal utility varies over time. Assume that workers and firms are risk neutral, infinitely lived, and discount future payoffs in continuous time at rate ρ.

In steady state, the Bellman equations for both employed and unemployed workers before idiosyncratic shocks are realized are given by:

$$\rho J_{jk}^{E} = w_{jk}\left(r_j\right)^{\alpha-1} d_{jk}^{-1} + \delta\left(J_j^{U} - J_{jk}^{E}\right), \tag{A.2a}$$

[1] Modeling commuting costs as a reduction in effective units of labor is equivalent to introducing these instead as a reduction in utility as in standard models.

$$\rho J_j^U = b(r_j)^{\alpha-1} + \sum_{k\in\mathcal{L}} \lambda_{jk}(J_{jk}^E - J_j^U)\,, \tag{A.2b}$$

where δ is the job separation rate and λ_{jk} is the probability that a worker living in j finds a job in k. In the continuous time case, δ is interpreted as shocks occurring according to a Poisson process. Replacing (A.2a) in (A.2b) gives:

$$\rho J_j^U = \frac{(r_j)^{\alpha-1}}{(\rho+\delta+\bar{\lambda}_j)}\left((\rho+\delta)b + \sum_{k\in\mathcal{L}} \frac{\lambda_{jk} w_{jk}}{d_{jk}}\right). \tag{A.3}$$

where $\bar{\lambda}_j \equiv \sum_{k\in\mathcal{L}} \lambda_{jk}$ is the probability that a worker living in j finds a job anywhere in the city

Housing supply. In contrast to the stylized model presented in Section 2, each location j is endowed with a fixed supply of housing space (H_j) owned by absentee landlords.[2] Utility maximization implies that workers spend a fraction $(1-\alpha)$ of income on housing. The housing market clearing condition expresses the rent price as a function of the supply of housing space:

$$r_j = (1-\alpha)\frac{\bar{x}_j N_{Rj}}{H_j}. \tag{A.4}$$

where $\bar{x}_j$ is the average income of residents living in location j.

Production. Firms post V_k vacancies in each neighborhood. The Bellman equations for a filled vacancy with match output y_{jk} and an unfilled vacancy with fixed cost $\mathscr{g}$ in workplace k are:

$$\rho J_{jk}^F = y_{jk} - w_{jk} + \delta(J_k^V - J_k^F), \tag{A.5a}$$

$$\rho J_k^V = -\mathscr{g} + f_k(s,u,v)(J_k^F - J_k^V)\,, \tag{A.5b}$$

where $f_k(s,u,v) = \frac{\sum_j U_j \lambda_{jk}(s)}{V_k}$ is the job filling probability. By assumption, firms enter freely up to the point where the value of opening a vacancy is zero in each workplace location, $J_k^V = 0$. From Eqs. (A.5a), (A.5b), and the free-entry condition, the demand for labor at workplace location k is given by:

[2] This assumption greatly simplifies the quantitative implementation as otherwise one would require data on housing supply elasticities across locations. Monte *et al.* (2018) show that results are robust to introducing positive developed land supply elasticities.

$$\frac{\mathbb{E}_{|j}\left[y_{jk} - w_{jk}\right]}{\rho + \delta} = \frac{g}{f_k(s,u,v)} \qquad \forall k \in \mathcal{L}. \tag{A.6}$$

Wage setting. In each period, wages are set using a generalized Nash bargaining over the match surplus, so that:

$$\beta\left(J_{jk}^F - J_k^V\right) = (1-\beta)\left(J_{jk}^E - J_j^U\right), \tag{A.7}$$

where β is the worker's share of the match surplus. And the wage-setting condition is given by:

$$w_{jk} = \frac{\beta y_{jk} + (1-\beta)\Omega_j}{\beta + (1-\beta)\left(r_j\right)^{\alpha-1} d_{jk}^{-1}}. \tag{A.8}$$

where Ω_j is the value of workers outside option (i.e., value of unemployment) which is defined as in Eq. (A.3).

Job acceptance. Workers will accept any job offer in which the expected utility that they would get by taking the job is higher than the utility they would get by remaining unemployed. Therefore, the reservation wage, w_{jk}^R, is implicitly defined by the equality:

$$J_{jk}^E\left(w_{jk}^R\right) = J_j^U. \tag{A.9}$$

That is, an unemployed worker will take a job if and only if $w_{jk} \geq w_{jk}^R$, with the reservation wage given by solving for w_{jk}:

$$\frac{w_{jk}}{d_{jk}} = \frac{(\rho+\delta)b + \sum_{k\in\mathcal{L}} \lambda_{jk} w_{jk} d_{jk}^{-1}}{\left(\rho + \delta + \bar{\lambda}_j\right)}. \tag{A.10}$$

Under the optimal job acceptance rule, the flow value of accepting a job must equal the opportunity cost of remaining unemployed. The payoff of being employed is equal to the wage rate net of commuting costs, while the opportunity cost is a function of the value of home production and the expected payoff from future matches.

Search and matching. Workers are assumed to only search when unemployed.[3] Workers are assumed to hear about job vacancies both through formal and informal methods (i.e., referrals). Thus, unemployed workers can also learn about a job

[3] This implies that there is no on-the-job search. While this assumption does not greatly affect the analysis, as job-to-job transitions do not directly change the level of unemployment, it keeps the model tractable and facilitates the calibration.

opportunity through their employed friends. The probability that an unemployed worker living in j finds a job in k is given by:

$$\lambda_{jk}\left(s, u_j, \pi_{jk}, v_k\right) = v_{jk} + \left(1 - v_{jk}\right) \cdot q_{jk}\left(s_j, u_j, \pi_{jk}, v_k\right) \tag{A.11}$$

where $v_{jk} = \Theta_{jk} v_k$ is the probability that an unemployed worker living in j finds a job in k directly, v_k is the vacancy rate, which is defined as $v_k = V_k / (L_{Ek} + U)$, Θ_{jk} is a measure of direct access to information about available jobs , and $q_{jk}\left(s_j, u_j, \pi_{jk}, v_k\right)$ is the probability that an unemployed worker finds a job through the network, which is given by Eq. (16) in Section 2.2.[4]

When searching for jobs across locations, workers may find vacant jobs directly, but since workers are endowed with a unit of time, the further a job is located from the worker's place of residence, the less likely he is to learn about it using formal search methods. Thus, job opportunities are lost at a rate that depends on the distance between locations.

Assume also that networks are delimited by place of residence, such that an unemployed worker is in direct contact with $s_j < N_{Rj}$ other workers, who can be either employed or unemployed.[5]

In labor markets with search frictions, flows in and out of employment must equal in steady state:

$$\lambda_{jk} U_j = \delta L_{jk} \tag{A.12}$$

And the number of workers living in j must be consistent with:

$$N_{Rj} = U_j + \sum_{k \in \mathcal{L}} L_{jk} \tag{A.13}$$

Which can be rewritten as:

$$N_{Rj} = \frac{\delta + \bar{\lambda}_j}{\delta} U_j \tag{A.14}$$

Mobility. As before, workers are imperfectly mobile. Unemployed workers are free to locate anywhere but, after finding a job, workers cannot change their place of residence to reduce commuting costs. After choosing a residential location, workers enjoy the mean value of amenities A_j, which is assumed to enter multiplicatively in the value function. Therefore, after the idiosyncratic shock is realized, unemployed

[4] The measure of direct access to job information can be interpreted as information loss (for values lower than 1) and may be a function of distance between locations or search efficiency. I calibrate this directly using the structure of the model.

[5] Since I don't observe individual networks in my data, I define networks based on geographical boundaries. Now, this is consistent with the empirical evidence.

workers pick their neighborhood of residence j to maximize their intertemporal value function and mean value of amenities:

$$\max_{j\in\mathcal{L}} \rho J_j^U A_j + e_{ij} \,. \qquad \text{(A.15)}$$

Idiosyncratic preferences for living in different locations (e_{ij}) are drawn from a Gumbel distribution, $F(e_{ij}) = \exp\left(-\exp(-e_{ij})\right)$. Using the properties of the extreme value distributions, the probability that an unemployed worker chooses to live j is given by:

$$\pi_{Rj}^U = \frac{\exp\left[A_j(r_j)^{\alpha-1}\left(\rho+\delta+\bar{\lambda}_j\right)^{-1}\left((\rho+\delta)b+\sum_{k=1}^{\mathcal{L}}\lambda_{jk}w_{jk}d_{jk}^{-1}\right)\right]}{\sum_{\ell=1}^{\mathcal{L}}\exp\left[A_\ell(r_\ell)^{\alpha-1}\left(\rho+\delta+\bar{\lambda}_\ell\right)^{-1}\left((\rho+\delta)b+\sum_{k=1}^{\mathcal{L}}\lambda_{\ell k}w_{jk}d_{\ell k}^{-1}\right)\right]} \equiv \frac{\Phi_j}{\Phi}. \qquad \text{(A.16)}$$

Since all workers face the same choice when unemployed, the residential choice probability must equal the fraction of workers who choose to live in neighborhood j, $\pi_{Rj}^U = \pi_{Rj} = N_{Rj}/\bar{N}$.

In equilibrium, the population mobility condition implies that the expected value of unemployed workers must be equalized across all locations with positive population and equal to the constant $\bar{\mathcal{J}}$. Given the Gumbel distributional assumption, the expected value of unemployed workers can be written as:

$$\mathbb{E}\left[\max_{j\in\mathcal{L}}\left(\rho J_j^U A_j + e_{ij}\right)\right] = \gamma + \ln\left(\sum_{j=1}^{\mathcal{L}} \Phi_j\right) = \bar{\mathcal{J}} \qquad \text{(A.17)}$$

where $\mathbb{E}$ is the expectations operator and the expectation is taken over the distribution of the idiosyncratic component of utility, and γ is the Euler-Mascheroni constant.

Commuting. The commuting clearing condition requires that the number of workers employed in neighborhood k (L_{Ek}) equals the sum across all residential neighborhoods j of the number of employed workers living in j (L_{Rj}) times their conditional probability of commuting to k $(\pi_{jk|j}^E)$:

$$L_{Ek} = \sum_{j=1}^{\mathcal{L}} \pi_{jk|j}^E L_{Rj} \,. \qquad \text{(A.18)}$$

Using the steady state labor market equilibrium (A.12), the probability that an employed worker commutes to work to neighborhood k conditional on living in neighborhood j can be written as:

$$\pi^E_{jk|j} = \frac{L_{jk}}{L_{Rj}} = \frac{(\lambda_{jk}/\delta)U_j}{\sum_k(\lambda_{jk}/\delta)U_j} = \frac{\lambda_{jk}}{\bar{\lambda}_j}. \tag{A.19}$$

Expected income. The average income of workers living in j, as introduced in (A.4), is equal to the average income of employed workers times the probability of being employed plus the income of unemployed workers times the probability of being unemployed:

$$\bar{x}_j = (1 - u_j)\left(\sum_{k\epsilon\mathcal{L}} \pi^E_{jk|j}(w_{jk}/d_{jk})\right) + u_j b. \tag{A.20}$$

The average income for employed workers living in j is equal to the sum of wages net of commuting costs in all possible workplace locations weighted by the conditional commuting probabilities.

Equilibrium. Having characterized the equilibrium behavior of agents within the city, I can proceed to define the equilibrium of the model. In what follows, I use bold math font to denote vectors or matrices.

DEFINITION A.1. Given the model's parameters $\{\alpha, \kappa, \rho, \delta, \beta, s_j\}$, the constant $\bar{\mathcal{J}}$, unemployment benefit b, exogenous location-specific characteristics $\{\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{y}, \boldsymbol{g}, \boldsymbol{\tau}\}$, and total city population $\bar{N}$, the steady-state general equilibrium of the model is characterized by the vector $\{\boldsymbol{r}, \boldsymbol{V}, \boldsymbol{w}, \boldsymbol{\lambda}, \boldsymbol{N_R}, \boldsymbol{\pi^U_R}, \boldsymbol{\pi^E}\}$.

The seven elements of the equilibrium vector are determined by the following system of seven equations: housing market clearing (A.4), equilibrium entry condition (A.6), wage-setting condition (A.8), job-finding probability (A.9), labor market clearing (A.12), residential choice probability (A.14), commuting probability (A.17). And populations add up to the city total, i.e., $\bar{N} = \sum_j^{\mathcal{L}} N_{Rj}$.

A.2 Determining $\{\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{y}, \boldsymbol{g}, \boldsymbol{\Theta}\}$ from $\{\alpha, \kappa, \rho, \delta, \beta, s_j\}$ and the Observed Data

The model can be calibrated to recover unobserved location-specific characteristics $\{\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{y}, \boldsymbol{g}, \boldsymbol{\Theta}\}$ given known values of the model's parameters $\{\alpha, \kappa, \rho, \delta, \beta, s_j\}$ and the observed data $\{\boldsymbol{N_R}, \boldsymbol{U_R}, \boldsymbol{L}, \boldsymbol{r}, \boldsymbol{w}, \boldsymbol{\tau}\}$. Unobserved location-specific characteristics can be recovered as structural residuals of the model as there is a one-to-one mapping from the parameters and the observed data to these unobserved location characteristics. To do so, I use the recursive structure of the model.

1. Given δ and the observed data $\{\boldsymbol{N_R}, \boldsymbol{U_R}\}$, the job finding probabilities $\{\boldsymbol{\bar{\lambda}_R}\}$ can be uniquely determined from the labor market clearing condition.

2. Given job finding probabilities $\{\bar{\boldsymbol{\lambda}}_{\boldsymbol{R}}\}$ and the observed data $\{\boldsymbol{L}\}$, residence-workplace-specific job finding probabilities $\{\boldsymbol{\lambda}\}$ can be uniquely determined from the commuting probabilities.

3. Given $\{\alpha, \kappa\}$, unemployment benefit b, the observed data $\{\boldsymbol{r}, \boldsymbol{N_R}, \boldsymbol{U_R}, \boldsymbol{w}, \boldsymbol{\tau}\}$, and residence-workplace-specific job finding probabilities $\{\boldsymbol{\lambda}\}$, the stock of housing space $\{\boldsymbol{H}\}$ can be uniquely determined from the housing market clearing condition.

4. Given $\{\alpha, \kappa, \rho, \delta, \beta\}$, unemployment benefit b, the observed data $\{\boldsymbol{r}, \boldsymbol{w}, \boldsymbol{\tau}\}$, and residence-workplace-specific job finding probabilities $\{\boldsymbol{\lambda}\}$, the match output $\{\boldsymbol{y}\}$ can be uniquely determined from the wage-setting condition.

5. Given $\{\rho, \delta\}$, the observed data $\{\boldsymbol{N_R}, \boldsymbol{U_R}, \boldsymbol{L}, \boldsymbol{w}, \boldsymbol{\tau}\}$, residence-workplace-specific job finding probabilities $\{\boldsymbol{\lambda}\}$, and the match output $\{\boldsymbol{y}\}$, the fixed entry cost $\{\mathcal{g}\}$ can be uniquely determined from the equilibrium entry condition.[6]

6. Given $\{\rho, \delta, s_j\}$, the observed data $\{\boldsymbol{U_R}, \boldsymbol{L}, \boldsymbol{w}\}$, residence-workplace-specific job finding probabilities $\{\boldsymbol{\lambda}\}$, the match output $\{\boldsymbol{y}\}$, and the fixed entry cost $\{\mathcal{g}\}$, the measure of direct job information access $\{\boldsymbol{\Theta}\}$ can be uniquely determined from the equilibrium entry condition and the job finding probability condition.

7. Given $\{\alpha, \kappa, \rho, \delta\}$, unemployment benefit b, the observed data $\{\boldsymbol{r}, \boldsymbol{w}, \boldsymbol{\tau}, \boldsymbol{N_R}\}$, and residence-workplace-specific job finding probabilities $\{\boldsymbol{\lambda}\}$, residential amenities $\{\boldsymbol{A}\}$ can be uniquely determined from the residential choice probabilities.

A.3 Algorithm for Computing Equilibrium and Counterfactuals

I outline the iterative algorithm used to solve for the model's equilibrium and compute counterfactuals taking as given location-specific characteristics $\{\boldsymbol{A}, \boldsymbol{H}, \boldsymbol{y}, \boldsymbol{\mathcal{g}}\}$, parameters $\{\alpha, \kappa, \rho, \delta, \beta, s_j\}$, and observed data on travel times between locations $\{\boldsymbol{\tau}\}$.

1. Guess matrices $\boldsymbol{\lambda^0}, \boldsymbol{w^0}$ and vector $\boldsymbol{\pi_R^0}$

2. Given matrices $\boldsymbol{\lambda^t}, \boldsymbol{w^t}$ and vector $\boldsymbol{\pi_R^t}$

   (a) Compute the spatial distribution of the population: $N_{Rj}^t = \pi_{Rj}^t \bar{N}$

[6] I target $\mathcal{g}$ to match an aggregate vacancy rate of 5%, consistent with estimates for Colombia. Specifically, I derive the following expression from Eq. (A.6):

$$\mathcal{g} = \frac{1}{0.05(\rho+\delta)(L+U)} \sum_k \left(\mathbb{E}[y_{jk} - w_{jk}] \left(\sum_j U_j \lambda_{jk} \right) \right).$$

(b) Compute commuting probabilities: $\pi_{jk|j}^{E,t} = \lambda_{jk}^t / \sum_k \lambda_{jk}^t$

(c) Compute unemployment rates: $u_j^t = \delta / \left(\delta + \sum_k \lambda_{jk}^t\right)$

(d) Compute workplace employment: $L_{Ek}^t = \sum_j \pi_{jk|j}^{E,t} \left(1 - u_j^t\right) N_{Rj}^t$

(e) Compute the probability of hearing of a job directly: $v_{jk}^t = \Theta_{jk} V_k^t / \left(L_{Ek}^t + \sum_j u_j^t N_{Rj}^t\right)$, where $V_k^t = \left(\mathbb{E}_{|j}\left[y_{jk} - w_{jk}^t\right]\right)\left(\sum_j u_j^t N_{Rj}^t \lambda_{jk}^t\right) / g(\rho + \delta)$

(f) Compute rents:

$$r_j^t = (1 - \alpha) \frac{N_{Rj}^t}{H_j} \left[\left(1 - u_j^t\right) \left(\sum_j \pi_{jk|j}^{E,t} \left(w_{jk}^t / d_{jk}\right) \right) + u_j^t b \right]$$

(g) Compute reservation wages:

$$w_{jk}^{R,t} = \frac{d_{jk}}{\left(\rho + \delta + \sum_k \lambda_{jk}^t - \lambda_{jk}^t\right)} \left((\rho + \delta) b + \sum_{k' \in \mathcal{L}} \frac{\lambda_{jk'}^t w_{jk'}^t}{d_{jk'}} \right)$$

(h) Update the main variables:

$$\tilde{\lambda}_{jk} = v_{jk}^t + \left(1 - v_{jk}^t\right) \left[1 - \left(1 - v_k^t \left(1 - u_j^t\right) \pi_{jk|j}^{E,t} \frac{1 - \left(1 - u_j^t\right)^{s_j^t}}{s_j^t u_j^t} \right)^{s_j^t} \right],$$

$$\widetilde{w}_{jk} = \max \left\{ \frac{\beta y_{jk} + (1 - \beta) \frac{\left(r_j^t\right)^{\alpha - 1}}{\left(\rho + \delta + \sum_k \lambda_{jk}^t\right)} \left((\rho + \delta) b + \sum_k \frac{\lambda_{jk}^t w_{jk}^t}{d_{jk}} \right)}{\beta + (1 - \beta) \left(r_j^t\right)^{\alpha - 1} d_{jk}^{-1}}, w_{jk}^{R,t} \right\}$$

$$\tilde{\pi}_{Rj} = \frac{\exp\left[A_j \left(r_j^t\right)^{\alpha - 1} \left(\rho + \delta + \sum_k \lambda_{jk}^t\right)^{-1} \left((\rho + \delta) b + \sum_{k=1}^{\mathcal{L}} \lambda_{jk}^t w_{jk}^t d_{jk}^{-1} \right) \right]}{\sum_{\ell=1}^{\mathcal{L}} \exp\left[A_\ell \left(r_\ell^t\right)^{\alpha - 1} \left(\rho + \delta + \sum_k \lambda_{\ell k}^t\right)^{-1} \left((\rho + \delta) b + \sum_{k=1}^{\mathcal{L}} \lambda_{\ell k}^t w_{\ell k}^t d_{\ell k}^{-1} \right) \right]}$$

3. Iterate until convergence. If $\left\| \left(\boldsymbol{\lambda}^t, \boldsymbol{\pi}_R^t, \boldsymbol{w}^t\right) - \left(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\pi}}_R, \widetilde{\boldsymbol{w}}\right) \right\| < \epsilon_{tol}$, then stop. Otherwise, set $\left(\boldsymbol{\lambda}^{t+1}, \boldsymbol{\pi}_R^{t+1}, \boldsymbol{w}^{t+1}\right) = \zeta\left(\boldsymbol{\lambda}^t, \boldsymbol{\pi}_R^t, \boldsymbol{w}^t\right) + (1 - \zeta)\left(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\pi}}_R, \widetilde{\boldsymbol{w}}\right)$ for some $\zeta \in (0,1)$ and return to step 2.

A.4 Description of Main Variables and Data Sources

Labor force. The labor force was estimated using a weighted count of all working-age individuals either working or looking for a job. It is important to note that the EMB defines the working age as 15 years or older.

Total unemployment. The total number of unemployed workers was estimated using a weighted count of all unemployed individuals, defined as those of working age that, if given the opportunity to work the week before being surveyed, were available to work.

Commuting flows. A weighted count of the number of individuals that commute between two points in the city (the place of residence and the workplace location). While all surveyed individuals report their residential location, only 35% of all employed workers report their place of work. Therefore, commuting flows are obtained for those individuals with complete information and the remainder is imputed using observed commuting probabilities.

Rents. Since the EMB does not ask individuals to state the total area rented by the household but does report the number of rooms occupied, a proxy was generated to create a homogenous measurement for rental price across different living arrangements. Therefore, the average housing rental price was estimated as a weighted average of the price per room reported by surveyed individuals who declared renting or subletting a home within each residential location.

Wage matrix. Wages include labor income (including overtime pay) reported by wage and salary workers and monthly income of self-employed workers. When the income reported in the EMB differs from the one reported in the RELAB, I impute the information from the REALB to all observations for which reported wages in the survey is lower or do not report any information at all. Information from the RELAB comes from social security contributions. Wages are then estimated as a weighted average using the information on all individuals commuting to work between pairs (residence-workplace locations).

Commuting time matrix. I use the STATA command *osrmtime* to estimate the travel times between different residential and workplace locations. This command uses OpenStreetMap data and the Open Source Routing Machine (OSRM) to perform time and distance calculations between geographical locations. I compute the travel time for the shortest route between all possible tracts (sectors) and then average travel times for all sectors within each larger neighborhood (UPZs), thus rendering an average commute between all possible combinations of UPZs.