Local Labor Markets and Aggregate Productivity *

Paolo Martellini†

Federal Reserve Bank of Minneapolis

February 24, 2021

Abstract

I propose a dynamic spatial equilibrium model that accounts for the heterogeneous labor market experience of workers across US cities. Productivity differentials between large and small cities emerge as an equilibrium outcome due to spatial sorting, increasing returns to scale in job search, and knowledge diffusion through local peer effects. The model delivers testable predictions with respect to selection into and returns to migration, which are supported by the data. I use this framework to quantify the aggregate implications of relaxing zoning regulations in large cities. I show how the resulting relocation of workers affects the size and composition of cities, the return from local interactions, and the spatial distribution of productivity. An alternative scenario in which local productivity was invariant to zoning policy would overstate the magnitude of the equilibrium income gains by a factor of 3.

JEL Codes: E24, J24, J31, R12

Keywords: Knowledge Diffusion, Search Frictions, Life-Cycle Income Dynamics, Urban Wage Premium

*This paper was part of my dissertation at the University of Pennsylvania. I am grateful to Guido Menzio, Dirk Krueger, José-Víctor Ríos-Rull, and Jesús Fernández-Villaverde for their support throughout all stages of this project. I also thank Francesco Agostinelli, Anmol Bhandari, Alessandro Dovis, Andrew Shephard, Chris Tonetti, Mike Waugh, and seminar participants at the University of Pennsylvania, EIEF, NYU Search Theory Workshop, CEMFI, UCSD, Richmond Fed, UW Madison, Online Urban Economics Seminar, Minneapolis Fed, Urban Economic Association Conference, Southern Economic Association Conference, Econometric Society Winter Meeting. I thank the U.S. Bureau of Labor Statistics for providing access to the NLSY79 confidential geocode data. A previous version of this paper circulated under the title "The City-Size Wage Premium: Origins and Aggregate Implications".

The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. All errors are my own.

†Contact: pa.martellini@gmail.com.

1 Introduction

Within most countries around the world, there exist vast wage differentials between cities, particularly between large and small ones. This observation naturally gives rise to two questions: First, what are the aggregate implications of local policies that trigger the reallocation of workers across cities? Second, how are the outcomes of such policies affected by the nature of wage differentials themselves? On one extreme, if the heterogeneous performance of cities is due solely to sorting of unequally productive workers, local policies have no effect on aggregate productivity. On the opposite extreme, if productivity is an exogenous characteristic of cities, the income gains from worker reallocation are potentially enormous. In this paper, I investigate the nature of labor productivity differentials between large and small US cities, and their implications for the evaluation of local policies. I show that spatial wage premia are due both to sorting and to *endogenous* features of cities, which are in turn determined by city size and human capital composition. As a consequence, local policies that trigger worker relocation across cities do affect aggregate productivity, but their quantitative effect depends on the equilibrium response of local productivity itself.

I address the above-mentioned questions by building a spatial equilibrium life-cycle model of location choice and labor market dynamics that incorporates three sources of productivity differences between large and small cities. First, I allow for increasing returns to scale in the matching process between workers and firms. Increasing returns represent the idea that the concentration of a large number of workers and firms inside a city reduces the transportation and information frictions that are associated with the process of job search. Second, I model the idea that geographic proximity promotes the diffusion of knowledge between workers and that it is particularly beneficial in a city that is rich in human capital. Accordingly, I allow for workers in large cities to experience more frequent interactions with one another and, to the extent that large cities host a greater fraction of high-skilled workers, to learn from better peers. Third, the so called city-size wage premium may be driven by sorting, which occurs through endogenous migration decisions over the life cycle. While larger cities are more productive, they also feature higher house prices, because of imperfectly elastic housing supply.

To identify the importance of these three channels, I exploit heterogeneity in labor market outcomes between workers in large and small cities. In the data, a city-size wage premium is present at any level of labor market experience, but it significantly increases over the life cycle. Inspired by this evidence, the parameters of the knowledge diffusion technology are identified by cross-city differences in wage *growth*. Empirically, unemployment and job-to-job transition rates do not vary systematically with city size. This observation imposes a restriction on the exogenous distribution of potential matches such that increasing returns to scale in the search process induce workers and firms to be more selective with respect to the quality of matches they are willing to create. Under such restriction, matches in larger cities are endogenously better, and proportionally so at all stages of labor market experience. Hence, the presence of lower search frictions in larger cities is identified by the average *level* of the city-size wage premium. Last, sorting is determined by the optimal mobility choice of workers, given the identified difference in knowledge diffusion and search efficiency between large and small cities.

I estimate the model using the National Longitudinal Survey of Youth 1979 (NLSY79), which is a panel of US workers who entered the labor market in 1980s and that I observe for 20 consecutive years since they started working. I define large (small) cities as commuting zones with more (less) than 750,000 people in 1990. I estimate that exchanges of ideas through interactions among workers are 20% more frequent in large cities. In addition, the higher rates of knowledge diffusion in large cities leverage an equilibrium human capital distribution of peers that first order stochastically dominates the distribution in small cities. As learning is allowed to vary by workers' education and human capital level, I find that the diffusion of ideas is particularly beneficial to college graduates and, conditional on education, to workers with lower initial human capital. Confirming the hypothesis of increasing returns to scale in the labor market, I find that the contact rate between workers and firms in large cities is 73% higher than in small ones. Such estimate is quantitatively in line with direct evidence I provide on the elasticity of the number of applications per vacancy received by firms—and applications to vacancies submitted by workers—with respect to city size. That is, under an empirically plausible measure of higher search efficiency in large cities, the model does not need to further rely on the traditional, exogenous, TFP difference between cities in order to replicate the observed city-size wage premium.

In the model, sorting is induced entirely by the (heterogeneous) benefits from matching and knowledge diffusion, in the face of a higher cost of living in large cities. Importantly, the wage of movers is not explicitly targeted in the estimation, which uses only information on aggregate wage differences between large and small cities.[1] Therefore, I validate the model according to its ability to reproduce the micro evidence on the wage of movers with respect to that of stayers in the pre-migration city and of incumbents in the destination one. I find that the model quantitatively replicates the wage path of workers who move from small to large cities, which is steeper compared with that of stayers. Newcomers into large cities earn significantly less than incumbents, although the difference partially shrinks in the years after migration. Focusing on the flow in the opposite direction, I show that movers to small cities are negatively selected in terms of pre-migration wages and that their income prospects further decline after moving, compared with those of stayers in large

[1] The vast majority of workers never move from a small to a large city, nor vice versa. See the discussion in section 3.5, where I show that both in the model and in the data, the identifying moments barely change when computed on the whole sample or on the sample of non-movers only.

cities. Taken together, these results highlight how the model successfully addresses the issue of selection into—and returns to—migration, which has proved to be a challenge in existing empirical work on the nature of wage differentials between cities.

Equipped with a quantitative theory of the city-size wage premium, I explore the aggregate implications of local policies that change the size and composition of cities. I study the equilibrium response of the economy to a relaxation of zoning regulations in large cities, which takes the form of an increase in housing supply elasticity. This policy counterfactual is motivated by the recent debate on whether place-based policies might increase aggregate income by triggering the relocation of workers toward some of the largest, most productive cities in the US. This paper incorporates two new features to the debate on this topic: fully endogenous local productivity differentials and dynamic bene ts from experience in large cities. I highlight the importance of these margins by contrasting the *equilibrium* response to the policy change with an alternative scenario in which productivity was an *exogenous* characteristic of cities. I show that the latter scenario overstates the percentage growth in total income by a factor of 3. When large cities further grow in size, more workers gain access to their productivity bene t. However, the workers who are attracted to large cities are, on average, both less skilled than those who choose to live there under strict housing regulation and more skilled than those who remain in small cities after the policy change. As a result, the human capital composition of workers deteriorates in both types of cities, with the additional negative impact on local productivity due to a lower quality of peers.

Related Literature

This paper is closely related to the literature that studies the origins of spatial wage differentials. [2] Glaeser and Maré (2001) document the existence of an urban-rural wage premium in the US, which has both a level and a growth component (see also De La Roca and Puga (2017)). [3, 4] Combes et al. (2012) nd that agglomeration economies, in contrast to selection due to stronger competition, are responsible for the higher productivity of large cities, while Davis and Dingel (2019) build a model in which large cities emerge as the location in which high-ability workers cluster in order to share their knowledge. [5] While both papers

[2] In their seminal handbook chapter, Duranton and Puga (2004) list matching and learning as two of the three main sources of agglomeration economies, with input sharing as the third one.

[3] Eckert, Hejlesen, and Walsh (2020) address the endogeneity of workers' initial location decision by considering a sample of refugees who were randomly assigned to Danish cities. They nd that while entry-level wages do not differ across space, accumulating experience in Copenhagen is associated with higher wage growth.

[4] Combes, Duranton, and Gobillon (2011) review the issues involved in the identi cation of the city-size wage premium using reduced-form speci cations, particularly with regard to the dif culties implied by mobility over the life cycle. Estimating the foundations of the city-size wage premium is empirically challenging when there exists a systematic life-cycle component in both the frequency of labor market episodes—like job-to-job transitions and human capital accumulation—irrespectively of the worker's location, and in the decision to move in or out of large cities. In this regard, a life cycle model is key in order to recover the fundamental parameters that govern the pattern of wages and migration episodes observed in the data.

[5] See also Behrens, Duranton, and Robert-Nicoud (2014), who jointly model sorting, selection, and agglomeration in order to account

address cross-sectional heterogeneity between cities, their static nature makes them silent on the life-cycle profile of the wage premium.

To the best of my knowledge, this paper introduces the first empirical equilibrium model with dynamic knowledge diffusion in cities. The mechanism I adopt is related to the theoretical contribution by Glaeser (1999), who builds a two-period model in which homogeneous young workers learn from (skilled) old ones. By allowing for heterogeneity in human capital and learning ability in a quantitative life-cycle model, this paper speaks to the evidence on selection into migration and on the short- and medium-term return to moving.

Exploring the matching channel, Baum-Snow and Pavan (2012) find that search frictions are not significantly different between small and large cities, while they estimate a steeper wage profile for workers in large cities. [6] Martellini and Menzio (2020) provide restrictions on the shape of the match quality distribution, under which increasing returns to scale in the search process are consistent with the observed lack of variation in labor market flows with respect to city size. While the model in this paper satisfies those restrictions, this is not the case for Baum-Snow and Pavan (2012), hence the different findings. I then provide suggestive evidence on the existence of lower search frictions in larger labor markets, using worker- and firm-level information on how the number of job applications varies along the city-size distribution. In addition, I replace the exogenous wage profile in Baum-Snow and Pavan (2012) with an equilibrium learning model that can be used for policy analysis.

A related literature takes the existence of spatial wage differentials as given and focuses on its aggregate implications. Hsieh and Moretti (2019) and Herkenhoff, Ohanian, and Prescott (2018) study the effect of relaxing land-use regulation in some large US cities (or states) on total output, through to the relocation of workers toward more productive locations. Importantly, they abstract from worker heterogeneity and human capital accumulation. In those papers, the key determinant of spatial productivity differentials is given by locations' exogenous TFP levels. Compared with this literature, I show that endogenizing the sources of spatial wage differentials significantly reduces the income gains following a similar policy experiment.

Methodologically, this paper nests into the literature that studies productivity gains through knowledge diffusion (Luttmer (2007), Lucas and Moll (2014), Perla and Tonetti (2014)), with applications in various economic contexts like learning from coworkers inside a firm (Herkenhoff et al. (2018), Jarosch, Oberfield, and Rossi-Hansberg (Forthcoming)), trade (Buera and Oberfield (2020), Perla, Tonetti, and Waugh (2021)),

for heterogeneity in productivity between and within cities.

[6] Schmutz and Sidibé (2019) abstract from worker heterogeneity and human capital, and build a model of the French economy with frictional labor markets and migration. They find that job-to-job transitions are a major source of wage growth within large cities. Using German data, Dauth et al. (2019) provide evidence that larger cities are characterized by higher assortative matching between workers and firms. Importantly, their findings are robust to controlling for heterogeneity in the firm and worker composition of cities. See also Gould (2007) for a dynamic model in which a worker's life-cycle wage profile is a function of his location.

peer effects in neighborhoods (Fogli and Guerrieri (2019)). Contributing to this body of work, I explore the role of knowledge diffusion in accounting for the remarkable difference in life-cycle wage growth across cities. This paper shares some key features with Lucas (2004), who models the long-run transition from a rural to an urban economy, the latter being characterized by a human capital-intensive technology and by learning from others. In contrast, I focus on an equilibrium in which small and large cities coexist, and I provide a quantification of the "external effect" of other workers on the process of human capital accumulation in a way that is empirically consistent with the labor market experience of workers in US cities.[7]

Last, this paper is closely related to the literature on the determinants of wage growth over the life cycle, both within and between jobs. Burdett and Mortensen (1998) and Postel-Vinay and Robin (2002) model search and matching frictions in the process of "climbing the job ladder", while Bagger et al. (2014) and Menzio, Telyukova, and Visschers (2016) augment a search model with human capital accumulation through learning by doing. This paper explores how search efficiency and human capital accumulation are affected by a worker's location through non-constant returns to scale in the matching function and local knowledge spillovers, both of which are affected by the size and composition of the city where the worker lives.

2 The Model

2.1 Overview

I consider a continuous-time economy made of two types of locations: small and large cities. Cities are inhabited by a continuum of workers with different ages, levels of education, and human capital, and by a continuum of identical firms. Inside each city, or local labor market, workers can be either employed or unemployed. Local labor markets are characterized by search frictions and heterogeneity in firm-worker match quality. Workers search both on and off the job. When they contact a firm, they observe the quality of the potential match and decide whether to form a new employment relationship. Workers also search across cities but must pay a moving cost if they decide to migrate. Workers accumulate human capital through learning by doing and through interactions with other workers, thanks to knowledge diffusion (or imitation). While learning by doing is unaffected by the worker's location, I assume that interactions require geographic proximity, so that workers learn exclusively from those who are located in their city.

[7] The main specification in Lucas (2004) is such that everyone in the city learns only from the most skilled worker. Hence, he conjectures that a social planner would want only the "leader" to invest time in learning, in order to maximize the extent of knowledge spillovers. However, he concludes that "if one is to gain the ability to use the theory to discover ways to improve on the equilibrium, a better description of the social character of the learning process will be needed."

Both migration decisions over the life cycle and learning determine the equilibrium size and human capital composition of cities. City size is allowed to affect the amount of search frictions and the frequency of interaction between workers through increasing returns to scale in the labor market and in the process of knowledge diffusion, respectively. The human capital composition of a city determines the quality of peers. Workers consume a non-tradeable local good (e.g. housing) whose equilibrium price might differ across cities. Local prices operate as a congestion force in workers' location choice.

2.2 Environment

2.2.1 Geography

The economy is made of N small cities and 1 large city. Cities, or locations, are denoted by $i \in \{small, large\}$. Small and large cities are heterogenous with respect to the meeting rate in the labor market, the rate of interaction among workers, the equilibrium distribution of peers, and house prices. These endogenous features result from the existence of increasing returns to scale in the search and knowledge diffusion processes and sorting on unobservable human capital, given fundamental heterogeneity between cities with respect to their housing supply elasticity, contact rate with firms in other cities, and vacancy creation cost, all of which are explained in detail below.

2.2.2 Workers

Demographics. The economy is populated by a measure M of workers. Workers are indexed by the tuple (h, a, e). Human capital $h \in H = \{h_1, h_2, ..., h_L\}$ is discrete and evolves endogenously over the life cycle. Age is denoted by $a \in \{y, o\}$, where y stands for young and o for old. The worker's education type is denoted by e, which is permanent throughout his life and is equal to either hs (high school) or col (college). Workers inelastically supply one indivisible unit of labor and maximize the present value of net flow income discounted at rate r. Net flow income is equal to $b_i h - q^e p_i$ if the worker is unemployed or $wzh - q^e p_i$, if the worker is employed at a job of match quality z and receives wage wzh, described below. The gross flow payoff per unit of human capital from being unemployed in city i is b_i, and $q^e p_i$ is the flow cost of non-tradeable goods in city i. Workers are born young and grow old at (Poisson) rate ψ_y. When old, they exit the economy at rate ψ_o and are replaced by new young workers in their same city. Newborns draw their education and initial human capital from a distribution with cdf $G_{0,i}(h, e)$ and probability mass function $g_{0,i}(h, e)$. I assume that the probability that a worker enters the economy with given education depends on the educational attainment of the worker he is replacing. Conditional on e, the initial human capital distribution is independent of the worker's location. This implies that $G_{0,i}(h, e) = G^h_0(h|e) G^e_{0,i}$, where

$G^e_{0,i} = G^e_0(p_i^{col,o})$ and $p_i^{col,o}$ is the equilibrium fraction of old college graduates in city i.[8]

Human Capital Accumulation. Young workers accumulate human capital through two channels.

First, they experience so-called "learning by doing", which captures the additional skills a worker gains by performing a given set of tasks while employed. This form of learning represents the form of human capital accumulation that is unaffected by the worker's location. From learning by doing, the human capital of an employed worker with education e improves from h_ℓ to $h_{\ell+1}$ at the rate

$$h^e \exp(\ h h_\ell).$$

This formulation generates a decline in learning probability with respect to the worker's current human capital, while it allows the level of the learning rate to vary by education.

The second form of human capital accumulation—which I call "knowledge diffusion", or "imitation"—is affected by geographic proximity with other workers (Jaffe, Trajtenberg, and Henderson 1993; Akcigit et al. 2018). For example, highly populated cities may be characterized by more frequent interactions and exchanges of ideas. I define the rate of interaction between workers's dependence on city size as the "flow of ideas" channel. In addition, the amount of learning that occurs through imitation depends on the type of individuals a worker interacts with. Therefore, the human capital composition of a city—or peer effects—is also a key determinant of the gains from knowledge diffusion. I model a flexible imitation technology by assuming that a worker with education e in city i experiences an increase in human capital from h_ℓ to $h_{\ell+1}$ at rate $s_i k(G_i(h_{\tilde\ell}), h_\ell, e)$, where $s_i \quad s(M_i)$, for some function s. The dependence of the learning rate on city size, M_i, and on the equilibrium distribution of human capital in city i, $G_i(h_{\tilde\ell})$, represent the flow of ideas channel and peer effects, respectively. Concretely, the function k takes the form

$$k(G_i(h_{\tilde\ell}), h_\ell, e) = E_{G_i(h_{\tilde\ell})}[h^e_{ver}\max\{h_{\tilde\ell} \quad h_\ell, 0\} + h^e_{hor} h_{\tilde\ell}].$$

The first term on the right-hand side captures the kind of learning that occurs exclusively when interacting with more-skilled workers (*vertical imitation*), in the tradition of Lucas and Moll (2014), and Perla and Tonetti (2014). The second term represents the fact that workers might learn from everyone else, since even less-skilled workers have some knowledge to transfer (*horizontal imitation*). Taken together, the technology of knowledge diffusion can be interpreted as follows. A worker of type (h_ℓ, e) meets another worker of type $h_{\tilde\ell} \quad G_i(h_{\tilde\ell})$ at rate s_i, and he becomes of type $h_{\ell+1}$ with probability $h^e_{hor} h_{\tilde\ell}$, if $\tilde\ell \quad \ell$ or with probability

[8] Workers in the NLSY79 sample were administrated a test of cognitive ability known as the Armed Force Qualification Test (AFQT). The assumption I make on the workers' initial distribution is motivated by the observation that conditional on education, the average AFQT score upon entering the labor market is virtually identical in small and large cities—as already documented by De La Roca, Ottaviano, and Puga (2014) and Baum-Snow and Pavan (2012).

$h^e_{hor} h_{\tilde{\ell}} + h^e_{ver}(h_{\tilde{\ell}} - h_{\ell})$, if $\tilde{\ell} > \ell$.[9]

Although the focus of this paper is on how human capital accumulation is affected by a worker's location, education-specific learning by doing accounts for the spatial heterogeneity in learning that is due solely to the educational composition of cities. In the absence of this channel, the role of knowledge diffusion in large cities is overstated whenever college graduates are more likely to both live in large cities and experience faster wage growth irrespectively of their location.

2.2.3 Firms

Each city is also populated by a positive measure of homogeneous firms. Firms operate a constant returns to scale technology that transforms one unit of labor into zh units of output. The variable z denotes the quality of the match, which is the component of productivity that is specific to the firm-worker pair, while h is the human capital of the worker. Firms maximize the present value of their profits, $(1-w)zh$, discounted at rate r.

2.2.4 Local Labor Market

The labor market is characterized by search frictions. Workers can be either employed or unemployed.[10] An unemployed worker who lives in city i contacts a firm, also located in city i, at rate $\lambda_{0,i} \equiv \lambda_0(M_i)$. The dependence of λ_0 on city size is what I define the "matching channel". This channel captures the idea that because of lower information and transportation frictions, workers in larger cities have access to a broader set of potential matches. Employed workers in city i contact firms in their same city at rate $\lambda_{1,i} \equiv \lambda_1(M_i) = \rho\lambda_{0,i}$, where the parameter $\rho \in [0,1)$ captures the relative search efficiency on the job. Upon meeting, the firm-worker pair draws a match quality $\hat{z} \sim F(\hat{z})$, where $\hat{z} \in [\underline{z}, \overline{z}]$, $\underline{z} > 0$, and $\overline{z} \leq \infty$. If the pair decides to form a match, it starts producing, and the worker receives a fraction β of the gains from trade; otherwise, they keep searching. Jobs are exogenously destroyed at rate δ^e.

While in the first part of the paper I take the meeting rates λ as a given parameter, in section 4, I introduce a zero-profit condition in the market for vacancies, in order to allow firms' hiring behavior to endogenously respond to policy changes.

---

[9]Since the support of the human capital distribution is bounded from above, workers of type h_L experience neither learning by doing nor knowledge diffusion. In practice, the set H is chosen so that only a negligible measure of workers ever achieve the human capital level h_L over the life cycle.

[10]The model abstracts from flows into and out of the labor force. In the empirical section, I pool unemployed workers and those out of the labor force into a single category that I simply refer to as unemployment. As stated in the quantitative section, I estimate the model on a representative sample of young males, for whom fluctuations in labor force participation are less likely to be a concern.

2.2.5 Location Choice

Workers in city i are also contacted by rms located in a different city, which I refer to as city i. The meeting rate between an unemployed worker in city $i = \text{small}$ and a rm in city $i = \text{large}$ is denoted by $l_{0,\ \ i} \quad l_0(M \ \ _i) = r \ l_{0,\ \ i}$. The meeting rate between an unemployed worker in city $i = \text{large}$ and a rm in each one of the N cities $i = \text{small}$ is denoted by $\frac{1}{N} l_{0,\ \ i} \quad \frac{1}{N} l_0(M \ \ _i) = \frac{1}{N} r \ l_{0,\ \ i}$, so that the overall cross-city meeting rate for workers in the large city is equal to $r \ l_{0,\text{small}}$. The parameter r captures the extent of search frictions across—rather that within—cities. This speci cation corresponds to the intuition that if a larger city generates more job offers for its current residents, it does so for workers in a different city as well, although possibly at a proportionally lower rate.

Upon meeting, the rm-worker pair observes the quality of the match $\hat{z} \quad F(\hat{z})$. In contrast to meetings that occur inside a given city, the worker also draws a migration cost $c \quad D(c)$, which is iid distributed across migration episodes. According to the realization of $(\hat{z}, c)$, one of the following events occurs: i) the worker migrates and the match is formed; ii) the worker migrates, but the rm-worker pair does not form the match, in which case the worker becomes unemployed in city i; iii) the worker does not migrate and remains in his current city and employment status.

It is worth emphasizing the reason why the economy is populated by $N > 1$ small cities, even though I focus on an equilibrium in which they are all identical to one another. Under the classi cation I adopt in the quantitative section, the number of small cities in the US economy is much larger than the number of large cities. Therefore, while there are only two types of cities in equilibrium, setting $N > 1$ allows me to aggregate city-level outcomes and evaluate the macroeconomic impact of local policies.

2.2.6 Housing Market

Each city i is characterized by a supply function for housing,

$$p_i = p_{0,i} Q_i^{g_i} \qquad p_{0,i} > 0, g_i \ 2 \ R,$$

where g_i is the (inverse) elasticity of housing supply. [11] I assume that the housing stock is owned by absentee landlords outside the economy. A worker of education e, located in city i, inelastically demands q^e units of housing; hence, he incurs a ow payment $q^e p_i$. While I refer to p_i as house price, or local price, for simplicity, p_i is meant to represent both non-tradeable consumption and the share of tradeable goods whose price is affected by local house prices.

[11] The assumption of constant elasticity housing supply at the city level is common in the urban literature (see, among others, Hsieh and Moretti (2019)), and it allows me to map the effect of land use regulation on house prices through a single parameter, g_i.

2.2.7 Contracts

I assume that the contracts offered by firms to workers are sufficiently flexible that the outcome of the matching process is bilaterally efficient, in the sense that the joint value of a match—that is, the sum of the present discounted value of the firms's profit and the worker's utility—is maximized. As a consequence, the allocation of workers across jobs and cities does not depend on the history of wages. Many contractual environments satisfy this convenient property (see Menzio and Shi (2011) for some examples). In this paper, I follow the approach in Bagger et al. (2014) and assume that a worker is paid a fraction w of his productivity—a so-called "piece rate"—so that his wage is equal to wzh. The wage setting mechanism is described in detail in Appendix A, but it can be summarized as follows. When a worker is hired, either from unemployment or through a job-to-job transition, the equilibrium piece rate is the unique value of w that solves the Nash bargaining problem. Over the course of the employment relationship, even if an outside offer does not trigger a job-to-job transition—because the joint value of the current match is higher than the joint value of the new potential match—the piece rate is revised upward if the value of the new potential match is higher than the value the worker is currently receiving. Notice that a worker who migrates incurs a one-time cost c. Hence, such a cost needs to be subtracted from the value of any new potential match that requires a change of city.[12]

2.3 Definition of a Stationary Equilibrium

In this section, I define a stationary equilibrium for this economy. As previously mentioned, I restrict attention to equilibria in which the N small cities are all identical to each other in terms of size and composition. In order to define an equilibrium, I introduce the following notation. Let $U(h_\ell, a, e, i)$ be the present discounted value of income of an unemployed worker of human capital h_ℓ, age a, and education level e who lives in city i. Let $V(h_\ell, a, e, i, z)$ be the sum of the present discounted value of utility to the worker and profit to the firm if the worker has type (h_ℓ, a, e), the firm-worker pair is located in city i, and the employment relationship has match quality z. I refer to $V(h_\ell, a, e, i, z)$ as the joint value of a match.

[12] To keep the model simple, I rule out the possibility of quitting a job while remaining inside the same city. This assumption is quantitatively innocuous, as I find that only 0.03% of workers are employed in a match they would not have formed to begin with. Without this assumption, quitting would be observed in equilibrium after a worker becomes old or he accumulates human capital whenever i) the reservation match quality is increasing in age or in human capital for at least certain values of h, and ii) the worker's match quality is sufficiently close to its reservation value. Besides, if quitting is allowed, unemployment represents a credible outside option, that might trigger an increase in the piece rate upon aging or learning. This possibility would complicate the analysis without adding any relevant insight. What makes this restriction quantitatively negligible is the fact that the reservation match quality is almost constant in age and in the level of human capital and that workers leave marginal matches at a sufficiently high rate.

The value of unemployment $U(h_\ell, a, e, i)$ satis es the following Hamilton-Jacobi-Bellman equation, HJBE:

$$
\begin{aligned}
rU(h_\ell, a, e, i) = \; & b_i h_\ell \quad q^e p_i + \\
& s_i k(G_i, h_\ell, e)[U(h_{\ell+1}) \quad U]1\{a = y\} + \; y_a[U(o)1\{a = y\} \quad U] + \\
& l_{0,i} E_F[\max\{ b(V(\hat{z}) \quad U), 0\}] + \\
& l_{0, \; i} E_{F,D}[\max\{ b(V(\; i, \hat{z}) \quad U(\; i)) + \; U(\; i) \quad U \quad c, \\
& b(V(\; i, \hat{z}) \quad U \quad c), U(\; i) \quad U \quad c, 0\}].
\end{aligned}
\tag{2.1}
$$

For ease of notation, I omit the dependence of the value and policy functions on the right-hand side of the HJBE from the individual states that are the same as on the left-hand side. The LHS of equation (2.1) is the annuitized value of unemployment. The rst line on the RHS is the ow payoff of unemployment, net of the house price. The second line shows the gains from knowledge diffusion for the young, and the transition to old age or outside the economy. The third line shows the option value of searching in the worker's current city, which is equal to the rate at which an unemployed worker meets a rm, multiplied by the fraction of surplus that accrues to the worker if the match is formed. Since the joint value of a match is strictly increasing in its quality, the decision to create a match gives rise to a cutoff $R(h_\ell, a, e, i)$, which is implicitly de ned by

$$V(h_\ell, a, e, i, R(h_\ell, a, e, i)) = U(h_\ell, a, e, i). \tag{2.2}$$

The last two lines of Equation (2.1) describe the event in which the worker contacts a rm in city $\; i$. According to the realization of $(\hat{z} \quad F, c \quad D)$, the gain from a cross-city meeting is equal to the maximum between four terms, from left to right: i) moving as employed to city $\; i$, having unemployment in city $\; i$ as outside option in the bargaining protocol (high $\hat{z}$, low c); ii) moving as employed to city $\; i$, having unemployment in city i as outside option (high $\hat{z}$, high c); iii) moving as unemployed to city $\; i$ (low $\hat{z}$, low c); iv) not moving (low $\hat{z}$, high c). The max operator in the fourth line pins down the migration cost threshold, $x(h_\ell, a, e, i, z, z \;)$, that a worker of type (h_ℓ, a, e) in city i, with match quality z is willing to pay to move to city $\; i$ in a job of quality $z \;$. The cutoff migration costs are given by

$$x(h_\ell, a, e, i, 0, 0) = U(h_\ell, a, e, \; i) \quad U(h_\ell, a, e, i), \tag{2.3}$$

$$x(h_\ell, a, e, i, 0, z \;) = V(h_\ell, a, e, \; i, z \;) \quad U(h_\ell, a, e, i), \tag{2.4}$$

where a value of 0 in place of z or $z \;$ is intended to represent unemployment.

The HJBE (2.5) describes the joint value of a firm-worker pair:

$$
\begin{aligned}
rV(h_\ell, a, e, i, z) = {} & zh_\ell - q^e p_i + \\
& [s_i k(G_i, h_\ell, e) + h^e \exp(- h h_\ell)][V(h_{\ell+1}) - V]\mathbb{1}\{a = y\} + y_a[V(o)\mathbb{1}\{a = y\} - V] + \\
& d^e(U - V) + l_{1,i} E_F[\max\{b(V(\hat{z}) - V), 0\}] + \\
& l_{1,-i} E_{F,D}[\max\{b(V(-i, \hat{z}) - U(-i)) + U(-i) - V - c, \\
& b(V(-i, \hat{z}) - V - c), U(-i) - V - c, 0\}].
\end{aligned} \tag{2.5}
$$

As most events are common to employed and unemployed workers, I highlight only the differences between the terms in Equation (2.5) and their counterparts in Equation (2.1). In the first line on the RHS, the flow payoff is given by the output produced by the firm-worker pair, net of the house price. The human capital accumulation process in the second line is analogous to the process for unemployed workers, except for the presence of learning by doing. The third line shows the change in value that follows an exogenous job destruction and the expected gain from searching on the job in city i. Since V is strictly increasing in z, such gain is positive if and only if the worker draws a match quality $\hat{z} > z$. The last two lines are identical to those in Equation (2.1), except for the fact that V replaces U as the current value. Therefore, the migration cost thresholds in Equations (2.3) and (2.4) are respectively replaced by

$$x(h_\ell, a, e, i, z, 0) = U(h_\ell, a, e, -i) - V(h_\ell, a, e, i, z), \tag{2.6}$$

$$x(h_\ell, a, e, i, z, z') = V(h_\ell, a, e, -i, z') - V(h_\ell, a, e, i, z). \tag{2.7}$$

Labor market and location decisions, learning, and aging induce a distribution of workers over the state space, f. The distribution determines the frequency of meetings between workers and firms (matching channel), the rate of knowledge diffusion (flow of ideas channel), the human capital distributions of cities (peer effects), and house prices.

The Kolmogorov Forward equation, KFE, (2.8) describes the law of motion of the measure of unemployed workers, $f(h`, a, e, i, 0)$:

$$
\begin{aligned}
0 = \; & s_i[k(G_i, h` \;\; _1, e) f(h` \;\; _1) \quad k(G_i, h`, e) f(h`)] 1 f\, a = \; yg + \\
& y_y f(y) 1 f\, a = \; og \quad y_a f \; + \; y_o g_{0,i}(h`, e) M_i(o) 1 f\, a = \; yg \\
& l_{0,i} f\,[1 \quad F(R)] + \; d \int_{\underline{z}}^{\bar{z}} f(\hat{z}) d\hat{z} + \\
& \frac{1}{N_i} F(R) \;\; l_{0,i} D(x(\;\; i, 0, 0)) f(\;\; i, 0) + \; l_{1,i} \int_{\underline{z}}^{\bar{z}} D(x(\;\; i, \hat{z}, 0)) f(\;\; i, \hat{z}) d\hat{z} \\
& '' \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \# \\
& l_{0, \;\; i} f \quad F(R(\;\; i)) D(x(0, 0)) + \int_{R(\;\; i)}^{\bar{z}} D(x(0, \hat{z})) dF(\hat{z}) \quad .
\end{aligned}
\tag{2.8}
$$

The LHS is the derivative of the distribution with respect to time, which is equal to 0 in a steady state. The rst line on the RHS states that learning through imitation induces an out ow of young workers of human capital $h`$ and an in ow of young workers of human capital $h` \;\; _1$. The second line shows the in ow and out ow of workers due to aging and the entry of young workers into the labor market. Newborns are distributed according to $g_{0,i}(h`, e)$. They replace old workers in city i—whose measure is denoted by $M_i(o)$—at rate y_o. The third line is related to local labor market ows. It shows the ow out of unemployment of workers who have accepted a job offer and the in ow into unemployment of workers whose jobs are destroyed. The fourth line represents the in ow of workers from city i, who are either unemployed or employed at some match quality $\hat{z}$, before moving to city i as unemployed. Such workers draw a suf- ciently low migration cost that induces them to migrate, but they also draw a match quality below the reservation value in city i. The term $1/N_i$ accounts for the fact that while all the mobility from small cities happens toward the single large city in the economy, workers in large cities are equally likely to receive an offer from any of the N small cities. Hence, $N_i = 1$ if $i = \text{large}$, while $N_i = N$ if $i = \text{small}$. The fth line describes the migration ow in the opposite direction: unemployed workers moving into city i, either as unemployed or as employed at match quality $\hat{z}$.

Equation (2.9) is the KFE for the measure of employed workers at match quality z, $f(h`, a, e, i, z)$:

$$
\begin{aligned}
0 = \; & f\,[s_i k(G_i, h_{`-1}, e) + h^e \exp( \quad hh_{`-1})]\, f\,(h_{`-1}) \quad [s_i k(G_i, h_`, e) + h^e \exp(\quad hh_`)]\, f\,(h_`)\, g 1 f\, a = \; y g + \\
& y_y f\,(y) 1 f\, a = \; o g \quad y_a f + \\
& l_{0,i} f\,(0)\, f(z) 1 f\, z \quad R g + l_{1,i} \int_{\underline{z}}^{z} f\,(\hat{z}) d\hat{z}\, f(z) \quad l_{1,i} f\,(1 \quad F(z)) \quad d^e f + \\
& \frac{1}{N_i} f(z) 1 f\, z \quad R g \;\; l_{0,i} D(x(\quad i, 0, z))\, f\,(\quad i, 0) + l_{1,i} \int_{\underline{z}}^{\bar{z}} D(x(\quad i, \hat{z}, z))\, f\,(\quad i, \hat{z}) d\hat{z} \; \# \\
& l_{1, \;\, i} f \quad F(R(\quad i))\, D(x(z, 0)) + \int_{R(\quad i)}^{\bar{z}} D(x(z, \hat{z}))\, dF(\hat{z}) \quad .
\end{aligned}
\tag{2.9}
$$

Equation (2.9) is analogous to Equation (2.8), except for few differences. First, employed workers accumulate human capital through learning by doing, not just through knowledge diffusion. Second, the measure of employed workers at match quality z increases because of both hiring from unemployment and on-the-job search from $\hat{z} < z$ (third line, rst two terms). The measure decreases because of transitions to jobs with quality above z and to unemployment (third line, last two terms). The fourth line shows the in ow of workers from city i into jobs of match quality z in city i. The last line represents the out ow of workers toward city i, either as unemployed or as employed at some match quality $\hat{z}$.

Integrating over the steady-state distribution f, we can compute the measure of workers of age a in city i, $M_i(a)$; the share of workers of age a in city i with a college degree, $COL_i(a)$; and the equilibrium human capital distribution in city i, $G_i(h_`)$. Last, the equilibrium price in the housing market is given by

$$
p_i = p_{0,i} Q_i^{g_i} = [(q^{col} COL_i + q^{hs}(1 \quad COL_i))\, M_i]^{g_i}. \tag{2.10}
$$

The equations in this section lead to the following de nition.

De nition. A stationary equilibrium is a tuple $f\, V, U, R, x, f\,, p g$ of value and policy functions, equilibrium distribution, and house prices such that i) U solves the HJBE (2.1); ii) V solves the HJBE (2.5); iii) R satis es the optimality condition (2.2), iv) x satis es the optimality conditions (2.3), (2.4); (2.6) and (2.7); v) f solves the system of KFEs (2.8)-(2.9); and vi) p satis es (2.10).

2.4 Discussion

The model includes a number of simplifying assumptions that are motivated by both data availability and computational tractability. Before turning to the quantitative analysis of the model, I brie y provide motivating evidence in support of two assumptions that are directly related to the nature of productivity differences between cities. The results in this section are based on data from the decennial Census and the

CPS. Further details are reported in Appendix B.

First, one might conjecture that heterogeneity in occupational composition across cities might play a substantial role in accounting for the city-size wage premium and its life-cycle dynamics. An intuitive test of this hypothesis consists of verifying whether the coef cient on city-size in a canonical wage regression is smaller once the worker's (3-digit) occupation is controlled for. I nd that such a drop in the correlation between wages and city size (both in level and growth rate) is, in fact, present in the data, but it is only slightly larger than the drop associated with controlling for workers' education. This nding, together with the limited size of the panel used in the estimation, lends support to the choice of a simple(r) characterization of worker heterogeneity based on education and general human capital.

Second, the learning technology features randomness in the type of workers that interact with one another within a city. While it is reasonable to believe that workers can partially direct their search for peers, I provide suggestive evidence against a perfect segregation by, say, education group. I nd that college graduates experience higher wages (and wage growth) in larger cities, but this is also the case in cities with a higher share of college graduates over total population. [13] To appreciate the role of the educational composition of cities, I compute the marginal effect on the wage of college graduates from the addition of one high school graduate to their city (not to be interpreted in a casual sense). Such in ow increases city size but reduces the college share, with an overall negative effect on the wage of college graduates in that city. Through the lens of the model, if the knowledge diffusion process were perfectly excludable across education types, the above-mentioned marginal effect would likely be positive, or null.

3 Quantitative Analysis

The model is not amenable to a closed-form solution, owing to the interaction between labor market dynamics, human capital accumulation, and location choice. To solve it numerically, I adapt the nite difference method in Achdou et al. (Forthcoming) to a spatial economy with search frictions and knowledge diffusion. After describing the data (Section 3.1) and specifying the parametric assumptions employed in the estimation (Section 3.2), I discuss the identi cation of the model parameters (Section 3.3). The model is estimated via the method of simulated moments. [14] The results of the estimation are presented in Section 3.4. In Section 3.5, I validate the model by verifying its ability to replicate, without targeting, the wage difference between movers and stayers in the years before and after migrating. Last, I decompose the observed

[13] A static version of this result, that abstracts from wage growth, can be found in Rossi-Hansberg, Sarte, and Schwartzman (2019).

[14] Newborns in the model are all unemployed, while they are employed in the data. To bring the model closer to the data, I follow an approach similar to Lise and Postel-Vinay (2020). I simulate a pre-sampling period of six months in which I shut down all events that can happen to a worker except meeting a rm in the local labor market. I then drop the pre-sampling period and consider the worker's state at the end of it as the initial condition.

wage premium into the contribution of sorting, matching, ow of ideas, and peer effects (Section 3.6).

3.1 Data

The main source of data for this paper is the NLSY79. The NLSY79 is a survey of young men and women who were between 14 and 21 years old on December 31, 1978. The survey comprises a "cross-sectional" subsample that is representative of the US population and other subsamples that target speci c demographic groups. Interviews took place annually from 1979 until 1994 and biennially thereafter. For each respondent, the NLSY79 contains information on highest educational attainment, weekly employment status, job transitions, wages, and location. In order to have a homogeneous sample and avoid dealing with issues related to labor force participation, I use only information on men from the cross-sectional subsample. Further sample restrictions involve dropping individuals who entered the labor force before January 1, 1978, which is the rst date for which labor market information is available, or have not worked for at least 20 years by 2012. I also drop individuals with a signi cant amount of missing information on education, job history or location.[15] The survey contains information on respondents' county of residence, which I uniquely assign to a commuting zone (CZ), following the methodology developed by David Dorn.[16] Because of the small number of observations associated with each CZ, I then group all the CZs in the sample into two size categories according to their total population in 1990. Each group contains CZs with populations less or more than 750,000 individuals. I refer to these groups as small and large cities, respectively. The threshold is chosen in order to guarantee both substantial heterogeneity between groups and similar group size in the NLSY79. The nal sample contains information on the labor market experience of 1,532 men, who are observed for 20 consecutive years since the month they started working after completing formal education.

3.2 Parametric Assumptions

In order to estimate the model, I impose the following parametric assumptions.

Local Labor Market. The meeting rate between unemployed workers and rms, both located in city i, is given by the constant elasticity function

$$l_{0,i} \equiv l_0(M_i) = c M_i^{c_M}, \qquad c > 0,\ c_M \in R.$$

The parameter c_M represents the degree of returns to scale in the search process.

[15] Additional details about the sample are provided in Appendix C.

[16] https://www.ddorn.net/data/Dorn_Thesis_Appendix.pdf

Match Quality Distribution. Upon meeting, a rm-worker pair samples a match quality

$$z \quad \text{Pareto}(\underline{z}, a) \qquad \underline{z} > 0, \ a > 1,$$

with cdf $F(z)$, where $\underline{z}$ is the lower bound of the support of the distribution and a is the tail coef cient. The choice of a Pareto distribution is motivated in Section 3.3.

Human Capital Accumulation . The rate of interaction between workers in city i is equal to

$$s(M_i) \quad t M_i^{t_M}, \qquad t > 0, \ t_M \ 2 \ R.$$

Analogously to c_M, t_M captures the degree of returns to scale in knowledge diffusion.

Migration Cost. Upon migrating, workers pay a cost c $\text{Logistic}(m_c, s_c)$, with cdf $D(c)$. The parameters m_c and s_c represent the mean and standard deviation of the distribution, respectively. Similarly to the Normal, the Logistic distribution is symmetric about the mean and can be described by two parameters that have a direct mapping into the characteristics of the distribution itself. In a model with endogenous migration decisions, the property of having a closed-form conditional expectation makes the Logistic distribution computationally convenient.

Initial Human Capital Distribution. Let $\hat{h}je$ $\text{LogN}(m^e, s^e)$ be a random variable with pdf $\hat{g}_0^h(\hat{h}je)$. Let p^e be the fraction of workers in the economy with education e and p^{e,e^0} be the probability that a worker with education e is replaced by one with education e^0. Newborns' human capital and education are distributed according to

$$g_{0,i}(h_{`}, e) = g_0^h(h_{`} je) g_{0,i}^e,$$

where

$$g_0^h(h_{`} je) = \frac{\hat{g}_0^h(h_{`} je)}{å_{\varsigma=1}^{L} \hat{g}_0^h(h_{\varsigma} je)}, \tag{3.1}$$

$$g_{0,i}^{col} = [p^{col,col} COL_i(o) + p^{hs,col}(1 \quad COL_i(o))] \hat{p}. \tag{3.2}$$

Equation (3.1) states that the conditional distribution of initial human capital is assumed to be approximately log-normal (Huggett, Ventura, and Yaron 2006) on a nite set of points, with parameters m^e and s^e, which are allowed to vary by education. According to Equation (3.2), for each old worker who exits the economy in city i, the education of the newborn in the same city is governed by the transition probability p^{e,e^0}. The normalizing factor $\hat{p}$ guarantees that the total fraction of college graduates in the economy is equal to p^{col}.[17]

3.3 Identi cation

The model is estimated at monthly frequency. The choice of a ne partition of workers' experience allows me to better replicate the behavior of some high-frequency events, like unemployment-to-employement and job-to-job transitions. Even though all parameters are either externally calibrated or jointly estimated from simulated data, I provide an intuitive identi cation argument that clari es how the empirical moments inform the model parameters. Overall, there are 25 internally estimated parameters, using 29 targeted moments.

Human Capital Accumulation and Aging. The rate of human capital accumulation through knowledge diffusion is given by $s(M_i)k(G_i(\tilde{h}), h, e)$, where $s(M_i) = t\, M_i^{t_M}$ and $k(G_i(\tilde{h}), h, e) = E_{G_i(\tilde{h})}[h^e_{ver} \max\{\tilde{h} - h, 0\} + h^e_{hor}\tilde{h}]$. Since k is a constant returns to scale function of the parameters h^e_{hor} and h^e_{ver}, I normalize $t = 1$. Hence, the set of parameters related to human capital accumulation is given by $\{t_M, h^{hs}_{hor}, h^{col}_{hor}, h^{hs}_{ver}, h^{col}_{ver}, h^{hs}, h^{col}, h\}$. In order to estimate these eight parameters, I de ne eight groups of workers. First, I split the sample by educational group (hs and col). Then, within each category, workers are divided into those with wages in the top and bottom half of the wage distribution in the rst year of employment. Last, at any point of the life cycle, each of the four groups is partitioned into workers located in large cities and those located in small ones.

The average growth rate of wages for workers in each group provides a moment that is used in the estimation. Learning by doing is described by the parameters h^{hs}, h^{col}, and h. Intuitively, h^{hs} and h^{col} determine the average wage growth by education group, and h reproduces the curvature of wage pro les, irrespectively of a worker's location. The knowledge diffusion technology is described by h^{hs}_{ver}, h^{col}_{ver}, h^{hs}_{hor}, and h^{col}_{hor}. Vertical imitation, h^{hs}_{ver} and h^{col}_{ver}, is identi ed from the faster growth rate of wages—heterogeneous across cities—for workers who are in the bottom half of the initial wage distribution compared with those in the top half. Horizontal imitation, h^{hs}_{hor} and h^{col}_{hor}, is pinned down by differences across cities in wage growth for workers who start in the top half of the wage distribution. The parameter t_M leverages h^e_{ver} and h^e_{hor} in determining the overall higher wage growth in large cities. Notice that the mapping between the learning parameters and wage growth is affected by the equilibrium human capital distribution in large and small cities. The better the quality of peers, the faster wage growth is, even under constant returns in meeting

[17] The normalizing factor is given by

$$\hat{p} = \frac{p^{col}\sum_i N_i M_i(o)}{\sum_i N_i M_i(o)[p^{col,col}COL_i(o) + p^{hs,col}(1 \quad COL_i(o))]}.$$

($t_M = 0$).

Since young (old) workers age (retire) at rate y_y (y_o), imposing $\frac{1}{y_y} + \frac{1}{y_o} = 40 \quad 12$ guarantees that workers retire after 40 years of work on average. The share of life spent as young, $\frac{y_o}{y_o + y_y}$, is also the time when human capital is accumulated, and it is pinned down by the ratio between the average wage in the second 10 years versus that in the first 10 years of labor market experience. Intuitively, for a given total wage growth, the ratio is higher the longer workers are young.

Local Labor Market. The job destruction rate, d^e, is taken directly from the data. It is equal to the average probability of becoming unemployed in a given month across the entire sample period. I pin down the elasticity of the meeting rate in the local labor market with respect to city size, c_M, by targeting the average wage premium between large and small cities. The intuition is that a higher meeting rate makes firm-worker pairs more selective with respect to the type of matches they are willing to form, which implies that the reservation match quality is higher in larger cities. Notice that workers' types need not be—and, in fact, they are not—the same across cities. Hence, higher match quality is responsible only for the portion of the average wage premium that is not accounted for by difference in human capital composition.

One could expect the existence of lower search frictions in larger cities to be ruled out by the lack of systematic differences in unemployment and job-to-job transitions rates along the city-size distribution. However, in an economy with homogeneous workers and a single location, Martellini and Menzio (2020) prove that if and only if $z \quad \text{Pareto}(\underline{z}, a)$, the presence of lower search frictions in a larger market, that is, $c_M > 0$, is consistent with observing the same labor market flows in cities of different size. Quantitatively, I find that the intuition in Martellini and Menzio (2020) carries through to the much richer environment of this paper. An alternative hypothesis that could also account for the observed labor market flows in large and small cities is the existence of constant returns to scale in the search process, $c_M = 0$, combined with an arbitrary match quality distribution. As I show below, if $c_M = 0$, the model understates the average wage premium in the economy, and it needs to resort to other mechanisms outside the ones proposed in this paper. More importantly, constant returns to scale would be inconsistent with direct estimates of c_M that I introduce in Section 3.4, and that I obtain from measuring the heterogeneity in the job search behavior of workers and firms along the city-size distribution. Reassuringly, the estimated positive value of c_M that allows the model to replicate the average wage premium in the economy is well in line with such direct measures.

Since c determines the level of the firm-worker meeting rate, this parameter is identified by the average unemployment rate in the economy. Conditional on the values of c and c_M, the relative efficiency of search on the job, r, is identified by the average number of jobs held by workers over the life cycle.

Match Quality Distribution and Bargaining Power. It is well known that if $z \quad \text{Pareto}(\underline{z}, a)$, $\underline{z}$ cannot be identified separately from c.[18] Hence, I normalize $\underline{z} = 1$. The slope of the distribution, a, and workers' bargaining power, b, are jointly identified by the average wage growth in a job-to-job transition and by returns to tenure at a firm. As a measure of returns to tenure, I use the wage difference between equally experienced workers who spent more—compared with less—than a fifth of their working years at a given firm (very similar results are found using other cutoffs). The intuition is that the model can match a high average wage growth in job-to-job transitions by posing either a thick tail of the match quality distribution, that is, a small a, or a high value of workers' bargaining power, b. However, while in the former case returns to tenure would be large—since the dispersion in match quality would be large—in the latter returns to tenure would be small, since workers would receive most of the job surplus upon joining the firm.

Location Choice. Workers in city i contact firms in city i at a rate that is r times as large as that of workers who live in city i. Migrating entails a cost c Logistic(m_c, s_c). The values of r, m_c, and s_c are jointly identified by relative city size, the hazard rate of moving from a small to a large city (or vice versa), and the share of workers in large cities with a college degree.[19] Since idiosyncratic migration motives—captured by s_c—are symmetric across cities, high values of s_c tend to equalize the equilibrium relative city size but also increase the probability that the realized migration cost is sufficiently low for a worker to choose to migrate. Both decreasing m_c and increasing r positively affect the migration hazard rate. However, since migration entails a lump-sum cost and college graduates earn higher income, lower values of m_c disproportionately facilitate the access of high school graduates to large cities.

Initial Human Capital Distribution. I normalize the mean of initial log-human capital for high school graduates to $m^{hs} = 0$. Given m^{hs}, m^{col} is determined by the initial mean wage gap between college and high school graduates in the economy. The standard deviations of the initial human capital distributions, s^{hs} and s^{col}, are pinned down by the inter-quartile range of initial wages (for high school and college graduates, respectively). The value of p^{col} is equal to the fraction of college graduates in the data. Using information on parental education for workers in the sample, p^{e,e^0} is given by the probability that a worker has education e^0, conditional on his father's educational attainment e.

Housing Cost. The house price function in city i is described by a level, $p_{0,i}$, and an elasticity parameter, g_i. Workers with education e consume q^e units of housing. The values of g_i are obtained from the estimates in Saiz (2010).[20] I normalize $q^{hs} = 1$ and choose values of $p_{0,i}$ and q^{col} so that each educational group has

[18] Let $F(z) = 1 \quad (\underline{z}/z)^a$. The rate at which an unemployed worker of type (h,a,e) accepts a job in city i is equal to $c M_i^{cM}[1 \quad F(R(h,a,e,i))] = c M_i^{cM}\,\underline{z}/R^a(h,a,e,i)$. Clearly, labor market flows can identify only $c\,\underline{z}$.

[19] For given relative city size and hazard rate of moving from a small to a large city, the hazard rate of moving in the opposite direction does not provide additional identifying information, since it is pinned down by the law of motion for city size, combined with the assumption of stationarity.

[20] Saiz (2010) reports the elasticity of housing supply for 269 MSAs in 2000. I assign to g_{large} (g_{small}) the population-weighted inverse elasticity for the 65 (204) MSAs with population above (below) 750 thousand. Splitting the sample in Saiz (2010) according to the size of the commuting zones that mostly overlap with those MSAs—when such CZs can be clearly identified, as is the case for virtually all

the average expenditure share of non-tradeable goods computed by Moretti (2013).

Remaining Parameters. The monthly discount rate, r, is set to 1.25%, which corresponds to an annual value of 15%, as in Herkenhoff et al. (2018). This value is higher than what is usually adopted in traditional macroeconomic models. In contrast to standard concave utility functions, a linear utility has an in nite elasticity of substitution. Hence, it accomodates wide uctuations in ow payoffs, which might lead workers to temporarily accept negative wages in exchange for signi cant wage growth. Therefore, the value of r in this environment is meant to capture not only the rate of time preference but also, at least in part, the inverse of the workers' intertemporal elasticity of substition.

The parameters b_i are chosen to match a ow value from unemployment equal to 60% of the average wage in city i. This target lies within the (somewhat wide) range of traditional estimates, such as 40% in Shimer (2005) and 70% in Hall and Milgrom (2008).

Last, the (relative) number of small cities, N, is equal to 9.5, which is the ratio between the number of small cities and the number of large ones in the 1990 US Census.

3.4 Estimation Results

3.4.1 Parameter Values

In this section, I discuss how the key estimated parameters shed light on the proposed mechanisms behind the city-size wage premium and workers' location choice. The full list of parameters is reported in Table D.1 in Appendix.

The degree of returns to scale in the technology of knowledge diffusion, t_M, is equal to 0.067. This value implies that at the equilibrium relative city size, $M_{large}/M_{small} = 15.7$, workers in large cities experience $15.7^{t_M} = 1.20$ times more interactions with one another than those in small cities. A higher frequency of meetings in large cities leverages an equilibrium distribution of peers that rst order stochastically dominates the distribution in small cities. The relative importance of vertical and horizontal imitation is embodied in the parameters h^e_{ver} and h^e_{hor}. The estimation implies that the human capital of a college graduate at the 75th percentile of the human capital distribution in large cities grows at 2.8% per year, compared with 1.8% for a worker in the same position of the human capital distribution of a small city. The same comparison at the 25th percentile is 6% versus 4.5%. Notice that this is a conservative representation of the heterogeneity in learning opportunity across cities, since for any quantile of the distribution, the associated human capital value is higher in large cities.[21] A similar pattern applies to high school graduates, though

large cities—generates almost identical results.

[21]For example, the human capital of a worker at the 25th percentile of the small city distribution would increase at 9.6% per year if he were in a large city.

both h^{hs}_{ver} and h^{hs}_{hor} take lower values than h^{col}_{ver} and h^{col}_{hor}, respectively. College graduates also accumulate more human capital than high school graduates irrespective of their location, owing to a superior learning-by-doing technology ($h^{col} > h^{hs}$). Recall that learning is a prerogative of young workers. According to the estimated value of $y_o / (y_y + y_o) = 0.46$, workers learn during their first 0.46 40 = 18.4 years of working life on average.

Turning to the characteristics of the search-and-matching technology, I find an elasticity of the contact rate between workers and firms with respect to city size, c_M, equal to 0.2, which corresponds to increasing returns to scale ($c_M = 0$ under constant returns). The value of c_M suggests that the frequency of meetings between workers and firms, both located in a large city, is $15.7^{c_M} = 1.73$ times higher than the meeting rate inside a small city. At the same time, the estimated tail coefficient of the Pareto distribution of match quality is $a = 3.6$, which implies that a match at the 75th percentile of the distribution is 36% more productive than a match at the 25th percentile. As explained in the identification section, lower search frictions leverage the dispersion in match quality and allow for the formation of more productive jobs. Recall that not all the productivity advantages of creating better matches are passed on to the workers through higher wages, since workers capture an estimated $b = 20\%$ of the gains from trade.

The degree of returns to scale in the search technology is estimated by targeting the average wage premium between small and large cities at the equilibrium level of sorting on human capital. Therefore, it is of particular interest to compare the estimated value of c_M with direct evidence on the heterogeneity in search frictions across cities. Interpreting a firm-worker contact in the model as an application to an open vacancy in the data, I compare the estimated value of c_M with two empirical measures of the same statistic. First, Martellini and Menzio (2020) report an estimate of the elasticity of the number of applications per vacancy received by firms with respect to the population of the commuting zone where the firm is located. The estimated elasticity, provided to us by Ioana Marinescu, is equal to 0.52. On the other side of the labor market, the 1982 wave of the NLSY79 collected information on the most recent job search experience of all the workerswho were employed at the moment they were surveyed. Using information on workers' commuting zones and the number of employers they had contacted divided by the number of weeks they had been looking for a job, I compute an elasticity of 0.12. To the best of my knowledge, this estimate is new to the literature. Interestingly, the value of $c_M = 0.2$ obtained from the model lies within the range of direct estimates derived from the worker, 0.12, and the firm side, 0.52.

With regard to location choice, the estimated value of r implies that workers contact firms in a different city at a rate that is 8% of the within-city meeting rate. The average migration cost, m_c, is equal to about two months of median income in the economy. In their seminal paper on interstate migration, Kennan and Walker (2011) estimate a much larger average cost equal to more than $ 300,000. Beyond comparing two different types of location choice, an additional difference between their paper and mine is that I replace a frictionless labor market with one characterized by search frictions. As pointed out by Schmutz and Sidibé (2019), the presence of search frictions greatly reduces the size of the migration cost that is necessary in order to replicate the observed mobility rate. Besides, the model in this paper replicates, without targeting, the average wage gain of about 9% experienced by workers upon changing location.[22] [23]

3.4.2 Model Fit

Table D.2 in Appendix reports the model-generated moments next to their empirical counterparts. The model ts the data well along all dimensions.

Through the lens of the model, the characteristics of the knowledge diffusion process generate variations in wage growth across locations, educational category, and relative position in the initial wage distribution. In line with the existence of vertical imitation, college graduates with low initial wages experience faster wage growth in large cities, 6.3% per year, than in small cities, 4.7%. Learning from better peers also bene ts college graduates with high initial wages, who experience a 4% wage growth per year, against 3% in small cities. Even though high school graduates are characterized by a much atter experience gradient, they display a similar pattern in terms of variations in wage growth across cities and initial wages (see Figure 8 in Appendix).

The average wage growth in a job-to-job transition, which is key to identify the tail of the match quality distribution, is closely matched at about 11%, and it is also very similar across cities. In addition, the model successfully replicates the declining experience pro le of job-to-job transitions, or EE rate (Figure 1). After "climbing the job ladder" in early career stages, the probability of further improving on the current match declines, and so does the frequency of job-to-job transitions. Both in the model and in the data, the EE rate is remarkably similar across cities of different size. This nding is entirely consistent with the existence of lower search frictions in large cities. On the one hand, employed workers in larger cities contact rms more frequently, contributing to a higher observed number of transitions. On the other hand, the endogenous distribution of employed workers over match qualities is also better in large cities, so that, conditional on a meeting, workers have a lower probability of improving on their current match. Under a Pareto match quality distribution, these two opposing forces cancel each other out, delivering a pattern of job-to-job

[22]I compute the percentage wage change in the three months after migration compared with the three months before. The average value in the model is 9%, which is remarkably close to the empirical 9.1%. However, in the data, moving from a small to a large city is associated with a wage gain virtually identical to the one from moving in the opposite direction. In the model, wage gains are equal to 16% and 6%, respectively. An asymmetric mobility cost, or a systematic preference for living in large cities, would help close this gap.

[23]The reason why the estimates in Kennan and Walker (2011) are also consistent with the observed wage gain at migration is that, contrary to this paper, they allow for heterogeneity in location-speci c preferences across people. They show that correcting for the utility gain from living in a preferred location generates actual moving costs paid by movers that are close to zero, if not slightly negative

Figure 1. Job-to-Job Transition Rate

Job-to-job transition rate within large (thick blue line) and small (red thin line) cities, in the model (solid line) and in the data (dotted line).

transitions that is line with the data. A similar intuition can explain why the unemployment rate is virtually identical in large and small cities (6%): the equilibrium increase in the reservation match quality in large cities prevents the higher meeting rate between rms and workers from turning into a higher transition rate out of unemployment.

Turning to the location choice, we rst nd that the average hazard rates of moving from a small to a large city and vice versa are equal to 0.31% and 0.18% per month, respectively. The life-cycle pro le of migration patterns is depicted in Figure 2. The model is consistent with the fact that migration ows into large cities are considerably larger than those in the opposite direction. Large cities offer two main advantages: a higher rate of human capital accumulation and lower search frictions in the labor market. Furthermore, the model replicates—without targeting—the steady decline in migration rates over the life cycle: learning happens only at a "young" age, and older workers—often employed in good matches—are less likely to start a new job, including in a different city.

Importantly for the study of the aggregate implications of local policy, the model closely matches the empirical relative city size. The large city in the model represents US commuting zones with a population of more than 750,000 people in 1990 and an average of 2.2 million, compared with 140,000 for small cities. The model also replicates the spatial distribution of education. The share of college graduates in large cities is tightly matched at 29%, against 19% in small cities. College graduates bene t relatively more from the

Figure 2. Migration Hazard Rate

Hazard rate of moving out of a large (thick blue line) and a small (thin red line) city, in the model (solid line) and in the data (dotted line).

presence of better peers and, because of production complementarities between human capital and match quality, from lower search frictions.

3.5 Model Validation: Migration Episodes

In the previous section, I showed that the model is able to account for the labor market experience of workers in small and large cities. The estimation highlights the existence of increasing returns to scale in labor search and, to a lesser extent, in the rate of knowledge diffusion. Furthermore, I show that larger cities provide superior learning opportunities due to better peers.

Notice that the wage of movers is not explicitly targeted in the estimation, which uses only information on aggregate wage differences between large and small cities. In addition, the identifying information is overwhelmingly provided by workers who never move from a small to a large city nor vice versa (78% of the NLSY79 sample). To make this point more transparent, I re-compute the targeted moments on the subsample of workers who never move, both in the data and in the model. The empirical moments obtained from the set of non-movers are very close to their counterparts computed on the entire sample and are also remarkably similar to those generated by non-movers in the model (see Appendix D). Yet small and large cities are not isolated from one another, and some workers do in fact change location over the life cycle. Therefore, it is natural to ask whether the migration incentives generated by the matching and knowledge diffusion channels are consistent with the observed labor market experience of movers. In this section, I test the model's ability to replicate the relative wage of movers with respect to both stayers in the pre-migration city and incumbents in the destination one.

To do so, I apply the methodology introduced in the context of job displacements by Jacobson, LaLonde, and Sullivan (1993) to the study of migration episodes between small and large cities. I estimate the following specification:

$$\log(w_{it}) = \beta X_{it} + \sum_{k=-10^{+},..,-1,1,..,10^{+}} \delta_k d_{itk} + \varepsilon_{it}.^{24} \tag{3.3}$$

The dependent variable is the inflation-adjusted log-hourly wage of worker i in year t. The controls include a quadratic term in experience, an indicator for college education, and year dummies. The key coefficients of interest are given by the vector of δ_k associated to the dummy variables d_{itk}, which are equal to 1 if worker i in time t is in his k^{th} year after migrating to a city of different size. Negative values of k stand for years before migration takes place.[25] I run four different versions of Regression 3.3, which are distinguished by the type of city a worker migrates from and by the selected control group, that is, the set of observations where $d_{itk} = 0$. In Figure 3, I plot the coefficients generated from the NLSY79 sample (green circles) and those obtained using model-generated data (orange diamonds).

The top left shows the wage of workers who move from a large into a small city, compared with the wage of those who remain in large cities. Leaving a large city is associated with lower pre-migration wages and a much larger, but substantially flat, wage drop in the years after moving into a small city. The top right panel repeats the same exercise with respect to the opposite migration flow. Migrants do not particularly differ from stayers until the moving year. After that, the wage of those who move to large cities starts to diverge, and it reaches a 28% premium after 10 years or more. Except for somewhat overestimating the extent of sorting out of small cities, the predictions from the model are remarkably close to the empirical evidence over the entire time window around migration episodes.

Recall that the estimation targets the aggregate wage difference between small and large cities. Since the model is able to replicate the relationship between the wages of movers and stayers, it should not be surprising that the model-based comparison between movers and incumbents is fairly in line with the data (bottom panels of Figure 3). In the bottom left panel, I show that migrants to small cities earn somewhat less than the incumbents, even if they were earning significantly more while working in large cities. This finding supports the existence of a productivity benefit from being in large cities that workers lose upon migrating

[24] Glaeser and Maré (2001) introduced a similar regression in the context of migration into and out of urban areas. One of the contributions of this paper is providing a micro-founded equilibrium theory that is empirically consistent with such reduced-form evidence.

[25] The value of $\delta_{10^{+}}$ ($\delta_{-10^{+}}$) stands for 10 or more years after (before) the migration episode.

Figure 3. Wage of Movers and Non-Movers

Wage of movers vs. stayers (top panels) and vs. incumbents (bottom panels). Left: from small city. Right: from big city. Data: green circles. Model: orange diamonds.

out. From the bottom right panel, we observe how future migrants out of small cities earn less than those who already live in large cities and continue to do so even after migration takes place. This evidence is in line with the idea that most of the bene t from living in a large city does not accrue on impact. Consistently with the dynamic return to experience in large cities, the gap with respect to the incumbents slowly shrinks in the 10 years or more after migration.

Figure 4. Job Tenure and Migration

Wage of movers to small cities vs stayers in large cities. Dark stars: baseline. Light triangles: control for tenure. Left: data. Right: model.

In both the data and the model, there exists a positive correlation between job tenure and wages. Recent migrants are relatively new to their job, hence part of the discrepancy between their wage and non-movers' is, in fact, due to heterogeneity in job tenure. To illustrate this point, I augment Regression 3.3 with a quadratic term in the tenure at the current job, and compute new values for the coef cients d_k. The updated plots—in Appendix D—highlight how the results differ from the baseline speci cation, but they do so in a very similar fashion in the model as in the data. As an example, here I report the comparison between movers to small cities and stayers in large cities (Figure 4). The left panel shows the coef cients estimated on NLSY79 data. While the relative wage loss after leaving a large city is essentially at in the baseline regression (dark green stars), it is increasing over the years, when tenure is controlled for (light green triangles). A very similar pattern emerges from the model (right panel, orange shades). The intuition is simple. Since recent movers have relatively little tenure at their current job, controlling for job tenure reduces the size of the wage gap with respect to stayers, but only in the years right after migration. In subsequent years, the gap between stayers and movers gradually increases, as the latter no longer bene ts from the superior knowledge diffusion in large cities.

3.6 Quantifying the Sources of the Wage Premium: Decomposition

In the previous sections, I estimated the model by matching the labor market outcomes of workers in large and small cities, and I validated it with respect to the wage of movers in the years before and after migration. I now quantify the contribution to the city-size wage premium over the life-cycle of matching, knowledge diffusion, and sorting.

The decomposition proceeds as follows. I set the aggregate characteristics of cities at their baseline equilibrium values, treating them as fixed parameters, and I remove the possibility of migration between cities. I define $Q_i = \{l_{0,i}, l_{1,i}, s_i, G_i\}$, for $i = \{small, large\}$, to be the set of aggregate characteristics of city i in the baseline economy, and I create four scenarios s, each denoted by $Q^s_i = \{l^s_{0,i}, l^s_{1,i}, s^s_i, G^s_i\}$.[26] Because of the assumption of no migration across cities, the policy functions in city i depend only on workers' idiosyncratic states (and match quality, if employed) and on Q^s_i, but they do not depend on Q^s_{-i}. In other words, large and small cities are treated as separate economies. In all scenarios, I set $Q^s_{small} = Q_{small}$, and alternately vary the characteristics of large cities, Q^s_{big}. The thought experiment consists of comparing the life-cycle city-size wage premium in a world in which only one of the aggregate characteristics of large cities differs from its counterpart in small cities.

In the first scenario, I isolate the contribution of sorting. I solve the decision problem of workers and firms and simulate life-cycle wage paths under the parametrization $Q^{NS}_{big} = Q_{big}$. I label this scenario no sorting, NS, as the only departures from the baseline economy are imposing the same initial distribution of workers across cities—in terms of both human capital and education—and removing all migration flows between them. The left panel of Figure 5 shows the city-size wage premium in the data (black dotted line), the baseline economy (thick black line), and the no sorting scenario (thin blue line). The vertical distance between the baseline and the no sorting scenario captures the contribution of sorting over the life cycle.

The remaining city-size wage premium is represented by the blue thin line reported on both panels of Figure 5, and it is due to heterogeneity in the equilibrium characteristics of small and large cities. The red dotted line on the right panel of Figure 5 corresponds to an economy in which only the matching channel is operative, $Q^M_{large} = \{l_{0,large}, l_{1,large}, s_{small}, G_{small}\}$. The pink dashed and green dashed-dotted lines repeat the same exercise with regard to the flow of ideas, $Q^F_{large} = \{l_{0,small}, l_{1,small}, s_{large}, G_{small}\}$, and peer effects, $Q^P_{large} = \{l_{0,small}, l_{1,small}, s_{small}, G_{large}\}$, respectively.[27]

[26] I omit house prices from the aggregate variables that characterize a city. If migration is ruled out, house prices are equivalent to a lump-sum tax paid by every worker in the economy—although potentially different across cities—but they have no impact on wages.

[27] In the flow of ideas and peer effect scenarios, I also set the value of b in large cities equal to its value in small cities. Recall that the flow value from unemployment in city i is $b_i h$. In the model, wages are determined by the human capital of the worker and the match quality of the job. Given that the value of b_i aims at replicating a flow value of unemployment that is a constant fraction of wages, it is natural to set it equal across cities in those scenarios in which the only systematic difference between workers in small and large cities is in terms of human capital.

Figure 5. Decomposition of the City-Size Wage Premium

Percentage wage premium. Left panel: data (black dotted line), baseline model (black thick line), `no sorting' scenario (blue thin line). Right panel: `no sorting' (blue thin line), `matching' (red dotted line), ` ow of ideas' (pink dashed line), `peer effect' (green dashed-dotted line) scenario.

Starting from the left panel of Figure 5, we observe how the contribution of sorting slowly rises with experience, and it reaches 12% after 20 years. I further isolate the contribution of sorting on observable education, up to 7%, from sorting on unobservable human capital, up to 5%. To perform this additional decomposition, not shown in the graphs, I re-simulate the no sorting scenario, weighting the observations in each city and at each experience level according to the experience-speci c educational composition of cities in the baseline economy. The modest role of sorting on human capital is consistent with the ndings by, among others, Eeckhout, Pinheiro, and Schmidheiny (2014) and De La Roca and Puga (2017). While in the static environment in Eeckhout, Pinheiro, and Schmidheiny (2014) lack of sorting is due to production complementarities between workers at the top and the bottom of the skill distribution, in this paper, it is the result of two opposing forces. On the one hand, high-skill workers bene t more from locating in large cities, because of complementarity between human capital and match quality. On the other hand, low-skill workers have more room for learning in large cities, thanks to vertical imitation in the process of knowledge diffusion.

The right panel of Figure 5 shows that matching has a level effect on the wage premium of about 10 pp, but it plays no role in the growth of the premium over the life cycle. Once again, if the distribution of match qualities is Pareto, increasing returns to scale in the search process generate better matches at any level of experience. At the same time, a Pareto distribution is the only one that is consistent with increasing returns to scale, even in the absence of size-dependent differences in unemployment and job-to-job transition rates. To the contrary, the ow of ideas and peer effects are negligible for newborns, as human capital accumulation occurs over time, but they are responsible for the entire growth of the wage premium that is not accounted for by sorting, up to 15 pp, with peer effects being by far the predominant force. [28]

It is worth commenting on two features of the decomposition presented in this section. First, the aggregate variables in the economy are not allowed to respond, in equilibrium, to changes in workers' decisions: they are treated as xed parameters. The decomposition described above is meant to highlight the role of city characteristics—at their equilibrium value—in generating the city-size wage premium and not the role of increasing returns and knowledge diffusion in shaping the characteristics of cities themselves. Second, besides isolating the role of sorting, the absence of migration ows is also required in order to correctly measure the contribution of any speci c treatment (matching, ow of ideas, and peer effects). To see why, consider the matching scenario. If workers were allowed to migrate, the resulting wage premium would necessarily represent a combination of both the actual object of interest—that is, higher average match quality in large cities—and the sorting behavior induced by matching itself.

4 Endogenous Market Tightness

The stationary equilibrium of the model includes an exogenous speci cation for the rate at which a worker contacts a rm in the local labor market and across cities, l and l , respectively. In evaluating the aggregate consequences of local policies, it is natural to allow rms to adapt their behavior to changes in the pro tability of creating vacancies in different cities. In order to accommodate this margin, I explicitly introduce a competitive market for vacancies, in the tradition of Mortensen and Pissarides (1994).

I assume that the number m_i of meetings between a rm and an unemployed worker, both located in city i, is given by the product between $M_i^{c_M}$ and a constant returns to scale function m,

$$m_i = M_i^{c_M} m(s_i, v_i) = M_i^{c_M} s_i^{1 \; z} v_i^{z},$$

where v_i is the number of vacancies, and s_i is the actual measure of workers seeking jobs in city i. Hence, the meeting rate between a rm and an employed worker when both are located in city i is equal to $r\, m_i$, while it is equal to $r \; m_i$ ($rr \; m_i$) if the unemployed (employed) worker is in city i at the moment of the

[28] Recall that ow output is equal to $h \cdot z$. A residual wage premium of about 1 pp is due to the interaction between matching and knowledge diffusion.

contact. It follows that

$$s_i = [u_i + r(1 \quad u_i)]M_i + r \quad [(u_{\;i} + r(1 \quad u_{\;i}))]M_{\;i}, \tag{4.1}$$

where u_i is the unemployment rate in city i.

Since search is random and m has constant returns to scale, an unemployed worker in city i contacts a rm, also in city i, at rate

$$\frac{m_i}{s_i} = M_i^{c_M} q^z, \tag{4.2}$$

where $q = v_i / s_i$ is assumed to be the same across cities. From the reduced-form speci cation introduced in Section 3, the contact rate (4.2) is also equal to $_{0,i} = M_i^{c_M} c$. Petrongolo and Pissarides (2001) document a wide range of empirical estimates of z and suggest a value of approximately 0.5, which allows me to recover the value of $q = c^{\frac{1}{z}}$. Last, the vacancy creation cost, k_i, is pinned down by the zero pro t condition in the market for vacancies

$$k_i = M_i^{c_M} q^{z \quad 1}(1 \quad b)S(i), \tag{4.3}$$

where $S(i)$ is the expected job surplus from contacting a worker and $(1 \quad b)$ is the share that accrues to the rm. The expected surplus can be easily obtained combining the joint value of matches V and unemployment U, both described in Section 2. A formal derivation of $S(i)$ is provided in Appendix A.2.

Applying Equation (4.3) to both large and small cities, I nd $k_{large}/ k_{small} = 1.6$. Because of increasing returns to scale in the search process and a better human capital composition, rms in large cities are willing to pay a 60% higher vacancy creation cost. A higher value of k reduces rms' incentives to create vacancies in a given city, in a fashion similar to how higher house prices represent a congestion force in workers' location choice.

5 Housing Policy and Aggregate Productivity

Extensive research has investigated the sources of the remarkable dispersion in productivity across US cities. Perhaps less explored is how the nature of such spatial heterogeneity is central to evaluating the aggregate implications of local policies that trigger the relocation of workers across cities and, in particular, toward more productive ones. I illustrate this point by focusing on a type of policy that has been at the forefront of the academic and political debate in recent years: relaxing housing regulation in some large US cities. In Section 5.1, I describe the policy experiment and show how this paper is related to the existing literature on this topic. In Section 5.2, I propose an identity that allows the decomposition of the change in total income after the policy implementation into pre-policy observables and equilibrium responses of local productivity in small and large cities. Section 5.3 presents the quantitative results.

5.1 Housing Regulation: Background

City-level (inverse) elasticities of housing supply, g_i, have been estimated by Saiz (2010). Averaged across large and small cities, they are equal to 1/1.47 and 1/2.46, respectively. A growing literature has documented that part of the higher inverse elasticity in some large US cities—the fact that house prices grow more rapidly as the city expands—is not due to physical constraints but is the outcome of tighter land use regulation.[29]

Hsieh and Moretti (2019), henceforth HM, explore the aggregate implications of strict housing regulation in the most productive (large) US MSAs. They nd that if house price dispersion in 2009 had been as low as it was in 1964, the increase in productivity of some large cities would have triggered a much higher increase in employment, and the growth rate of US GDP would have been twice as large during the same time period. In their model, housing regulation leads to higher house prices, which prevent productive cities from expanding and generating higher GDP growth. Herkenhoff, Ohanian, and Prescott (2018), henceforth HOP, develop a neoclassical growth model with multiple regions. They nd that GDP would be 12% higher if land use regulation in all US states moved halfway toward that of the least restrictive state, Texas. According to their estimates, California and New York—where some of the largest US cities are located—are among the most restrictive states. Both HM and HOP model spatial heterogeneity in productivity as an exogenous difference in TFP between locations.[30] In addition, they impose homogeneity between workers and abstract from life-cycle considerations. Parkhomenko (2018) builds a model with sorting on ability and endogenous regulation. He shows that equalizing housing regulation across cities reduces the extent of sorting in large cities, but the ensuing output gains are generated exclusively by the relocation of workers toward exogenously more productive locations. [31]

This paper introduces two new margins to the debate. First, exogenous heterogeneity in productivity is replaced by agglomeration forces in the form of increasing returns to scale in the labor market and in the process of knowledge diffusion. Second, in line with the increase in the city-size wage premium over the life cycle, part of the productivity differential between large and small cities is dynamic and stems from differences in the rate of human capital accumulation. I also allow for heterogeneity in human capital between workers to contribute to the process of knowledge diffusion in cities through spatial sorting. All these as-

[29] Saiz (2010) estimates the regulation's contribution to the housing supply elasticity of 269 US cities. Glaeser and Gyourko (2018) compute the magnitude of house prices in excess of construction costs in US MSAs and nd that it is negatively correlated with the number of building permits issued in the same MSA. See also the discussion in Glaeser, Gyourko, and Saks (2005) and Gyourko, Saiz, and Summers (2008).

[30] HOP consider an extension of their model in which TFP is an exogenous increasing function of local output, but they do not take a stand on the foundations of such agglomeration force.

[31] The model in Parkhomenko (2018) includes constant elasticity production externalities that depend on the size but not the composition of a city, in the spirit of Kline and Moretti (2014). This type of externality has no aggregate implications, since moving a worker from one location to another reduces the output of the rst by the same amount by much it increases the output of the latter.

pects—increasing returns, knowledge diffusion, sorting—are endogenous to the policy environment. While this paper exhibits a more stylized geography than the traditional urban literature, the new margins considered in this paper aim at complementing the current debate on the aggregate implications of housing and, more generally, place-based policies.[32]

5.2 Interpreting the Policy Outcome

I introduce an identity that highlights the role of endogenous local productivity in shaping the equilibrium response to a policy change. Throughout the analysis, I focus on the steady-state total income in the economy before and after the implementation of the policy, assuming a working life of 40 years for all workers.

I denote by j^0 (j^1) the value of any variable j before (after) the policy implementation and by $\bar{j}$ the mean of j. Let $DM = M^1_{big} - M^0_{big} = N(M^0_{small} - M^1_{small})$ be the equilibrium change in the measure of workers located in large cities. The difference between total income, denoted by Y, after, compared with before, the change in policy is equal to

$$\begin{aligned} Y^1 - Y^0 &= [M^1_{big}\bar{y}^1_{big} + NM^1_{small}\bar{y}^1_{small}] - [M^0_{big}\bar{y}^0_{big} + NM^0_{small}\bar{y}^0_{small}] \\ &= [(M^0_{big} + DM)\bar{y}^1_{big} + (NM^0_{small} - DM)\bar{y}^1_{small}] - [M^0_{big}\bar{y}^0_{big} + NM^0_{small}\bar{y}^0_{small}] \\ &= DM(\bar{y}^1_{big} - \bar{y}^1_{small}) + M^0_{big}(\bar{y}^1_{big} - \bar{y}^0_{big}) + NM^0_{small}(\bar{y}^1_{small} - \bar{y}^0_{small}), \end{aligned} \tag{5.1}$$

where y_i represents income in city i. It is easy to see that if average income in small and large cities is policy invariant, that is, $\bar{y}^1_{big} = \bar{y}^0_{big}$ and $\bar{y}^1_{small} = \bar{y}^0_{small}$, the last line of Identity (5.1) simplifies to

$$Y^1 - Y^0 = DM(\bar{y}^0_{big} - \bar{y}^0_{small}). \tag{5.2}$$

It follows that the change in total income is given by the average productivity gap currently present in the data, multiplied by the additional measure of workers located in large cities after the policy implementation. I define the term on the RHS of Identity (5.2) as the direct effect (D).

In the general case, in which such invariance is not satisfied, the last line of Identity (5.1) can be further

[32] HM and Parkhomenko (2018) model an economy with more than 200 MSAs, while HOP consider seven groups of states. One advantage of those environments is the ability to account for richer heterogeneity in productivity and house prices across cities (or between states)

re-arranged as

$$Y^1 - Y^0 = \underbrace{\Delta M(\bar{y}^0_{big} - \bar{y}^0_{small})}_{(D)} + \underbrace{M(\bar{y}^1_{small} - \bar{y}^0_{small})}_{(S)} + \underbrace{M^1_{big}[(\bar{y}^1_{big} - \bar{y}^1_{small}) - (\bar{y}^0_{big} - \bar{y}^0_{small})]}_{(P)} . \quad (5.3)$$

In order to interpret the terms in Identity (5.3), we can think of the income produced by a worker as being the sum of two components: the income he would produce in a small city and, for a worker located in a large city, the income gap between large and small cities. Hence, the change in total income can be decomposed into the sum of the direct effect (D) and two new terms. First, $M(\bar{y}^1_{small} - \bar{y}^0_{small})$ is the change in the average income of small cities multiplied by the entire population in the economy, (S). Second, $M^1_{big}[(\bar{y}^1_{big} - \bar{y}^1_{small}) - (\bar{y}^0_{big} - \bar{y}^0_{small})]$ is the change in the city-size income premium, multiplied by the size of large cities after the policy, (P).

5.3 Results

I consider the steady-state equilibrium response of the economy to a 1% and a 2% increase in the elasticity of housing supply in large cities, while I leave the same parameter in small cities unchanged. Specifically, I replace $g_{large} = -1/1.474$ with $g_{large} = \{-1/1.489, -1/1.504\}$. Intuitively, under looser housing regulation, large cities gain in population: they become 8.8% and 17.3% larger—well within the range considered in existing work. The increase in size is accompanied by a change in their educational and human capital composition. The share of workers in large cities who hold a college degree drops from 29% in the baseline economy to 28% and 27% as g_{large} shrinks. The equilibrium human capital distributions deteriorate in both large and small cities. In Figure 6, I show the deviation of those distributions from the distributions in the baseline economy. The green line with circles (purple line with triangles) is associated with a 1% (2%) change in policy. In both the left and the right panel—large and small cities, respectively—a less restrictive housing policy causes a leftward shift of the human capital distribution. This finding can be explained by the fact that workers who live in large cities under lower housing restrictions are more skilled than those who remain in small cities, but less skilled than those who live in large cities in the baseline economy. Such change in sorting further affects human capital accumulation through its effect on the knowledge diffusion process within cities. Next, I explore the aggregate implications of the change in city size and composition, as measured by variations in total income.

Figure 6. Housing Regulation and Human Capital

Difference between the human capital distribution under $Dg_{big} =$ 0.01% (green line with circles) and $Dg_{big} =$ 0.02% (purple line with triangles), with respect to its value in the baseline economy. Left panel: large city. Right panel: small city.

The outcome of the housing policy can be analyzed by using Identity (5.4),

$$\underbrace{\frac{Y^1 - Y^0}{Y^0}}_{\text{Total}} = \underbrace{\frac{DM(\bar{y}^0_{big} - \bar{y}^0_{small})}{Y^0}}_{(D)} + \underbrace{\frac{M(\bar{y}^1_{small} - \bar{y}^0_{small})}{Y^0}}_{(S)} + \underbrace{\frac{M^1_{big}[(\bar{y}^1_{big} - \bar{y}^1_{small}) - (\bar{y}^0_{big} - \bar{y}^0_{small})]}{Y^0}}_{(P)}, \qquad (5.4)$$

which includes the same terms as Identity (5.3), each divided by Y^0 in order to obtain percentage deviations from the baseline. Figure 7 shows the overall effect (black bars) and its components (grey bars). The direct effect is equal to +1.8% (+3.6%), while the total effect is +0.6% (+1.1%), when g_{large} is lowered by 1% (2%). The direct effect (D) is clearly positive: since large cities are more productive, their expansion increases total income. However, because of endogenous changes in the productivity of cities, the other two terms create a wedge between the direct and the total effect of the policy. As previously mentioned, the composition of small cities deteriorates compared with the baseline economy, and so do peer effects in those cities. In addition, increasing returns to scale have an adverse effect on the average match quality and on the frequency of interactions in a city that shrinks in size. All these forces contribute to the negative sign of the second term, (S). Because of weaker sorting and worse peers, income in large cities declines as well. The contribution of agglomeration forces (i.e. matching and flow of ideas) is positive in a city that expands. Since such forces do not fully offset the human capital loss in large cities, the income premium (P) between large and small cities is slightly lower in the after-policy—compared with the pre-policy—equilibrium.

Figure 7. Housing Regulation and Aggregate Income Gains

Percentage change in total income (black bar), and in its components (grey bars). Left panel: $Dg_{big} =$ 1%. Right panel: $Dg_{big} =$ 2%.

All in all, even accounting for the equilibrium response of agglomeration forces, peer effects, and sorting, relaxing housing restrictions in large cities does generate non-negligible income gains. Yet, assuming that local productivity was policy invariant would overstate the percentage increase in total income by a factor of 3.

6 Conclusions

The geography of economic activity in the US is characterized by signi cant and persistent productivity differentials between cities, particularly between large and small ones. In this paper, I build a model in which this productivity gap emerges as an equilibrium outcome due to spatial sorting, increasing returns to scale in the labor search process, and heterogeneity in the (endogenous) composition of peers. The model replicates the observed labor market experience of workers in large and small cities, and it is tested with respect to its predictions on the experience of movers in the years before and after a change of location.

In the presence of spatial dispersion in productivity, a recent debate has drawn attention to the aggregate implications of local policies that restrict workers' access to large and productive cities. By performing a policy experiment in which housing regulation in large cities is relaxed, I show that micro-founding the nature of local productivity is key in order to evaluate the policy outcome. While increasing housing supply in large cities would further expand their size, it would also change their composition and the productivity bene t from living in those cities, mainly through a weakening of sorting and peer effects. As a result, assuming that local productivity was invariant to the policy would overstate the actual income gain by a factor of 3.

This paper leaves at least three important open questions for future research. First, I abstract from rm heterogeneity and rm sorting across locations. Through the lens of this paper, a high-quality rm is one with better workers, hence better peers. The distinction between worker composition and exogenous rm types is salient insofar as rms' location decisions might feature a different elasticity to policy changes compared with their workers'. Second, I model human capital as being one-dimensional, with workers always bene ting from interacting with one another and increasing their stock of knowledge. In reality, cities are heterogeneously specialized in different occupations, which presumably require a certain degree of occupation-speci c human capital. Capturing this richer heterogeneity while preserving the dynamic structure of the model is a challenging task, but it could provide insights into the differential impact of technological and policy innovations across cities and sectors of the economy. Third, the geography in this paper is restricted to the binary state of living or not living in a given city. At a smaller scale, the internal structure of a city is also likely to affect the frequency and quality of interactions that occur between workers and rms and the opportunities for the exchange of knowledge. Further improving our understanding of the nature of local interactions will help inform the internal design of cities, land use regulation, and transportation infrastructures.

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. Forthcoming. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach". Review of Economic Studies.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. 2018. "Dancing with the Stars: Innovation through Interactions". NBER Working Paper No. 24466.

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin. 2014. "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics". American Economic Review 104 (6): 1551–1596.

Baum-Snow, Nathaniel, and Ronni Pavan. 2012. "Understanding the City Size Wage Gap". Review of Economic Studies 79 (1): 88–127.

Behrens, Kristian, Gilles Duranton, and Frederic Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration". Journal of Political Economy 122 (3): 507–553.

Buera, Francisco J., and Ezra Oberfield. 2020. "The Global Diffusion of Ideas". Econometrica 88 (1): 83–114.

Burdett, Kenneth, and Dale T Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment". International Economic Review 39 (2): 257–273.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin. 2006. "Wage Bargaining with On-the-Job Search: Theory and Evidence". Econometrica 74 (2): 323–364.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2011. "The Identification of Agglomeration Economies". Journal of Economic Geography 11 (2): 253–266.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sebastien Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection". Econometrica 80 (6): 2543–2594.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum. 2019. "Matching in Cities". Working Paper.

Davis, Donald R, and Jonathan I Dingel. 2019. "A Spatial Knowledge Economy". American Economic Review 109 (1): 153–170.

De La Roca, Jorge, Gianmarco I P Ottaviano, and Diego Puga. 2014. "City of Dreams". Working Paper.

De La Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities". Review of Economic Studies 84 (1): 106–142.

Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies". In Handbook of Regional and Urban Economics, edited by J. V. Henderson and J. F. Thisse, 4:2063–2117.

Eckert, Fabian, Mads Hejlesen, and Conor Walsh. 2020. "The Return to Big City Experience: Evidence from Refugees in Denmark". Working Paper.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014. "Spatial Sorting". Journal of Political Economy 122 (3): 554–620.

Fogli, Alessandra, and Veronica Guerrieri. 2019. "The End of the American Dream? Inequality and Segregation in US cities". NBER Working Paper No. 26143.

Glaeser, Edward L. 1999. "Learning in Cities". Journal of Urban Economics 46 (2): 254–277.

Glaeser, Edward L., and Joseph Gyourko. 2018. "The Economic Implications of Housing Supply". Journal of Economic Perspectives 32 (1): 3–30.

Glaeser, Edward L., Joseph Gyourko, and Raven Saks. 2005. "Why Is Manhattan So Expensive? Regulation and the Rise in Housing Prices". Journal of Law and Economics 48 (2): 331–369.

Glaeser, Edward L., and David Maré. 2001. "Cities and Skills". Journal of Labor Economics 19 (2): 316–342.

Gould, Eric D. 2007. "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium". Review of Economic Studies 74 (2): 477–506.

Gyourko, Joseph, Albert Saiz, and Anita Summers. 2008. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index". Urban Studies 45 (3): 693–729.

Hall, Robert E, and Paul R Milgrom. 2008. "The Limited In uence of Unemployment on the Wage Bargain". American Economic Review 98 (4): 1653–1674.

Herkenhoff, Kyle F., Jeremy Lise, Guido Menzio, and Gordon M. Phillips. 2018. "Production and Learning in Teams". NBER Working Paper No. 25179.

Herkenhoff, Kyle F., Lee E. Ohanian, and Edward C Prescott. 2018. "Tarnishing the Golden and Empire States: Land-use Restrictions and the U.S. Economic Slowdown". Journal of Monetary Economics 93:89–109.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation". American Economics Journal: Macroeconomics 11 (2): 1–39.

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2006. "Human Capital and Earnings Distribution Dynamics". Journal of Monetary Economics 53 (2): 265–290.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers". American Economic Review 83 (4): 685–709.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations". Quarterly Journal of Economics 108 (3): 577–598.

Jarosch, Gregor, Ezra Ober eld, and Esteban Rossi-Hansberg. Forthcoming. "Learning from Coworkers". Econometrica

Kennan, John, and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions". Econometrica 79 (1): 211–251.

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

Lise, Jeremy, and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation". American Economic Review 110 (8): 2328–2376.

Lucas, Robert E. Jr. 2004. "Life Earnings and Rural-Urban Migration". *Journal of Political Economy* 112 (S1): S29–S59.

Lucas, Robert E. Jr, and Benjamin Moll. 2014. "Knowledge Growth and the Allocation of Time". *Journal of Political Economy* 122 (1): 1–51.

Luttmer, Erzo G. J. 2007. "Selection, Growth, and the Size Distribution of Firms". *Quarterly Journal of Economics* 122 (3): 1103–1144.

Martellini, Paolo, and Guido Menzio. 2020. "Declining Search Frictions, Unemployment and Growth". *Journal of Political Economy* 128 (12): 4387–4437.

Menzio, Guido, and Shouyong Shi. 2011. "Efficient Search on the Job and the Business Cycle". *Journal of Political Economy* 119 (3): 468–510.

Menzio, Guido, Irina A Telyukova, and Ludo Visschers. 2016. "Directed Search over the Life Cycle". *Review of Economic Dynamics* 19:38–62.

Moretti, Enrico. 2013. "Real Wage Inequality". *American Economic Journal: Applied Economics* 5 (1): 65–103.

Mortensen, Dale T, and Christopher A Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment". *Review of Economic Studies* 61 (3): 397–415.

Parkhomenko, Andrii. 2018. "The Rise of Housing Supply Regulation in the U.S.: Local Causes and Aggregate Implications". *Working Paper*.

Perla, Jesse, and Christopher Tonetti. 2014. "Equilibrium Imitation and Growth". *Journal of Political Economy* 122 (1): 52–76.

Perla, Jesse, Christopher Tonetti, and Michael Waugh. 2021. "Equilibrium Technology Diffusion, Trade, and Growth". *Accepted at American Economic Review* 111 (1): 73–128.

Petrongolo, Barbara, and Christopher Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function". *Journal of Economic Literature* 39 (2): 390–431.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity". *Econometrica* 70 (6): 2295–2350.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman. 2019. "Cognitive Hubs and Spatial Redistribution". NBER Working Paper 26267.

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

Schmutz, Benoît, and Modibo Sidibé. 2019. "Frictional Labour Mobility". *Review of Economic Studies* 86 (4): 1779–1826.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies". *American Economic Review* 95 (1): 25–49.

Appendix

A Additional Details on Equilibrium Values

A.1 The Value of Wage Contracts

Let $W(h`, a, e, i, z, w)$ be the present discounted value of a worker of human capital $h`$, age a, education e, who lives in city i, is employed at match quality z, and is being paid a piece rate w. It is insightful to highlight the discrepancy between W and the joint value of the match at which the worker is employed, V. To this aim, the HJBE that characterizes the value W can be written as

$$\begin{aligned} rW(h`, a, e, i, z, w) = {} & rV \quad (1 \quad w)zh` + \\ & [s_i k(G_i, h`, e) + h^e exp( \quad hh`)][(W(h`_{+1}) \quad W) \quad (V(h`_{+1}) \quad V)]1 f a = yg + \\ & y_h[(W(o) \quad V(o))1 f a = yg \quad (W \quad V)] + d^e(V \quad W) + W(i) + W \ (i), \end{aligned} \tag{A.1}$$

where it is understood that individual state variables are omitted from the RHS if they take the same value as on the LHS.

The LHS is the annuitized value of W. The first line on the RHS has two terms. The first term is the annuitized joint value from the worker's current match. The second term captures the fact that the worker only receives a fraction w of the flow product $zh`$. The second line shows the change in value due to human capital accumulation, net of the change in the joint value of the match that is already included in the term rV. An analogous intuition applies to the aging process described by the first term on the third line. The joint value of a match incorporates the event of job destruction, but workers only lose the portion of the match value that they were receiving. Since firms lose $(V \quad W)$ when the job is destroyed, $d^e(V \quad W)$ needs to be added back to the worker's value. The last two terms represent the additional value the worker obtains when receiving outside offers and/or moving.

The value $W(i)$ is given by

$$W(i) = l_{1,i} \left" (V \quad W)(1 \quad F(z)) + \int_{z^f(w)}^{z} (V(\hat{z}) \quad W) dF(\hat{z}) \right\# .$$

If a worker transitions to a new job inside the same city, i.e. to a job with match quality $\hat{z} > z$, he receives the portion of the value of the old match that was accruing to the previous employer before the workers' job-to-job transition (first term). This is because the previous employer is always willing to deliver at most the joint value of the match in order to retain the worker, while competing with the poaching firm. The second term shows the gain in value from receiving an outside offer, inside the same city, which triggers a wage renegotiation, but not a job-to-job transition. This event occurs when the worker and the new potential employer draw a match quality $\hat{z} < z$ such that $V(\grave{h}, a, e, i, \hat{z})$ is higher than the current present discounted value to the worker, $W(\grave{h}, a, e, i, z, w)$. The minimum match quality that triggers a wage increase is given by the value $z^f(\grave{h}, a, e, i, z, w)$, which is implicitly defined by

$$V(\grave{h}, a, e, i, z^f(\grave{h}, a, e, i, z, w)) = W(\grave{h}, a, e, i, z, w).$$

A worker who draws a match quality $z^f < \hat{z} < z$ receives the joint value of the new potential match, $V(\hat{z})$. This assumption is slightly different from Bagger et al. (2014), who use the bargaining protocol in Cahuc, Postel-Vinay, and Robin (2006). In Bagger et al. (2014), the worker would additionally receive a fraction b of the difference between the joint value of his current match and the new potential one.

The last term in Equation (A.1) shows the gain from meeting firms in city $-i$, and it is equal to

$$\begin{aligned} W_{-i}(i) = {} & \lambda_{1,-i} \mathbb{E}_D[(U(-i) - W)1\{c < x(z,0)\} + (U(-i) - W - c)1\{x(z,0) < c < x(z,0,w)\}]F(R(-i)) + \\ & \lambda_{1,-i} \int_{R_{-i}}^{\bar{z}} \mathbb{E}_D[(V - W)1\{c < x(z,\hat{z})\} + (V(-i,\hat{z}) - W - c)1\{x(z,\hat{z}) < c < x(z,\hat{z},w)\}]dF(\hat{z}). \end{aligned} \tag{A.2}$$

The first line on the RHS of Equation (A.2) represents the situation in which the worker draws a match quality below the reservation value in city $-i$. The first term inside the expectation describes the case in which the worker draws a migration cost $c < x(z, 0)$, hence he moves to city $-i$, and obtains the additional value $U(-i) - W$. The second term encompasses those draws that are both higher than the migration threshold, and sufficiently low that the worker can credibly threaten to quit and move unless his promised value is increased. The highest such cost, which equates the value from quitting to the current promised value, is equal to

$$x(\grave{h}, a, e, i, z, 0, w) = U(\grave{h}, a, e, -i) - W(\grave{h}, a, e, i, z, w).$$

The last two lines follow the same logic described above, applied to the circumstances in which the new potential match quality $\hat{z}$ lies above the reservation value in city $-i$. In this case, the maximum moving cost such that the threat to leave the current employer is credible, and the piece rate is renegotiated, is given by

$$x(h_\ell, a, e, i, z, \hat{z}, w) = V(h_\ell, a, e, -i, \hat{z}) - W(h_\ell, a, e, i, z, w).$$

A.2 Firm Expected Surplus

Firms in city i receive a fraction $(1-\beta)$ of the expected surplus $S(i)$ of creating a match with a worker. The expectation is taken with respect to the type of worker the firm meets, the match quality sampled by the firm-worker pair, and the migration cost—if the worker moves from city $-i$. I define the surplus of a match $S(h_\ell, a, e, i', z', i, z) = V(h_\ell, a, e, i, z) - V(h_\ell, a, e, i', z')$, where $i' \in \{i, -i\}$ and a value of $z' = 0$ stands for the case in which the worker is hired from unemployment. The expected surplus in Equation (4.3) is then given by

$$\begin{aligned} S(i) = \sum_{h_\ell, a, e} \int_{R_i}^{\bar{z}} \Big[& f(i,0) S(i,0,i,\hat{z}) + \mathbb{E}_{f(i,\tilde{z})}[S(i,\tilde{z},i,\hat{z})]\rho + \\ & f(-i,0)\mathbb{E}_D[S(i,0,i,\hat{z}) \mathbb{1}\{c < x(-i,0,i,\hat{z})\} + \\ & (S(-i,0,i,\hat{z}) - c)\mathbb{1}\{x(-i,0,i,0) < c < x(-i,0,i,\hat{z})\}]\rho + \\ & \mathbb{E}_{f(-i,\tilde{z})}[\mathbb{E}_D[S(i,0,i,\hat{z}) \mathbb{1}\{c < x(-i,\tilde{z},i,0)\} + \\ & (S(-i,\tilde{z},i,\hat{z}) - c)\mathbb{1}\{x(-i,\tilde{z},i,0) < c < x(-i,\tilde{z},i,\hat{z})\}]]\rho\rho \Big] dF(\hat{z}).[33] \end{aligned} \tag{A.3}$$

The first line on the RHS corresponds to the surplus from hiring an unemployed (first term) or employed (second term) worker in the local labor market. Recall that the mass of unemployed workers in city i is equal to $f(i,0)$, while the equilibrium distribution of employed workers over match quality $\tilde{z}$ is $f(i,\tilde{z})$. The parameter ρ in the second term captures the relative search efficiency of employed workers. The second and third line report the expected surplus from hiring a currently unemployed worker from city $-i$. If the workers draws a migration cost $c < x(-i,0,i,0)$, his bargaining threat point is the value of unemployment in city i. That is, from the point of view of bargaining with the firm, his migration cost is sunk. If instead $x(-i,0,i,0) < c < x(-i,0,i,\hat{z})$, the worker's outside option consists in not moving, and the migration is factored into the bargaining process. The parameter ρ captures the relative search efficiency across cities. The interpretation of the last two lines follows the same logic as the two lines above, applied to hiring currently employed workers from city $-i$, who are distributed over match qualities according to $f(-i,\tilde{z})$.

[33]Once again, the state variables (h, a, e) are omitted from S and x.

B Motivating Evidence on Model Assumptions

B.1 Occupational and Educational Composition of Cities

I compute the size of the wage premium that can be accounted for by compositional differences between cities, and, specifically, by the over-representation in large cities of those workers with characteristics—i.e. education or occupation—that are correlated to higher wage growth.

In regression B.1, I project worker's i wage on indicators for living in a large city c, being a college graduate, and working in a given 3-digit Census occupation, each interacted with the worker's age $a_i \in$ {21-25,26-30,31-35,36-40}. Table B.1 reports the estimated values of β_{large}, which are the key objects of interest. Column (1) shows the estimates of a regression without controls, or, equivalently, in which I restrict $\beta_{college,a} = \beta_{occ,a} = 0$. Column (2) and (3) allow either $\beta_{college,a}$ or $\beta_{occ,a}$ to be different from zero, respectively. The main takeaway from this exercise is that the contribution of occupational sorting to the wage premium (reflected in the smaller coefficient β_{large} in Column (3) compared to Column (1)) is only slightly larger than that of education (Column (2) vs Column (1)), which is explicitly modeled in this paper.

$$log(w_i) = \beta_0 + \beta_{large,a}\mathbb{1}\{c_i \text{ is large}\} \times a_i + \beta_{col,a}\mathbb{1}\{\text{college}\}_i \times a_i + \beta_{occ,a}\mathbb{1}\{\text{occupation}\}_i \times a_i + \epsilon_i \quad \text{(B.1)}$$

Table B.1. *Educational and Occupational Composition of Cities*

Source: decennial Census. In order to mimic the NLSY79 sample, I consider male workers that are between age 20-30 (Census 1980), age 26-35 (Census 1990), age 31-40 (Census 2000). An indicator variable for each Census year is included in the regression. The sample is further restricted to workers earning at least $5k per year and $1 per hour, working at least 10 hours per week and 45 weeks per year. I also exclude movers in 'this or last year' for whom it is not possible to assign a CZ. All the standard errors (not reported) are of order 1e-3 and the p-values<0.001.

$\beta_{large,a}$		(1)	(2)	(3)
large city	age 20-25	0.08	0.07	0.06
	age 26-30	0.18	0.15	0.13
	age 31-35	0.24	0.18	0.16
	age 36-40	0.26	0.19	0.17
control for education		N	Y	N
control for occupation		N	N	Y
R^2		0.090	0.158	0.268
n. of obs.		2.1M	2.1M	2.1M

B.2 City Size and City Composition

Some might wonder whether the random nature of the knowledge diffusion process presented in this paper should be replaced by a view in which college graduates *a*) only learn from each other and *b*) are perfectly able to locate each other within a city. In such a scenario, the wage—and especially wage growth—experienced by college graduates would be unaffected by the number of high school graduates that live in their same city. To test this hypothesis, in B.2 I regress the wage of a college graduate, indexed by i, on city size, $log(pop_{c_i})$, and on the fraction of city population with a college degree, $log(col\ share_{c_i})$. Both variables are interacted with the worker's age a_i 2 {21-25,26-30,31-35,36-40}.

The results in Table B.2 show that college graduates experience higher wage (growth) in larger cities, but even more so in cities in which there is a high concentration of college graduates. Notice that $col\ share_{c_i}$ is equal to the ratio between the sheer number of college graduates and city size. Hence, college graduates of age a earn $(b_{pop,a} \quad b_{col\ share,a})$% higher wages in a city that is 1% larger, but has the same number of college graduates. It is easy to see that $b_{pop,a} \quad b_{col\ share,a}$ is *always negative and strictly decreasing in a*. While it is reasonable to think that workers are able to partially direct their search for peers, the findings in this section speak in favor of the existence of a random component in the knowledge diffusion process.

$$log(w_i) = b_0 + b_{pop,a} log(pop_{c_i}) \quad a_i + b_{col\ share,a} log(col\ share_{c_i}) \quad a_i + e_i \tag{B.2}$$

Table B.2. *City Size and City Composition*

Source: decennial Census. In order to mimic the NLSY79 sample, I consider male workers that are between age 20-30 (Census 1980), age 26-35 (Census 1990), age 31-40 (Census 2000). An indicator variable for each Census year is included in the regression. The sample is further restricted to workers earning at least \$5k per year and \$1 per hour, working at least 10 hours per week and 45 weeks per year. I also exclude movers in 'this or last year' for whom it is not possible to assign a CZ. Last, I restrict the sample to college graduates only.

		$b_{pop,a}$			$b_{col\ share,a}$
log(pop)	age 20-24	0.032 (0.007)	log(col share)	age 20-24	0.045 (0.032)
	age 25-29	0.055 (0.003)		age 25-29	0.081 (0.016)
	age 30-34	0.068 (0.004)		age 30-34	0.141 (0.025)
	age 35-39	0.070 (0.005)		age 35-39	0.186 (0.034)
n. of obs.		357684	$R^2 = 0.124$		

C Data

The representative cross-sectional sample of the NLSY79 includes 3003 men. I exclude workers that never entered the labor force, were not employed for more than 5 consecutive years, or were already in the labor force when they started being surveyed. Workers enter the labor force the first quarter in which they spend 390 hours either employed or unemployed, where the number of hours spent as unemployed is equal to the number of weeks of unemployment multiplied by 20. Once they enter the labor force, workers enter the estimation sample the first time they are employed. In order to keep a balanced panel, but also avoid issues of non-random sample selection, I keep only workers that by 2012 had completed 20 full years since entering the estimation sample.

I build a monthly panel by sampling the interview week, whenever available, and the third week of the month, otherwise. For each monthly observation, I observe labor market status, working hours, hourly wage, and whether the worker experienced a change of employer. I keep only wage observations that are associated to jobs at which workers spend at least 10 hours per week. I consider workers as still employed at their last job if they are observed not to be working for a certain period of time, but then return to their last employer. In fact, it would be hard to justify the existence of search frictions and unknown match quality with the most recent past employer.

Location information, in the form of county of residence, is available at interview dates, and between interviews during 1978-1982 and after 2000. When location is not observed, I adopt the following assignment procedure. I assume that workers stay at their current location for the entire spell of a job, and I assign each job to the modal location in case I observe more than one location for the same job. I assign each county to a commuting zone (CZ) using the cross-walk provided by David Dorn.[34] I assign an observation to the CZ observed in the previous month if such CZ is the same as the CZ where the worker lives one or two months after the missing observation. The remaining periods with missing location information are split into two parts: the first half is assigned to the last observed CZ and the second half to the first CZ observed after the period with missing information. The underlying assumption is that the actual migration date is uniformly distributed over the period in which location information is not available. Last, I assign CZs to the 'large city' category if they have population of more than 750 thousand in 1990, so guaranteeing both substantial heterogeneity in size between large and small cities, and a comparable number of observations from each category in the NLSY79.

The final sample has 240 monthly observations for each of the 386 (1146) workers with a college (high school) degree.

[34] https://www.ddorn.net/data.htm

D Additional results

Table D.1. *Parameter Values*

Internally Estimated	Description	Value
s_{hs}	inital hc hs, std	0.28
m^{col}, s^{col}	inital hc col, mean and std	0.3, 0.47
c	meeting rate	0.07
c_M	returns to scale matching	0.20
r	relative search eff. on the job	0.4
r	relative search eff. across cities	0.08
m_q	migration cost (mean)	2 median monthly wage
s_q	migration cost (std)	1.6 median monthly wage
a	shape match quality distrib	3.6
b	workers' bargaining power	0.2
t_M	returns to scale knowledge	0.07
h_h^{hs}, h_h^{col}	horizontal imitation	0, 0.0014
h_v^{hs}, h_v^{col}	vertical imitation	0.003, 0.007
h^{hs}, h^{col}, h	learning by doing	0.005, 0.008, 0.850
$y_o/(y_y + y_o)$	share of young	0.46
$p_{0,large}$, $p_{0,small}$	house price (intercept)	4.6e-5, 0.007
p_0^{col}	housing demand (col vs hs)	1.68
b_{large}, b_{small}	flow payoff unempl.	(1.27,1.13) $\underline{z}$
Externally Calibrated	**Description**	**Value**
m_{hs}	inital hc hs, mean	0
d^{hs}, d^{col}	job destruction	0.0135, 0.005
r	discount rate	0.0125
p^{col}	college share	0.2521
$p^{hs,col}$, $p^{col,col}$	intergen. transition rate to college	0.175, 0.62
g_{large}, g_{small}	(inverse) house supply elasticity	1.47^{1}, 2.46^{1}
N	number of small cities	9.5
$1/y_y + 1/y_o$	average life span	40 12 months

Table D.2. *Model Fit: Empirical and Model-generated Moments*

Moment (large/small)	Data	Model
Labor Market		
unemp. rate (%)	5.8/5.9	5.8/5.9
n. of jobs	6.9/7.2	7.1/7.2
mean wage gap (%)	29.3	29.2
EE wage growth (%)	10.7/11.8	10.7/11.2
return to tenure (%)	23.7	25.2
UI replacement rate (%)	60	61.6/60
Cities		
city size	15.7	15.6
college share (%)	29/19	29/20
hazard rate (small ! big, %)	0.31	0.31
hazard rate (big ! small, %)	0.18	0.17
exp share non-trad. (col, %)	58	55/55
exp share non-trad. (hs, %)	58	58/57
Wage growth (%)		
bottom-col-large	6.3	6.7
bottom-col-small	4.7	4.6
top-col-large	4.0	4.0
top-col-small	3.1	3.0
bottom-hs-large	4.3	4.4
bottom-hs-small	3.6	3.0
top-hs-large	2.6	2.7
top-hs-small	1.3	1.8
wage 11-20 vs. 1-10 yr. of exp.	30.6	30.4
Initial Wage		
col wage premium (%)	38.7	37.3
$75^{th}/25^{th}$ pctile, hs	1.63	1.62
$75^{th}/25^{th}$ pctile, col	1.88	1.90

Table D.3. *Non-Movers Sample: Empirical and Model-generated Moments*

Targeted Moments Computed on the Set of Non-Movers

Moment (large/small)	Data	Model
Labor Market		
unemp. rate (%)	5.1/5.4	5.0/5.0
n. of jobs	6.38/6.14	6.63/6.47
mean wage gap (%)	32.7	34.1
EE wage growth (%)	11/11.8	10.9/11.5
return to tenure (%)	22.5	25.8
college share (%)	31/19	31/20
Wage growth (%)		
bottom-col-large	6.6	6.8
bottom-col-small	4.8	4.6
top-col-large	4.0	4.0
top-col-small	2.5	3.0
bottom-hs-large	4.3	4.4
bottom-hs-small	3.8	3.1
top-hs-large	2.5	2.8
top-hs-small	1.3	2.0
wage 11-20 vs. 1-10 yr. of exp.	30.6	32.0
Initial Wage		
col wage premium (%)	41	38
$75^{th}/25^{th}$ pctile, hs	1.65	1.64
$75^{th}/25^{th}$ pctile, col	1.86	1.95

Figure 8. *Life-Cycle Wage Profiles in Large and Small Cities*

Wage profile of workers in the sample according to their education (top vs. bottom panels), city (left vs. right), and position in the wage distribution in the first year of employment (top vs. bottom lines inside each subplot). Data: dotted line. Model: solid line. The shaded area represents the 95% empirical bootstrap confidence interval.