The City-Size Wage Premium:
Origins and Aggregate Implications [*]

Paolo Martellini[†]

University of Pennsylvania

Job Market Paper

November 19, 2019

Abstract

Why do workers earn higher wages in larger cities, and why does the city-size wage premium increase over the life cycle? What are the aggregate implications of spatial wage differentials? To answer these questions, I introduce the first dynamic equilibrium model of knowledge diffusion and labor market frictions in cities. I estimate the model by matching aggregate differences in labor market outcomes between small and large US cities. To validate the model, I verify its performance in replicating micro evidence on selection into—and the return to—migration. I find that knowledge diffusion becomes increasingly important over the life cycle, as it accounts for up to forty percent of the wage premium at twenty years of labor market experience. Higher firm-worker match quality has a level effect of about a third of the average wage premium. The residual wage premium is explained by life-cycle sorting on observable education and unobservable human capital. I then show that the optimal spatial allocation of workers displays less dispersion in city size and stronger educational sorting than the laissez-faire equilibrium, due to heterogeneity across workers in the ability to generate knowledge spillovers. Last, I use my framework to study the consequences of relaxing housing regulation in large cities. I show how the resulting relocation of workers across cities affects local productivity. A hypothetical scenario in which productivity is policy-invariant overstates the magnitude of the equilibrium income gains by more than a factor of 2.5.

[†]Contact: paoloma@sas.upenn.edu.

1   Introduction

Why do workers in larger cities earn higher wages? What are the aggregate implications of spatial wage differentials? Figure 1 shows the average hourly wage of workers who live in large cities (blue thick line) and in small cities (red thin line) plotted against labor market experience.[1] It is easy to observe the existence of a city-size wage premium that is equal to about 15% at labor market entry, and grows up to 38% after 20 years. This paper is about understanding the origins
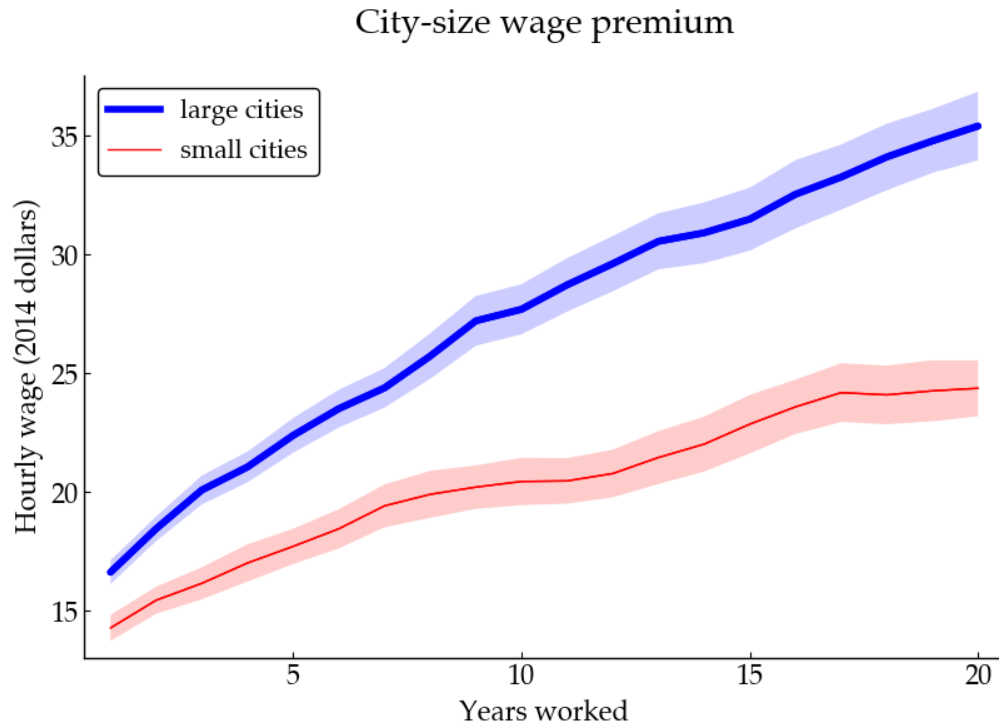
City-size wage premium

Figure 1: Data from the National Longitudinal Survey of Youth 1979 (NLSY79). Large cities (pop. $> 750k$, thick blue). Small cities (pop. $< 750k$, thin red). Shaded areas: 95% bootstrap confidence interval.

and aggregate implications of the city-size wage premium documented in Figure 1. In particular, I consider three possible sources of heterogeneity between small and large cities, that I embed into a spatial equilibrium life-cycle model of workers' location choice and wage growth.

First, I treat each city as a local labor market characterized by search frictions, and I allow the rate at which workers sample job offers to be affected by city size. The argument for the existence of non-constant returns to scale is that the concentration of a large number of workers and firms inside a narrow geographic unit, like a city, reduces the transportation and information frictions that are associated with the process of job search. If firm-worker matches are heterogeneous, sam-

---

[1]Throughout the paper, large (small) cities refer to commuting zones with more (less) than 750 thousand people in 1990.

pling jobs at a faster rate generates more selectivity in the quality of accepted jobs, hence higher wages.

The second mechanism I analyze is the advantage of large cities in promoting the exchange of ideas between workers. Geographic proximity promotes the transmission of those forms of knowledge that require direct interactions between people. In the model, this intuition takes the form of a human capital accumulation process that varies with the worker's location. I allow workers in large cities to experience more frequent interactions with each other, and, to the extent that large cities host a greater fraction of high-skilled workers in equilibrium, to learn from better peers. Stochastic aging, from young to old, ends the period of life during which workers learn.

Third, the city-size wage premium may be driven by sorting on observed education—which determines learning ability—and on unobserved human capital. Sorting occurs through endogenous migration decisions upon receiving a job offer from another city. Differences in the propensity to migrate are generated by heterogeneity, across workers, in the benefit from forming better matches and being exposed to a better process of knowledge diffusion. The productivity benefits of living in a large city are traded off against higher house prices, which operate as a congestion force that constraints the equilibrium relative city size.

To measure the importance of these three channels, I use information on workers' wage profiles and labor market flows in small and large cities. The parameters of the knowledge diffusion technology are identified by the difference in wage *growth* within large and small cities, for given worker's education and ranking in the wage distribution at the time of labor market entry. I impose a restriction on the match quality distribution such that arbitrary returns to scale in the search process can be consistent with the observed lack of variation in unemployment and job-to-job transition rates across cities. Under such restriction, increasing returns to scale lead to the formation of proportionally better matches 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. Sorting is determined by the optimal mobility choice of workers, given the identified difference in knowledge diffusion and search speed in large and small cities.

The model is estimated using a panel of workers from the National Longitudinal Survey of Youth started in 1979 (NLSY79). Workers in the sample are observed for the first 20 consecutive years since labor market entry. 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 cities. Similarly, exchanges of ideas through interactions among workers are 20% more frequent in large cities. I refer to increasing returns in the search process and in the frequency of interactions as agglomeration forces. In addition, I find that higher rates of knowledge diffusion in large cities leverage an equilibrium human capital distribution of peers that first order stochas-

tically dominates the distribution in small cities. I show that the diffusion of ideas is particularly beneficial to college graduates and, conditional on education, to workers with lower initial levels of human capital.

The model is validated according to its ability to reproduce the micro evidence on the wage of movers with respect to stayers in the pre-migration city, and incumbents in the destination one. The wage of movers are not esplictly used in the estimation of the model, which only targets aggregate differences between wages in large and small cities. Since 78% of workers never move from a small to a large city, nor vice versa, non-movers provide the vast majority of the information that identifies the model parameters.[2] I find that the model quantitatively replicates the steeper wage path of workers who move to large cities, compared to stayers. Newcomers into large cities earn significantly less than incumbents, although the difference partially shrinks with experience. With respect to the opposite migration flow, 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 stayers in large cities. Furthermore, the post-migration wage of movers is not significantly different from the wage of incumbents in small cities.

Controlling for the quality of the current match—proxied by job tenure—affects the comparison between movers, stayers, and incumbents, but it does so in a remarkably similar fashion in the model as in the data. The contribution of heterogeneity in job tenure in accounting for the wage dynamics, and selection into migration, highlights the importance of modeling a frictional labor market with match-specific job quality. In addition, I provide external validation for the estimated elasticity of the firm-worker meeting rate with respect to city size, which is equal to 20% in the model. Empirically, I interpret the firm-worker meeting rate as the average number of applications per vacancy per unit of time. The model estimate lies well inside the range of direct elasticities obtained from the search behavior of firms, that I documented in previous work (52%), and novel evidence on the heterogeneity in the number of job applications across cities, obtained from workers in the NLSY79 (12%).

To answer the first research question, I decompose the life-cycle city-size wage premium into the contribution of increasing returns to search, knowledge diffusion, and sorting. I show that, because of increasing returns to search, wages in large cities are 11% higher than in small cities, but this channel has only a level effect on the city-size wage premium. To the contrary, sorting and knowledge diffusion become increasingly important for more experienced workers, and they generate a 12% and 15% wage premium after 20 years, respectively. More than three quarters of

---

[2]I verify that the moments obtained from the set of non-movers are almost identical to their counterparts computed on the entire sample, and are also remarkably similar to those generated by non-movers in the model. Re-estimating the model on the sample of non-movers delivers very similar parameter values. See the discussion in Section 3.5.

the contribution of knowledge diffusion is due to heterogeneity in peer effects across cities, while differences in the rate at which ideas are exchanged between workers play a much smaller role. The decomposition highlights a sharp qualitative difference between the life-cycle contribution of lower search frictions and higher learning rates to the city-size wage premium. In fact, while in principle it is reasonable to expect lower search frictions to also have a growth effect, through faster search on the job, I show that this conjecture would be inconsistent with the empirical evidence on labor market flows.

In the second part of the paper, I explore the aggregate implications of the estimated heterogeneity in productivity between small and large cities.

I study the equilibrium response of the economy to an exogenous increase in the housing supply elasticity of large cities—for example, due to the relaxation of land-use regulation. The existence of vast and persistent wage differentials across cities has raised the question of whether place-based policies might improve aggregate outcomes by triggering the relocation of workers toward more productive places. Specifically, the existence of local land-use regulation that constraints the amount of available housing in some of the most productive large US cities has spurred a recent academic and policy debate about the potential gains associated with the relaxion of such constraints. Although the geography in this paper is admittedly more stylized than in the traditional urban literature, I bring two new features to the debate on this topic: endogenous agglomeration forces and dynamic benefits from experience in large cities. I highlight the importance of these margins by contrasting the equilibrium response to a change in policy, with an alternative scenario in which productivity was an exogenous characteristic of cities. I show that such alternative scenario overstates the percentage growth in total labor income by more than a factor of 2.5. As large cities grow in size, due to an expansion in housing supply, the presence of increasing returns to scale in the search process and in the rate of knowledge diffusion leads to higher productivity in those cities. However, this gain is more than offset by a reduction in the extent of sorting of high-skilled workers, which, in turn, is responsible for the deterioration in the quality of peers. In equilibrium, large cities experience a decline in productivity (and wages). At the same time, small cities are also negatively affected by this policy. As they shrink, they suffer from both the outmigration of their high-ability workers, and the reversal of agglomeration forces.

Last, motivated by the existence of agglomeration and knowledge spillover externalities, I compute and characterize the constrained-efficient allocation of workers across cities and jobs. In the optimal allocation, large cities shrink in size, but have a much higher concentration of college graduates than in the equilibrium. The solution to the planner's problem can be implemented using a set of flow transfers, indexed by workers' characteristics (and location), and financed through lump-sum taxes. Under the optimal policy, college graduates receive a bigger transfer

if they locate in large rather than small cities, while the opposite is true for high school graduates. The externalities in this paper are endogenous and dynamic in nature: workers who learn from others generate a stronger externality themselves. This force favors the agglomeration of workers with high human capital and high learning ability—i.e. college graduates—into large cities. High school graduates have a negative impact on peer effects, but they do not particularly gain from interacting with other workers. Hence, the planner would subsidize their relocation to small cities.

Related Literature

The first research question in this paper is closely related to the literature that studies the origins of spatial wage differentials.[3] Glaeser and Maré (2001) document the existence of a urban-rural wage premium in the US, which has both a level and a growth component. The empirical literature that followed has highlighted the presence of spatial sorting (Combes, Duranton, and Gobillon (2008), in the context of France) and faster wage growth associated to longer tenure in large cities (De La Roca and Puga (2016), using a sample of Spanish workers).[4]

There have been examples of structural models that attempt to formalize and discipline the various mechanisms that could account for the observed spatial wage heterogeneity.[5] Davis and Dingel (2019) assume that workers can divide their time between working and interacting with others. Large cities emerge as the location in which high-ability workers cluster in order to share their knowledge. Combes et al. (2012) estimate that agglomeration economies, in contrast to selection due to stronger competition, are responsible for the higher productivity of large cities, but they do not take a stand on the sources of such agglomeration forces.[6] While these papers address the cross-sectional heterogeneity between cities, their static nature makes them silent with respect to the life-cycle profile of the wage premium.

To the best of my knowledge, this paper introduces the first equilibrium model with dynamic

---

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

[4]Eckert, Hejlesen, and Walsh (2019) address the endogeneity of workers' initial location decision by considering a sample of refugees that were randomly assigned to Danish cities. They find that, while entry-level wages do not differ across space, accumulating experience in Copenhaghen significantly increases the probability of working in high-skill occupations and high-wage firms.

[5]Combes, Duranton, and Gobillon (2010) review the issues involved in the identification of the city-size wage premium using reduced-form specifications, in particular with regard to the difficulties 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.

[6]See also Behrens, Duranton, and Robert-Nicoud (2014), who jointly model sorting, selection, and agglomeration in order to account for variations in productivity between and within cities.

knowledge diffusion in cities that can be empirically estimated. The mechanism I adopt is related to the theoretical contribution by Glaeser (1999). Glaeser (1999) builds a two-period model, where homogeneous young workers learn from the (skilled) old. By allowing for heterogeneity in human capital and learning ability in a quantitative life-cycle model, this paper can speak to the evidence on selection into migration, and on the short and medium run return to a change in location.[7]

Exploring the matching channel, 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 faster job-to-job transitions are a major source of wage growth within large cities. Baum-Snow and Pavan (2012) use NLSY79 data to estimate a search and matching model equipped with exogenous returns from experience in cities of different size. They find that search frictions are not significantly different between small and large cities, while working in large cities is associated with a steeper wage profile.[8] The theoretical contribution in Martellini and Menzio (2018) helps understand why the results in Baum-Snow and Pavan (2012) differ from mine. Martellini and Menzio (2018) 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). I also provide external evidence that supports the existence of lower search frictions in larger labor markets, based on the heterogeneity in search behavior of workers and firms along the city-size distribution. In addition, I replace the exogenous difference in the return from experience across cities 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 some 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 to this literature, I show that endogenizing the sources of spatial wage differentials—and their equilibrium response to a change in policy—might significantly reduce the resulting income gains.

---

[7] In a second version of his model, Glaeser (1999) allows for multiple skill types, but, in order to maintain tractability, he assumes away migration across cities.

[8] 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.

An additional benefit of building an equilibrium model with endogenous productivity differences between locations is that it allows me to characterize the efficient allocation in the economy, given the presence of agglomeration forces and knowledge spillovers. Fajgelbaum and Gaubert (2019) compute the efficient allocation in a static setting with multiple locations, and positive production externalities within and between two sets of workers—college and high school graduates. They find that the optimal educational sorting is weaker than in the laissez-faire equilibrium. Rossi-Hansberg, Sarte, and Schwartzman (2019) build a model that shares many features with Fajgelbaum and Gaubert (2019), but they recover negative cross-type externalities. The optimal allocation in their paper features an increase in sorting of high-skilled workers into large cities. The dynamic setting in this paper shares key features with both those models. Quantitatively, I find that the congestion created by high school graduates in large cities—as in Rossi-Hansberg, Sarte, and Schwartzman (2019)—plays a dominant role in determining the optimal allocation, compared with the benefits from relocating college graduates into (human capital-poor) small cities—in the spirit of Fajgelbaum and Gaubert (2019).

Methodologically, this paper nests into the literature that studies productivity gains through knowledge diffusion. This literature has built on the theoretical contributions by, among others, Luttmer (2007), Lucas (2009), Lucas and Moll (2014), Perla and Tonetti (2014), who developed models in which the engine of growth is the imitation of other workers', or firms', productivity. Recent empirical work has attempted to quantify the importance of this channel in various economic environments. Herkenhoff et al. (2018) and Jarosch, Oberfield, and Rossi-Hansberg (2019) measure the amount of learning from coworkers inside a firm; Buera and Oberfield (2019) consider a world in which the stock of knowledge of a country depends on the productivity of its international trade partners; Fogli and Guerrieri (2019) study how neighborhood quality affects the process of children's skill formation. Differently from the existing work, I explore the life-cycle contribution of knowledge diffusion in accounting for the remarkable difference in wage profiles across cities. Compared to the common assumption of vertical imitation, i.e. workers only learn from those who are more skilled than they are, I adopt a flexible knowledge diffusion technology that also allows for horizontal imitation, i.e. workers learn from everyone. I then use heterogeneity in wage growth for workers in different positions of the entry wage distribution to discipline the contribution of these two types of knowledge diffusion, and I find that they are both quantitatively significant. This paper also shares some key features with Lucas (2004). His model aims at accounting for the long-run transition from a rural to an urban economy, where the latter is characterized by a human capital intensive technology and by learning from others. In contrast, I focus on a stationary equilibrium in which small and large cities coexist and host heterogenous workers in terms of productivity and life-cycle stages. At the cost of losing tractability, I provide a

quantification of the 'external effect' of other workers on the process of human capital accumulation proposed by Lucas, in a way that is empirically consistent with the labor market experience of workers in US cities.[9]

Last, this paper is also closely related to the literature on the determinants of wage growth over the life cycle, both within and between jobs. In a seminal paper, Topel and Ward (1992) show that a large fraction of the wage growth of young workers is associated with transitions to different employers. A more recent tradition of papers has used models with search and matching frictions to study the reallocation of workers toward better jobs, a process known as 'climbing the job ladder' (Burdett and Mortensen (1998), Postel-Vinay and Robin (2002)). Bagger et al. (2014), Menzio, Telyukova, and Visschers (2016) augment a search model with human capital accumulation through learning by doing, and measure the contribution of both job transitions and learning to life-cycle wage growth. The present paper contributes to this literature by exploring how search efficiency and human capital accumulation are affected by a worker's location. In this paper, the intensity of search frictions is allowed to depend on city size through the presence of non-constant returns to scale in the matching function. To the traditional skill acquisition through learning by doing, I add a process of knowledge diffusion, which is affected by the size and composition of the city where the worker is located.

2   The Model

I consider an economy made of two types of locations, small and large cities. Cities are inhabited by a continuum of workers with different age, 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 exclusively learn from those who are located in their same city. Both migration decisions

---

[9]The main specification in Lucas (2004) is such that everyone in the city only learns 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".

over the life cycle and human capital accumulation determine the equilibrium size and composition of cities. In turn, these city characteristics feed back into workers' decision problems. 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 must consume a scarce non-tradeable local good whose equilibrium price is also affected by city size and composition. Local, or house, prices operate as a congestion force in workers' location choice.

In the remainder of this section, I first present a rigorous description of the economic environment, then I define a stationary equilibrium for this economy.

2.1 Environment

Time is continuous and goes from 0 to ∞.

2.1.1 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 between workers, the equilibrium distribution of peers, and house prices. These features are rationalized by the existence of increasing returns to scale in the search and knowledge diffusion processes, sorting on unobservable human capital, and consumption of a scarce non-tradable good. In equilibrium, these differences are the endogenous outcomes that result from a fundamental heterogeneity between cities with respect to their housing supply function, migration opportunities to other cities, and vacancy creation cost, all of which I explain in detail below.

2.1.2 Workers

Demographics. The economy is populated by a measure M of workers. Workers are indexed by the tuple (h, a, e). Human capital h is a discrete variable that belongs to the set $\mathcal{H} = \{h_1, h_2, ..., h_L\}$ and evolves endogenously over the life cycle. Age is denoted by $a \in \{y, o\}$, where y stands for young, and o for old. e denotes the worker's education type, 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 $\omega z h - q^e p_i$, if the worker is employed at a job of match quality z and receives a piece rate ω. b_i is the gross flow payoff per

unit of human capital from being unemployed in city i, and $q^e p_i$ is the flow cost from living in city i, discussed below. Workers are born young and turn into old at Poisson rate ψ_y.[10] When old, they leave the economy (or retire) at rate ψ_o, and are replaced by a new young worker in their same city. Newborns draw their education and initial human capital level 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 $G_{0,i}(h, e) = G_0^h(h|e)G_{0,i}^e$, where $G_{0,i}^e = G_0^e(COL_i(o))$ and $COL_i(o)$ is the equilibrium fraction of old college graduates in city i.[11]

Human capital accumulation. Workers accumulate human capital through two channels.

First, they experience the 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 those aspects of the human capital accumulation process that are unaffected by the worker's location. Due to learning by doing, the human capital of an employed worker with education e improves from h_ℓ to $h_{\ell+1}$ at rate

$$\eta^e exp(-\eta h_\ell).$$

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.

Second, I model how interactions between workers spur the exchange of ideas. This form of learning—that I define knowledge diffusion, or imitation—is potentially influenced by the economic environment in which the worker is located. For example, the rate at which ideas flow between workers may depend on city size, as highly populated cities may be characterized by more frequent social interactions. I define the dependence of the meeting rate between workers on city size, i.e. the existence of increasing returns to scale in the frequency of interactions, as the 'flow of ideas' channel. In addition, the amount of learning that occurs through imitation is likely to depend on the human capital of the individuals a worker interacts with. Therefore, if imitation requires geographic proximity (as suggested, among others, by Jaffe, Trajtenberg, and Henderson 1993, Akcigit et al. 2018), so that workers only learn from those who are located in their same city, the composition of a city is also a crucial determinant of the gains from knowledge diffusion.

---

[10]For the remaining of this paper, the word 'rate' stands for 'Poisson rate', unless otherwise specified.

[11]Workers in the NLSY79 sample were administrated a test of cognitive ability, known as the Armed Force Qualifying Test (AFQT). The assumption I make on the workers' initial distribution is motivated by the observation that, conditional on education, the average AFQT score is almost identical in small and large cities, as already documented by De La Roca, Ottaviano, and Puga (2019) and Baum-Snow and Pavan (2012). In the quantitative section, I also show that the model replicates the life-cycle distribution of college graduates across cities, including at labor market entry.

I define the contribution of the human capital composition of a city to learning as peer effects, or 'peers'. In practice, the actual shape of the imitation process is theoretically ambiguous, and it is ultimately an empirical question. I model a flexible imitation technology by assuming that a worker in city i experiences an increase in human capital from h_ℓ to $h_{\ell+1}$ at rate $\sigma_i \kappa(G_i(h_{\tilde{\ell}}), h_\ell, e)$, where $\sigma_i \equiv \sigma(M_i)$, for some function σ. The dependence of the learning rate on city size, M_i, and on the equilibrium distribution of human capital in city i, $G_i(h_{\tilde{\ell}})$, is meant to capture the flow of ideas channel and peer effects, respectively. Concretely, the function κ takes the form,

$$\kappa(G_i(h_{\tilde{\ell}}), h_\ell, e) = \mathbb{E}_{G_i(h_{\tilde{\ell}})}[\eta_v^e max\{h_{\tilde{\ell}} - h_\ell, 0\} + \eta_h^e h_{\tilde{\ell}}].$$

The first term on the right-hand side captures the form of learning that exclusively occurs when interacting with more-skilled workers (*vertical imitation*). This is the type of knowledge diffusion that has been most frequently assumed in the theoretical literature (Lucas 2009, Lucas and Moll 2014, 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*). The technology of knowledge diffusion can be interpreted as follows. A worker of type h_ℓ meets another worker of type $h_{\tilde{\ell}} \sim G_i(h_{\tilde{\ell}})$, and he becomes of type $h_{\ell+1}$ with probability $\eta_h^e h_{\tilde{\ell}}$, if $\tilde{\ell} \leq \ell$, or with probability $\eta_h^e h_{\tilde{\ell}} + \eta_v^e(h_{\tilde{\ell}} - h_\ell)$, if $\tilde{\ell} > \ell$.[12]

Although the focus of this paper is on how human capital accumulation is affected by the worker's location, accounting for learning by doing plays the role of controlling for selection of workers into different locations according to their educational level. Absent this channel, the role of imitation in large cities is overstated if college graduates are more likely to both live in large cities, and experience faster wage growth irrespectively of their location.

2.1.3 Firms

Each city is also populated by a positive measure of 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, $(1 - \omega)zh$, discounted at rate r.

---

[12]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, $\mathcal{H}$ is chosen so that only a negligible measure of workers ever achieves the human capital value h_L over the life cycle.

2.1.4 Local Labor Market.

The labor market is characterized by search frictions. Workers can be either employed or un-employed.[13] 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 employers. 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 $\rho \in [0,1)$ cap-tures the relative search efficiency on the job. Upon meeting, the firm-worker pair draws a match quality $\hat{z} \sim F(\hat{z})$, where $\hat{z} \in [\underline{z}, \bar{z}]$, $\underline{z} > 0$ and $\bar{z} \leq \infty$.[14] If the pair decides to form a match, it starts producing, and the worker receives a fraction β of the gains from trade. If they do not, they keep searching. Jobs are exogenously destroyed at rate δ^e.

In the first part of the paper, I take the function λ as exogenous. In Section 4, I endogenize the meeting rates using a zero-profit condition in the market for vacancies.

2.1.5 Location Choice

Workers in city i are also contacted by firms located in a different city, which I refer to as city $-i$. The meeting rate between an unemployed worker in city $i = small$ and a firm in city $-i = large$ is denoted by $\lambda_{0,-i}^* \equiv \lambda_0^*(M_{-i}) = \rho^*\lambda_{0,-i}$. The meeting rate between an unemployed worker in city $i = large$ and a firm in each one of the N cities $-i = small$ is denoted by $\frac{1}{N}\lambda_{0,-i}^* \equiv \frac{1}{N}\lambda_0^*(M_{-i}) = \frac{1}{N}\rho^*\lambda_{0,-i}$, so that the overall cross-city meeting rate for workers in the large city is equal to $\rho^*\lambda_{0,small}$. The parameter ρ^* captures the extent of search frictions across cities. Since ρ^* is not indexed by i, it follows that

$$\frac{\lambda_{0,large}^*}{\lambda_{0,large}} = \frac{\lambda_{0,small}^*}{\lambda_{0,small}}.$$

While it imposes a restriction on migration opportunities, this parsimonious specification corre-sponds 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. In the quantitative section, I show that not only the model matches the average hazard rate of migrating, but it is also able to tightly match the *life-cycle* behavior of migration flows in both directions.

---

[13]I do not model 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 sec-tion, 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.

[14]In the quantitative section, I empirically motivate the assumed lack of heterogeneity in ρ and F across cities.

Upon meeting, the firm-worker pair observes the quality of the match $\hat{z} \sim F(\hat{z})$. Differently from meetings that occur inside a given city, the worker also draws a migration cost $c \sim D(c)$, which is *iid* distributed across migration episodes. Given 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 firm-worker pair decides not to 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 state.

In the rest of this paper, I focus on an equilibrium in which small cities are all identical to each other, and it is then convenient to assume away any migration flow between them. Under the classification 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, setting $N > 1$ is necessary in order to aggregate observations at the city level into macroeconomic statistics.

2.1.6 Housing Market

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

$$p_i = p_{0,i} Q_i{}^{\gamma_i} \qquad p_{0,i} > 0, \gamma_i \in \mathbb{R},$$

where γ_i is the (inverse) elasticity of housing supply. This specification is analogous to Hsieh and Moretti (2019). I also assume that the housing stock is owned by absentee landlords, so that housing income does not accrue to the workers in the economy. A worker of education e, located in city i, consumes q^e units of housing, hence he pays a flow price $q^e p_i$. The parameter q^e captures how the amount, or quality, of housing services varies by education. 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 components of tradeable goods whose price is affected by local house prices.

2.1.7 Contracts

To conclude the description of the environment, 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—i.e. the sum of the presented 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). Specifically, a worker is paid a fraction ω of his productivity, a so-called 'piece rate', which implies that his wage is equal to $\omega z h$. The wage setting

mechanism is described in details 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 ω that solves the Nash bargaining problem, where the worker's bargaining power is equal to β. The firm's threat point is always equal to 0, since firms do not make any profit unless they hire a worker. The worker's threat point is the present discounted value of being unemployed, if he is hired from unemployment, or the joint value of the previous match, if he was already employed. Over the course of the employment relationship, the worker might receive three types of outside offers. First, as already mentioned, if an outside offer triggers a separation, the piece rate is pinned down by Nash bargaining with the new firm. Second, if the worker's present discounted value of utility implied by his current piece rate is higher than the maximum value the new potential employer can offer, neither the worker's employer nor his piece rate changes. A new potential employer would always offer, at most, the joint value of the new potential match. Third, even if a job-to-job transition does not occur—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 worker receives an outside offer that would deliver more value to the worker than what he is currently being promised by his employer.

Notice that a worker who migrates to a different city incurs a one-time cost c. Hence, in every unemployment-to-employment, or job-to-job, transition that requires a change of location, the realized value of c needs to be subtracted from the joint value in the new employment relationship. Similarly, an outside offer to a worker in city i from a firm in city $-i$ triggers a wage raise only if accepting the offer would provide more value to the worker than what he is currently being promised, even after subtracting the migration cost the worker would have to pay.[15]

2.2 Definition of a Stationary Equilibrium

In this section, I define a stationary equilibrium for this economy. The term stationary stands for the fact that the set of equilibrium objects is constant over time. As mentioned above, 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 $U(h_\ell, a, e, i)$

---

[15]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. Absent 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, which might trigger an increase in the piece rate. 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.

be the present discounted value of income of an unemployed worker of human capital h_ℓ, age a, 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.

The value of unemployment $U(h_\ell, a, e, i)$ satisfies the following Hamilton-Jacobi-Bellman Equation (HJBE),

$$
\begin{aligned}
rU(h_\ell, a, e, i) = b_i h_\ell - q^e p_i + \\
& \sigma_i \kappa(G_i, h_\ell, e)[U(h_{\ell+1}) - U]\mathbb{I}\{a = y\} + \psi_a[U(o)\mathbb{I}\{a = y\} - U] + \\
& \lambda_{0,i}\mathbb{E}_F[max\{\beta(V(\hat{z}) - U), 0\}] + \\
& \lambda_{0,-i}^*\mathbb{E}_{F,D}[max\{\beta(V(-i, \hat{z}) - U(-i)) + U(-i) - U - c, \\
& \beta(V(-i, \hat{z}) - U - c), U(-i) - U - c, 0\}].
\end{aligned}
\tag{2.1}
$$

For ease of notation, I omit the dependence of the value and policy functions on the right-hand side of the HJBEs in this section from those elements of the worker's individual state that are the same as in the value on the left-hand side. The LHS of Equation (2.1) is the annuitized value of unemployment. The first line on the RHS is the flow payoff of unemployment, net of the house price. The second line shows the gains from knowledge diffusion and the transition to old age (retirement) if the worker is young (old). 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 firm, multiplied by the fraction of surplus that accrues to the worker, if the match is formed. Since it is easy to show that the joint value of a match is strictly increasing in its quality, the decision to create a match gives rise to a cutoff match quality $R(h_\ell, a, e, i)$, which is implicitly defined 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 firm in city $-i$. According to the realization of $(\hat{z} \sim F, c \sim D)$, the gain from a cross-city meeting is equal to the maximum between four terms: 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, and with match quality z, is willing to pay to move to city $-i$ in a job of quality z^*. To avoid introducing additional

16

notation, I adopt the convention that a value of z and/or z^* equal to 0 stands for a worker who is unemployed in city i and/or moves as unemployed to city $-i$. The cutoff migration costs are given by

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

$$x(h_\ell, a, e, i, 0, z^*) = V(h_\ell, a, e, -i, z^*) - U(h_\ell, a, e, i). \tag{2.4}$$

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

$$
\begin{aligned}
rV(h_\ell, a, e, i, z) = zh_\ell - q^e p_i + \\
[\sigma_i \kappa(G_i, h_\ell, e) + \eta^e exp(-\eta h_\ell)][V(h_{\ell+1}) - V]\mathbb{I}\{a = y\} + \psi_a[V(o)\mathbb{I}\{a = y\} - V] + \\
\delta^e(U - V) + \lambda_{1,i}\mathbb{E}_F[max\{\beta(V(\hat{z}) - V), 0\}] + \\
\lambda^*_{1,-i}\mathbb{E}_{F,D}[max\{\beta(V(-i, \hat{z}) - U(-i)) + U(-i) - V - c, \\
\beta(V(-i, \hat{z}) - V - c), U(-i) - V - c, 0\}].
\end{aligned}
\tag{2.5}
$$

As most events are common to employed and unemployed workers, I only highlight 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 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 replaced by, respectively,

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

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

Labor market and location decisions, learning, and aging induce a distribution of workers over the state space. In equilibrium, this distribution feeds back into the agents' decisions. First, city size determines the frequency of meetings between workers and firms (the matching channel), and the rate of knowledge diffusion (the flow of ideas channel). Second, the human capital distributions of cities determine the magnitude of peer effects. Third, the size and educational composition of

cities affect house prices, hence workers' location choice.

The Kolmogorov Forward Equation, KFE, (2.8) describes the law of motion of the measure of unemployed workers, $\phi(h_\ell, a, e, i, 0)$,

$$
\begin{aligned}
0 = {} & \sigma_i [\kappa(G_i, h_{\ell-1}, e)\phi(h_{\ell-1}) - \kappa(G_i, h_\ell, e)\phi(h_\ell)]\mathbb{I}\{a = y\} + \\
& \psi_y \phi(y)\mathbb{I}\{a = o\} - \psi_a \phi + \psi_o g_{0,i}(h_\ell, e)M_i(o)\mathbb{I}\{a = y\} - \\
& \lambda_{0,i}\phi[1 - F(R)] + \delta \int_{\underline{z}}^{\bar{z}} \phi(\hat{z})d\hat{z} + \\
& \frac{1}{N_i}F(R)\left[\lambda_{0,i}^* D(x(-i, 0, 0))\phi(-i, 0) + \lambda_{1,i}^* \int_{\underline{z}}^{\bar{z}} D(x(-i, \hat{z}, 0))\phi(-i, \hat{z})d\hat{z}\right] - \\
& \lambda_{0,-i}^*\phi\left[F(R(-i))D(x(0, 0)) + \int_{R(-i)}^{\bar{z}} D(x(0, \hat{z}))dF(\hat{z})\right].
\end{aligned}
\tag{2.8}
$$

The LHS is the derivative of the distribution with respect to time, which is equal to 0 in a steady state. The first line on the RHS states that learning through imitation induces an outflow of young workers of human capital h_ℓ and an inflow of young workers of human capital $h_{\ell-1}$. The second line shows the inflow and outflow of workers due to aging, and the entry of young workers into the labor market. Newborns are distributed according to $g_{0,i}(h_\ell, e)$. They replace old workers in city i—whose measure is denoted by $M_i(o)$—at rate ψ_o. The third line is related to local labor market flows. It shows the flow of workers out of unemployment because of an accepted job offer, and the inflow into unemployment of workers whose jobs are destroyed. The fourth line represents the inflow 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 sufficiently low migration cost that induces them to migrate, but 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 one of the N small cities. Hence,

$$
N_i = \begin{cases} 1 & \text{if} \quad i = \text{large} \\ N & \text{if} \quad i = \text{small}. \end{cases}
$$

The fifth line describes the migration flow in the opposite direction. It shows the flow of workers 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, $\phi(h_\ell, a, e, i, z)$,

$$0 = \{[\sigma_i \kappa(G_i, h_{\ell-1}, e) + \eta^e exp(-\eta h_{\ell-1})]\phi(h_{\ell-1}) - [\sigma_i \kappa(G_i, h_\ell, e) + \eta^e exp(-\eta h_\ell)]\phi(h_\ell)\}\mathbb{I}\{a = y\}+$$

$$\psi_y \phi(y)\mathbb{I}\{a = o\} - \psi_a \phi +$$

$$\lambda_{0,i}\phi(0)f(z)\mathbb{I}\{z \geq R\} + \lambda_{1,i} \int_{\underline{z}}^{z} \phi(\hat{z})d\hat{z}f(z) - \lambda_{1,i}\phi(1 - F(z)) - \delta^e \phi +$$

$$\frac{1}{N_i}f(z)\mathbb{I}\{z \geq R\} \left[\lambda_{0,i}^* D(x(-i, 0, z))\phi(-i, 0) + \lambda_{1,i}^* \int_{\underline{z}}^{z} D(x(-i, \hat{z}, z))\phi(-i, \hat{z})d\hat{z} \right] -$$

$$\lambda_{1,-i}^* \phi \left[F(R(-i))D(x(z, 0)) + \int_{R(-i)}^{\bar{z}} D(x(z, \hat{z}))dF(\hat{z}) \right].$$

(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, and 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 by workers with match quality $\hat{z} < z$ (third line, first two terms). It also decreases because of transitions to jobs with quality above z, and to unemployment (third line, last two terms). The fourth line shows the inflow of workers from city $-i$ into jobs of match quality z in city i. The first term represents workers who are hired from unemployment, the second term represents those who are already employed in city $-i$ at match quality $\hat{z}$. The fifth line shows the outflow of workers toward city $-i$, either as unemployed, or as employed at some match quality $\hat{z}$.

From the steady-state distribution ϕ, it is possible to compute the measure of workers of age a in city i,

$$M_i(a) = \sum_{e=hs,col} \sum_{\ell=1}^{L} \left[\int_{\underline{z}}^{\bar{z}} \phi(h_\ell, a, e, i, z)dz + \phi(h_\ell, a, e, i, 0), \right],$$

(2.10)

and total population in city i,

$$M_i = M_i(y) + M_i(o).$$

(2.11)

Similarly, the share of college graduates of age a in city i, and the overall share of college graduates in city i are given by, respectively,

$$COL_i(a) = \frac{\sum_{\ell=1}^{L} \left[\int_{\underline{z}}^{\bar{z}} \phi_i(h_\ell, a, col, i, z)dz + \phi(h_\ell, a, col, i, 0) \right]}{M_i(a)},$$

(2.12)

and

$$COL_i = \frac{COL_i(y)M_i(y) + COL_i(o)M_i(o)}{M_i}.$$

(2.13)

The human capital distribution in city i is

$$G_i(h_\ell) = \frac{\sum_{a=y,o} \sum_{e=hs,col} \sum_{\tilde{\ell}=1}^{\ell} \left[\int_{\underline{z}}^{\bar{z}} \phi(h_{\tilde{\ell}}, a, e, i, z) dz + \phi(h_{\tilde{\ell}}, a, e, i, 0) \right]}{M_i}. \tag{2.14}$$

Last, the equilibrium price in the housing market is given by

$$p_i = p_{0,i} Q_i^{\gamma_i} = [(q^{col} COL_i + q^{hs}(1 - COL_i)) M_i]^{\gamma_i}. \tag{2.15}$$

Equations (2.10)-(2.14) show how the law of motion for the measure ϕ in Equations (2.8) and (2.9) can be re-written in a compact form using the function $\tilde{\Gamma}$, defined as

$$0 = \tilde{\Gamma}(\phi, R, x, \lambda, \lambda^*). \tag{2.16}$$

Equation (2.16) states that the evolution of ϕ is a function of its current value and of the policy functions R and x. I also highlight the dependence of $\tilde{\Gamma}$ on the meeting rates λ and λ^*. This turns out to be convenient when I define the planner's problem in Section 6, in which the reduced-form meeting rates are replaced by a function of the endogenous labor market tightness.

The content of the above equations can be summarized in the following definition.

Definition. A stationary equilibrium is a tuple $\{V, U, R, x, \phi, p\}$ of value and policy functions, equilibrium distributions, and house prices such that i) U solves the HJBE (2.1) ii) V solves the HJBE (2.5), iii) R satisfies the optimality condition (2.2), iv) x satisfies the optimality conditions (2.3), (2.4), (2.6) and (2.7), v) ϕ solves the system of KFEs (2.8)-(2.9), where $M_i(a)$, $COL_i(a)$ and $G_i(h_\ell)$ are given by (2.10), (2.12) and (2.14), respectively, vi) p satisfies (2.15), where COL_i is given by (2.13), and M_i by (2.11).

3   Quantitative Analysis

The model is not amenable to a closed-form solution, due to the interaction between labor market dynamics, human capital accumulation and location choice. Therefore, I solve it numerically, by adapting the finite difference method introduced by Achdou et al. (2017) 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 identification of the model parameters (Section 3.3). The model is estimated via the method of

simulated moments.[16] 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 address the first research question of this paper by decomposing the wage premium into the contribution of sorting, matching, flow of ideas, and peer effects (Section 3.6).

3.1 Data

The main source of data for this paper is the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 is a survey of young men and women that 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 supplemental samples that represent minorities, econonomically disadvantaged, and population serving in the US military. 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 only use information on men from the cross-sectional subsample. Further sample restrictions involve dropping individuals that entered the labor force before January 1, 1978—i.e. the first date for which labor market information is available—or completed 20 years of experience after 2012. I also drop individuals with a significant amount of missing information on education, job history or location.[17] The survey contains information on the respondents' county of residence, which I uniquely assign to a commuting zone (CZ), following the methodology developed by David Dorn.[18] Because of the small number of observations associated to 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 population less, or more, than 750 thousand 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 final sample contains information on the labor market experience of 1532 men, for 20 consecutive years since the first month they are observed as employed after completing formal education. In order to assess the representativeness of the sample under consideration, I compute the value of some key summary statistics in the NLSY79, and their counterparts in larger publicly available

---

[16]Newborns in the model are all unemployed, while they are employed in the data. To bring the model closer to the data, I follow a similar approach to Lise and Postel-Vinay (2019). I simulate a pre-sampling period of 6 months in which I shut down all events that can happen to a worker, except meeting a firm 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.

[17]Additional details about sample selection are provided in Appendix B.

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

datasets, the Current Population Study (CPS) and the US Decennial Census. The share of men with at least a 4-year college degree in the NLSY79, 25.21%, is close to the analogous statistics for the US economy in the same time period, 26.8%, obtained from the CPS. The fraction of population living in large cities in the NLSY79 is 61.9%, against a value of 62.5% in the 1990 US Census. Last, average hourly earnings, measured in 2014 dollars, are equal to $24.7, and they are somewhat higher than the value of $21.6 obtained from the March Supplement of the CPS.

3.2 Parametric Assumptions

The estimation of the model requires imposing parametric assumptions on the relationship between city size and meeting rates, and on the distributions of match quality, migration cost, and newborns' human capital.

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

$$\lambda_{0,i} \equiv \lambda_0(M_i) = \chi M_i^{\chi_M}, \qquad \chi > 0, \ \chi_M \in \mathbb{R}.$$

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

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

$$z \sim Pareto(\underline{z}, \alpha) \qquad \underline{z} > 0, \ \alpha > 1,$$

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

Human Capital Accumulation. The rate at which ideas flow inside city i is equal to

$$\sigma(M_i) \equiv \tau M_i^{\tau_M}, \qquad \tau > 0, \ \tau_M \in \mathbb{R}.$$

Analogously to χ_M, τ_M captures the degree of returns to scale in the process of knowledge diffusion.

Migration Cost. Upon migrating, workers pay a cost $c \sim Logistic(\mu_c, \sigma_c)$, with cdf $D(c)$. The parameters μ_c and σ_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 $\hat{h}|e \sim LogN(\mu^e, \sigma^e)$ be a random variable with pdf $\hat{g}_0^h(\hat{h}|e)$. Let π^e be the fraction of workers in the economy with education e, and $\pi^{e,e'}$ be the probability that a worker with education e is replaced by one with education e'. Newborns' human capital and education are distributed according to

$$g_{0,i}(h_\ell, e) = g_0^h(h_\ell|e)g_{0,i}^e,$$

where

$$g_0^h(h_\ell|e) = \frac{\hat{g}_0^h(h_\ell|e)}{\sum_{\hat{\ell}=1}^{L} \hat{g}_0^h(h_{\hat{\ell}}|e)}, \tag{3.1}$$

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

Equation (3.1) states that the conditional distribution of initial human capital is assumed to be approximately log-normal (as in Huggett, Ventura, and Yaron 2006) on a finite set of points, with parameters μ^e and σ^e that are allowed to vary by education.[19] According to Equation (3.2), for each old worker that exits the economy in city i, the education of the newborn in the same city is governed by the transition probability $\pi^{e,e'}$. The normalizing factor $\hat{\pi}$ guarantees that the total fraction of college graduates in the economy is equal to π^{col}.[20]

3.3 Identification

The model is calibrated at monthly frequency. The choice of a fine partition of workers' experience allows to better replicate the behavior of some high-frequency events, like unemployment-to-employement and job-to-job transitions. Even though all the parameter are either externally calibrated or jointly estimated from simulated data, I provide an intuitive identification argument that clarifies how specific empirical facts are informative about certain 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 $\sigma(M_i)\kappa(G_i(h_{\hat{\ell}}), h_\ell, e)$, where $\sigma(M_i) = \tau M_i^{\tau_M}$. Since κ is a constant returns to scale function of the parameters η_h^e and η_v^e, I normalize $\tau = 1$. Hence, the set of parameters related to human capital accumulation is given by $\{\tau_M, \eta_h^{hs}, \eta_h^{col}, \eta_v^{hs}, \eta_v^{col}, \eta^{hs}, \eta^{col}, \eta\}$. In order

---

[19]The value of h_L is such that $\int_0^{h_L} \hat{g}_0^h(x|e)dx \approx 1, \forall e$.

[20]The normalizing factor is given by

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

to estimate these 8 parameters, I define 8 groups of workers. First, I split the sample by educational categories (*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 first year of employment. This operation delivers four groups. Last, at any point of the life cycle, each of the four groups is partitioned into workers located in large and small cities.[21]

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 η^{hs}, η^{col}, and η. Intuitively, η^{hs} and η^{col} determine the average wage growth by educational category, and η reproduces the curvature of wage profiles, *irrespectively of a worker's location*. The knowledge diffusion technology is described by η_v^{hs}, η_v^{col}, η_h^{hs}, and η_h^{col}. Vertical imitation, η_v^{hs} and η_v^{col}, is identified from the faster growth rate of wages—potentially *heterogeneous across cities*—for workers that are in the bottom half compared to the top half of the initial wage distribution. Horizontal imitation, η_h^{hs} and η_h^{col}, is pinned down by *differences across cities* in wage growth for workers that start in the top half of the wage distribution. The parameter τ_M leverages η_v^e and η_h^e 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 in a city, the faster wage growth inside that city, even under a hypothetical scenario in which $\tau_M = 0$.

Since young (old) workers age (retire) at rate ψ_y (ψ_o), imposing $\frac{1}{\psi_y} + \frac{1}{\psi_o} = 40 \times 12$ guarantees that workers retire after 40 years of work, on average. Only young workers are assumed to accumulate human capital. The share of life spent as young, $\frac{\psi_o}{\psi_o + \psi_y}$, is then pinned down by the ratio between the average wage in the second 10 years vs. the first 10 years of labor market experience. Intuitively, for given total wage growth, such ratio is higher the more time workers spend as young, i.e. the higher $\frac{\psi_o}{\psi_o + \psi_y}$.

Local Labor Market. The job destruction rate, δ^e, is taken directly from the data. It is equal to the average conditional probability of becoming unemployed in a given month, across the entire sample period. I now turn to the identification of meeting rates in the local labor market. Because of the small sample size, the remarkable variation in city size observed in the data is absorbed into two categories. Therefore, the model is only able to identify the average meeting rate, conditional on living in a large or a small city. Nevertheless, the choice of a parametric form allows changes in city size to affect meeting rates in counterfactual experiments. I pin down the elasticity of the number of meetings with respect to city size, χ_M, by targeting the average wage premium between

---

[21]If I first split the sample by education, then by city size, and last by position in the within-city initial wage distribution, I recover virtually identical growth rates of wages for each of the 8 groups.

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 the human capital composition need not be—and, in fact, it is not—the same across cities. Hence, higher match quality is only responsible for the portion of the average wage premium that is not accounted for by equilibrium sorting.

One could expect lower search frictions in larger cities to also generate higher transition rates out of unemployment, or across jobs. However, in the data, the unemployment rate and the average number of jobs held over the life cycle are virtually identical across cities of different size. This empirical regularity imposes a restriction on the shape of the match quality distribution. In an economy with homogenous workers and a single location, Martellini and Menzio (2018) prove that, if and only if $z \sim Pareto(\underline{z}, \alpha)$, the presence of lower search frictions in a larger market, i.e. $\chi_M > 0$, is consistent with the observed similarity in labor market flows in cities of different size. Quantitatively, I find that the intuition in Martellini and Menzio (2018) carries through to the much richer environment of this paper. An alternative hypothesis that could also account for the observed labor market flows within large and small cities is the existence of constant returns to scale in the search process, $\chi_M = 0$, combined with an arbitrary match quality distribution. As I show below, under constant returns to scale, the model would underpredict the average wage premium in the economy, and it would need 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 χ_M that I document 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 value of χ_M that allows the model to replicate the average wage premium in the economy is well in line with such direct measures.

Since χ 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 χ and χ_M, the relative efficiency of search on the job, ρ, is identified by the average number of jobs that a worker holds during his first 20 years of work experience.

Match Quality Distribution and Bargaining Power. It is well known that if $z \sim Pareto(\underline{z}, \alpha)$, $\underline{z}$ cannot be separately identified from χ.[22] Hence, I normalize $\underline{z} = 1$. The slope of the distribution, α, and workers' bargaining power, β, are jointly identified by the average wage growth in a job-to-job transition and from returns to tenure at a firm. As a measure of returns to tenure, I use the

---

[22]The rate at which an unemployed worker of type (h, a, e) accepts a job in city i is equal to $\chi M_i^{\chi_M} F(R(h, a, e, i)) = \chi M_i^{\chi_M} \frac{\underline{z}}{R^\alpha(h,a,e,i)}$. Clearly, labor market flows can only identify $\chi \underline{z}$.

wage difference between workers that spent more—compared to less—than a fifth of their work experience at a given firm (very similar results are found using other cutoffs). The argument goes that both a thick tail of the distribution, i.e. a low value of α, and a high value of β are consistent with high wage growth in a job-to-job transition. This is because a thick tail is associated with high dispersion in match quality, hence high productivity gains from changing jobs. At the same time, if workers have high bargaining power, they immediately reap the benefit of a higher match quality, and see their wage increase on impact. However, for given wage growth in a job-to-job transition, low values of α and high β have opposite predictions in terms of returns to tenure. If there is high heterogeneity in matches, and workers have low bargaining power (low α, low β), returns to tenure are high, since outside offers are likely to trigger an upward revision of the piece rate. Under this parametric configuration, workers receive the productivity gains from the creation of a better match over the course of the employment relationship. In the opposite case (high α, high β), returns to tenure are depressed, since the actual dispersion in match quality is low, and workers receive most of the benefits from a job-to-job transition upon joining the new firm.

Location Choice. Workers in city i contact firms in city $-i$ at a rate that is ρ^* times as large as workers that live in city $-i$. Migrating entails a cost $c \sim Logistic(\mu_c, \sigma_c)$. The values of ρ^*, μ_c and σ_c are jointly identified by relative city size, the hazard rate of moving from a small to a large city, and the share of workers in large cities with a college degree.[23] Since idiosyncratic migration motives—captured by σ_c—are symmetric across cities, high values of σ_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 μ_c and increasing ρ^* positively affects the migration hazard rate. However, since migration entails a lump-sum cost, and college graduates earn higher income, on average, lower values of μ_c would disproportionally 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 $\mu^{hs} = 0$. Given μ^{hs}, μ^{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, σ^{hs} and σ^{col}, are pinned down by the inter-quartile range of initial wages (for high school and college graduates, respectively). The value of π^{col} is equal to the fraction of college graduates in the data. Using information on parental education for workers in the sample, $\pi^{e,e'}$ is given by the probability that a worker has education e', conditional on his father's educational

---

[23]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 is pinned down by the law of motion for city size, combined with the assumption of stationarity.

attainment e.

Housing Cost. The house price function in city i is described by a level, $p_{0,i}$, and an elasticity parameter, γ_i. Workers with education e consume q^e units of housing. The values of γ_i are taken from Saiz (2010).[24] I normalize $q^{hs} = 1$, and choose values of $p_{0,i}$ and q^{col} so that each educational group has the average expenditure share on 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 infinite elasticity of substitution. Hence, it accomodates wide fluctuations in flow payoffs, which might lead workers to temporarily accept negative wages, in exchange for significant 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 flow value from unemployment equal to 60% of the average wage in city i. This target lies between the traditional value of 40% in Shimer (2005) and the more recent estimate of 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 and large cities 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 1.

The degree of returns to scale in the technology of knowledge diffusion, τ_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 $15.7^{\tau_M} = 1.20$ times more interactions with each other than in small cities. The relative importance of vertical and horizontal imitation is embodied in the parameters η_v^e and η_h^e. 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 2.76% per year, compared to 1.8% for a worker

---

[24]The online appendix in Saiz (2010) reports the elasticity of housing supply for 269 MSAs in 2000. I assign to γ_{large} (γ_{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 it is the case for virtually all large cities—provides very similar results.

Internally Estimated	Description	Value
σ_{hs}	inital hc hs, std	0.28
μ^{col}, σ^{col}	inital hc col, mean and std	0.3, 0.47
χ	meeting rate	0.06
χ_M	returns to scale matching	0.20
ρ	relative search eff. on the job	0.4
ρ^*	relative search eff. across cities	0.08
μ_q	migration cost (mean)	$2 \times$ median monthly wage
σ_q	migration cost (std)	$1.6 \times$ median monthly wage
α	shape match quality distrib	3.6
β	workers' bargaining power	0.2
τ_M	returns to scale knowledge	0.067
$\eta_h^{hs}, \eta_h^{col}$	horizontal imitation	0, 0.0013
$\eta_v^{hs}, \eta_v^{col}$	vertical imitation	0.0026, 0.007
$\eta^{hs}, \eta^{col}, \eta$	learning by doing	0.005, 0.008, 0.85
$\psi_o/(\psi_y + \psi_o)$	share of young	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.	$(1.27, 1.13) \times \underline{z}$

Externally Calibrated	Description	Value
μ_{hs}	inital hc hs, mean	0
$\delta^{hs}, \delta^{col}$	job destruction	0.0135, 0.005
r	discount rate	0.0125
π^{col}	college share	0.2521
$\pi^{hs,col}, \pi^{col,col}$	intergen. transition rate to college	0.175, 0.62
$\gamma_{large}, \gamma_{small}$	(inverse) house supply elasticity	$1.47^{-1}, 2.46^{-1}$
N	number of small cities	9.5
$1/\psi_y + 1/\psi_o$	average life span	40×12 months

Table 1: Parameter Values.

in the same position of the human capital distribution of a small city. The same comparison at the 25th percentile is 6% vs 4.56%. Clearly, college graduates learn more in large than in small cities, and particularly so the less skilled they are. 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.[25]

Figure 2 provides a graphical intuition of the contribution of horizontal and vertical imitation to the wage profiles of college graduates, given differences across cities in the equilibrium distribution of human capital, G_i, and in the meeting rate between workers, σ_i. In the first row of the Figure, the top (bottom) lines of each graph represent the life-cycle profiles of workers whose wage was in the top (bottom) half of the wage distribution of college graduates in the first year of work. The solid lines represent the wage paths obtained from the model using the baseline parameter values. The baseline profiles are remarkably similar to their empirical counterparts (see the comparison in Appendix D). The dashed lines are obtained by simulating the model under the following restrictions: $\eta_v^{col} = 0$ and $\sigma_{large} = \sigma_{small}$. It is easy to see that the resulting profiles are flatter than the baseline, and particularly so for workers located in large cities, and with initially lower wages. Intuitively, initially less-skilled workers would lose the most from the absence of vertical learning. Allowing η_v^{col} to take its estimated value largely fills the gap with respect to the baseline wage profiles, but the model would still slightly underpredict the overall wage growth in large cities (dashed-dotted lines in the top left panel). Once the estimated σ_{large} replaces σ_{small} in the simulation of wages in large cities, the model is able to fully replicate the empirical wage profiles. Interestingly, even under the restriction $\sigma_{large} = \sigma_{small}$, the simulated paths go a long way toward accounting for the higher wage growth in large cities. This result is due to the presence of better peers. The equilibrium human capital distribution in large cities first order stochastically dominates the distribution in small cities (bottom panels of Figure 2).

A similar wage pattern characterizes the labor market experience of high school graduates, though both η_v^{hs} and η_h^{hs} take lower values than η_v^{col} and η_h^{col}, respectively. In fact, I find that $\eta_h^{hs} = 0$, which implies that initially high-skilled high school graduates have virtually no room for further learning from others. College graduates also accumulate more human capital than high school graduates irrespectively of their location, due to a superior learning-by-doing technology ($\eta^{col} > \eta^{hs}$). Recall that learning is a prerogative of young workers. According to the estimated value of $\psi_o / (\psi_y + \psi_o) = 0.46$, 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, χ_M, equal to 0.2, which implies increasing returns to scale ($\chi_M = 0$ corresponds to constant returns). The estimated value of χ_M suggests that the frequency of meetings between workers and firms, both located in a large city,

---

[25]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 was in a large city.
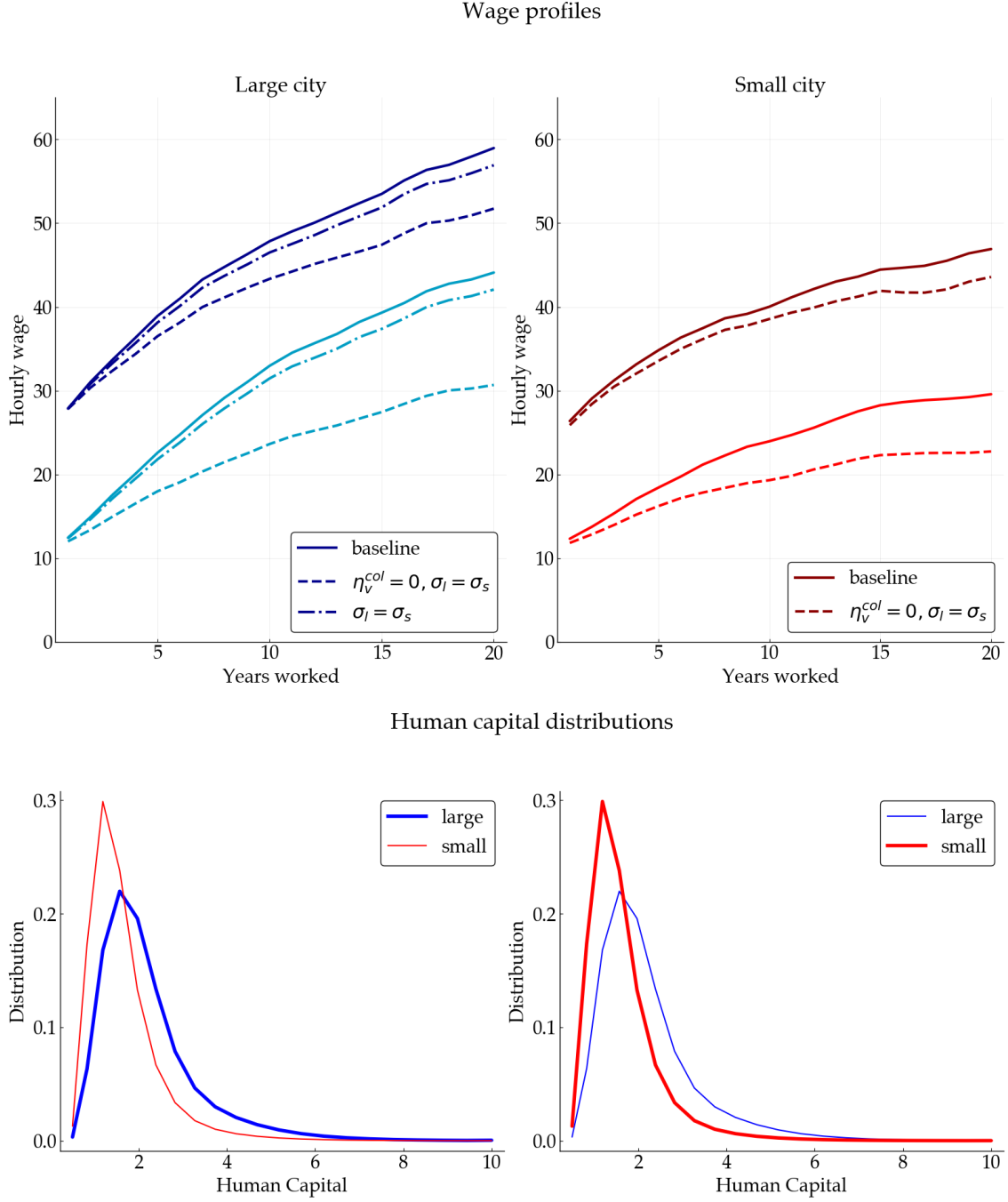
Wage profiles

Human capital distributions

Figure 2: Top panels: wage profiles of college graduates in large (left) and small (right) cities. Bottom panels: equilibrium human capital distributions. The legends in the top panels characterize the top 3 lines in each graph. The same mapping between line styles and parameter configurations applies to the three light-colored lines in the bottom of the same graphs.

is 15.7$^{\chi_M}$ = 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 $\alpha = 3.6$, meaning that a match at the 75th percentile of the distribution is 35.7% more productive than a match at the 25th per-

centile. As explained in the identification section, lower search frictions leverage the dispersion in match quality, and allow for the formation of more productive jobs. However, not all the productivity advantages of creating better matches are passed to the workers through higher wages, since workers capture an estimated $\beta = 20\%$ 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 χ_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 χ_M with two empirical measures of the same statistic. First, Martellini and Menzio (2018) 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 *workers* who were employed at the moment they were surveyed. Using information on the number of employers the workers had contacted, and the number of weeks they had been looking for a job, I compute the elasticity of the number of applications per week of search with respect to the population of the commuting zone where they were living. To the best of my knowledge, this estimate is new in the literature. I find an elasticity of 0.12. Interestingly, the value of χ_M obtained from the model, i.e. 0.2, lies inside the range of direct estimates derived from the worker, 0.12, and the firm side, 0.52.

With regard to the location choice, the estimated value of ρ^* implies that workers contact firms in a different city at a rate that is only 8% the rate of workers who already live in that local labor market. At the same time, the average migration cost, μ_c, is equal to about 2 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 thousand dollars. 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 experienced by workers upon changing location, about 9%.[26] If the observed level of mobility was due to higher values of μ_c,

---

[26]I compute the percentage wage change in the three months after migration, compared to 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 virtually identical wage gain as 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

the model would generate a counterfactually high value of this statistic.[27]

3.4.2 Model Fit

Table 2 reports the model-generated moments next to their empirical counterparts. While the number of targeted moments is greater than the number of parameters, the model fits the data very well along all dimensions. In all the Figures of this section, the thick blue (thin red) lines describe variables related to large (small) cities, while the corresponding dotted lines show the analogous observations in the data, surrounded by the 95% bootstrap confidence interval.[28]

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, workers with lower initial wages (i.e. in the bottom half of each education-city category) experience faster wage growth. This is particularly true for college graduates in large cities, whose wage grows at an average 6.3% per year, over 20 years, compared with 4.7% in small cities. The benefit of large cities also applies to 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 flatter experience gradient, they display a similar pattern in terms of variations in wage growth across cities and initial position in the wage distribution.

On the search and matching side, 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 experience profile of the frequency of job-to-job transitions (or EE rate), although the profile from the model is slightly flatter than the data at high levels of labor market experience (Figure 3). At early career stages, workers frequently change jobs, as they keep looking for a better match. After some years in which they 'climb the job ladder', the probability of further improving on their current matches 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, and, as a consequence, so is the number of jobs held after 20 years in the labor market (about 7). This finding is entirely consistent with the existence of significantly lower search frictions in large cities. On the one hand, employed workers

---

living in large cities, would help close this gap.

[27]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-specific preferences across people. Once the utility gain from living in a preferred location is accounted for, they show that the migration cost paid by movers is actually slightly *negative*.

[28]The label 'Years worked' on the x-axis of the graphs is short for total labor market experience, which includes the unemployment spells between two jobs.

Moment (large/small)	Data	Model
Labor Market		
unemp. rate (%)	5.8/5.9	6.0/6.1
n. of jobs	6.9/7.2	7.0/7.2
mean wage gap (%)	29.3	29.3
EE wage growth (%)	10.7/11.8	10.9/11.2
return to tenure (%)	23.7	23.0
UI replacement rate (%)	60	61.3/60.4
Cities		
city size	15.7	15.6
college share (%)	29/19	29/19
hazard rate (small→big, %)	0.31	0.32
hazard rate (big→small, %)	0.18	0.18
exp share non-trad. (col, %)	58	56/55
exp share non-trad. (hs, %)	58	58.4/57.2
Wage growth (%)		
bottom-col-large	6.3	6.3
bottom-col-small	4.7	4.4
top-col-large	4.0	3.8
top-col-small	3.1	2.9
bottom-hs-large	4.3	4.3
bottom-hs-small	3.6	3.0
top-hs-large	2.6	2.7
top-hs-small	1.3	1.7
wage 11-20 vs. 1-10 yr. of exp.	30.6	29.3
Initial Wage		
col wage premium (%)	38.7	39.2
$75^{th}/25^{th}$ pctile, hs	1.63	1.64
$75^{th}/25^{th}$ pctile, col	1.88	1.89

Table 2: Targeted moments in the data and in the model.

in larger cities contact firms 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. This reduces the conditional probability of drawing a match quality that further improves on the existing one, contributing to lowering the EE rate. If and only if the sampling match quality distribution is Pareto, these two opposing forces cancel each other out, de-
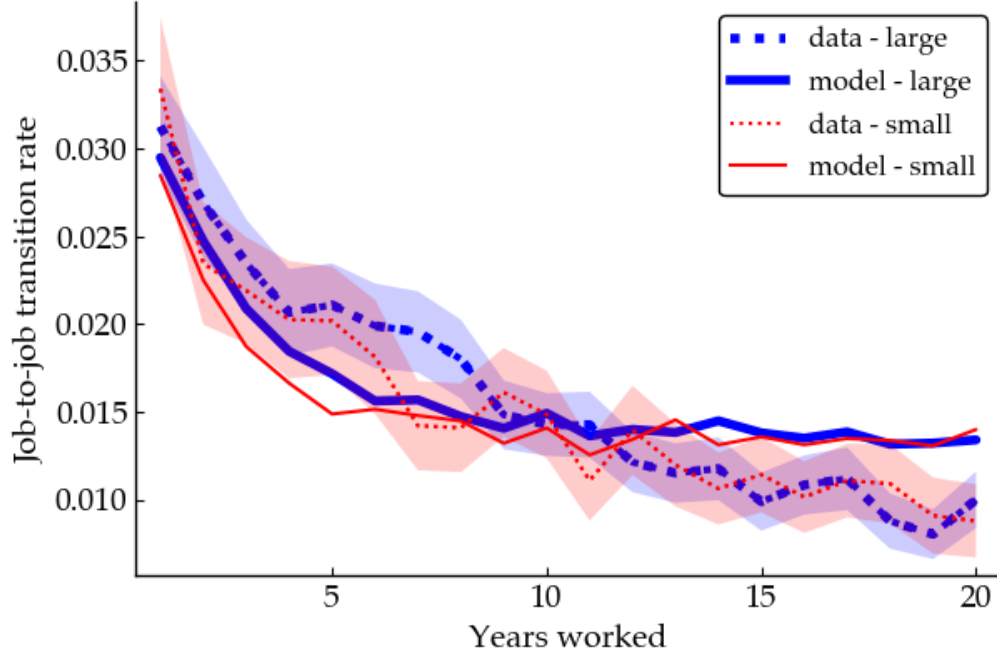
Figure 3: Job-to-job transition rate within large (thick blue line) and small (red thin line) cities, in the model (solid line) and in the data (dotted line).

livering a pattern of job-to-job transitions that is line with the data. A similar intuition can explain why the unemployment rate is virtually identical across cities ($\approx 6\%$): the equilibrium increase in the reservation match quality in large cities prevents the higher meeting rate between firms and workers from turning into a higher frequency of unemployment-to-employment transitions.

Turning to the location choice, 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 profile of migration patterns is depicted in Figure 4. The model is consistent with the fact that the hazard rate of migrating from small into large cities is considerably larger than the hazard rate in the opposite direction. Furthermore, the model replicates—without targeting—the steady decline in migration rates over the life cycle. Large cities offer two main advantages: a higher rate of human capital accumulation and lower search frictions in the labor market. As a consequence, they attract more workers compared with small cities. The age profile of migration is also consistent with the assumption that learning only happens at a 'young' age, and that older workers, who are more often employed at better matches, are less likely to start a new job, including in a different city.

The model generates an equilibrium relative city size that is almost identical to the one observed in the data. The endogenously larger (smaller) city in the model represents the average US commuting zone with a population of more (less) than 750 thousand people in the 1990 Census. These two sets of cities had average population of 2.2 milion and 140 thousand people, respectively. Figure 5 shows that the model-generated data resembles the mildly hump-shaped life-cycle
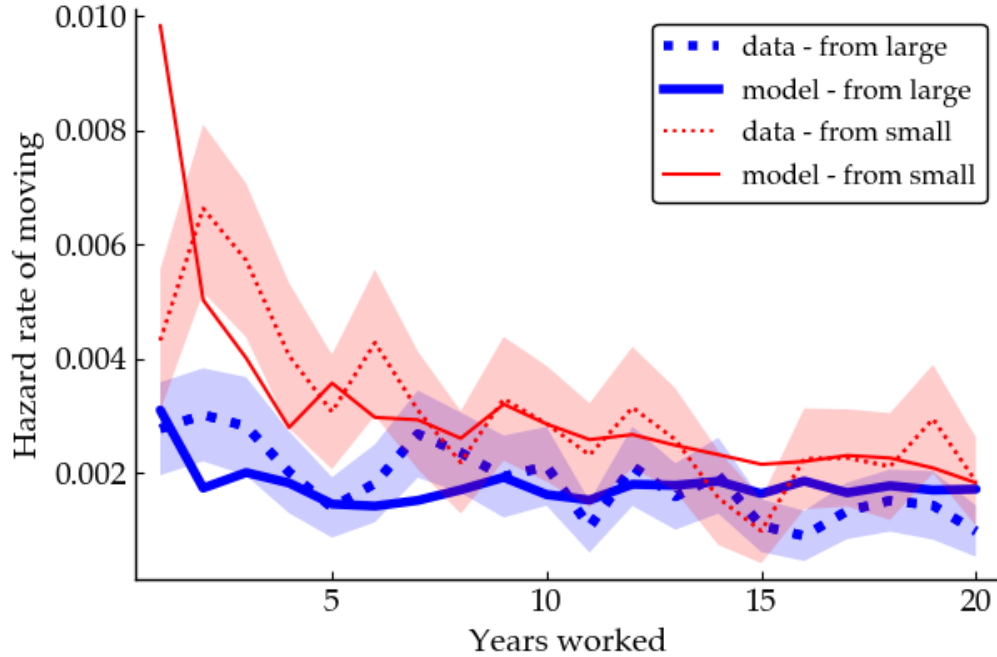
Figure 4: Hazard rate of moving out of a large (thick blue line) and a small (thin red line) city, in the model (solid line) and in the data (dotted line).

behavior of the share of workers located in large cities in the NLSY79 sample, although the confidence interval cannot rule out that such share is constant. In addition, the model replicates the educational composition of cities. Figure 6 shows the life-cycle profile of the fraction of workers in large cities that have at least a 4-year college degree. The average college share in large cities is tightly matched at 29%, compared with 19% in small cities. These plots are intuitive. Workers move to large cities in order to accumulate human capital and to find better jobs. As they age, some of them move back to small cities, which are on average 27% less expensive. Since college graduates have higher levels of human capital—which is complement to match quality in the production function—they benefit relatively more from lower search frictions in large cities. College graduates are also more likely to learn from others through knowledge diffusion. This explains why not only the size but also the educational composition of large cities generated from the model is broadly consistent with its empirical counterpart.

3.5 Model Validation: Patterns of Migration

In the previous section, I showed that the model is able to give an accurate representation of the labor market experience of workers in small and large cities. The estimation highlights the existence of increasing returns to scale in the labor search process and, to a lesser extent, in the rate of knowledge diffusion. Furthermore, I show that larger cities provide better learning
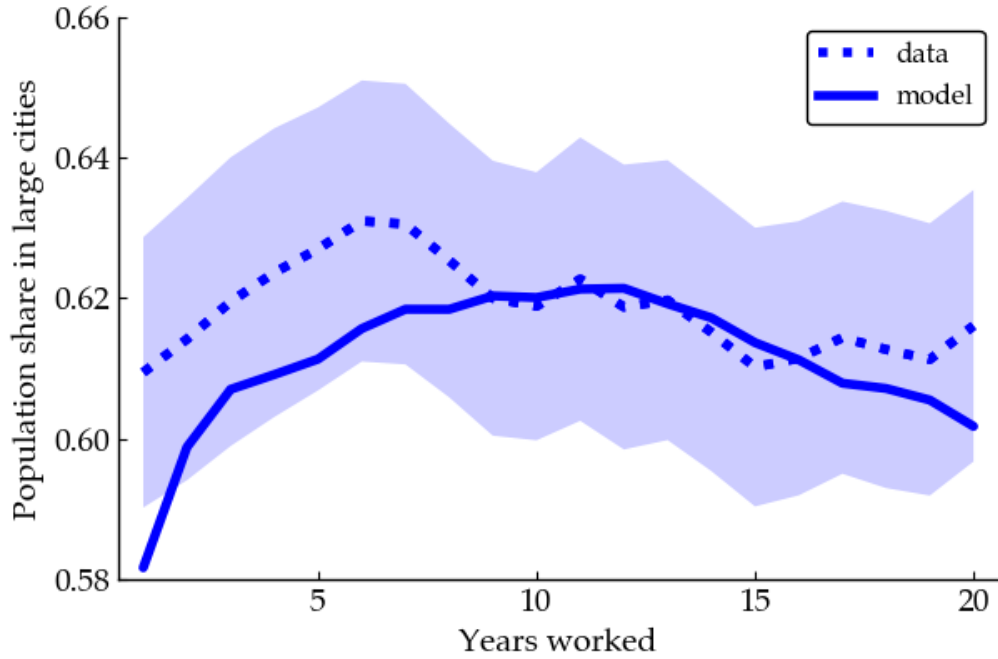
Figure 5: Fraction of population located in large cities, in the model (solid line) and in the data (dotted line).
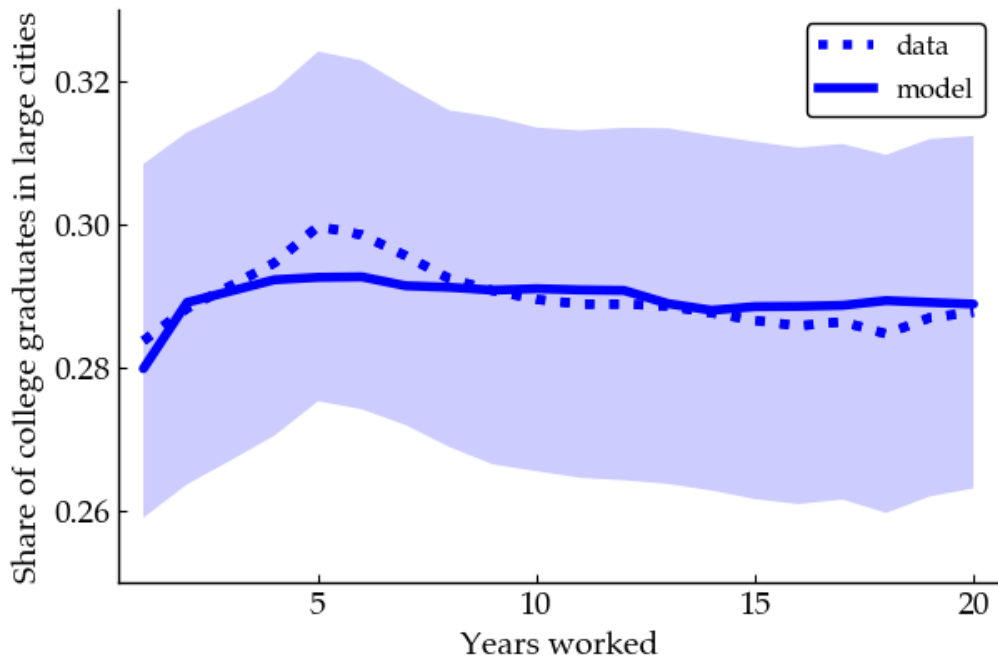

Figure 6: Fraction of population in large cities with a college degree, in the model (solid line) and in the data (dotted line).

opportunities due to peer effects.

While the model targets the frequency of migration, and the education composition of cities, none of the wage information used in the estimation is specific to migrants. Since 78% of workers in the NLSY79 sample never move from a small to a large city, nor vice versa, the identification of

the heterogeneity in the productivity gains from living in large cities is mostly driven by the labor market experience of non-movers. To make this point more transparent, I re-compute the targeted moments on the subsample of workers that never move, both in the data and in the model. The empirical moments obtained from the set of non-movers are almost identical 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).[29] At the same time, small and large cities are not isolated from each other, 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 validate the model by testing its 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 migration episodes. Jacobson, LaLonde, and Sullivan (1993) perform an event study in which they compare the wage of a worker in the years right before an unemployment spell, to the wage in the years that follow. In a very similar fashion, I adopt the following specification:[30]

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

The outcome variable is the inflation-adjusted logged hourly wage of worker i in year t. The controls include a quadratic term in experience, an indicator variable that is equal to 1 if worker i has at least a 4-year college degree, 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.[31] I run four different versions of Regression 3.3, that are distinguished by the type of city a worker migrates from, and by the selected control group, i.e. the set of observations where $d_{itk} = 0$. In Figure 7, I plot the coefficients generated from the NLSY79 sample (green circles) and those obtained using model-generated data (orange diamonds).

The top-left panel shows the wage of workers around the time when they move from a small into a large city, compared to the wage of those who remain in small cities. It is easy to observe

---

[29]I also re-estimate the model on the subsample of non-movers, and I obtain almost the same parameter values as in the baseline estimation. However, in order to perform policy analysis, and compute the optimal allocation, it is desirable to use a configuration of parameters that is the most consistent with the aggregate behavior of the US economy.

[30]See also Glaeser and Maré (2001) for an analogous empirical specification applied to migration into and out of urban areas.

[31]The value of δ_{10} (δ_{-10}) stands for 10 *or more* years after (before) the migration episode.
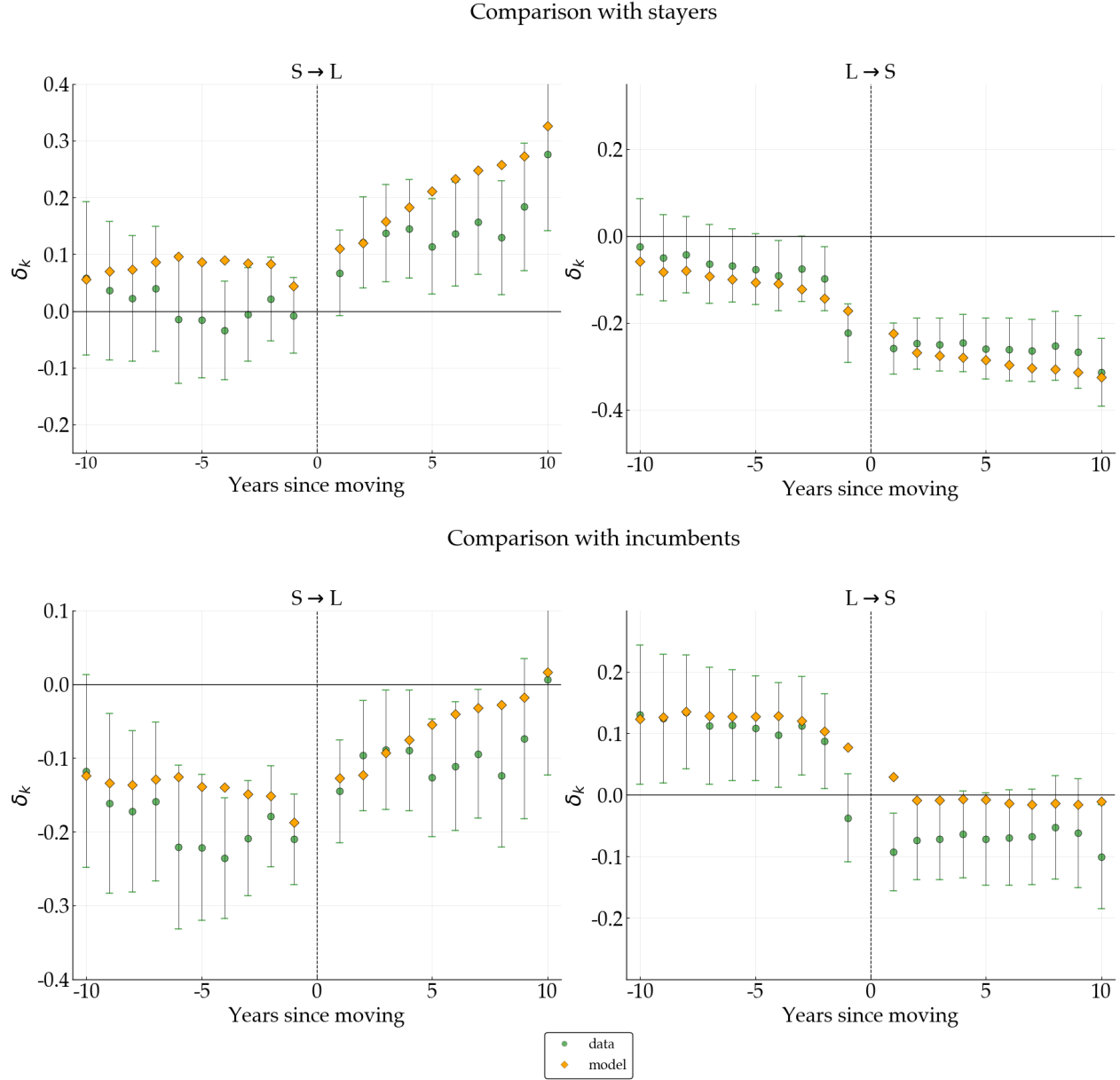
Figure 7: Wage of movers vs. stayers (top panels) and vs. incumbents (bottom panels). Left: from small city. Right: from big city. Data: green circles. Model: orange diamonds.

that migrants do not particularly differ from stayers until the moving year. After that, the wage of migrants starts to diverge, and it reaches a 28% premium after 10 years or more. The top-right panel repeats the same exercise with respect to the opposite migration flow. Leaving a large city is associated with lower pre-migration wages, and a much larger, but substantially flat, wage discount in the years after moving into a small city. Except for slightly 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 the 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 also matches the wage heterogeneity between movers and incumbents. Such comparison is represented in the bottom panels of Figure 7. While they live in small cities, migrants earn less then those who are already living in large cities before their arrival, and continue to do so even after migration takes place (bottom-left panel). This evidence is in line with the idea that most of the benefit from living in a large city does not accrue on impact. Consistently with the dynamic nature of the return to experience in large cities, the gap with respect to the incumbents slowly shrinks in the 10 years after the migration episode. Last, in the bottom-right 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 out.

Through the lens of the model, the variation in Figure 7 originates from differences across workers in terms of human capital and match quality, while the migration decision is also affected by idiosyncratic preferences. These components are not separately observed. The model predicts that higher match quality is associated with longer job tenure, as better matches tend to last longer. At the same time, recent migrants have lower job tenure, and, both in the model and in the data, jobs with longer tenure pay significantly higher wages—after controlling for worker observable characteristics.

Motivated by this evidence, I attempt to measure the contribution of heterogeneity in match quality to wage differentials between movers, stayers, and incumbents. I augment Regression 3.3 with a quadratic term in the tenure at the current job, and re-draw the new values of the coefficients δ_k. The updated plots—in Appendix D—highlight how the results differ with respect to the baseline specification, but they do so in a remarkably similar fashion in the model as in the data. To gain intuition, here I show and describe one of the four updated plots (Figure 8). I construct this plot by repeating the regression represented on the top-right panel of Figure 7—workers who move to small cities vs. stayers in large cities—and I compare the results that are obtained from excluding or including controls for job tenure. The left panel of Figure 8 shows the coefficients estimated on NLSY79 data. While the relative wage loss after leaving a large city is essentially flat in the baseline regression (dark green stars), it is much steeper when tenure is controlled for (light green triangles). An almost identical pattern emerges from the model (right panel, orange shades). Since recent movers to small cities have been at their current job for a shorter period of time, controlling for job tenure reduces the size of their wage discount with respect to stayers in large cities, but only in the years right after migration (from -26% to -17% in the first year). After removing
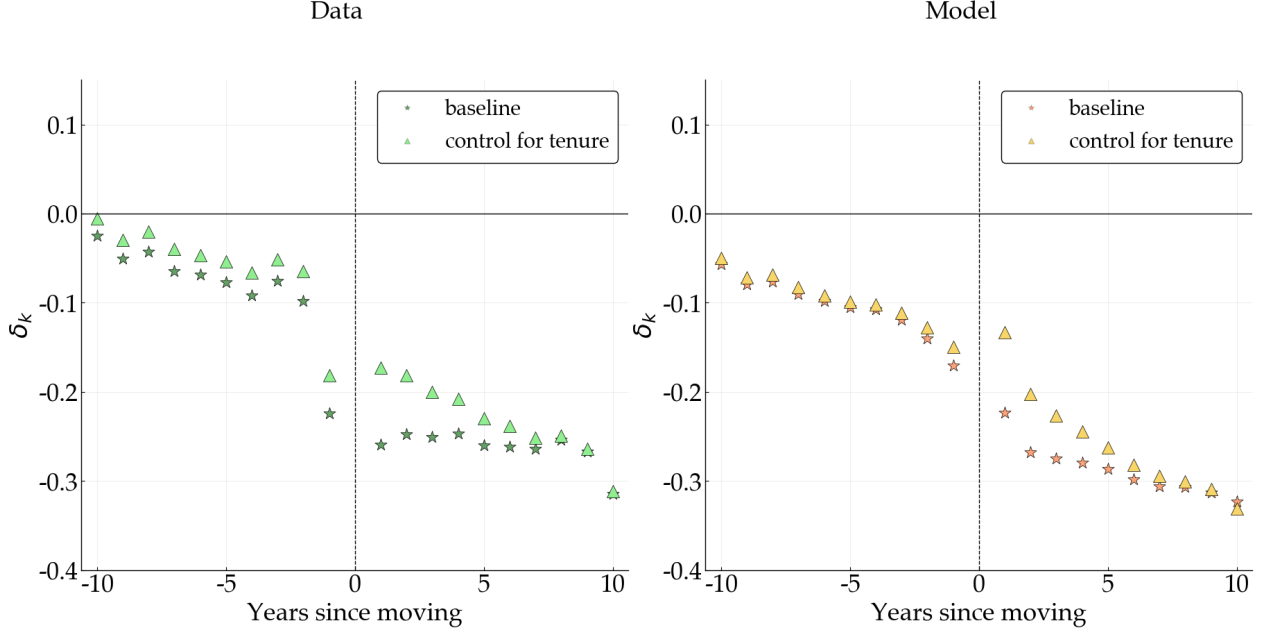
L → S, comparison with stayers

Figure 8: Wage of movers to small cities vs stayers in large cities. Dark stars: baseline. Light triangles: control for tenure. Left: data. Right: model.

the tenure effect, a stronger diverging wage pattern emerges, since movers to small cities are no longer exposed to the superior knowledge diffusion of large cities.

3.6 Inside the Agglomeration Black Box: Decomposition

The validation exercise in the previous section confirmed the model ability to replicate the wage experience of migrants with respect to stayers and incumbents. The unobservable worker characteristics in the model—human capital, match quality, and moving cost—are quantitatively consistent with the motives that drive location choices in the data. I am now in the position to measure the contribution of each of the proposed sources of the city-size wage premium and its life-cycle behavior.

Throughout the decomposition, I fix the set of aggregate variables at their baseline equilibrium values, treating them as parameters in the worker's problem. I also impose that workers cannot migrate at any moment of their life. In order to perform the decomposition, I consider four scenarios, which I model as follows. Let $\tilde{\Theta}_i = \{\tilde{\lambda}_{0,i}, \tilde{\lambda}_{1,i}, \tilde{\sigma}_i, \tilde{G}_i\}$, for $i = \{small, large\}$, be the set of aggregate characteristics of city i in the baseline economy, and let $\Theta_i^s = \{\lambda_{0,i}^s, \lambda_{1,i}^s, \sigma_i^s, G_i^s\}$ be the set of aggregate variables that characterize city i in scenario s.[32] Because of the assumption of no

---

[32]I omit house prices from the aggregate characteristics of cities. If migration is ruled out, house prices are equivalent
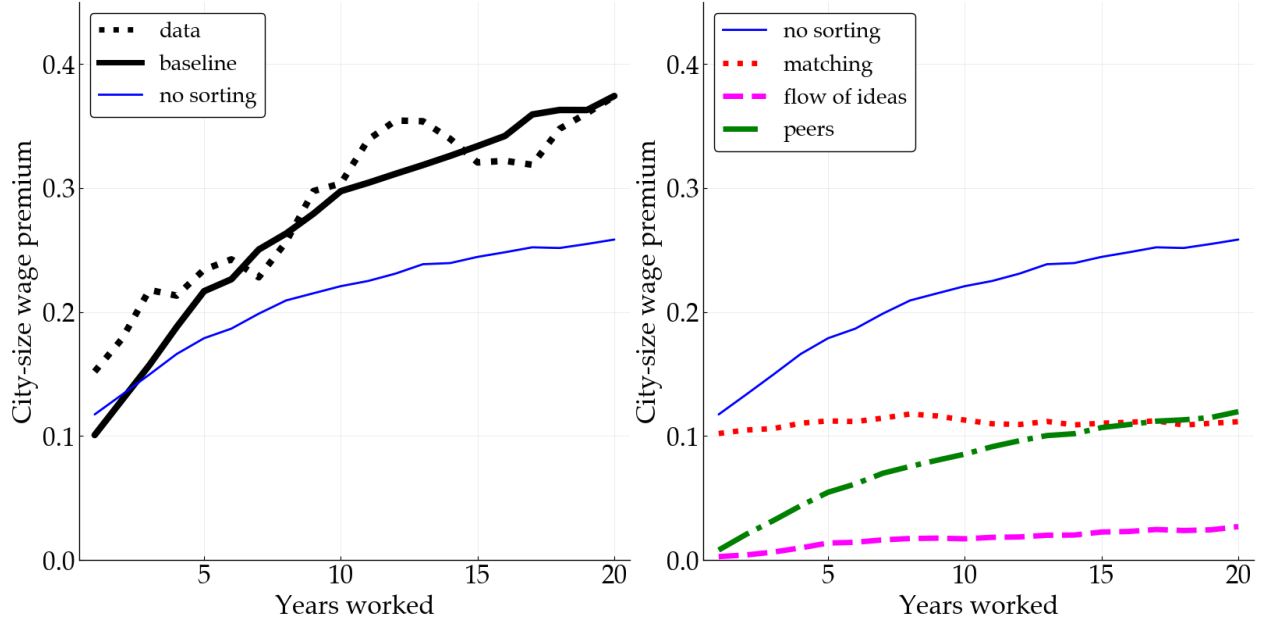
Figure 9: Percentage wage premium. Left panel: data (black dotted line), baseline model (black thick line), 'no sorting' scenario (blue thin line). Right panel: 'no sorting' (blue thin line), 'matching' (red dotted line), 'flow of ideas' (pink dashed line), 'peer effect' (green dashed-dotted line) scenario.

migration across cities, the joint value of an employment relationship and the value of unemployment in city i only depend on the worker's idiosyncratic state (and match quality, if employed) and on Θ_i^s, but do not depend on Θ_{-i}^s. In other words, large and small cities are treated as separate economies. Across all scenarios s, I set $\Theta_{small}^s = \tilde{\Theta}_{small}$, and only vary the characteristics of large cities, Θ_{big}^s. The thought experiment is comparing the life-cycle 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 paths of wages, under the parametrization $\Theta_{big}^{NS} = \tilde{\Theta}_{big}$. 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 9 shows the city-size wage premium in the data (black dotted line), the baseline economy (thick black line) and in 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—i.e. the wage premium net of the role of sorting—is represented by the blue thin line reported on both panels of Figure 9, and it is due to heterogeneity

---

to a lump-sum tax paid by every worker in the economy—although potentially different across cities—but they have exactly no impact on wages.

in the equilibrium aggregate characteristics of small and large cities. In order to quantify the contribution of each treatment, I compute the wage premium under three additional scenarios, illustrated in the right panel of Figure 9. The red dotted line corresponds to an economy in which only the matching channel is operative, $\Theta_{large}^{M} = \{\tilde{\lambda}_{0,large}, \tilde{\lambda}_{1,large}, \tilde{\sigma}_{small}, \tilde{G}_{small}\}$. The pink dashed line repeats the same exercise with respect to the flow of ideas, $\Theta_{large}^{F} = \{\tilde{\lambda}_{0,small}, \tilde{\lambda}_{1,small}, \tilde{\sigma}_{large}, \tilde{G}_{small}\}$. Last, the green dashed-dotted line shows an economy in which the only heterogeneity between small and large cities is due to differences in peer effects, $\Theta_{large}^{P} = \{\tilde{\lambda}_{0,small}, \tilde{\lambda}_{1,small}, \tilde{\sigma}_{small}, \tilde{G}_{large}\}$.[33]

Starting from the left panel of Figure 9, we observe how the contribution of sorting slowly rises with experience, and it reaches 12% after 20 years.[34] 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 actual experience-specific educational composition of cities in the baseline economy.[35] The modest role of sorting on unobservable human capital in accounting for the city-size wage premium is consistent with the findings by, among others, Eeckhout, Pinheiro, and Schmidheiny (2014) and De La Roca and Puga (2016). While in the static environment by 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-skilled workers benefit more from locating in large cities, because of complementarity between human capital and match quality. On the other hand, low-skilled workers have more room for learning in large cities, thanks to vertical imitation in the process of knowledge diffusion.

The right panel of Figure 9 shows the contribution of each of the three proposed treatments. Consistently with the identification strategy presented above, matching has a level effect on the wage premium of about 11%, but it plays no role in the growth of the premium over the life cycle.

---

[33]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 jobs. 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 the values of b equal across cities, in a scenario in which the only systematic difference between workers in small and large cities is in terms of human capital.

[34]At labor market entry, the wage premium is slightly higher in the no sorting scenario than in the baseline economy. I verify that this is not due to the presence of negative initial sorting in the economy, but to a slightly higher average piece rate in large compared to small cities. When workers are allowed to migrate, as in the baseline economy, firms pay workers part of the future surplus they might obtain from a job-to-job transition across cities. Such influence of future offers on wages is muted in the no sorting scenario. Since, in the baseline economy, very young workers are more likely to move to large cities than in the opposite direction, workers that are born in small cities suffer relatively more from the absence of cross-city offers, in terms of the piece rate they are able to bargain. This is why the wage premium in the no sorting scenario is slightly bigger than in the baseline economy, at labor market entry.

[35]The difference between the resulting wage profile and the *no sorting* in Figure 9 isolates the contribution of sorting on education. The residual difference with respect to the baseline profile captures the wage premium due to sorting on human capital.

	5 years	10 years	20 years
Sorting	4	8	12
Matching	11	11	11
Flow of Ideas	1.5	2	3
Peers	5.5	9	12
Total	22	30	38

Table 3: Decomposition of the city-size wage premium (%).

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 the fact that increasing returns to scale do not trigger differences in unemployment and job-to-job transition rates across cities of different size. To the contrary, the flow 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%. Heterogeneity in the equilibrium composition of peers makes up three quarters of the higher rate of learning in large cities, leaving a much smaller role to increasing returns to scale in the flow of ideas. Table 3 summarizes these findings by breaking down the wage premium at 5, 10 and 20 years of work experience.

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. That is, they are treated as fixed parameters. The decomposition described above is meant to highlight the role of city characteristics—at their current equilibrium value—in determining the city-size wage premium, and not the role of increasing returns and knowledge diffusion in shaping the characteristics of cities. Second, besides isolating the role of sorting, the absence of migration flows is also required in order to correctly measure the contribution of any specific treatment (matching, flow of ideas, and peer effects). Consider, for example, 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—i.e. a higher average match quality in large cities—and the sorting behavior induced by matching itself.

4   Endogenous Market Tightness

So far, I focused on a stationary equilibrium for the economy, which is an allocation in which city size and composition are constant over time. Hence, I dispensed with modeling vacancy cre-

ation, and I adopted a reduced-form specification for the rate at which a worker contacts a firm in the local labor market and across cities, λ and λ^*, respectively. In solving for the constrained-efficient allocation and performing a counterfactual policy experiment, it is natural to allow firms to adapt their behavior to changes in the profitability of creating vacancies in different cities. In order to accomodate this margin, I explicitly introduce a perfectly competitive market for vacancies, in the tradition of Mortensen and Pissarides (1994).

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

$$m_i = M_i^{\chi_M} m(s_i, v_i) = M_i^{\chi_M} s_i^{1-\zeta} v_i^\zeta,$$

where v_i is the number of vacancies, and s_i is the actual measure of workers seeking jobs in city i.[36] In line with the assumption of proportional search efficiency across employment states and cities, the number of meetings per unit of time between a firm and an employed worker, both located in city i, is equal to ρm_i. The meeting rate between an unemployed (employed) worker in city $-i$ and a firm in city i is equal to $\rho^* m_i$ ($\rho \rho^* m_i$). It follows that

$$s_i = [u_i + \rho(1 - u_i)]M_i + \rho^*[(u_{-i} + \rho(1 - u_{-i}))]M_{-i}, \tag{4.1}$$

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 firm, also in city i, at rate

$$\frac{m_i}{s_i} = M_i^{\chi_M} \theta^\zeta,$$

where $\theta = v_i/s_i$ is assumed to be the same across cities.[37] From the reduced-form specification introduced in Section 3, such contact rate is also equal to $\lambda_{0,i} = M_i^{\chi_M} \chi$. Petrongolo and Pissarides (2001) document a wide range of empirical estimates of ζ, and suggest a value of approximately 0.5. However, while in this paper it does not guarantee efficiency, applying the Hosios (1990) condition simplifies the nature of the optimal policy I introduce in Section 6.[38] Hence, I set $\zeta =$

---

[36]I use the term 'meeting'—instead of the standard 'matching'—function to highlight the fact that not all meetings in this model turn into matches.

[37]The assumption of equal market tightness across cities can be relaxed using either direct measures of vacancies at the city level, combined with local unemployment data, or variations in the average duration of vacancies across cities. While data on both measures is scant, this kind of empirical evidence is certainly worth exploring in the study of search and matching frictions in local labor markets.

[38]In an economy with on-the-job search and multiple, connected, labor markets, the Hosios condition would not guarantee efficiency even in the absence of the other externalities that are peculiar to this paper (agglomeration and knowledge diffusion). Replacing random with directed search is a natural solution in order to remove the inefficiency that originates from the search process.

$1 - \beta = 0.8$. Given ζ, I recover the value of $\theta = \chi^{\frac{1}{\zeta}}$. Last, the vacancy creation cost, k_i, is pinned down by the zero profit condition

$$k_i = M_i^{\chi_M} \theta^{\zeta - 1} (1 - \beta) \mathcal{S}(i), \tag{4.2}$$

where $\mathcal{S}(i)$ is the expected job surplus from contacting a worker from any city, or in any employment status, and $(1 - \beta)$ is the share that accrues to the firm.

I define the gross surplus from a match, $S(h_\ell, a, e, i', z', i, z) \equiv V(h_\ell, a, e, i, z) - V(h_\ell, a, e, i', z')$, where it is understood that $V(h_\ell, a, e, i, 0)$ stands for $U(h_\ell, a, e, i)$, and that i' might take the value of either i or $-i$. The specification 'gross' is due to the presence of a migration cost that the worker has to pay whenever he moves from city $-i$. The expected surplus $\mathcal{S}(i)$ is described in detail in Appendix C, but it is possible to gain intuition by using the following stylized representation,

$$\begin{aligned} \mathcal{S}(i) = \; & \mathbb{E}_{\phi(i,0),F}[S(i,0,i,\hat{z})] + \mathbb{E}_{\phi(i,z),F}[S(i,z,i,\hat{z})] + \\ & \mathbb{E}_{\phi(-i,0),F,D}[S(-i,0,i,\hat{z}) - c] + \mathbb{E}_{\phi(-i,z),F,D}[S(-i,z,i,\hat{z}) - c].^{39} \end{aligned} \tag{4.3}$$

The terms on the first line of Equation (4.3) correspond to the expected surplus from meeting an unemployed, or employed worker, in city i. The expectation is taken with respect to the probability of meeting each type of worker, and to the sampling distribution of match quality, F. The second line is analogous to the first, but it is related to hiring a worker who lives in city $-i$ before accepting a job in city i. If forming a new match involves migration, the expectation also takes into account the random component of the migration cost, $c \sim D(c)$.

Applying Equation (4.2) to both large and small cities, I find $k_{large}/k_{small} = 1.6$. Because of increasing returns to scale in the search process, and a better human capital composition, firms in large cities are willing to pay a 60% higher vacancy creation cost. A higher value of k reduces firms' incentives to create vacancies in a given city, in a similar fashion as higher house prices represent a congestion force in workers' location choice.

5 Large Cities and Housing Regulation

Understanding the sources of the remarkable difference in wages across cities is a long standing endeavour in the literature, and an interesting question in and of itself. Yet, the nature of the city-size wage premium has also important implications for the aggregate and distributional consequences of those place-based policies that trigger the relocation of workers across cities. I illustrate this point by computing the equilibrium of the economy after the implementation of a policy that has been at the center of the academic and political debate: 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 to decompose the change in total income, after the policy implementation, into pre-policy observables and predicted equilibrium responses. The decomposition highlights how the outcome of the policy is crucially affected by the endogenous behavior of productivity in small and large cities, which is in turn determined by the sources of the city-size wage premium. Section 5.3 presents the results.

5.1 Housing Regulation: Background

In the estimation of the model, the (inverse) elasticity of housing supply is constructed using the estimates from Saiz (2010). I find that, on average, it is equal to 1/1.47 in large cities, but only 1/2.46 in small cities. A growing literature has documented that part of the higher inverse elasticity in some large US cities—i.e. the fact that house prices grow more rapidly as the city expands—is not due to physical constraints, but it is the outcome of tighter regulation.[40]

Hsieh and Moretti (2019), HM henceforth, explore the aggregate implications of tighter regulation in the most productive (large) US MSAs. They find that if house price dispersion in 2009 had been as low as in 1964, the increase in productivity of some large cities would have triggered a much higher increase in employment, and a less staggering rise in wages, in those cities. In equilibrium, the growth rate of US GDP would have been twice as large as its actual value during the last 50 years. The intuition is that an increase in house prices, due to regulation, prevents a productive city from further expanding, hence from generating higher GDP growth. Herkenhoff, Ohanian, and Prescott (2018), HOP henceforth, develop a neoclassical growth model with multiple regions. They find that GDP would be 12% higher in a steady state in which current land-use regulation in all US states moved halfway toward that of the least restrictive state, i.e. Texas. According to their estimates, California and New York are among the most restrictive states. Notably, they are also the states where some of the largest cities in the country are located.

Both HM and HOP model spatial heterogeneity in productivity as an exogenous difference in TFP between locations.[41] Besides, they study economies populated by homogeneous workers, endowed with a constant amount of efficiency units. In contrast, this paper introduces two new margins to this literature. First, exogenous heterogeneity in productivity is replaced by agglom-

---

[40]Saiz (2010) estimates the contribution of regulation 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 find 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).

[41]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.

eration 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 in nature, and stems from differences in the rate of human capital accumulation. I also allow for heterogeneity in human capital between workers, so that spatial sorting might contribute to measured differences in productivity between cities. Crucially, all these aspects—agglomeration, knowledge diffusion, sorting—are endogenous to the size and composition of cities, and might alter the assessment of the aggregate consequences of changes in local policies.

In fact, consider two opposite theories of spatial wage differentials. On one extreme, productivity is an intrinsic characteristic of cities, and regulation just limits access to productive locations. This is the main view of HM and HOP. On the opposite extreme, heterogeneity in productivity between cities is exclusively explained by workers' sorting on ability. According to such view, regulation may arise from the preference of high-skilled individuals in certain (large) cities for excluding other workers from accessing those cities. Parkhomenko (2018) shares some features with this second hypothesis. He builds a model with sorting on ability and endogenous regulation—motivated by the desire to prevent newcomers from creating congestion in the use of local amenities. However, he also assumes exogenous productivity differences between cities. He shows that equalizing housing regulation across cities reduces the extent of sorting in large cities, but the ensuing output gains are exclusively generated by the relocation of workers toward exogenously more productive locations.[42]

While this paper exhibits an exogenous housing supply elasticity, hence exogenous regulation, it presents a theory of the city-size wage premium that embeds some key components of both the opposite views introduced above. Spatial wage differentials are partially due to sorting, and partially to *endogenous* differences in productivity between cities. Thus, although I adopt a much more stylized geography than the traditional urban literature,[43] the new margins considered in this paper complement the current debate on the consequences of housing, and, more generally, place-based policies.

---

[42]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 former by the same amount as it increases the output of the latter. See the discussion in Section 6.

[43]HM and Parkhomenko (2018) model an economy with more than 200 MSAs. HOP consider 7 groups of states. Those more realistic environments can capture the heterogeneity in productivity and house prices also *within* the set of large cities—or between states—in the US.

5.2 Interpreting the Policy Outcome

I introduce an identity that highlights the role of the estimated heterogeneity between small and large cities in shaping the equilibrium response to a policy change. Throughout the analysis, I compare the steady state of the economy, before and after the implementation of the policy. As the estimation of this paper is based on information on the wage of workers in their first 20 years of work experience, I evaluate the policy outcome in terms of total labor income of this set of workers.[44]

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 $\Delta M = M^1_{big} - M^0_{big} = N(M^0_{small} - M^1_{small})$ be the equilibrium change in the measure of workers located in large cities. The difference between total labor income, denoted by *Wage*, after and before the change in policy is equal to

$$
\begin{aligned}
Wage^1 - Wage^0 &= [M^1_{big}\bar{w}^1_{big} + NM^1_{small}\bar{w}^1_{small}] - [M^0_{big}\bar{w}^0_{big} + NM^0_{small}\bar{w}^0_{small}] \\
&= [(M^0_{big} + \Delta M)\bar{w}^1_{big} + (NM^0_{small} - \Delta M)\bar{w}^1_{small}] - [M^0_{big}\bar{w}^0_{big} + NM^0_{small}\bar{w}^0_{small}] \\
&= \Delta M(\bar{w}^1_{big} - \bar{w}^1_{small}) + M^0_{big}(\bar{w}^1_{big} - \bar{w}^0_{big}) + NM^0_{small}(\bar{w}^1_{small} - \bar{w}^0_{small}).
\end{aligned}
$$

$$(5.1)$$

It is easy to verify that, if average wages in small and large cities are policy invariant, i.e. $\bar{w}^1_{big} = \bar{w}^0_{big}$ and $\bar{w}^1_{small} = \bar{w}^0_{small}$, the last line of Identity (5.1) simplifies to

$$
Wage^1 - Wage^0 = \Delta M(\bar{w}^0_{big} - \bar{w}^0_{small}).
$$

$$(5.2)$$

It follows that the change in total labor income is given by the average city-size wage premium that is currently observed 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) the direct effect (D).

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

---

[44]Extending the policy analysis to total labor income in the economy—assuming a working life of exactly 40 years for all workers—delivers very similar results.

be further re-arranged as,

$$
Wage^1 - Wage^0 =
$$

$$
\Delta M(\bar{w}^1_{big} - \bar{w}^1_{small}) + M^0_{big}(\bar{w}^1_{big} - \bar{w}^0_{big}) + NM^0_{small}(\bar{w}^1_{small} - \bar{w}^0_{small}) +
$$

$$
\underbrace{\Delta M(\bar{w}^0_{big} - \bar{w}^0_{big})}_{=0} - \underbrace{\Delta M(\bar{w}^0_{small} - \bar{w}^0_{small})}_{=0}
$$

$$
= \Delta M(\bar{w}^0_{big} - \bar{w}^0_{small}) + M^1_{big}(\bar{w}^1_{big} - \bar{w}^0_{big}) + NM^1_{small}(\bar{w}^1_{small} - \bar{w}^0_{small}) \tag{5.3}
$$

$$
= \Delta M(\bar{w}^0_{big} - \bar{w}^0_{small}) + M^1_{big}(\bar{w}^1_{big} - \bar{w}^0_{big}) + (\underbrace{NM^1_{small} - M^1_{big} + M^1_{big}}_{=0})(\bar{w}^1_{small} - \bar{w}^0_{small})
$$

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

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

5.3   Results

I describe 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 $\gamma_{large} = 1/1.474$ with $\gamma_{large} = \{1/1.489, 1/1.504\}$. Intuitively, under looser housing regulation, large cities gain in population, as they become 10.4% and 20.7% larger, respectively. 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 goes from 29%, in the baseline economy, to 28% and 27.1%, as γ_{large} shrinks. Interestingly, the equilibrium human capital distributions deteriorate in *both* large and small cities. In Figure 10, I show the deviation of those distributions from the distributions in the baseline economy. The green line with circles (purple 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 locate in large cities under lower housing restrictions are more skilled than those who remain in small cities, but less than those who locate in large cities in the baseline economy. Next, I explore
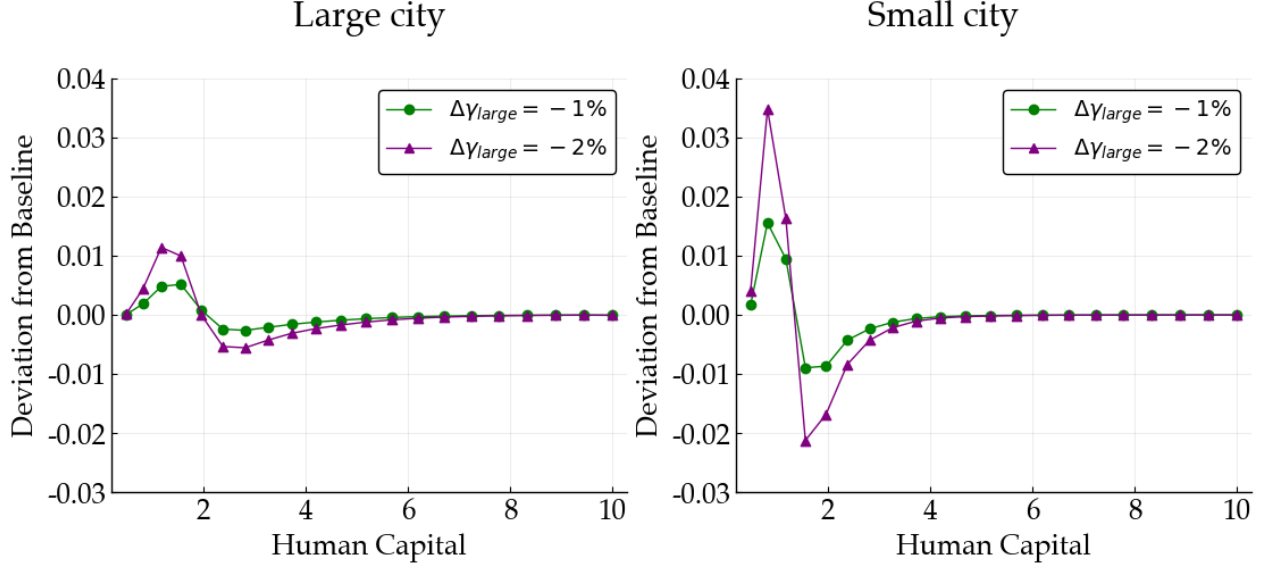
Figure 10: Difference between the human capital distribution under $\Delta\gamma_{big} = -0.01\%$ (green line with circles) and $\Delta\gamma_{big} = -0.02\%$ (purple line with triangles), with respect to its value in the baseline economy. Left panel: large city. Right panel: small city.

the aggregate implications of the change in city size and composition, as measured by variations in total labor income.

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

$$\underbrace{\frac{Wage^1 - Wage^0}{Wage^0}}_{\text{Total}} =$$

$$\underbrace{\frac{\Delta M(\bar{w}^0_{big} - \bar{w}^0_{small})}{Wage^0}}_{(D)} + \underbrace{\frac{M^1_{big}[(\bar{w}^1_{big} - \bar{w}^1_{small}) - (\bar{w}^0_{big} - \bar{w}^0_{small})]}{Wage^0}}_{(P)} + \underbrace{\frac{M(\bar{w}^1_{small} - \bar{w}^0_{small})}{Wage^0}}_{(S)}, \qquad (5.4)$$

which includes the same terms as Identity (5.3), divided by $Wage^0$ in order to obtain percentage deviations from the baseline. Figure 11 shows the overall effect (black bars) and its components (grey bars). The direct effect is equal to +1.8% (+3.55%), while the total effect is +0.65% (+1.25%), when γ_{large} is lowered by 1% (2%). Intuitively, the direct effect (D) is necessarily positive: since large cities currently pay higher wages, increasing their size contributes to raising total labor 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 to the baseline economy, and so do peer effects in those cities. In addition, increasing returns to scale in the search process and in the flow of ideas have an adverse effect on the average match quality, and on the frequency of interactions, in a
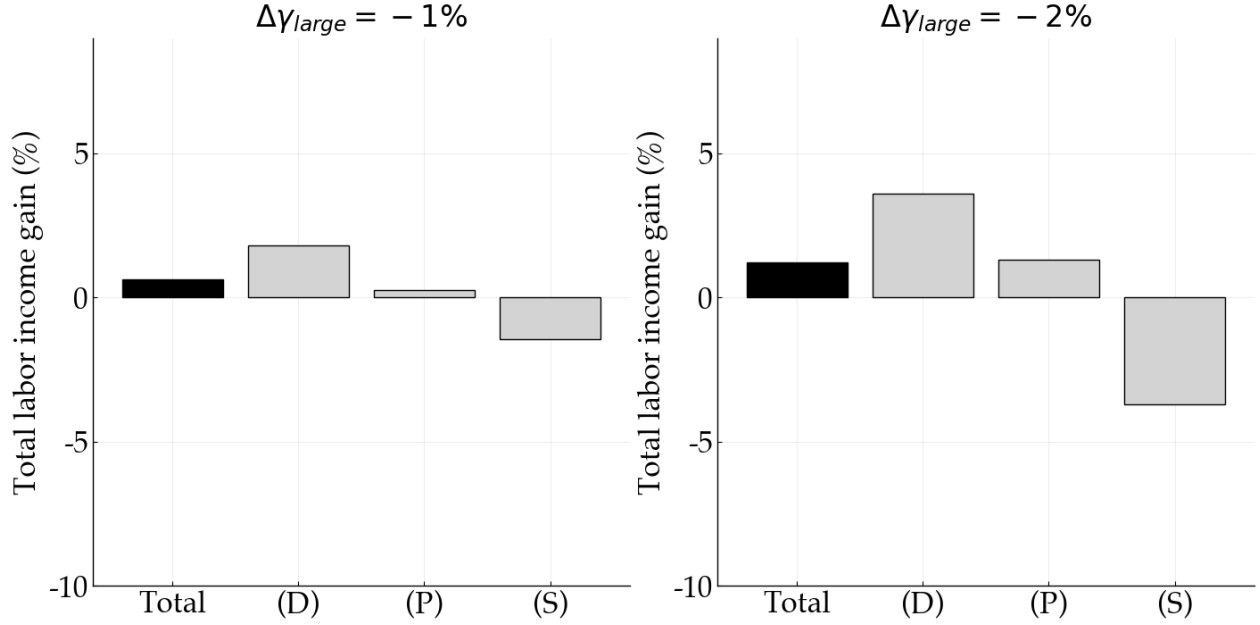
Figure 11: Percentage change in total labor income (black bar), and in its components (grey bars). Left panel: $\Delta\gamma_{big} = -1\%$. Right panel: $\Delta\gamma_{big} = -2\%$.

city that shrinks in size. All these forces contribute to the negative sign of the term denoted by (S). Because of weaker sorting and worse peers, wages in large cities *decline* as well. The contribution of agglomeration forces—matching and flow of ideas—is positive in a city that expands. Hence, the average wage in large cities declines relatively less than in small ones, driving up the city-size wage premium (P). All in all, even accounting for the equilibrium response of agglomeration forces, peer effects, and sorting, relaxing housing restrictions in large cities generates non-negligible income gains. Yet, assuming that productivity dispersion across locations was invariant to the policy would overstate the percentage increase in income by more than a factor of 2.5.

6   Optimal Allocation

In the previous section, I focused on the income gains generated by a change in housing regulation, but avoided taking a stand on its welfare implications. A proper welfare analysis of this type of policy would require the inclusion of housing wealth and political economy considerations that go beyond the scope of this paper. However, even abstracting from housing policy—i.e. taking housing supply elasticities as fixed parameters—one might wonder whether a social planner could still improve on the equilibrium allocation. In fact, the equilibrium of this economy may not be optimal because of the existence of three potential sources of inefficiencies.

First, increasing returns to scale in the search process and in the technology of knowledge

diffusion create a feedback effect—that workers do not internalize—between location decisions and the meeting rates, λ, λ^* and σ. This process gives rise to an *agglomeration externality*. Since λ, λ^* and σ are constant elasticity functions, the increase in the number of meetings in a city is associated with a proportional reduction in another city. Kline and Moretti (2014) show that if the externality in productivity has constant elasticity, the gains from allocating an additional worker to one location are exactly offset by losses in another location. Such neutrality does not need to hold in the present environment, since constant elasticity with respect to city size is a property of the meeting rates, but not necessarily of productivity. Besides, Fajgelbaum and Gaubert (2019) show that the result in Kline and Moretti (2014) does not hold in an economy characterized by either compensating differentials across locations—as in the presence of non-tradeable goods—or sorting of heterogenous agents.

Second, workers do not internalize the benefit they have on each other through the process of knowledge diffusion. The *knowledge spillover externality* in this paper has the same nature as the externality in Lucas and Moll (2014). In their paper, agents learn from everyone in the economy, and choose how to allocate their time between producing and learning. The inefficiency takes the form of workers' underinvestment in their own human capital. In contrast, in the spatial economy I consider, the externality originates from workers' location choice, so that even those who cannot learn anymore, i.e. 'old' workers, might behave suboptimally.

Third, because of the assumption of random search in the labor market, workers who live in a certain location, and choose which jobs to take, do not internalize the cost paid by firms to post the vacancies they might come into contact with (*vacancy creation externality*). Specifically, workers do not take into account the fact that such cost varies by location and, in efficiency units, employment status. Recall that an unemployed worker in city i, where a vacancy costs k_i, meets a firm in city $-i$, where a vacancy costs k_{-i}, at a rate that is only ρ^* times as large as in the local labor market. Similarly, the relative search efficiency of employed—compared to unemployed—workers is equal to ρ. To provide a concrete example, since $k_{large} > k_{small}$ and $\rho^* < 1$, the search technology implies that, everything else equal, the planner allocates workers to larger cities with a lower probability than in the equilibrium.

Motivated by the existence of these externalities, in this section I consider the problem of a social planner who aims at maximizing the present discounted value of total output in the economy, net of housing and migrations costs. In the present environment, under the assumption that the planner can redistribute resources using lump-sum taxes and transfers, this maximization delivers the constrained-efficient allocation of workers to cities and jobs.[45] The planner is subject to the

---

[45]Constrained-efficient allocations might still differ from each other with respect to the distribution of consumption.

same mobility and labor market frictions as the agents in the economy. In particular, the planner chooses the reservation match quality in a firm-worker meeting, the cutoff cost in workers' migration decisions, and the number of vacancies to post in each city. I also assume that the planner is a price taker in the housing market. Price taking can be rationalized by the standard assumption in the urban literature that the aggregate housing supply function—introduced in Section 2—is the result of profit maximization by a continuum of perfectly competitive *absentee* landlords.

In the remainder of this section, I first set up the planner's problem and transform it into an equivalent but tractable one: finding the marginal social value of an unemployed worker, and of an employment relationship of given quality. I then characterize the optimal steady-state allocation, show how it departs from the equilibrium, and compare it to the optimal spatial allocation found in the existing literature.

6.1 The Marginal Social Value of a Worker

I solve the problem of a social planner that maximizes the present discounted value of total net output. Let $\tilde{u}(h_\ell, a, e, i, z, t)$ be the net flow payoff of a worker in city i, employed at match quality z (or unemployed, if $z = 0$), with human capital h_ℓ, age a, and education e at time t:

$$\tilde{u}(h_\ell, a, e, i, z, t) = (b_i \mathbb{I}\{z = 0\} + z)h_\ell - q^e p_i(t) - \lambda^*_{\mathbb{I}\{z>0\}, -i}(t) \int_{\underline{z}}^{\bar{z}} \int_{-\infty}^{max\{x(i,z,\hat{z},t), x(i,z,0,t)\}} c \, dD(c) dF(\hat{z}).$$

(6.1)

The first term on the RHS of Equation (6.1) is the flow value from production, or unemployment. The second term is the cost of housing. The third term is the actual cost of moving from city i to city $-i$, which is given by the rate at which a worker receives an offer from city $-i$, $\lambda^*_{\mathbb{I}\{z>0\}, -i}(t)$, multiplied by the expected migration cost, conditional on migrating. Notice that the expectation is taken with respect to both the quality of the new potential match, $\hat{z}$, and the realization of the migration cost, c. The max operator in the integral is due to the worker's choice between moving as employed or unemployed to city $-i$.[46] In light of the microfoundation of the firm-worker meeting rates introduced above, $\lambda^*_{\mathbb{I}\{z>0\}, -i}(t) = (\mathbb{I}\{z = 0\} + \rho \mathbb{I}\{z > 0\})\rho^* \theta_{-i}(t)^\zeta M_{-i}(t)^{\chi_M}$. The total flow payoff, u, generated by a worker of type (h_ℓ, a, e, i, z) is equal to the sum of the net flow payoff in Equation (6.1), and the vacancy posting cost associated to hiring such worker. The cost of posting $v_i(t)$ vacancies in city i at time t is equal to $k_i v_i(t) = k_i \theta_i(t) s_i(t)$. The definition of the effective measure of workers seeking a job in city i at time t, $s_i(t)$, is given by the time-varying equivalent of Equation (4.1). Taking into account the relative search efficiency on and off the job, within and

---

[46]The usual omission of the state variables (h_ℓ, a, e) has been applied.

across cities, I obtain

$$u(h_\ell, a, e, i, z, t) = \tilde{u}(h_\ell, a, e, i, z, t) - (k_i \theta_i(t) + \rho^* k_{-i} \theta_{-i}(t))(\mathbb{I}\{z = 0\} + \rho \mathbb{I}\{z \neq 0\}). \quad (6.2)$$

The planner solves the following problem,

$$\mathcal{W}(\phi(\cdot, t)) = \max_{R^p(\cdot, \tau), x^p(\cdot, \tau), \theta^p(\cdot, \tau)} \int_t^\infty e^{-r\tau} \mathbb{E}_{\phi(\cdot, \tau)}[u(\cdot, \tau)] d\tau$$

$$\text{s.t. } \frac{\partial \phi(\cdot, \tau)}{\partial t} = \Gamma(\phi(\cdot, \tau), R^p(\cdot, \tau), x^p(\cdot, \tau), \theta^p(\cdot, \tau)), \quad (6.3)$$

where the expectation is taken with respect to the cross-sectional distribution of workers at each $\tau \geq t$. The policy functions in the planner's problem are characterized by the superscript $'p'$ in order to distinguish them from their equilibrium counterparts. Notice that the measure of workers and house prices in each city can be derived from $\phi(\cdot, t)$ using Equations (2.10) and (2.15). Hence, the state variable in Problem (6.3) can be parsimonously described by the distribution ϕ. The constraint in the maximization problem is given by the law of motion of ϕ, $\Gamma(\phi, R^p, x^p, \theta^p) \equiv \tilde{\Gamma}(\phi, R^p, x^p, \tilde{\lambda}(\theta^p), \tilde{\lambda}^*(\theta^p))$, where $\tilde{\Gamma}$ is defined by Equation (2.16), and $\tilde{\lambda}(\theta)$ and $\tilde{\lambda}^*(\theta)$ can be obtained using the microfoundation of market tightness introduced in Section 4.[47]

In order to solve for the optimal allocation, I follow the approach proposed by Lucas and Moll (2014) and transform the infinite-dimensional planner's problem into a standard system of Hamilton-Jacobi-Bellman and Kolmogorov Forward Equations. The details of the derivation and the formal characterization of the HJB and KF Equations can be found in Appendix C. The KFEs are the same as Equations (2.8) and (2.9), which define the law of motion of ϕ in the decentralized equilibrium. The HJBEs describe the marginal social value of each type of agent in the economy. They differ from Equations (2.1) and (2.5) due to the presence of additional terms that capture the externalities in the economy. In what follows, I show a stylized representation of the HJBE that characterizes the marginal social value of an unemployed worker, and highlight how it differs from the corresponding HJBE in the laissez-faire equilibrium.

Let $U^p(h_\ell, a, e, i, t)$ and $V^p(h_\ell, a, e, i, z, t)$ be the marginal social value of an unemployed worker, and of a firm-worker match, respectively, of type (h_ℓ, a, e, i, z) at time t. U^p satisfies the following

---

[47]For example, $\lambda_{0,i} = \chi M_i^{\chi_M} = \theta_i^{\frac{1}{\zeta}} M_i^{\chi_M}$. Hence, $\tilde{\lambda}_{0,i}(\theta_i) = \theta_i^{\frac{1}{\zeta}} M_i^{\chi_M}$. Also $\tilde{\lambda}_{0,i}^*(\theta_i) = \rho^* \theta_i^{\frac{1}{\zeta}} M_i^{\chi_M}$.

HJBE,

$$
rU^p(h_\ell, a, e, i, t) = \max_{R^p, x^p, \theta^p} RHS_U(U^p(h_\ell, a, e, i, t), \beta = 1) +
$$

$$
\frac{\partial \lambda_{0,i}}{\partial M_i} \mathbb{E}_{\phi(i)}[\Delta_z U^p] + \frac{\partial \lambda^*_{0,i}}{\partial M_i} \mathbb{E}_{\phi(-i)}[\Delta_z U^p] + \frac{\partial \sigma_i}{\partial M_i} \mathbb{E}_{\phi(i)}[\Delta_{h_\ell} U^p] +
$$

$$
\sigma_i \frac{\partial \mathbb{E}_{\phi(i), G_i}[\Delta_{h_\ell} U^p(y)]}{\partial \phi(h_\ell, a, e, i, t)} - (k_i \theta_i^p + \rho^* k_{-i} \theta_{-i}^p) +
$$

$$
\psi_o \mathbb{E}_{g^h_{o,i}(\tilde{h}|\tilde{e})} \left[\frac{\partial g^e_{0,i}(\tilde{e})}{\partial \phi(h_\ell, a, e, i, t)} U^p \right] \mathbb{I}\{a = o\} + \frac{\partial U^p(h_\ell, a, e, i, t)}{\partial t} \cdot p
$$

(6.4)

The LHS of Equation (6.4) is the annuitized marginal social value of an unemployed worker. The first line on the RHS is given by the RHS of Equation (2.1), and it is equal to the annuitized private value of an unemployed worker, if the worker captures the entire gains from trade from the formation of a match ($\beta = 1$). The second line describes the agglomeration externality. An additional worker in city i increases the rate at which firms and workers meet in the labor market, within and across cities, and the frequency at which workers interact with each other. The marginal changes in meeting rates multiply the average expected gains accruing to all the other workers that experience any of those events. The first term on the third line represents the knowledge spillover externality toward other (young) workers located in city i. This externality operates through the dependence of the rate of human capital accumulation, conditional on interacting with other workers, on the equilibrium human capital distribution in city i, G_i (see Equation 2.14).[48] The second term in the third line, $-(k_i \theta_i^p + \rho^* k_{-i} \theta_{-i}^p)$, corresponds to the vacancy creation externality. It is equal to the marginal cost of creating vacancies that might come into contact with an unemployed worker in city i. The first term in the last line represents the effect of the location of an old worker on the location—and education—of the newborn he is replaced by when he receives the retirement shock ψ_o.[49] The last term is the time derivative of U^p.

The HJBE that characterizes the marginal social joint value of a match is identical to Equation (6.4), once $RHS_U(U^p, \beta = 1)$ is replaced by $RHS_V(V^p, \beta = 1)$—that is the RHS of Equation (2.5)—and $(k_i \theta_i^p + \rho^* k_{-i} \theta_{-i}^p)$ is pre-multiplied by ρ, since employed workers are only ρ times as

---

[48]To avoid repeating conceptually similar terms, I only show terms with $\Delta_z U^p$ and $\Delta_{h_\ell} U^p$, on the second and third lines of Equation (6.4), but it is understood that an unemployed worker exerts the same type of agglomeration and knowledge spillover externalities on employed workers as well.

[49]This component of the marginal social value derives from the OLG structure of the economy, and not from the interaction between economic agents. This is why I do not list and discuss it among the externalities at the beginning of this section. Quantitatively, I find that it does not play a significant role in shaping the discrepancy between the optimal and equilibrium allocations.

likely to contact a firm as the unemployed.

The solution to the planner's problem is given by the set of policy functions (R^p, x^p, θ^p) that satisfy

$$V^p(h_\ell, a, e, i, R^p(h_\ell, a, e, i, t), t) = U^p(h_\ell, a, e, i, t) \tag{6.5}$$

$$x^p(h_\ell, a, e, i, z, z^*, t) = V^p(h_\ell, a, e, -i, z^*, t) - V^p(h_\ell, a, e, i, z, t) \tag{6.6}$$

$$k_i = \zeta \theta_i^p(t)^{\zeta - 1} \mathcal{S}^p(i, t). \tag{6.7}$$

Equations (6.6) and (6.5) are identical to the corresponding optimality conditions in the decentralized equilibrium. Equation (6.7) is analogous to Equation (4.2), except for the fact that the values U and V, and the surplus S, inside $\mathcal{S}$ are replaced by their optimal analogues, and the average meeting rate with a worker is replaced by the marginal one. However, the average and marginal rates are equal to each other, given the assumption $\zeta = 1 - \beta$. Clearly, the values of the optimal policy functions need not be the same as the equilibrium ones, since the marginal social and private vaues might differ from each other. Determining how the optimal allocation departs from the equilibrium one is what I turn to next.

6.2 Optimal Allocation and Policy

I solve the planner's problem and compute the optimal steady-state allocation for the economy. The denotation of a steady state implies that this is the time-invariant allocation the planner would choose $\forall \tau > t$ if the state of the economy was equal to the steady-state optimal distribution at a certain time t. Under the implicit assumption that the optimal allocation would converge to the same steady state for any value of the initial distribution, it is informative to compare the characteristics of the optimal allocation to the steady-state equilibrium of the economy.[50]

In the optimal allocation, large cities are 11.4% smaller than in the equilibrium, while small cities expand by 18.5%. The lower dispersion in city size is accompanied by a change in the optimal *composition* of large cities, which occurs along three main dimensions. First, the optimal fraction of college graduates is equal to 39%, compared with 29% in the equilibrium (R1). Second, educational sorting is particularly pronounced among old workers. In the equilibrium, the college share in large cities is almost constant with respect to age, while in the optimal allocation it is equal to 35% among young workers, and 45% among the old (R2). Third, 56% of workers in large cities are

---

[50]A milder requirement is that the optimal steady-state allocation would be part of the solution to a planner's problem in which the state variable $\phi(\cdot, t)$ was equal to the equilibrium steady-state distribution. Verifying this conjecture is certainly an important exercise to perform.
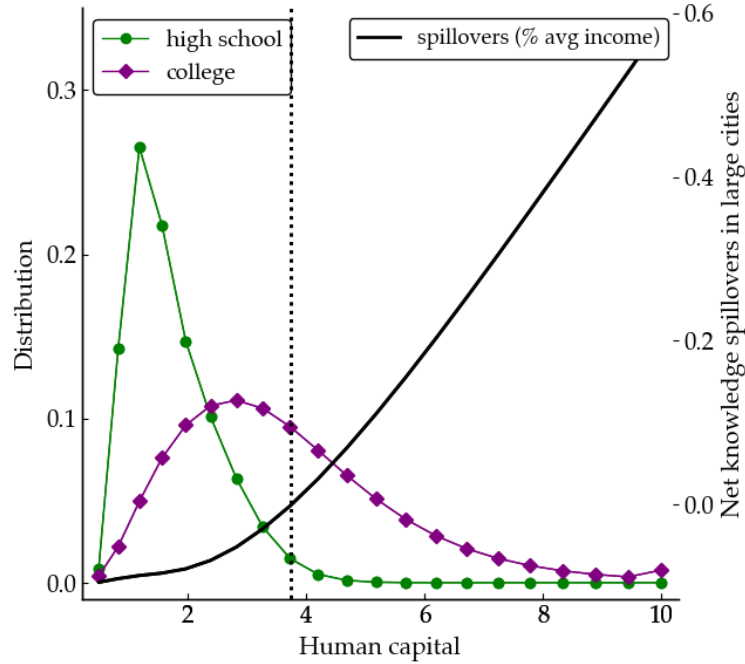
Figure 12: Left y-axis: human capital distribution of high school (gree line with circles) and college graduates (purple line with diamonds). Right y-axis: difference between knowledge spillover externalities in large and small cities as a percentage of average income in the economy (black solid line). The dotted line corresponds to the value of human capital at which the black solid line is equal to 0 (i.e. same externality in large and small cities).

young, compared with an equilibrium value of 51% (R3).

In order to understand the first result, R1, Figure 12 shows the human capital distribution of high school and college graduates in the economy, under the optimal allocation (left y-axis). On the same graph, I plot the difference between the knowledge spillover externality generated by a worker in a large and in a small city, as a function of his human capital (right y-axis).[51] I define this difference the net knowledge spillovers in large cities. Knowledge spillovers is the external-ity that varies the most with a worker's education—through the effect of learning ability on his equilibrium human capital.[52] Within each city, this externality is given by the combination of two ingredients. On the one hand, a worker transmits knowledge to the young. Such effect is positive and increasing in the worker's human capital. On the other hand, that same worker reduces the rate at which the young interact with—hence learn from—every other worker in the city. This congestion component is always negative and independent of the worker's type. Therefore, the

---

[51]See the first term on the third line of Equation (6.4). A formal definition of the knowledge spillover externality can be found in Appendix C.

[52]The other externality that varies by education is given by the dependence of newborns' education (hence initial human capital) on the education of the old workers they replace. Quantitatively, such intergenerational channel has a much smaller effect on the optimal allocation, compared with the effect of the knowledge spillover externality.

higher a worker's human capital, the higher the knowledge spillover externality he generates, *in every city*. However, the rate at which ideas are diffused, and the return to human capital through the formation of better labor market matches, are increasing in city size. It follows that high-skilled workers are disproportionally more beneficial to others when they are located *in large cities*. The dotted line corresponds to the human capital level at which workers generate the same knowledge spillover externality in both cities. It is easy to see that the human capital distribution of college graduates first order stochastically dominates the distribution of high school graduates, which implies that college graduates are more likely to be on the right side of the dotted line. That is, they are more likely to provide net positive knowledge spillovers in large cities.[53]

In light of the argument behind R1, R2 can be explained by the *life-cycle* increase in human capital inequality between education groups. Since college graduates learn at a higher rate than high school graduates, the educational gap in the ability to generate knowledge spillovers grows with experience. Last, notice that the optimal large city is much richer in human capital, compared with the equilibrium one. This creates stronger incentives for young workers to move to large cities, which explains R3.

To gain additional insights into the heterogeneous externality each worker generates, I compute a simple type-specific optimal policy that can be readily obtained by comparing the private to the social marginal value of a worker. Specifically, I find the unique flow transfer (up to a lump-sum tax levied on all workers) that equates the solutions to the HJBEs (2.1) and (6.4), and I perform the same exercise for employed workers. As illustrated by Equations (6.5)-(6.7), the equality between the equilibrium and the optimal value functions guarantees that the decentralized and the optimal policy functions take the same values, and so does the steady-state allocation. While the absolute level of the transfers is not particularly informative, it is instructive to observe how they vary across cities, for different types of workers. In particular, I compute the population-weighted average subsidy by worker's age, education, and location. I then compute the difference in such average transfers by age and education, between large and small cities, and express them as a percentage of the average per capita output in the economy.

Under the optimal policy, the transfer received by college graduates that locate in large cities is higher than in small cities by an amount equal to 12.5% of average per capita output. To the opposite, the planner would subsidize high school graduates that choose to locate in small cities, as they would receive an additional 4.1% of average output compared to those who live in large cities. Conditional on education, the size of the average transfer is unaffected by the age of the

---

[53]In turn, the presence of more college graduates, who have high learning ability, provides an additional explanation for the disproportionally larger spillovers generated by high-skilled workers in large cities.

worker.[54]

The policy prescriptions in this paper are closest to Rossi-Hansberg, Sarte, and Schwartzman (2019), and they somewhat differ from those in Fajgelbaum and Gaubert (2019).[55] Fajgelbaum and Gaubert (2019) consider a multiple location, heterogeneous agent, static economy, characterized by externalities in both production and amenities. They find that a social planner would in fact re-allocate workers, and, in particular, college graduates, toward smaller cities. This is because small cities host fewer college graduates, and such scarsity makes them particularly valuable in those locations. Rossi-Hansberg, Sarte, and Schwartzman (2019) build a model in the spirit of Fajgelbaum and Gaubert (2019), augmented with multiple industries and input-output linkages. They divide workers according to whether they do—or do not—perform cognitive non-routine tasks (CNR vs. non-CNR). They estimate negative cross-type externalities, and find that it is optimal to reallocate non-CNR workers outside of large cities. Similarly to the non-CNR workers in Rossi-Hansberg, Sarte, and Schwartzman (2019), high school graduates in my model have lower average levels of human capital, and reduce the intensity of interactions between college graduates.

The externalities in the present paper are only indirectly related to production—through increasing returns to scale in the search process—but they primarily involve the process of human capital accumulation. Since peer effects are endogenous and dynamic, a worker who learns more exerts a more positive externality on other workers. College graduates have higher learning ability and human capital, hence they benefit more from—and also contribute more to—knowledge spillovers in large cities. Notice that, because of vertical learning, in this paper, as in Fajgelbaum and Gaubert (2019), college graduates would be very beneficial to workers who live in a (human capital-poor) small city. Yet, this benefit is quantitatively dominated by the externality generated by college graduates on each other, consistent with the findings in Rossi-Hansberg, Sarte, and Schwartzman (2019).

7   Conclusions

The US, as many other economies, displays a significant and persistent positive correlation between city size and productivity, usually proxied by wages, and in particular for more experienced workers. In this paper, I contribute to the literature that studies the origins of the city-size

---

[54]An age-independent transfer is able to generate large cities that are younger than in the equilibrium. To see why, notice that part of the social benefits from learning are in fact internalized by young workers—since they maximize the presented discounted value of their income. Hence, they do not need to be subsidized in order to have stronger incentives for locating in large cities, once those cities are richer in human capital, compared with the equilibrium.

[55]While they have different implications in terms of *composition*, both these papers—as well as mine—point toward less dispersion in the optimal distribution of city *size*, compared with the equilibrium.

wage premium. I build a life-cycle equilibrium model that jointly allows for heterogeneity in sorting behavior, increasing returns to scale in the labor search process, and spatial differences in the frequency and quality of knowledge diffusion between workers. I find that lower search frictions in the labor market facilitate the formation of better matches, which in turn generate a positive level effect on wages in large cities. The contribution of knowledge diffusion to the city-size wage premium emerges over the life cycle, mainly because of a better composition of peers, and to a lesser extent because of more frequent interactions between workers. By determining workers' migration decisions, better match quality and learning opportunities generate positive sorting of workers into large cities. In turn, sorting contributes to the equilibrium heterogeneity in the quality of peers across cities. Quantitatively, I find that matching generates a constant wage premium of 11%, while the role played by sorting and knowledge diffusion grows over the life cycle, up to 12% and 15%, respectively.

Throughout the paper, I stress how the mapping from heterogeneity in city characteristics into observed wage profiles is the result of the interaction between multiple channels that are dynamic in nature, and cannot be readily recovered from the data. Furthermore, a worker's location decision in the model is affected by a number of unobservable variables, like his level of human capital, quality of current and perspective matches, and idiosyncratic moving cost. Since the estimation targets only the *aggregate* differences in labor market outcomes between small and large cities, I use *micro evidence* on wages of movers and stayers to validate the proposed mechanisms behind the city-size wage premium. I show that the model generates incentives to move to—and from—large cities that are consistent with the empirical evidence on selection into, and the return to, migration. Controlling for job tenure affects the comparison between movers and stayers, but it does so in a virtually identical fashion in the model as in the data. This finding highlights the importance of accounting for labor market frictions, and match heterogeneity, in measuring the city-size wage premium, and understanding workers' location choice.

I conclude by pointing out the traditional disconnect between the studies that investigate the nature of spatial wage differentials, and those that explore their aggregate implications. As I microfound and measure the sources of the city-size wage premium in an equilibrium environment, I address and combine both of these literatures.

First, I contribute to the debate on the aggregate consequences of enacting place-based policies that trigger the relocation of workers toward more productive locations. In the context of an increase in housing supply—e.g. through a change in land-use regulation—I show that accounting for endogenous productivity differences between cities is crucial in order to asses the equilibrium response of the economy. A hypothetical scenario, characterized by exogenous productivity differences between locations, would overstate the labor income gains associated to an expansion of

large cities by more than a factor of 2.5.

Second, I compute and characterize the extent to which the optimal spatial allocation of workers differs from the equilibrium one. I highlight how heterogeneity between workers in private and social benefits from knowledge spillovers would induce a social planner to increase the concentration of high-skilled, college educated, workers into large cities. This result requires a note of caution, in light of the implicit assumption that the gains from improving on the equilibrium allocation can be redistributed without distorsions. It is certainly worth exploring how a more realistic set of policies would address the trade-off between the aggregate benefits from spatial sorting and the potential increase in inequality between places (and people).

Nesting the dynamic aspects of this paper into a conventional urban system, with a larger number of cities, is certainly an intriguing but challenging task. From a modeling perspective, the fact that the human capital distribution of cities is an equilibrium object poses non-trivial computational challenges. On the empirical side, estimating a dynamic model with a realistic system of cities would require a large longitudinal sample of workers, with an extensive cross-sectional dimension. Nonetheless, I believe that the margins introduced in the admittedly stylized geography of this paper are likely to apply to richer environments.

Relatedly, even though larger cities pay higher wages, size captures only a portion of the observed dispersion in local productivity. Differences in sectoral composition, for example, have been associated with the diverging fate of cities, like the manufacturing area of Detroit, and the innovative tech hub of San Francisco. In a non-stationary environment, in which places are differentially affected by aggregate trends, the same agglomeration forces behind the expansion of a city might reverse, and accelerate its decline. In such a context, the nature of human capital accumulation, and its transferability across cities and occupations, would have significant distributional implications for workers with different types and levels of experience. Extending the framework in this paper to account for the recent evolution in the geography of productivity across US cities represents a fruitful venue for future research.

References

Achdou, Yves, et al. 2017. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach". *NBER Working Paper 23732*.

Akcigit, Ufuk, et al. 2018. "Dancing with the Stars: Innovation Through Interactions". *NBER Working Paper No. 24466*.

Bagger, Jesper, et al. 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". *The 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. 2019. "The Global Diffusion of Ideas". *Econometrica (forthcoming)*.

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. 2008. "Spatial wage disparities: Sorting matters!" *Journal of Urban Economics* 63 (2): 723–742.

— . 2010. "The identification of agglomeration economies". *Journal of Economic Geography* 11 (2): 253–266.

Combes, Pierre-Philippe, et al. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection". *Econometrica* 80 (6): 2543–2594.

Dauth, Wolfgang, et al. 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. 2019. "City of Dreams". *Working Paper*.

De La Roca, Jorge, and Diego Puga. 2016. "Learning by Working in Big Cities". *The Review of Economic Studies* 84 (1): 106–142.

Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies". *Handbook of Regional and Urban Economics* 4:2063–2117.

Eckert, Fabian, Mads Hejlesen, and Conor Walsh. 2019. "The Return to Big City Experience: Evidence from Danish Refugees".

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

Fajgelbaum, Pablo, and Cecile Gaubert. 2019. "Optimal Spatial Policies, Geography and Sorting". *NBER Working Paper 24632*.

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 Perspective* 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". *The 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". *The 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 Influence of Unemployment on the Wage Bargain". *American Economic Review* 98 (4): 1653–1674.

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.

Herkenhoff, Kyle F., et al. 2018. "Production and Learning in Teams". *NBER Working Paper No. 25179*.

Hosios, Arthur J. 1990. "On the Efficiency of Matching and Related Models of Search and Unemployment". *The Review of Economics Studies* 57 (2): 279–298.

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". *The 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". *The Quarterly Journal of Economics* 108 (3): 577–598.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg. 2019. "Learning from Coworkers". *NBER Working Paper No. 25418*.

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". *The Quarterly Journal of Economics* 129 (1): 275–331.

Lise, Jeremy, and Fabien Postel-Vinay. 2019. "Multidimensional Skills, Sorting, and Human Capital Accumulation". *HHEI working paper 131*.

Lucas, Robert E. Jr. 2009. "Ideas and Growth". *Economica* 76 (301): 1–19.

— . 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". *The Quarterly Journal of Economics* 122 (3): 1103–1144.

Martellini, Paolo, and Guido Menzio. 2018. "Declining Search Frictions, Unemployment and Growth". *NBER Working Paper No. 24518*.

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". *The 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.

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". *The Quarterly Journal of Economics* 125 (3): 1253–1296.

Schmutz, Benoît, and Modibo Sidibé. 2019. "Frictional Labour Mobility". *The 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.

Topel, Rober H, and Michael P Ward. 1992. "Job Mobility and the Careers of Young Men". *The Quarterly Journal of Economics* 107 (2): 439–479.

Appendix

A Wage contracts

Let $W(h_\ell, a, e, i, z, \omega)$ be the present discounted value of a worker of human capital h_ℓ, age a, education e, who lives in city i, is employed at match quality z, and is being paid a piece rate ω. It is instructive to highlight the discrepancy between W and the joint value of the match at which the worker is employed, V. Hence, the HJBE that characterizes the value function W can be written as

$$rW(h_\ell, a, e, i, z, \omega) = rV(h_\ell, a, e, i, z) - (1 - \omega)zh_\ell +$$
$$[\sigma_i \kappa(G_i, h_\ell, e) + \eta^e exp(-\eta h_\ell)][(W(h_{\ell+1}) - W) - (V(h_{\ell+1}) - V)]\mathbb{I}\{a = y\} +$$
$$\psi_h[(W(o) - V(o))\mathbb{I}\{a = y\} - (W - V)] + \delta^e(V - W) + \mathcal{W}(i) + \mathcal{W}^*(i).$$
$$(A.1)$$

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 represents the fact that the worker only receives a fraction ω of the flow product zh_ℓ. 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 $rV(h_\ell, a, e, i, z)$. An analogous intuition applies to the aging process described by the first term on the third line. The joint value of a match accounts for the event of job destruction, but workers only lose the portion of the match value that they were receiving. Since firms lose $(V - W)$ when the job is destroyed, $\delta^e(V - W)$ is 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 $\mathcal{W}(i)$ is given by

$$\mathcal{W}(i) = \lambda_{1,i}\left[(V - W)(1 - F(z)) + \int_{z^f(\omega)}^z (V(\hat{z}) - W)dF(\hat{z})\right].$$

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 value of the old match that was accruing to the previous employer before the workers' job-to-job transition. 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 $\hat{z} < z$, but such that $V(h_\ell, a, e, i, \hat{z})$ is higher

than the current present discounted value to the worker, $W(h_\ell, a, e, i, z, \omega)$. The minimum match quality that triggers a wage increase is given by the function $z^f(h_\ell, a, e, i, z, \omega)$, which is implicitly defined by

$$V(h_\ell, a, e, i, z^f(h_\ell, a, e, i, z, \omega)) = W(h_\ell, a, e, i, z, \omega).$$

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 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 also receive a fraction β of the difference between the joint value of his current and the new potential match.

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

$$
\begin{aligned}
\mathcal{W}^*(i) = \lambda_{1,-i}^* \{ & D(x(z,0))[U(-i) - W] + \\
& (D(x(z,0,\omega)) - D(x(z,0)))[U(-i) - W - \mathbb{E}(c | x(z,0) < c < x(z,0,\omega))] \} F(R(-i)) + \\
& \lambda_{1,-i}^* \int_{R_{-i}}^{\bar{z}} (D(x(z,\hat{z})) - D(x(z,0)))(V(-i,\hat{z}) - W) + D(x(z,0))(V - W) + \\
& (D(x(z,\hat{z},\omega)) - D(x(z,\hat{z})))[(V(-i,\hat{z}) - W) - \mathbb{E}(c | x(z,\hat{z}) < c < x(z,\hat{z},\omega))]dF(\hat{z}).
\end{aligned}
$$

(A.2)

The first two lines on the RHS of Equation (A.2) represent the additional value the worker receives if moving as unemployed to city $-i$ (first line), and the gain from staying at his current firm, but credibly threatening to quit as unemployed to city $-i$ (second line). Conditional on drawing $\hat{z} < R(h_\ell, a, e, -i)$, a worker moves as unemployed if he draws a migration cost below $x(h_\ell, a, e, i, z, 0)$, defined by Equation (2.6). The threat of moving as unemployed is credible—hence it triggers an increase in wage—if the migration cost is above $x(h_\ell, a, e, i, z, 0)$, but below the cutoff $x(h_\ell, a, e, i, z, 0, \omega)$, given by

$$x(h_\ell, a, e, i, z, 0, \omega) = U(h_\ell, a, e, -i) - W(h_\ell, a, e, i, z, \omega).$$

The last two lines of Equation (A.2) apply the same logic to the case in which the match quality $\hat{z}$ with the new potential employer is high enough to trigger either a job-to-job transition, or an upward revision of the piece rate. In this case, the cutoff cost that induces the worker to credibly ask for a wage renegotiation is equal to

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

B 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 the time periods 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.[56] 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, CZs are assigned to the category "large city" ("small city") if they have population of more (less) than 750 thousand in 1990. This categorization guarantees both substantial heterogeneity in size between large and small cities, and a comparable number of observations from each type of city in the NLSY79.

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

---

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

(high school) degree.

C   Expected surplus and marginal social value of a worker

C.1   Expected surplus

Firms in city i receive a fraction $(1 - \beta)$ of the expected surplus $\mathcal{S}(i)$ 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 $-i$. The formal representation of the expected surplus introduced in Equation (4.3) is given by[57]

$$
\mathcal{S}(i) =
$$

$$
\sum_{h,a,e} \left\{ \int_{R_i}^{\bar{z}} S(i,0,i,\hat{z}) dF(\hat{z}) \phi(i,0) + \int_{\underline{z}}^{\bar{z}} \int_{\tilde{z}}^{\bar{z}} S(i,\tilde{z},i,\hat{z}) dF(\hat{z}) \phi(i,\tilde{z}) d\tilde{z} \rho + \right.
$$

$$
\int_{R_i}^{\bar{z}} \{ (D(x(-i,0,i,\hat{z})) - D(x(-i,0,i,0))) [S(-i,0,i,\hat{z}) -
$$

$$
\mathbb{E}(c|x(-i,0,i,0) < c < x(-i,0,i,\hat{z}))] + D(x(-i,0,i,0)) S(i,0,i,\hat{z}) \} dF(\hat{z}) \phi(-i,0) \rho^* +
$$

(C.1)

$$
\int_{\underline{z}}^{\bar{z}} \int_{R_i}^{\bar{z}} \{ (D(x(-i,\tilde{z},i,\hat{z})) - D(x(-i,\tilde{z},i,0))) [S(-i,\tilde{z},i,\hat{z}) -
$$

$$
\left. \mathbb{E}(c|x(-i,\tilde{z},i,0) < c < x(-i,\tilde{z},i,\hat{z}))] + D(x(-i,\tilde{z},i,0)) S(i,0,i,\hat{z}) \} dF(\hat{z}) \phi(-i,\tilde{z}) d\tilde{z} \rho \rho^* \right\}.
$$

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. Currently employed workers at match quality $\tilde{z}$ would move to a new job inside city i if and only if they draw a match quality $\hat{z} > \tilde{z}$. Notice how the parameter ρ in the second term captures the relative search efficiency of employed workers. The second and third lines show the expected surplus from hiring a currently unemployed worker from city $-i$. Recall that if the worker draws a migration cost $c < x(-i,\tilde{z},i,\hat{z})$, he migrates even without accepting a job. In that case, his threat point in the Nash bargaining protocol is being unemployed in city i. The last two lines follow the same logic, applied to hiring a currently employed worker from city $-i$.

C.2   Derivation of the marginal social value of a workers

The derivation of the marginal social value of a worker closely follows the methodology introduced by Lucas and Moll (2014). In order to numerically compute the optimal allocation, I

---

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

discretize the match quality distribution on a finite set of points. Recall that human capital, age, education, and location are discrete variables as well. Therefore, here I present the discretized planner's problem, and refer the reader to Lucas and Moll (2014) for the treatment of a dynamic programming problem in which the state variable is a continuous distribution. The economic intuition is the same for both approaches, and so is the mathematical derivation—except for the use of functional derivatives in the continuous-state problem.

Let j (or j') index the state of a worker, (h, a, e, i, z), under the usual assumption that a value of $z = 0$ means that the worker is unemployed. Let $\phi(j)$ be the measure of workers of type j. The planner's problem (6.3) in Section 6 can be written in recursive form as

$$rW(\phi) = \max_{R^p(\cdot), x^p(\cdot), \theta^p(\cdot)} \quad \sum_j u(j)\phi(j) + \sum_{j'} \frac{\partial W}{\partial \phi(j')}\Gamma(j'), \tag{C.2}$$

where $\Gamma(j') = \frac{\partial \phi(j')}{\partial t}$. The first order condition with respect to a generic policy function ξ, is given by

$$0 = \sum_j \frac{\partial u(j)}{\partial \xi}\phi(j) + \frac{\partial}{\partial \xi}\left(\sum_{j'} \frac{\partial W}{\partial \phi(j')}\Gamma(j')\right). \tag{C.3}$$

Differentiating Equation (C.2) with respect to $\phi(j)$, and using the envelope condition, I obtain

$$
\begin{aligned}
r\frac{\partial W(\phi)}{\partial \phi(j)} &= u(j) + \frac{\partial}{\partial \phi(j)}\left(\sum_{j'} \frac{\partial W}{\partial \phi(j')}\Gamma(j')\right) \\
r\frac{\partial W(\phi)}{\partial \phi(j)} &= u(j) + \sum_{j'} \frac{\partial^2 W}{\partial \phi(j')\partial \phi(j)}\Gamma(j') + \sum_{j'} \frac{\partial W}{\partial \phi(j')}\frac{\partial \Gamma(j')}{\partial \phi(j)},
\end{aligned} \tag{C.4}
$$

where I used the fact that

$$\frac{\partial^2 W}{\partial \phi(j)\partial \phi(j')} = \frac{\partial^2 W}{\partial \phi(j')\partial \phi(j)}.$$

Define the marginal social value of a worker $U^p(j, \phi) = \frac{\partial W(\phi)}{\partial \phi(j)}$. Plugging $U^p(j, \phi)$ into Equation (C.4) results in

$$r\tilde{U}^p(j, \phi) = u(j) + \sum_{j'} \frac{\partial \tilde{U}^p(j, \phi)}{\partial \phi_{j'}}\Gamma(j') + \sum_{j'} \tilde{U}^p(j', \phi)\frac{\partial \Gamma(j')}{\partial \phi(j)}. \tag{C.5}$$

The next step consists in transforming the high dimensional problem of solving for $\tilde{U}^p(j, \phi)$ into a tractable one, which has the same dimensionality as the individual decision problem in the laissez-faire equilibrium. I define the marginal social value along the optimal trajectory

$$U^p(j, t) = \tilde{U}^p(j, \phi(\cdot, t)) = \frac{\partial W(\phi(\cdot, t))}{\partial \phi(j, t)}. \tag{C.6}$$

The intuition is that, along the optimal trajectory, the information included in the entire cross-sectional distribution can be conveyed by the time dimension. Using the chain rule of differentiation, it follows that

$$\frac{\partial U^p(j,t)}{\partial t} = \sum_{j'} \frac{\partial \tilde{U}^p(j,\phi)}{\partial \phi(j')} \Gamma(j').$$

(C.7)

Combining Equations (C.7) and (C.5), I obtain the HJBE that describes the marginal social value of a worker of type j,

$$rU^p(j,t) = u(j) + \sum_{j'} U^p(j',t)\frac{\partial \Gamma(j')}{\partial \phi(j)} + \frac{\partial U^p(j,t)}{\partial t}.$$

(C.8)

The term $\sum_{j'} U^p(j',t)\frac{\partial \Gamma(j')}{\partial \phi(j)}$ captures how a change in the measure of workers of type j is associated with a change in the marginal value of all workers of type j', weighted by the effect of $\phi(j)$ on $\Gamma(j')$. Hence, this term includes the externalities that worker j creates on every other worker in the economy.

Using the law of motion of the distribution ϕ (Equations 2.8 and 2.9), I obtain the following HJBE,

$$
\begin{aligned}
rU^p(h_\ell, a, e, i, t) = {} & RHS_U(U^p, \beta = 1) + \\
& \text{agglomeration (matching)} + \text{agglomeration (flow of ideas)} + \\
& \text{knowledge spillovers} + \text{vacancy creation} + \text{OLG} + \\
& \frac{\partial U^p(h_\ell, a, e, i, t)}{\partial t},
\end{aligned}
$$

(C.9)

where $RHS_U(U^p, \beta = 1)$ is the RHS of Equation (2.1), i.e. the flow private value of an unemployed worker that captures all the gains from trade in the labor market ($\beta = 1$). The remaining terms on the RHS of Equation (C.9) are generated by the fact that $\phi(j)$ enters $\Gamma(j')$ not only directly, but also through its contribution to M_i (Equation 2.10), G_i (Equation 2.14), and COL_i (Equation 2.12).

The knowledge spillover externality is the key determinant of the wedge between the optimal

and the equilibrium composition of cities. Formally, it is equal to

$$\text{knowledge spillovers} \equiv$$

$$\frac{\sigma_i}{M_i} \sum_{\tilde{\ell} \neq L, \tilde{e}} \kappa^p(h_\ell, h_{\tilde{\ell}}, \tilde{e}) \int_{\underline{z}}^{\bar{z}} \phi(h_{\tilde{\ell}}, y, \tilde{e}, \tilde{z}, i)[V^p(h_{\tilde{\ell}+1}, y, \tilde{e}, i, \tilde{z}) - V^p(h_{\tilde{\ell}}, y, \tilde{e}, i, \tilde{z})]d\tilde{z} +$$

$$\phi(h_{\tilde{\ell}}, y, \tilde{e}, 0, i)[U^p(h_{\tilde{\ell}+1}, y, \tilde{e}, i) - U^p(h_{\tilde{\ell}}, y, \tilde{e}, i)] + \tag{C.10}$$

$$-\frac{\sigma_i}{M_i} \sum_{\tilde{\ell} \neq L, \tilde{e}} \kappa(G_i, h_{\tilde{\ell}}, e) \int_{\underline{z}}^{\bar{z}} \phi(h_{\tilde{\ell}}, y, \tilde{e}, \tilde{z}, i)[V^p(h_{\tilde{\ell}+1}, y, \tilde{e}, i, \tilde{z}) - V^p(h_{\tilde{\ell}}, y, \tilde{e}, i, \tilde{z})]d\tilde{z} +$$

$$\phi(h_{\tilde{\ell}}, y, \tilde{e}, 0, i)[U^p(h_{\tilde{\ell}+1}, y, \tilde{e}, i) - U^p(h_{\tilde{\ell}}, y, \tilde{e}, i)].$$

The sum of the terms in the first two lines of Equation (C.10) is always positive and strictly increasing in h_ℓ. A young worker of type $(h_{\tilde{\ell}}, \tilde{e})$ learns from a worker of type h_ℓ at rate $\kappa^p(h_\ell, h_{\tilde{\ell}}, \tilde{e})$. The function κ^p is the counterpart of κ, seen from the point of view of the worker who transmits knowledge to others. According to the technology of knowledge diffusion described in Section 2,

$$\kappa^p(h_\ell, h_{\tilde{\ell}}, \tilde{e}) = \eta_h^{\tilde{e}} h_\ell + \eta_v^{\tilde{e}} max(h_\ell - h_{\tilde{\ell}}, 0).$$

A worker h_ℓ has a positive externality on any other (young) worker in the same city, and, in particular, on workers with human capital $h_{\tilde{\ell}} < h_\ell$. The sum the tems in the last two lines of Equation (C.10) is clearly negative, and it is equal across all workers in city i. It represents the congestion in the learning process, due to the reduction in the rate at which young workers in city i learn from every other worker in the same city.[58]

The other terms of Equation (C.9) are given by,

$$\text{agglomeration (flow of ideas)} \equiv$$

$$\tau_M \frac{\sigma_i}{M_i} \sum_{\tilde{\ell} \neq L, \tilde{e}} \kappa(G_i, h_{\tilde{\ell}}, e) \int_{\underline{z}}^{\bar{z}} \phi(h_{\tilde{\ell}}, y, \tilde{e}, \tilde{z}, i)[V^p(h_{\tilde{\ell}+1}, y, \tilde{e}, i, \tilde{z}) - V^p(h_{\tilde{\ell}}, y, \tilde{e}, i, \tilde{z})]d\tilde{z} +$$

$$\phi(h_{\tilde{\ell}}, y, \tilde{e}, 0, i)[U^p(h_{\tilde{\ell}+1}, y, \tilde{e}, i) - U^p(h_{\tilde{\ell}}, y, \tilde{e}, i)],$$

---

[58]Recall that the knowledge spillover externality originates from the dependence of G_i on $\phi(j)$. Specifically, from Equation (2.14),

$$G_i(h_\ell) = \frac{\sum_{a=y,o} \sum_{e=hs,col} \sum_{\tilde{\ell}=1}^{\ell} \left[\int_{\underline{z}}^{\bar{z}} \phi(h_{\tilde{\ell}}, a, e, i, z)dz + \phi(h_{\tilde{\ell}}, a, e, i, 0) \right]}{\underbrace{\sum_{a=y,o} \sum_{e=hs,col} \sum_{\ell=1}^{L} \left[\int_{\underline{z}}^{\bar{z}} \phi(h_\ell, a, e, i, z)dz + \phi(h_\ell, a, e, i, 0), \right]}_{=M_i}}.$$

The first (last) two lines in Equation (C.10) correspond to the effect of an additional unit of $\phi(j)$ on the numerator (denominator) of the definition of $G_i(h_\ell)$.

$$\text{search} \equiv -(k_i \theta_i + \rho^* k_{-i} \theta_{-i}),$$

$$\text{OLG} \equiv \psi_o \sum_{\tilde{\ell}, \tilde{e}} U^p(h_{\tilde{\ell}}, y, \tilde{e}, i) \pi^{e, \tilde{e}} \hat{\pi} g_{0,i}^h(h_{\tilde{\ell}} | \tilde{e}),$$

agglomeration (matching) $\equiv$

$$\chi_M \frac{\lambda_{0,i}}{M_i} \int_{\underline{z}}^{\bar{z}} \sum_{\tilde{\ell} \neq L, \tilde{e}, \tilde{a}} \rho \int_{\underline{z}}^{\hat{z}} \phi(h_{\tilde{\ell}}, y, \tilde{e}, \check{z}, i)[V^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, i, \hat{z}) - V^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, i, \check{z})]d\check{z} +$$

$$\phi(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, 0, i)[V^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, \hat{z}, i) - U^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, i)]\mathbb{I}\{\hat{z} > R(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, i)\}d\hat{z} +$$

$$\chi_M \frac{\lambda_{0,i}^*}{M_i} \left[\int_{R_i}^{\bar{z}} \sum_{\tilde{\ell} \neq L, \tilde{e}, \tilde{a}} \rho \int_{\underline{z}}^{\bar{z}} \phi(h_{\tilde{\ell}}, y, \tilde{e}, \check{z}, -i) D(x(\check{z}, \hat{z})) \right.$$

$$[V^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, i, \hat{z}) - V^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, -i, \check{z}) - \mathbb{E}(c|c < x(\check{z}, \hat{z}))]d\check{z} +$$

$$\phi(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, 0, -i) D(x(0, \hat{z}))[V^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, \hat{z}, i) - U^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, -i) - \mathbb{E}(c|c < x(0, \hat{z}))]d\hat{z} +$$

$$\int_{\underline{z}}^{\bar{z}} D(x(\check{z}, 0))[U^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, i) - V^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, -i, \check{z}) - \mathbb{E}(c|c < x(\check{z}, 0))]d\check{z} F(R_i) +$$

$$\left. D(x(0,0))[U^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, i) - U^p(h_{\tilde{\ell}}, \tilde{a}, \tilde{e}, -i) - \mathbb{E}(c|c < x(0,0))]F(R_i) \right].$$

D Additional results

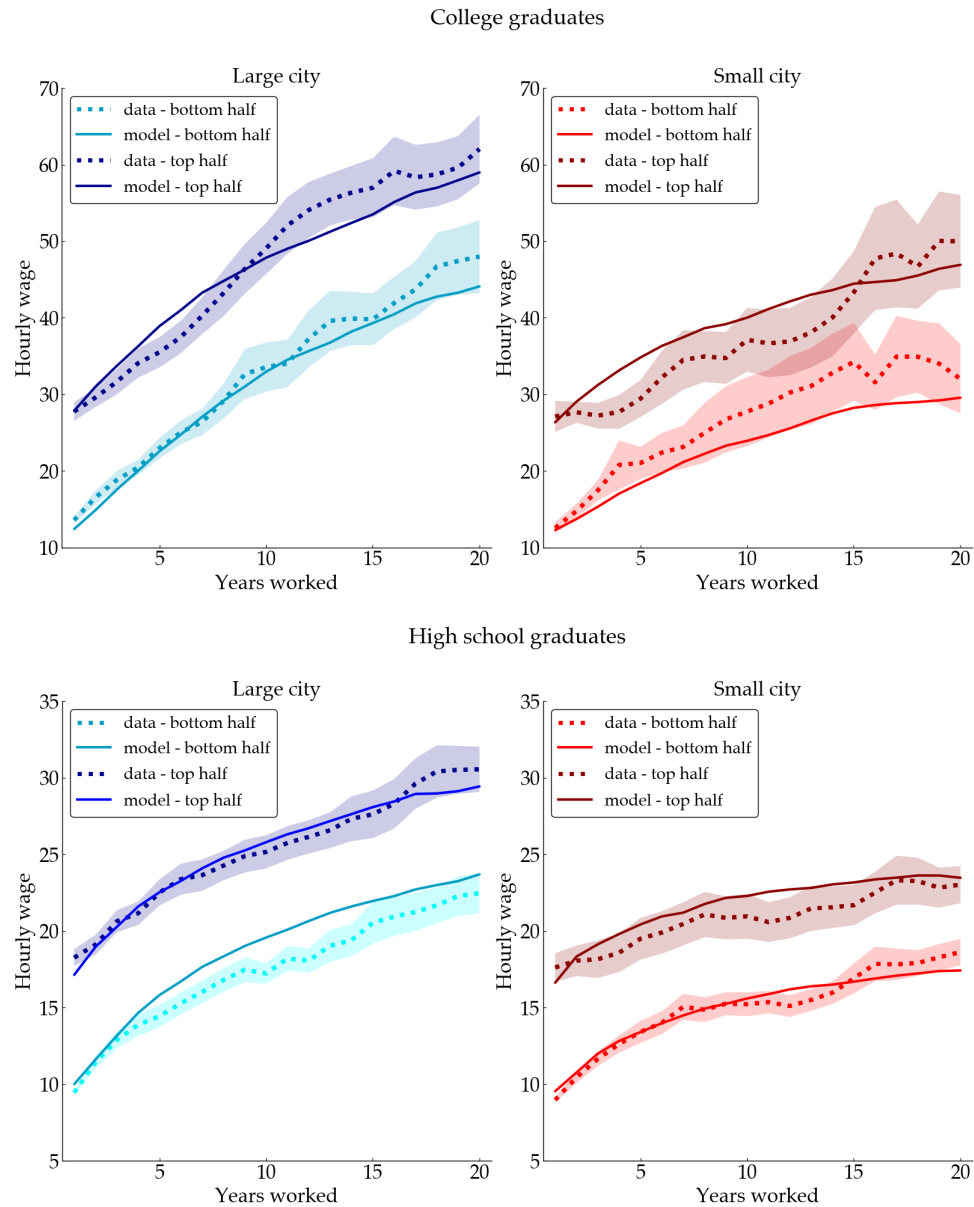## D.1 Identification: Empirical and model-generated wage profiles

Figure 13: 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.

D.2 Validation: Moments generated from the sample of non-movers

Moment (large/small)	Data	Model
Labor Market		
unemp. rate (%)	5.1/5.5	5.1/5.1
n. of jobs	6.38/6.14	6.54/6.56
mean wage gap (%)	32.7	34.5
EE wage growth (%)	11/11.8	11.1/11.3
return to tenure (%)	22.5	24.0
college share (%)	29.3/17.5	30.9/18.7
Wage growth (%)		
bottom-col-large	6.6	6.5
bottom-col-small	4.8	4.3
top-col-large	4.0	3.8
top-col-small	2.5	3.0
bottom-hs-large	4.4	4.4
bottom-hs-small	3.7	3.1
top-hs-large	2.5	2.8
top-hs-small	1.3	1.9
wage 11-20 vs. 1-10 yr. of exp.	30.6	31.2
Initial Wage		
col wage premium (%)	41	41.2
$75^{th}/25^{th}$ pctile, hs	1.65	1.66
$75^{th}/25^{th}$ pctile, col	1.86	1.91

Table 4: Targeted moments computed on the sample of non-movers

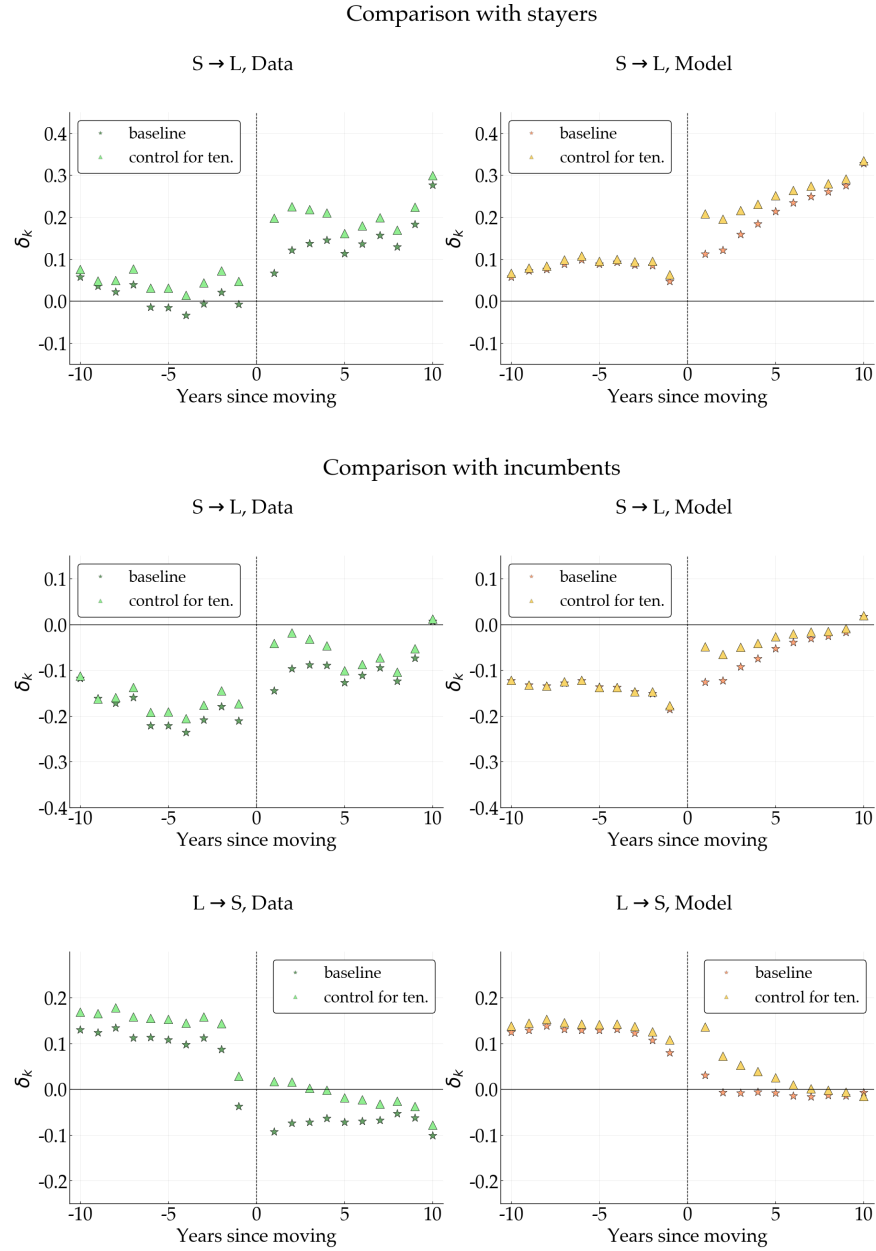## D.3 Validation: Control for tenure

Figure 14: Wage of movers with respect to stayers (top row) and incumbents (bottom two rows). Dark stars: baseline. Light triangles: control for tenure. Left: data. Right: model. The fourth comparison, i.e. movers to small cities compared to stayers in large cities, is reported in the main body of the paper (Section 3.5).