Occupational Matching and Cities

Theodore Papageorgiou
Boston College, Department of Economics*†

June 2020

Abstract

In this paper, I document that workers in larger cities have significantly more occupational options than workers in smaller ones. They are able to form better occupational matches and earn higher wages. I also note differences in the occupational reallocation patterns across cities. I develop a dynamic model of occupational choice that microfounds agglomeration economies and captures the empirical patterns. The calibration of the model suggests that better occupational match quality accounts for approximately 35% of the observed wage premium and a third of the greater inequality in larger cities.

Keywords: Occupations, Agglomeration Economies, Urban Wage Premium, Multi-armed Bandits, Geographical Mobility, Matching Theory, Wage Inequality, Job Vacancy Postings

JEL Classification: J24, J31, R23

* I am grateful to Joe Altonji, Apostolos Burnetas, Ed Coulson, Jan Eeckhout, Pablo Fajgelbaum, Manolis Galenianos, Ed Green, Yannis Ioannides, Boyan Jovanovic, John Kennan, Philipp Kircher, Matthias Kredler, Keith Kuester, Fabian Lange, Sanghoon Lee, Alex Monge, Diego Puga, Steve Redding, Richard Rogerson, Esteban Rossi-Hansberg, Rob Shimer, Kjetil Storesletten, Venky Venkateswaran as well as numerous seminar participants for useful comments. I am especially indebted to Bledi Taska for providing me with the Burning Glass data and his assistance in using them. This paper has previously circulated under the title "Worker Sorting and Agglomeration Economies".

† Maloney Hall 343, Chestnut Hill, MA, 02467, USA, theodore.papageorgiou@bc.edu

1 Introduction

Workers in larger cities are paid higher wages and produce more output. Since concentrating a large number of workers and firms in one region can be costly, several economists have argued that agglomeration economies exist. Agglomeration economies generally refer to any mechanism that makes economic agents more productive as the level of economic activity in their area increases. Over the years, economists have proposed several mechanisms such as human capital externalities and reduced transportation costs.[1] In a survey, however, Glaeser and Gottlieb (2009) note that even though there remains a robust consensus among urban economists that [agglomeration] economies exist, [...] the empirical quest to accurately measure such economies has proven to be quite difficult.

in each location to be determined endogenously. Cities with larger populations have larger markets and are therefore able to support more occupations. More occupations, in turn, attract more workers, both because of increased employment options but also because workers value consumption diversity. A larger city caters to more diverse consumer tastes, producing and hiring in a larger variety of services and products. Both the number of occupations and population are endogenously determined.

To my knowledge, this is the first paper to examine whether increased occupational availability leads to better matches and thus agglomeration economies, through a dynamic model. Following the classification of microfoundations of agglomeration economies by Duranton and Puga (2004), this paper falls under the category of better matching, as workers are able to form better occupational matches in larger cities. Also under the same category Helsley and Strange (1990) and Kim (1989, 1991) have proposed setups where heterogeneous workers and heterogeneous firms form better matches in large cities. Both papers consider static setups and therefore do not have predictions regarding worker reallocation. Bleakley and Lin (2012) document that young workers switch occupations more often in larger cities, which is related to my finding that recent movers in large cities are more likely to switch occupations. Gautier and Teulings (2009) find that large cities are more heterogeneous in terms of the job types (occupation/industry combinations) that are offered. Both papers interpret their findings as evidence of increasing returns to scale in the matching function between searching workers and vacant firms (see also Diamond, 1982 and Petrongolo and Pissarides, 2006). Indeed, as I discuss in Section 3.4, increasing returns to matching provide one potential explanation for the greater occupational availability in large cities. However, increasing returns to matching alone cannot match some of the patterns found in the data, such as the decline in wages prior to moving or switching occupations. I discuss further the related literature and whether the observed empirical patterns can be explained by one of the other mechanisms in Section 3.5, following the exposition of the model.

In addition, the mechanism is consistent with the findings of Baum-Snow and Pavan (2012) and De la Roca and Puga (2017), who decompose the wage premium into a static advantage that workers enjoy immediately upon arriving in a large city, a dynamic advantage that appears with time in a city, and sorting based on ability. Both papers find strong evidence in favor of a dynamic advantage, implying that the agglomeration mechanism becomes more important largely after a worker has arrived in a large city (see also Glaeser and Maré, 2001).[2,3]

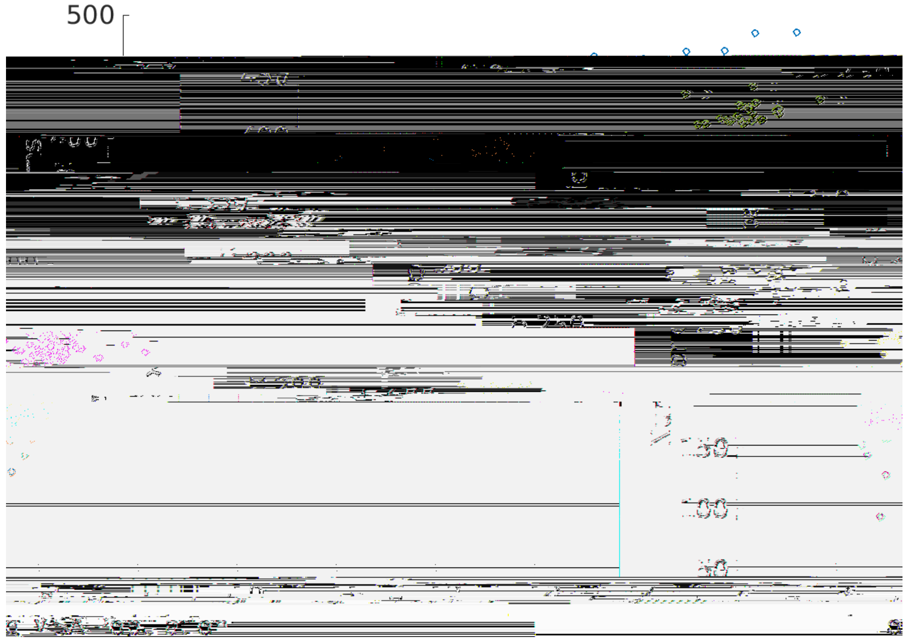

Figure 1: Number of Occupations vs. Log MSA Population - Burning Glass Vacancy Data

Fact 1. *There are more occupations available in large cities.*

I begin the empirical investigation by considering how occupational availability varies by city size. I nd that workers in larger cities have more occupations available to work in and this di erence is not driven by fringe occupations that would interest only few workers.

First, I use a unique database of job vacancies collected by Burning Glass Technologies (BG). BG collects information daily from more than 40,000 sources. The breadth of the coverage exceeds that of any one source, and in fact, BG claims that its database covers the near-universe of online job vacancies.[4]

The BG data contain information on the posting's detailed occupation (at the 6-digit Standard Occupation Classi cation (SOC) 2010 level), as well as whether it belongs to one of 381 metropolitan statistical areas (MSA). The rest of the analysis uses information on vacancies posted between February 1, 2016 and April 30, 2016. There are 6,103,537 postings during this period.

Figure 1 plots the number of 3-digit occupations (2002 Census Occupational Classi cation) in which there are vacancies in every MSA against its population as reported in the 2010 Census.[5] The relationship between the number of occupations with vacancies and city size is positive and approximately log-linear: a simple linear regression indicates that cities with double the size have approximately 70 more occupations.

[4] See also the discussion in Deming and Kahn (2018) and Hershbein and Kahn (2018), who are one of the rst to use the BG data.

[5] The gure uses the 2002 Census Occupational Classi cation, which has 508 occupations. Using the 2010 SOC codes (841 occupations) leads to very similar results.

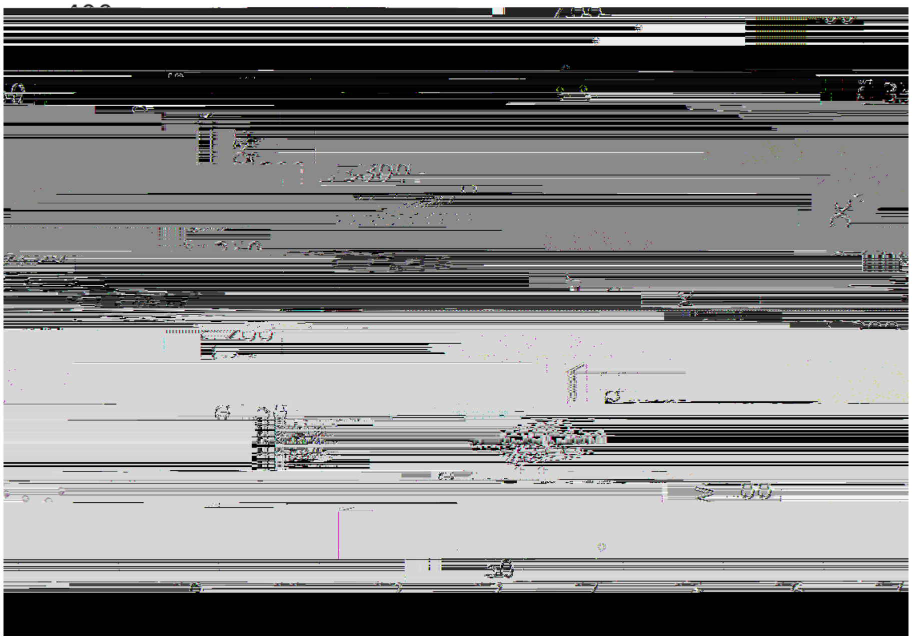

Figure 2: Weighted Number of Occupations vs. Log MSA Population - Burning Glass Vacancy Data

The vacancy data used so far are from online postings and while the data are comprehensive, they do not include postings that are not also posted online. In order to check whether there may be additional occupational opportunities beyond those reported in the BG data, I use employment data from the American Community Survey (ACS) from 2011 through 2015. If there are more occupations available to workers in a location beyond those captured in the BG data, then this would show up in employment outcomes, as one would expect workers to also be employed in occupations other than those reported in the BG data. However it turns out that 95.28% of employed workers in the ACS data are working in an occupation in which there is a local vacancy according to the BG data, suggesting that there are few occupational opportunities beyond those captured in the BG data.[7]

Finally, I confirm the same relationship using vacancy postings from the UK.[8] The strong positive relationship between city size and number of occupations is also present in a) the 2000 US Census data, b) the Occupational Employment Statistics, which report estimates of occupational employment in each metropolitan area using an establishment rather than a worker survey, and c) the Brazilian Annual Social Information Report (RAIS) for the state of São Paulo, which is a large matched employer-employee database that covers 97% of the formal market.[9] See Figures 1 through 4 in the Online Appendix.

	Initial	Moved<4 years	All
	ln(wage)	ln(wage)	ln(wage)
ln(current city pop)	0.0155	0.021	0.041
	(0.009)	(0.01)	(0.001)
Number of Obs	1261	4321	169536

Table 1: Wage Premium Evolution. Source: 1996 Panel of Survey of Income and Program Participation. Population data from 2000 Census. Controls include gender, race, education, marital status, rm size, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors in parentheses are clustered by individual.

	All years	Moved<4 years	Moved<4 years
	Occ. Switching	Occ. Switching	Occ. Switching
	Prob. (Probit)	Prob. (Probit)	Prob. (Probit)
ln(current city pop)	-6p)		

	Prob of Moving
	& Switching Occup (Probit)
ln(current city pop)	-0.0007
	(0.0002)
Number of Obs	144635

T05 Td [(T05 Tdon51 9.ble)-400(31 Tf 3.003 655.805 T45.686)]TJ ET PImpac(y)-48 ET q 47(Cc)1(it)27(y)-47(C28(o8))48 En q

	$\ln(\text{wage})_{t-1}$

In Appendix C, I relax this assumption and allow for the number of occupations in each location to be endogenously determined.

The basic environment is the following: di erent cities have a di erent number of occupations. Within a city, workers draw their productivity at each occupation. In a frictionless world, workers enter the occupation in which they are most productive. However I introduce the following friction, which induces city.[21,22] Moving from one city to another entails a cost

m occupations there. He then chooses one of the occupations and begins working there, or alternatively he can pay c and move to another city.

3.2 Behavior

arm, k, depends only on that arm's beliefs (in this case p^k).[25] I am able to use Gittins indices in this setup, because there is no cost to switching occupations in a city. Gittins indices cannot be used in the presence of even $\varepsilon > 0$ cost to switching (see Banks and Sundaram, 1994).[26]

Lemma 1. *The worker's optimal strategy takes the form of an index policy, whereby every period the worker chooses the occupation with the highest index.*

Proof. Gittins (1979). □

Here I follow the approach in Whittle (1982) and Karatzas (1984), whereby the transformed problem for every occupation is to assume that a worker has only two options: either work in that occupation or retire and obtain some retirement value. The retirement option is always available, so this is an optimal stopping problem where the worker needs to decide when and if to retire. The retirement value at which the worker is exactly indifferent between continuing with that arm or retiring corresponds to that occupation's Gittins index.

I first compute the optimal retirement policy for every occupation, k, with probability p^k of $\alpha^k = \alpha_G$ and the option of retiring with value W^k. In other words, a worker can either work in occupation k or retire and obtain value W^k.

In that case, the value function of a worker with posterior p^k and the option of retiring and obtaining value W^k, $V^k\left(p^k, W^k\right)$, satisfies the following Hamilton-Jacobi-Bellman equation

$$rV^k\left(p^k, W^k\right) = w\left(p^k\right) + \frac{1}{2}\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2 \left(p^k\right)^2 \left(1 - p^k\right)^2 V^k_{pp}\left(p^k, W^k\right) - \delta\left(V^k\left(p^k, W^k\right) - J\right),$$

where V^k_{pp} is the second derivative of V^k with respect to p. The flow benefit of the worker consists of his wage, plus a term capturing the option value of learning, which allows him to make informed decisions in the future. Finally, the worker leaves his current city exogenously at rate δ, pays cost c, and moves to a new one. J denotes the value of a worker about to move to another city

$$J = -c + \sum_{m=1}^{M} s_m E_p V(p_m),$$

25

where c is the moving cost, $E_{\mathbf{p}}V(\mathbf{p}_m)$ is the expected value of a worker who moves into a city with m occupations available for him to work in, s_m denotes the probability that the worker moves to a city with m occupations and

$$\mathbf{p}_m = \left(p^1\; p^2 \ldots\; p^m\right) \in \mathbb{R}^m,$$

is the vector of the posteriors for each occupation k in the city.

Guessing that V^k is increasing in p^k, the optimal stopping rule is to retire when p^k reaches $\underline{p}\left(W^k\right)$ such that the value matching and the smooth pasting conditions hold:

$$V^k\left(\underline{p}\left(W^k\right), W^k\right) = W^k \tag{3}$$

$$V^k_p\left(\underline{p}\left(W^k\right), W^k\right) = 0.$$

In other words, a worker chooses to stop experimenting and receive value W^k when his posterior reaches value $\underline{p}\left(W^k\right)$, defined above.

The solution to the above differential equation is given by

$$\begin{aligned} V^k\left(p^k, W^k\right) = {} & \frac{w\left(p^k\right) + \delta J}{r+\delta} \\ & + \frac{\alpha_G - \alpha_B}{r+\delta}\left(\underline{p}\left(W^k\right) + \frac{1}{2}d - \frac{1}{2}\right)^{-1} \underline{p}\left(W^k\right)^{-\frac{1}{2}+\frac{1}{2}d}\left(1-\underline{p}\left(W^k\right)\right)^{-\frac{1}{2}-\frac{1}{2}d} \\ & \left(p^k\right)^{\frac{1}{2}-\frac{1}{2}d}\left(1-p^k\right)^{\frac{1}{2}+\frac{1}{2}d}, \end{aligned}$$

where

$$\underline{p}\left(W^k\right) = \frac{(d-1)\left((r+\delta)W^k - \alpha_B - \delta J\right)}{(d+1)(\alpha_G - \alpha_B) - 2\left((r+\delta)W^k - \alpha_B - \delta J\right)}, \tag{4}$$

and $d = \sqrt{\frac{8(r+\delta)}{\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2} + 1}$.[27] V^k is increasing in p^k. Moreover, note that $\underline{p}\left(W^k\right)$ is strictly increasing in W^k.

The index of occupation k is the highest retirement value at which the worker is indifferent between working at occupation k or retiring with $W^k = W\left(p^k\right)$. Therefore, the Gittins index, $W\left(p^k\right)$, is implicitly defined by the following equation

$$W\left(p^k\right) = V^k\left(p^k, W^k\right), \tag{5}$$

[27] The interested reader should refer to the Online Appendix for a solution method to second-order, non-homogeneous differential equations.

where $W\left(p^k\right) = \max\left\{\tilde{W}^k\right\}$ and the set $\left\{\tilde{W}^k\right\}$ includes all possible retirement values, $\tilde{W}^k$, such that $\tilde{W}^k = V^k\left(p^k, \tilde{W}^k\right)$.

For equation (5) to hold, from equation (3), it must be the case that

$$p^k = \bar{p}\left(W^k\right). \tag{6}$$

Substituting condition (6) into the threshold condition, equation (4), obtains

$$p^k = \frac{(d-1)\left((r+\delta)\, W\left(p^k\right) - \alpha_B - \delta J\right)}{(d+1)\,(\alpha_G - \alpha_B) - 2\left((r+\delta)\, W\left(p^k\right) - \alpha_B - \delta J\right)} \tag{7}$$

$$W\left(p^k\right) = \frac{1}{r+\delta}\,\frac{(d+1)\,(\alpha_G - \alpha_B)\, p^k + \left(2p^k + d - 1\right)(\alpha_B + \delta J)}{2p^k + d - 1}. \tag{8}$$

In addition,

Lemma 2. *$W\left(p^k\right)$ is strictly increasing in p^k.*

Proof. See Appendix B. □

Given the above, the following proposition holds:

Proposition 1. *The optimal strategy of a worker in this setup is to work at occupation n, where*

$$n \in \arg\max_{k \in \{1,..m\}} \left\{p^k\right\}.$$

Proof. Follows from Lemma 1, Whittle (1982), equation (8), and Lemma 2. □

In other words, the Gittins index for each occupation reduces to the worker's beliefs, p^k, in that occupation. Workers always work in the occupation in which they believe they are best matched. This is true only when all occupations are identical. If, for instance, the signal-to-noise ratio, ζ, varies across occupations, then the Gittins index is given by equation (8).

Workers also have the option of moving to another city that provides known value, J. In the bandit problem, this is equivalent to a safe arm. Since J is trivially the retirement value associated with playing the safe arm, J also corresponds to the Gittins index of the safe arm. A worker will therefore play the safe arm, if and only if the retirement value (Gittins index) of all other arms is lower than J. In order to find the value of the posterior, $\underline{p}$, where the worker chooses to play the safe arm (i.e., move), I use equation (7) and substitute J for $W\left(p^k\right)$.

Proposition 2. A worker pays the fixed cost and moves when all his posteriors fall below a moving threshold $\underline{p}$ that is independent of m, the number of the city's available occupations. The moving threshold is given by

$$\underline{p} = \frac{(d - 1)(rJ - \alpha_B)}{(d + 1)(\alpha_G - \alpha_B) - 2(rJ - \alpha_B)}.$$

Summarizing, consider a worker who has just moved to a city. He immediately draws a prior p_0^k, out of a city is lower in cities with more occupations m (Fact 4). Intuitively, workers in larger cities are less likely to move both because they have more options and because they are better matched.

The above result implies that workers stay longer in cities with more occupations, m. Since the flow into a city is the same regardless of the number of occupations, this immediately implies that in equilibrium, cities with more occupations, m, have larger populations (Fact 1).[29]

I also examine the path of wages before moving. In the setup, workers move endogenously following a downward revision of their beliefs. This is also reflected in their wages, so workers experience wage decreases before moving and switching occupations, consistent with Fact 5.[30] One additional prediction of the model is that workers are switching occupations prior to the move, i.e., right before their posteriors hit $\underline{p}$. As mentioned at the end of Section 2, this prediction is true in the data as well, i.e., past occupational switching significantly increases the probability of a move.

I next turn to how the probability of switching occupations is affected by the number of occupations.

	Initial	All Years
Full Sample	ln(wage)	ln(wage)
ln(current city pop)	0.016	0.041
	(0.009)	(0.

	All years	Moved<4 years	Moved<4 years
	Occ. Switching	Occ. Switching	Occ. Switching
Full Sample	Prob. (Probit)	Prob. (Probit)	Prob. (Probit)
ln(current city pop)	-0.0025	0.0109	0.0255
	(0.0006)	(0.0067)	(0.0098)
ln(previous city pop)			-0.0081
			(0.0067)
Number of Obs	140842	3360	2047
Singles Only			
ln(current city pop)	-0.0041	0.0085	0.0263
	(0.0009)	(0.0087)	(0.0128)

Full Sample	Prob of Move & Switch Occup		ln(wage)$_{t\ 1}$
ln(current city pop)	-0.0007	Move$_t$ Occupation Switch$_t$	-0.024
	(0.0002)		(0.008)
Number of Obs	144635	Move$_t$ No Occupation Switch$_t$	-0.002
			(0.004)
		ln(wage)$_{t\ 2}$	0.847
			(0.001)
		Number of Obs	146462
Singles Only			
ln(current city pop)	-0.0013	Move$_t$ Occupation Switch$_t$	-0.018

I calibrate the setup to white males with a college education.[40] Moreover, because the setup does not allow for moving and remaining in the same occupation, I drop workers who move and keep the same occupation. There are two types of locations: areas with large populations and less populated areas. In the data this corresponds to locations with more than 500,000 inhabitants and those with less.

In my sample, workers who move to larger cities do not receive initially higher wages than their counterparts who move to less populated cities (p-value of 0.42).[41] This fact, viewed through the lens of the setup, implies that the distribution from which the initial beliefs are drawn, $g(.)$, has little variance. In my calibration, therefore, I set the prior belief for every occupation to be the same and equal to p_0, whose value needs to determined. Note that the above fact is consistent with higher occupational mobility for recent movers in larger areas (second column, Table 2): since they are not initially better matched than those who moved to smaller locations, they are more likely to take advantage of the increased options larger cities offer. Indeed, as shown below, the calibrated model replicates this feature of the data.

The calibration proceeds in three steps. First, I set the number of occupations in each of the two types of locations. I also set the discount rate to 5% annually (1.64% at the 4-month frequency). Second, I use worker reallocation moments to jointly pin down the key model parameters (s, δ, c, ζ and p_0, where s is the probability that a worker who moves goes to a large city). Third, I choose α_G and α_B to match the economy mean wage and the residual standard deviation of wages. In what follows, I discuss these three steps in detail. My setup is set in continuous time, but I sample the simulated data every 4 months to match the sampling in the SIPP. Appendix D contains more details.

Step 1: In order to set the number of available occupations in each location (large vs. small cities), I use two moments. First, from the BG data, I recover the ratio of available occupations in large cities over those in small cities. In particular, I compute that there are on average 322.6 occupations available to a worker residing in a large city, i.e., one with more than 500,000 inhabitants. Similarly, there are 141.6 occupations available on average to workers residing in small cities.[42] Therefore, the ratio of the number of occupations in large over small cities is 2.28, which is the first moment I target.[43]

[40] I focus on college graduates because Gould (2007) documents that the urban wage premium is larger for workers in white-collar jobs that are typically held by college graduates. Similarly, Davis and Dingel (2018) find that the college wage premia are higher in larger cities.

[41] As shown in Section 2, in the larger sample, workers who just moved to a new location, receive a higher wages if they moved to a highly populated area, but the coefficient is not large (first column of Table 1). This suggests that static advantages, whereby workers immediately become more productive upon arriving in larger cities, are not important in explaining the wage premium.

[42] For large cities I compute the population share of each large city and multiply it by the number of available occupations in that city and similarly for small cities. In addition, when computing the number of available occupations in each city, I weight each occupation by the number of workers who switch into it, as shown in Figure 2 (see the corresponding discussion in Section 2 and in footnote 6).

[43] If I change the threshold and require at least 5 vacancies for an occupation to be available, the ratio becomes 2.97.

α_G	28.93
α_B	8.29
s	54.96%
δ	0.00489
$\frac{\alpha_G \quad \alpha_B}{\sigma}$	0.1795
p_0	0.0937
c (implied $\underline{p}$)	91 (0.0304)

Table 9: Parameter Values

Moments:	Data	Model
Population Share in Large Cities	58.95%	58.93%
Moving Probability in Large Cities	0.50%	0.49%
Moving Probability in Small Cities	0.60%	0.58S Q BT /F50 10.9091 Tf 371.5p9r.0(

Figure 3: Sensitivity Analysis - Probability of Mover Going to Large City

and help to illustrate which features of the model deliver the quantitative results.

I begin by varying the probability that a worker who moves goes to a large city, s. This can be interpreted as increasing the share of large cities in the economy. As shown in Figure 3, not surprisingly this leads to an increase in workers' mean wages and therefore expected output in the economy: since the model predicts that workers in larger cities are more productive, the increase in the fraction of the population in large cities mechanically leads to an increase in average worker productivity. More interestingly, however, mean wages and therefore the productivity of workers in smaller cities also increase: as the share of large cities increases, the benefit of moving increases as well; as a result, workers in small cities are less willing to tolerate bad matches and are more likely to move. As shown in Figure 3, the moving threshold, $\underline{p}$, is indeed increasing in s. Workers try out more occupations -indeed, the probability of switching to a new occupation is also increasing in s- and are on average more productive.

Figure 4: Sensitivity Analysis - Productivity of a Good Match, α_G

I next consider the impact of increasing α_G.[48] Now, the benefit of being in a good match is higher. As a result, as shown in Figure 4, the moving threshold, $\underline{p}$, increases as the cost of moving, c, remains unchanged, while the potential benefits are now higher. Migration increases and workers now try out more occupations, leading to an increase in the probability of switching to a new occupation.

On the other hand, when c increases, as shown in Figure 5, the moving threshold, $\underline{p}$, falls, since migration is more costly. As a result, workers are more likely to be in a bad match and less likely to try out new occupations: both the mean wage and the probability of moving to a new occupation decline. Finally, the wage premium increases, since reducing migration across locations implies that it is even more beneficial to work in a large city that offers many choices.

I also consider the impact of allowing for dispersion in initial beliefs: rather than assuming that a worker's initial belief for all occupations is equal to p_0, I instead draw each occupation's prior from a beta

[48] When changing ${}_{G}$

Figure 5: Sensitivity Analysis - Cost of Moving

distribution with mean p_0 and consider various levels of its standard deviation. As shown in Figure 6, when initial belief dispersion increases, the di erence in the switching probability for recent movers to large cities relative to small increases: workers in large cities, who have more occupations available, are more likely to have belief draws that are close together, compared to workers in smaller cities. As the dispersion increases, the di erence disappears as workers try out fewer occupations and the average probability of switching to a new occupation falls. Interestingly, the initial wage premium increases substantially as initial belief dispersion goes up and can reach up to 40% (in my sample of highly educated workers, the initial wage premium is not statistically signi cant).

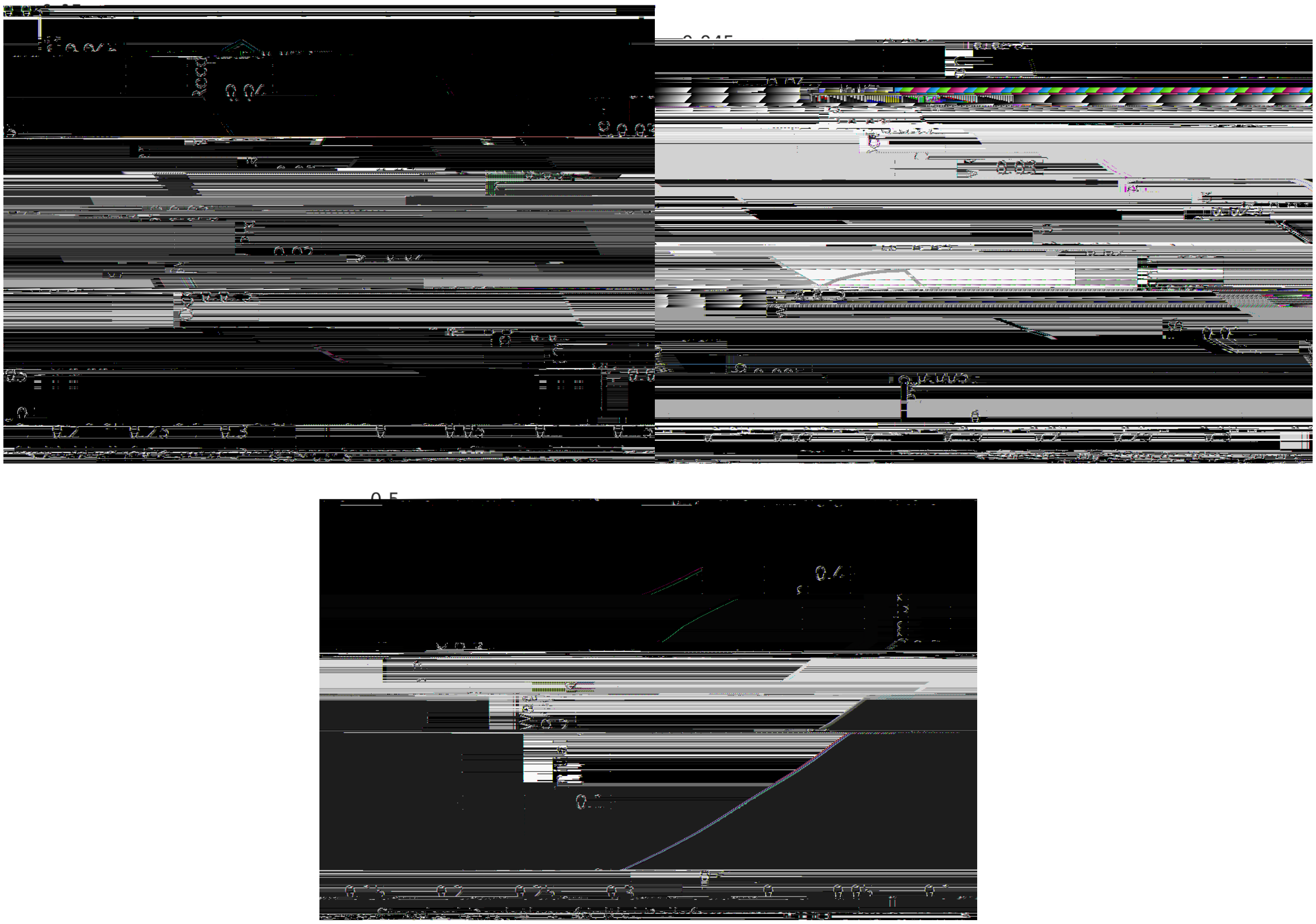

Figure 6: Sensitivity Analysis - Dispersion in Initial Beliefs

learning anyway, the value of having more occupations available is extremely small. As a result, workers in large cities are in similar quality matches as workers in smaller ones and the wage premium is close to zero. Conversely, when learning is fast, the wage premium again approaches zero for a di erent reason: workers spend far less time in low-quality matches and, as a result, sort quickly through their locations' occupations. Both the switching probability to new occupations and the average worker productivity increase, while the bene t of being in a large city with many occupations declines and so does the wage premium.

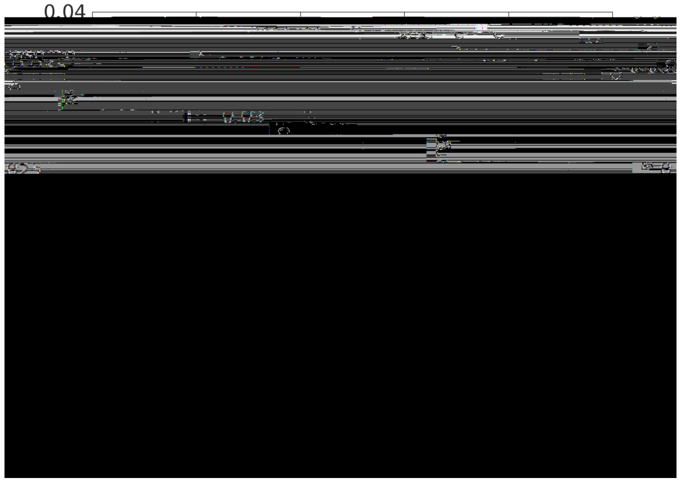

Figure 8: Higher occupational switching probability among recent movers to larger cities, compared to those who moved to smaller ones. Horizontal axis is percent di erence in speed of occupation-speci c human capital accumulation in large cities versus small.

The model is otherwise identical to that presented in Section 3: some cities o er more occupations than

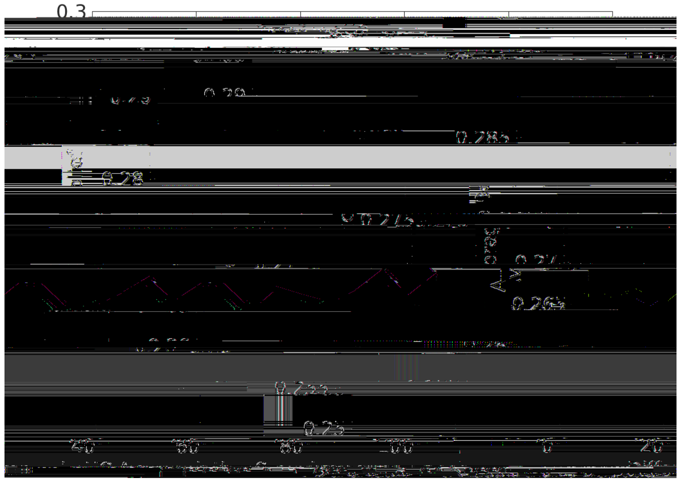

Figure 9: Average belief. Horizontal axis is percent di erence in speed of occupation-speci c human capital accumulation in large cities versus small.

I rst note that the implications of adding human capital accumulation for the reallocation moments are often the opposite of those of greater occupational availability. In particular, Figure 8 plots the higher occupational switching for recent movers in larger cities as a function of the di erence in the speed of human capital accumulation across cities of di erent sizes. The graph also plots the higher occupational switching for recent movers in the baseline calibration. The introduction of human capital leads to a lower di erence in occupational switching among recent workers, as the importance of nding a good

Figure 10: Wage premium between large and small cities. Horizontal axis is percent di erence in speed of

Targeted Moments:	Data	Baseline	Model with HC
Population Share in Large	58.95%	58.93%	55.26%
Moving Probability in Large Cities	0.50%	0.49%	0.48%
Moving Probability in Small Cities	0.60%	0.58%	0.48%
Higher Occup Sw Prob in Large	0.20%	0.93%	0.42%
Higher Occup Sw Prob in Large (recent)	3.34%	2.98%	1.95%
Mean Wage	$14.20	$14.20	$14.20
Residual Wage Standard Deviation	$5.97	$5.97	$5.95
Other Moments:			

6 Conclusion

This paper documents a number of facts related to the number of occupational opportunities in small and large cities and the relationship between city size, wages, occupational switching, and geographical mobility. Guided by these facts, I develop and calibrate a model where workers in larger cities have more occupations available and, as a result, form better matches. In my setup, agglomeration economies are not the result of larger cities exogenously having higher productivity. Rather, agglomeration economies are endogenously generated. I calibrate the model using moments related to occupational switching and geographical mobility. The calibrated model replicates approximately 35% of the observed wage premium and a third of the greater inequality in larger cities.

Both the data documented and the model introduced formalize the sentiment re ected in the press about certain jobs not being available in smaller cities and, as a result, workers choosing suboptimal matches. A career counselor gives the following advice: Be exible. Depending on just how small the city is in which you're looking for work, there may not be a wide range of specialty positions available - and certain jobs may not even exist in the area.[50] In addition, the premise of the paper -that cities are a great place to experiment- may be applicable in other areas beyond the labor market to other aspects of life, such as learning about one's ideal partner.

[50] http://www.glassdoor.com/blog/ nd-jobs-small-cities/

Appendix

A Data Description and Additional Results

The SIPP includes three variables that provide information regarding the geographical location of the respondents. The first identifies the worker's state. The second variable records whether the respondent is located in a metropolitan area. The third variable identifies one of 93 MSAs and CMSAs (consolidated metropolitan statistical areas), as defined by the Office of Management and Budget. I use the three location variables to identify whether a worker has moved. In my specification, a worker moves when (at least) one of the three location variables changes from one wave to the next.

Table 14 presents the cross-tabulation of workers switching occupations and moving. Most workers in the sample neither switch occupations nor move. A significant fraction of workers switch 3-digit occupations every period, consistent with estimates from other data sets (see Moscarini and Thomsson, 2007 for estimates from the CPS and Kambourov and Manovskii, 2008 for estimates from the Panel Study of Income Dynamics). Moreover, 6.78% of the sample moves every year, in line with the estimates from the CPS during the same period (6.72%)[51] and between a fifth and a quarter of those moves also involve an occupation switch.

In my investigation, I exclude workers in the armed forces. Hourly wages are deflated to real 1996 dollars using the Consumer Price Index. The measure of population in each metropolitan area is from the 2000 Census. Population in non-metropolitan areas is set to 200,000.[52]

Table 15 reports the destination occupations that occupational switchers enter by city size. Workers in large cities are more likely to switch to managerial and professional occupations, as well as administrative support occupations. Conversely, workers in smaller cities are more likely to switch to occupations such as handlers, machine operators, farming and service occupations.[53]

In addition, the ratio of net over gross occupational flows does not differ across cities of different sizes. In particular, it is equal to 0.1357 in cities with more than 500,000 inhabitants and 0.1376 in cities with

[51] The annual rate moving probability (not including moves inside the same county) was 6.72% for employed and unemployed people 16 and over in the 1998-1999 period.
http://www.census.gov/hhes/migration/files/cps/p20-531/tab07.txt

[52] In the SIPP the metro area with the lowest population had 252,000 residents.

[53] Note that the wage premium reported in Table 1 controls for major occupations, so it is not driven by workers in large cities working in high-paying occupations. In the model presented in Section 3, I purposely shut down occupational differences and highlight the role of increased occupational availability in larger cities. Allowing for more productive occupations in larger cities would, of course, lead to even higher predicted wage premia. The above fact is consistent with the findings of Eeckhout et al. (2014), who find that high-paying occupations are more prevalent in large cities, whereas there are more average-paying occupations in small cities.

	Switch Occupations:	
Move:	No	Yes
No	88.36%	9.38%
Yes	1.77%	0.49%

Table 14: Move and Occupational Switch. Source: 1996 Panel of Survey of Income and Program Participation. 4-month probabilities. 340,071 observations.

Occupation:	City	Non-City
A. Managerial (003-037)	13.85%	10.53%
B. Professional (043-199)	10.47%	8.36%
C. Technical Support (203-235)	3.68%	3.33%
D. Sales (243-285)	13.47%	13.46%
E. Administrative Support (303-389)	16.60%	13.76%
F. Private Household Occupations (403-407)	0.89%	0.94%
G. Protective Service (413-427)	1.43%	1.58%
H. Service (433-469)	12.29%	13.27%
I. Farming (473-499)	1.66%	3.20%
J. Precision Production (503-699)	8.59%	9.87%
K. Machine Operators (703-799)	5.79%	7.98%
L. Transportation (803-859)	3.89%	4.70%
M. Handlers (864-889)	7.40%	9.03%

Table 15: Fraction of Occupational Switchers That Enter Each Occupation. City is a location with more than 500,000 inhabitants. Source: 1996 Panel of Survey of Income and Program Participation. Population based on 2000 Census.

fewer than 500,000 inhabitants. This ratio for every occupation is computed as the absolute di erence between ows in and out of every occupation over their sum. The numbers reported are the weighted ratio

1.77%

C Endogenous Occupation Creation

In this appendix, I extend the model to allow for the number of occupations in each location to be endogenously determined. In equilibrium, cities with larger markets are able to support more occupations.

C.1 Environment

Time is continuous. There is a set of cities $l \in \{1, ..., L\}$. Each city, l, is characterized by the number of its occupations, $m \in \{1, ..., M\}$ and its population N, both of which are determined endogenously.

As before, there is a population of risk-neutral workers with discount rate r. There is one final good. Producing the final good requires intermediate goods. There is no trade across cities. Each intermediate good is produced by a different occupation.[54] In each location, workers derive utility from the consumption of the final good given by

$$C_t = \left(\sum_{k=1}^{m} c_{kt}^{\frac{\gamma - 1}{\gamma}} \right)^{\frac{\gamma}{\gamma - 1}},$$

where $\gamma > 1$ and c_{kt} is the consumption of good k at time t. The number of goods, m, may vary across locations.

Increased population causes a negative externality to workers (e.g., increased congestion and thus unknown, and let $p^{ik}_{0l} \in (0,1)$ be the worker's prior belief that $\alpha^{ik}_l = \alpha_G$. Priors are drawn independently from a known distribution with support $[0,1]$ and density $g(\cdot)$ when a worker enters a city. To reduce notational congestion, I drop the t, l, and i sub/superscripts in what follows.

A worker with posterior belief p^k, provides $\alpha_G p^k + \alpha_B \left(1 - p^k\right)$ (expected) units of effective labor per unit of time. If w_k is wage per effective unit of labor offered by occupation k, then the worker's wage income per unit of time is

$$w_k \left(\alpha_G p^k + \alpha_B \left(1 - p^k\right)\right).$$

As in the previous setup, a worker leaves his current city either endogenously or exogenously, according to a Poisson process with parameter $\delta > 0$. Moving from one city to another entails a cost $c > 0$. A difference from the previous model is that now workers move to any city they choose.

Total output of good k per unit of time, q_k, is linear in labor

$$q_k = l_k, \tag{10}$$

and there is also a fixed cost of production, f, in terms of the final good. l_k is the total labor input in occupation k and given by

$$l_k = \theta_k (w_k | w_{-k}) N \int \left(\alpha_G p^k + \alpha_B \left(1 - p^k\right)\right) h_k \left(p^k | w_k, w_{-k}\right) dp^k, \tag{11}$$

where N is total population in the particular location, $\theta_k(.)$ is the fraction of the labor force employed in occupation k, h_k is the distribution of beliefs of those workers who choose to be employed in occupation k, and w_{-k} is the vector of wages offered in all occupations in that location other than k.

Any profits, π_k, are split among city residents. There is free entry of intermediate good producers.

C.2 Behavior

Producer's k profits are given by

$$\pi_k = b_k q_k - w_k l_k - P f,$$

where P is defined in equation (13). Substituting in for equation (12), using equation (10) and taking the first-order conditions leads to the following price for good k

$$b_k = \frac{\gamma}{\gamma - 1 + \frac{dw(q_k|w_{-k})}{db_k}} w(q_k|w_{-k}). \tag{14}$$

The upward-sloping labor supply curve implies that when the producer increases his output, he must offer a higher wage to attract workers. The optimal price takes this effect into account through the term $\frac{dw(q_k|w_{-k})}{db_k} < 0$.

Free entry of intermediate goods implies that new goods will be created as long as they sustain non-negative profits. I next show that profits, π, are increasing in city population, N.

Since the price is affected by the wage, through the demand for labor, and using $q_k = l_k$, I obtain

$$\frac{dw(q_k|w_{-k})}{db_k} = \frac{dw(q_k|w_{-k})}{dl_k}\frac{dl_k}{db_k} = \frac{dw(q_k|w_{-k})}{dq_k}\frac{dq_k}{db_k}.$$

Using equation (12) and focusing on the symmetric equilibrium where $b_k = b$ for all k and all producers hire the same number of workers and make the same profits, I obtain

$$\frac{dq_k}{db_k} = -\frac{\gamma}{b}\left(\frac{wIN + \pi m}{bm} + m^{\frac{1}{1-\gamma}} f\right),$$

where

$$I = \int \left[\alpha_G p^k + \alpha_B \left(1 - p^k\right)\right] h\left(p^k\right) dp^k.$$

Moreover

$$q_k = l_k = \theta(w_k|w_{-k} = w) N I(w_k|w_{-k} = w),$$

where

$$I(w_k|w_{-k} = w) = \int \left[\alpha_G p^k + \alpha_B \left(1 - p^k\right)\right] h\left(p^k|w_k, w_{-k} = w\right) dp^k.$$

Therefore

$$\frac{dw(q_k)}{dq_k} = \frac{1}{\frac{dq_k}{dw_k}} = \frac{1}{N\frac{d\theta(w_k|w_{-k}=w) I(w_k|w_{-k}=w)}{dw_k}}.$$

Note that since $\frac{dw_k}{dq_k}$ 0 (because when demand for labor increases, that is a move up the labor supply curve), then

$$\frac{1}{N\frac{d\theta(w_k jw_k = w) l(w_k jw_k = w)}{dw_k}} > 0 \;)$$

$$\frac{d\theta(w_k jw_k = w)\, l\,(w_k jw_k = w)}{dw_k} > 0.$$

Given the above and normalizing $w_k = w = 1$ obtains

$$\frac{dw(q_k jw_k)}{db_k} = \frac{\gamma}{bN\frac{d\theta l}{dw_k}} \left(\frac{lN + \pi m}{bm} + m^{\frac{1}{1-\gamma}} f \right). \tag{15}$$

Furthermore

$$\pi = (b \quad 1)q \quad Pf.$$

Substituting in for q and W and solving leads to

$$\pi = \frac{(b \quad 1)\, lN}{m} \; m^{\frac{1}{1-\gamma}} fb. \tag{16}$$

Substituting in equation (15) for π leads to

$$\frac{dw(q_k jw_k)}{db_k} = \qquad \gamma \; in \; w$$

IN[d] 859

In equilibrium, each worker's consumption of the final good is given by[59]

$$C = \frac{(\alpha_G - \alpha_B)\, p^k + \alpha_B}{P(m)},$$

where $P(m) = m^{\frac{1}{1-\gamma}} b$ and $b_k = b$ for all k.

Following the same steps as in Section 3.2, I show that a worker moves to another city when the posterior of all his occupations reaches:

$$\underline{p}(N,m) = \frac{(d-1)\left[rJ - \frac{\alpha_B}{P(m)} + z(N)\right]}{(d+1)\frac{\alpha_G - \alpha_B}{P(m)} - 2\left[rJ - \frac{\alpha_B}{P(m)} + z(N)\right]},$$

where

$$d = \sqrt{\frac{8(r+\delta)}{\quad}}$$

;

standard equilibrium condition that all workers are always indifferent across locations is replaced by the condition that only the workers who move are indifferent.

D Model Simulation and Calibration Details

D.1 Model Simulation Details

In order to find values for s, δ, c, p_0 and ζ, I discretize the setup presented in Section 3 and simulate it. Each step is 60 days. I exploit the ergodicity of the setup and simulate a single worker for 5,000,000 periods.

More specifically, the increment of the Wiener process, dW, in the flow output equation (equation (1)) is approximated by ε where

$$\varepsilon = \sqrt{\Delta} \text{ with probability } \frac{1}{2}$$

and

$$\varepsilon = -\sqrt{\Delta} \text{ with probability } \frac{1}{2}$$

and Δ is the discretization step. Indeed, the variance of a Wiener process over a specific time interval is equal to the length of that time interval, since $W_t - W_s \sim N(0, t-s)$. The central limit theorem allows me here to approximate the normal distribution by the sum of the above Bernoulli trials.

Therefore, the evolution of beliefs for the case of a good match ($\alpha_l^{ik} = \alpha_G$) over a period of length Δ is given by

$$p_{t+\Delta} = p_t + p_t(1-p_t)\,\zeta\left[\frac{\alpha_G\Delta + \sigma\varepsilon - (p_t\alpha_G + (1-p_t)\,\alpha_B)\,\Delta}{\sigma}\right]$$

which simplifies to

$$p_{t+\Delta} = p_t + p_t(1-p_t)^2\zeta^2\Delta + p_t(1-p_t)\,\zeta\varepsilon$$

Similarly, in the case of a bad match ($\alpha_l^{ik} = \alpha_B$), the belief process is given by

$$p_{t+\Delta} = p_t - p_t^2(1-p_t)\,\zeta^2\Delta + p_t(1-p_t)\,\zeta\varepsilon$$

where ε is defined above.

Once beliefs are updated, the worker then picks his occupation for the following period by choosing the one with the highest belief, as dictated by Proposition 1. The occupational switching probability is computed by calculating how many workers in the simulation are employed in an occupation different from the one they were employed in 4 months ago.

The Poisson process of exogenous reallocation with parameter δ is approximated by a Poisson distribution with parameter $\delta \Delta$. A positive realization is equivalent to a reallocation shock.

D.2 Model Calibration Details

I calculate the number of occupations in areas with fewer than 500,000 inhabitants as follows: I first calculate the population-weighted number of occupations in metro areas with fewer than 500,000 inhabitants, which in this case is equal to 249.4. I then assume that non-metro areas have the least number of occupations observed in a metropolitan area (in this case 75). Since 14.73% of the sample lives in non-metropolitan areas and 26.33% lives in metro areas with population fewer than 500,000, I compute the population-weighted number of occupations in non-dense areas to equal 186.9.

The 4-month switching probability to new occupations is calculated as follows: the 4-month occupational switching probability for white males with a college degree is 7.32%. However, not all of these are switches to new occupations: 30% of workers return to their original occupation within 4 years.[61] This implies an annual rate of return switches of approximately 7.5%. In other words, a third of all annual switches are not switches to new occupations. Therefore, the 4-month switching probability to *new* occupations is 4.82%.

In my sample I have 7,452 wage observations. In order to calculate the residual standard deviation of wages, I use the sample of white college-educated males and run a regression of wages on marital status, quartic in age, firm size, and 13 occupational dummies. The R-square of that regression is 33.22%, implying that the residual standard deviation is $5.97.

I match the five moments described in the main text. The weighting matrix used is the inverse of the variance-covariance matrix of these moments, which is obtained by bootstrapping the sample 10,000 times. Rather than attempting to find directly the cost of moving c, I find the moving trigger $\underline{p}$ instead and then calculate the associated cost for which this trigger is optimal. In order to calculate the optimal moving trigger $\underline{p}$ for a particular value of the moving cost, I simulate the model using different triggers, compute the worker's utility at each one, and then select the trigger associated with the maximum utility.

The coefficients from the occupational switching probability regressions use the same controls as those presented in Table 2 for the subsample used. Moreover, the coefficients reported for both the simulation

[61] Kambourov and Manovskii (2008)

and the data are from a linear probability regression.

The moving cost, c, is found to equal 91. The average 4-month wage in the model equals \$14.20, so the annual wage equals \$42.60. Taking into account that the average hourly wage in the data is also \$14.20 and assuming that a worker works for 2000 hours a year, I translate the moving cost found in the setup to dollars as follows: 2000 14.20 91/(14.20 3) = \$60,667.

In the calibration of the model with human capital accumulation, in order to compute the Gittins index, I need to calculate numerically the value of moving, J, which I do by simulating 1000 workers over 800 periods. In order to calibrated the speed of human capital accumulation θ, I follow Kambourov and Manovskii (2009b), whose estimates suggest that it takes 5 years to become experienced, which corresponds to $\theta = 0.067$.[62] Assuming that the speed of human capital accumulation is 50% faster in large cities, I set θ in large and small cities so as to ensure that the average speed of human capital accumulation in the economy equals 0.067. Finally, I now can no longer calibrate α_G and α_B in a separate step, but they need to be calibrated jointly with the rest of the parameters.

E Robustness

[62] Given that the time period is 4 months, in order for inexperienced workers to be become experienced in 5 years on average, must equal 1=(3 5) = 0 :067.

	Data	Baseline				
# of Occupations in Large and Small		12 and 5	11 and 5	12 and 4	11 and 6	15 and 5
Ratio (# Occ Large over # Occ Small)	2.28	2.4	2.2	3	1.83	3
Targeted Moments:						
Population Share in Large Cities	58.95%	58.93%	58.95%	58.88%	58.92%	59.06%
Moving Probability in Large Cities	0.50%	0.49%	0.50%	0.47%	0.53%	0.47%
Moving Probability in Small Cities	0.60%	0.58%	0.58%	0.56%	0.61%	0.56%
Higher Occup Sw Prob in Large	0.20%	0.93%	0.48%	0.79%	0.42%	0.45%
Higher Occup Sw Prob in Large (recent)	3.34%	2.98%	2.20%	4.01%	1.14%	2.46%
Mean Wage	$14.20	$14.20	$14.20	$14.20	$14.20	$14.20
Residual Wage Standard Deviation	$5.97	$5.97	$5.97	$5.97	$5.97	$5.97
Other Moments:						
Prob of Switch to New Occupation	4.82%	4.29%	3.89%	3.69%	4.33%	4.51%
Initial Wage	$10.92	$10.23	$9.76	$9.76	$9.99	$9.48
Wage Premium	20.16%	6.86%	6.95%	10.45%	4.92%	9.58%
Wage Standard Deviation Premium	21.21%	7.21%	3.37%	2.95%	4.45%	3.41%
Parameters:						
α_G		28.93	25.36	24.76	28	24.68
α_B		8.29	7.93	8	7.98	7.64
s		54.96%	55.02%	54.25%	54.93%	54.96%
δ		0.00489	0.00488	0.0047	0.00513	0.00471
$\frac{\alpha_G \quad \alpha_B}{\sigma}$		0.1795	0.2302	0.2406	0.1886	

	Data					
# of Occupations in Large and Small		18 and 6	9 and 4	9 and 5	10 and 5	14 and 6
Ratio (# Occ Large over # Occ Small)	2.28	3	2.25	1.8	2	2.33
Targeted Moments:						

	Data		
# of Occupations in Large and Small		13 and 5	13 and 6
Ratio (# Occ Large over # Occ Small)	2.28	2.6	2.17
Targeted Moments:			
Population Share in Large Cities	58.95%	58.97%	59%
Moving Probability in Large Cities	0.50%	0.47%	0.51%
Moving Probability in Small Cities	0.60%	0.56%	0.59%
Higher Occup Sw Prob in Large			

References

Allen, T. (2014): Information Frictions in Trade, *Econometrica*, 82, 2041 2083.

Alvarez, F. and R. Shimer (2011): Search and Rest Unemployment, *Econometrica*, 79, 75 122.

Antonovics, K. and L. Golan (2012): Experimentation and Job Choice, *Journal of Labor Economics*, 30, 333 366.

Banks, J. S. and R. K. Sundaram (1994): Switching Costs and the Gittins Index, *Econometrica*, 62, 687 694.

Baum-Snow, N. and R. Pavan (2012): Understanding the City Size Wage Gap, *Review of Economic Studies*, 79, 88 127.

(2013): Inequality and City Size, *The Review of Economics and Statistics*, 95, 1535 1548.

Baumgardner, J. R. (1988): The Division of Labor, Local Markets, and Worker Organization, *Journal of Political Economy*, 96, 509 527.

Becker, G. S. and K. M. Murphy (1992): The Division of Labor, Coordination Costs, and Knowledge, *Quarterly Journal of Economics*, 107, 1137 1160.

Bergemann, D. and J. Valimaki

Costinot, A. and J. Vogel (2010): Matching and Inequality in the World Economy, *Journal of Political Economy*, 118, 747 786.

Davis, D. R. and J. I. Dingel (2018): A Spatial Knowledge Economy, *forthcoming, American Economic Review* 747 786.

Jovanovic, B. and R. Rob (1989): The Growth and Di usion of Knowledge, *Review of Economic Studies*, 56, 569 582.

Kambourov, G. and I. Manovskii (2008): Rising Occupational and Industry Mobility in the United States: 1968-1997, *International Economic Review*, 49, 41 79.

(2009a): Occupational Mobility and Wage Inequality, *Review of Economic Studies*, 76, 731 759.

(2009b): Occupational Speci city of Human Capital, *International Economic Review*, 50, 63 115.

Karatzas, I. (1984): Gittins Indices in the Dynamic Allocation Problem for Di usion Processes, *The Annals of Probability*, 12, 173 192.

Karlin, S. and H. Taylor (1981): *A Second Course in Stochastic Processes*, New York: Academic Press.

Kennan, J. and J. R. Walker (2011): The E ect of Expected Income on Individual Migration Decisions, *Econometrica*, 79, 211 251.

Kim, S. (1989): Labor Specialization and the Extent of the Market, *Journal of Political Economy*, 97, 692 705.

(1991): Heterogeneity of Labor Markets and City Size in an Open Spatial Economy, *Regional Science and Urban Economics*, 21, 109 126.

Kok, S. (2014): Town and City Jobs: How your Job is Di erent in Another Location, *Regional Science and Urban Economics*, 49, 58 67.

Krugman, P. (1991): Increasing Returns and Economic Geography, *Journal of Political Economy*, 99, 483 499.

Lee, S. (2010): Ability Sorting and Consumer City, *Journal of Urban Economics*, 68, 20 33.

Li, W. and R. Yao (2007): The Life-Cycle E ects of House Price Changes, *Journal of Money, Credit and Banking*, 39, 1375 1409.

Liptser, R. and A. N. Shiryaev (1977): *Statistics of Random Processes, Vol 2*, Berlin: Springer-Verlag.

Lucas, R. E. (1988): On the Mechanics of Economic Development, *Journal of Monetary Economics*, 22, 3 42.