Worker Sorting and Agglomeration Economies

Theodore Papageorgiou*

Penn State University, Department of Economics†

May 2013

Abstract

This paper contributes to our understanding of agglomeration economies in three ways: first it documents a number of novel facts relating to occupational switching patterns, moving patterns, and wages in large cities. Second, guided by these facts, it develops a model where larger cities have more occupations and as a result workers form better matches. Third, it calibrates the model to match moments relating to differences in moving probabilities and occupational switching probabilities and finds that better occupational match quality accounts for approximately 40% of the observed wage premium and a third of the greater inequality in larger cities.

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

JEL Classification: J24, J31, R23

*I am grateful to Joe Altonji, Nate Baum-Snow, Apostolos Burnetas, Ed Coulson, Pablo Fajgelbaum, Manolis Galenianos, Ed Glaeser, Ed Green, Yannis Ioannides, Boyan Jovanovic, Fabian Lange, Sanghoon Lee, Alex Monge, Steve Redding, Richard Rogerson, Esteban Rossi-Hansberg, Rob Shimer, Venky Venkateswaran as well as seminar participants at SED, CURE, HULM, CRETE, Princeton, St. Louis Fed, Cornell-Penn State Macro Workshop, Search and Matching Workshop at the University of Pennsylvania and Southwest Search and Matching Workshop at UC Davis for useful comments.

†502 Kern Building, University Park, PA 16802, USA. E-mail: *tpapageorgiou@psu.edu*. Web: *http://sites.google.com/site/theodorepapageorgiou/*

1 Introduction

Workers in more densely populated areas 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 that area increases. Over the years, economists have proposed several mechanisms such as human capital externalities and reduced transportation costs.[1] In a recent survey however Glaeser and Gottlieb (2009) note that "there remains a robust consensus among urban economists that [agglomeration] economies exist," but "the empirical quest to accurately measure such economies has proven to be quite difficult."

This paper contributes to our understanding of agglomeration economies in three ways. First we document a number of novel facts relating to moving patterns, occupational switching patterns and wages in large cities. These facts are not consistent with the standard urban theories. Second, guided by our findings, we develop a model where larger cities have more occupations which explains these facts. Third, we calibrate our model to match moments relating to moving probabilities and occupational switching probabilities and find that better occupational match quality accounts for approximately 40% of the observed wage premium and a third of the greater inequality in larger cities.

We first confirm the well-known regularity that workers in more densely populated cities earn higher wages. This wage difference however is not instantaneous, but instead appears with time in a location.[2] More specifically, when focusing on recent movers, workers who moved to a large city receive approximately the same wage as those who moved to a small city. At the same time recent movers to larger cities switch occupations at a higher rate than workers who moved to smaller cities. This difference reverses with time in the city and overall, the occupational switching rate is the same in large and small cities. Moreover, workers in larger cities are less likely to move to another location and switch occupations. We document that the patterns associated with moving and switching occupations are very different from those associated with moving and remaining in the same occupation. Finally, larger cities have more occupations and each doubling of a metropolitan area's population implies that approximately 100 more occupations appear. The above facts are not consistent with the standard urban theories[3] where

[1] See for instance Jacobs (1969), Lucas (1988), Jovanovic and Rob (1989), Krugman (1991), Glaeser et al. (1992) and Eaton and Eckstein (1997). See also Duranton and Puga (2004) for a literature survey.

[2] This is consistent with the findings of Glaeser and Maré (2001).

[3] Rosen (1979) and Roback (1982)

workers become immediately more productive upon arriving in larger cities and there is no worker reallocation in equilibrium.

Guided by these findings, we develop a spatial model with geographical mobility and occupational switching. The model's key features are: workers match with occupations and the quality of the match is uncertain and learned over time; there are more occupations in larger cities; it is costly to move across cities.

In equilibrium, increased options allow workers in larger cities to form better occupational matches compared to workers in smaller cities. Workers who recently moved to a large city do not initially form better matches than workers in smaller cities. As a result they do not receive higher wages. They have however more occupational options: this leads to higher occupational mobility for recent movers, who over time form better matches and obtain higher wages. Overall, occupational mobility is not higher in larger cities: on one hand workers have more options in larger cities; on the other they are on average better matched. These two effects roughly offset each other. Workers residing in larger cities are however unambiguously less likely to move, both because in equilibrium they are better matched and because they have more options. Workers who move experience wage declines before moving and wage gains upon moving, consistent with the data.

We calibrate the model using moments relating to differences in geographical mobility and occupational switching across different size cities. The model matches these moments well. It also matches the magnitude of occupational switching to new occupations, as well as the initial wage. We then look at the calibrated model's predictions regarding the wage premium and the greater wage inequality in larger cities: the model replicates approximately 40% of the observed wage premium and a third of the greater inequality in larger cities.

In our baseline setup the distribution of occupations across cities is exogenous. In the last part of the paper we extend the model to allow for the number of occupations 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 occupations, as well as population are endogenously determined.

This paper contributes to the literature that investigates the relationship between cities and labor market outcomes. It is related to papers that have argued that heterogeneous

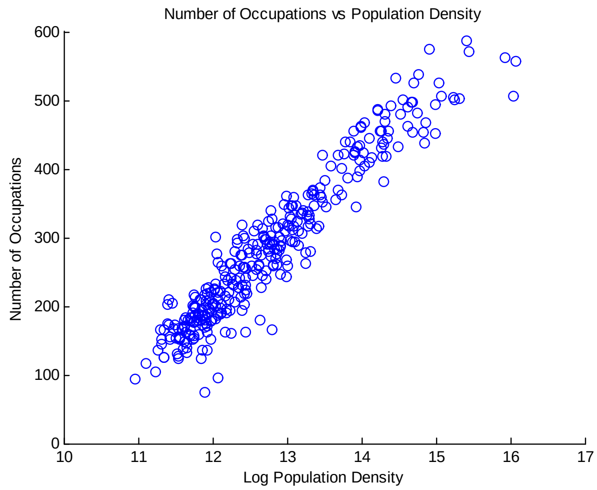

Figure 1: Number of Occupations vs. Population Density. Source: 2000 Occupational Employment Statistics. Population data from 2000 Census. 337 observations.

workers have more options in larger cities such as Helsley and Strange (1990), Kim (1991) and Andersson et al. (2007). These papers consider static setups and as a result do not have predictions regarding worker reallocation. The setup in the present paper however is dynamic and has a number of testable predictions. These predictions allow us to use the corresponding moments in the data and unlike the above-mentioned papers, calibrate our model.

The rest of the paper is organized as follows: Section 2 documents a number of facts on wages, moving patterns and occupational switching patterns in large cities. In Section 3 we introduce a model that is consistent with these facts and in Section 4 we calibrate it. Section 5 extends the model by endogenizing the number of occupations in each location. Section 6 concludes.

2 Facts

We first look at the number of occupations in each location and how it relates to population density. We then turn to micro level data to investigate: the city size wage premium and its evolution with time in the city; occupational switching and how it varies with city size; the patterns associated with moving and switching occupation and how they are different from those associated with moving and remaining in the same occupation.

We begin with the relationship between the number of occupations and population density. Using the 2000 Occupational Employment Statistics, in Figure 1, we plot the number of detailed occupations reported in each of 337 metropolitan areas against (logged) population in that area as reported in the 2000 Census.[4] As shown in the graph, the relationship between the two variables is positive and approximately log-linear. Each doubling of a city's population implies that approximately 100 more occupations appear.[5] The results here are also consistent with the findings of Duranton and Jayet (2011) who use French data to show that scarce occupations are more likely to be found in large cities.

For the remainder of this section, our main source of data is the 1996 Survey of Income and Program Participation (SIPP). In the 1996 SIPP, interviews were conducted every four months for four years and included approximately 36,000 households. It contains information about the worker's wage, three-digit occupation, three-digit industry, employer size, as well as the usual demographics, such as gender, age, race, education and marital status. The 1996 panel of the SIPP uses dependent interviewing, which is found to reduce occupational coding error (Hill (1994)). This makes it preferable to use when investigating occupational switching, compared to other panel datasets, such as the National Longitudinal Survey of Youth 1979. Furthermore, the SIPP follows original respondents when they move to a new address, unlike, for instance, the Current Population Survey which is an address-based survey. Appendix A contains more details about the data.

We first examine the evolution of the city size wage premium as a function of time in a city. The last column of Table 1, confirms the well-known empirical regularity that workers in more densely populated areas are paid significantly higher wages. The magnitude of the coefficient is in line with the results from other datasets.[6] In the first column of Table 1, we see that workers who just moved, also receive higher wages if they moved to a densely populated area, but the coefficient is smaller. Expanding the set to include workers who moved within the past four years leads to an increase of the urban wage premium equal to about half of that of the full sample.

This implies that the mechanism that generates these wage differences, appears to be

[4]Every year, the Bureau of Labor Statistics publishes Occupational Employment Statistics which report estimates of occupational employment in each metropolitan area. These estimates are based on a semiannual survey of nonfarm establishment selected from the list of establishments maintained by State Workforce Agencies for unemployment insurance purposes.

[5]When considering "broad" occupational groups instead of "detailed" groups (436 "broad" groups instead of 725 "detailed" ones), each doubling of a city's population implies that approximately 57 more occupations appear. For example "Anesthesiologists", "Family and General Practitioners", "Internists, General", "Obstetricians and Gynecologists", "Pediatricians, General" and "Surgeons" are different "detailed" occupations, but are all in the same "broad" occupation.

[6]See for instance column 1 of Table 4 in Glaeser and Gottlieb (2009) who use data from the Census Public Use Microdata Sample. See also Eeckhout et al. (2013).

	Initial	Moved <4 years	Full Sample
	ln(wage)	ln(wage)	ln(wage)
ln (population)	0.0155	0.021	0.041
	(0.009)*	(0.01)**	(0.001)***

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, firm size, quadratic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. 1,261, 4,321 and 169,536 observations respectively.

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

Table 2: Move and Occupational Switch. Source: Source: 1996 Panel of Survey of Income and Program Participation. 4-month probabilities. 340,071 observations.

relevant mostly *after* a worker arrives in a more densely populated area. Models where agglomeration economies are captured via an increasing function between TFP and city size cannot replicate this fact as they predict that workers moving to dense cities should immediately earn higher wages compared to those moving to less dense locations.

We next explore the patterns of occupational switching and geographical mobility. We begin with some basic facts: Table 2 presents the cross-tabulation of workers switching occupations and moving. Most workers in our sample neither switch occupations, nor move. A significant fraction of workers switch 3-digit occupations every period, consistent with recent estimates from other datasets (see Moscarini and Thomsson (2007) for estimates from the CPS and Kambourov and Manovskii (2008) for estimates from the PSID). Moreover, 6.78% of our sample moves every year, in line with the estimates from the CPS during the same period (6.72%)[7] and between a fifth and a quarter of those moves also involve an occupation switch.

We now turn to the relationship between occupational mobility and city size. In the first column of Table 3 we note that occupational mobility is somewhat lower in more densely populated areas. When focusing however on workers who moved to a location in the past 4 years, we notice that they are more likely switch occupations in denser locations.

[7]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

	Occ. Switching	Occ. Switching
	Prob. (Probit)	Prob. (Probit)
	All Residents	Moved <4 years
ln (population)	-0.0025	0.0109
	$(0.0006)^{***}$	$(0.0067)^{*}$

Table 3: Population Density Impact on Occupational Switching Probability. Source: Source: 1996 Panel of Survey of Income and Program Participation. Population data from 2000 Census. 4-month probabilities. Controls include gender, race, education, marital status, firm size, quadratic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.1016 (overall) and 0.1830 (recent). 140,842 and 3,360 observations respectively.

	Prob of Moving
	& Switching Occup (Probit)
ln (population)	-0.0004
	$(0.0001)^{***}$

Table 4: Population Density Impact on Probability of Moving and Switching Occupations. Source: 1996 Panel of Survey of Income and Program Participation. Population data from 2000 Census. 4-month probabilities. Controls include gender, race, education, marital status, firm size, quadratic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.0048. 242,666 observations.

This finding is consistent with the results in Bleakley and Lin (2012). Table 4 shows that the probability of moving and switching occupations is lower for residents of more densely populated areas. It's worth noting that the effect is quantitatively large: each doubling of the population reduces the probability of moving and switching occupations by 8.33%. On the other hand, from Table 3 it reduces the probability of an occupational switch by only 2.46%. In the setup described in the next section, larger cities offer workers more occupational options who in equilibrium form better occupational matches. As a result, city size has an ambiguous effect on occupational switching (on one hand workers are better matched, on the other they have more options); it predicts however that workers are unambiguously less likely to move from a large city (both because they are better matched and because they have more options).

We next show that the patterns associated with moves that are coupled with occupational switches are fundamentally different from moves where workers remain in the same

	Prob of Moving	Prob of Moving	Prob of Moving
	(Probit)	& Switching Occup (Probit)	& No Occup Switch (Probit)
Moved <4 years	0.0032	0.0035	-0.0022
	(0.0009)***	(0.0003)***	(0.0008)***

Table 5: Repeat Migration. Source: 1996 Panel of Survey of Income and Program Participation. 4-month probabilities. Controls include gender, race, education, marital status, firm size, quadratic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.0226, 0.0046 and 0.0171. 291,799, 271,597 and 271,597 observations respectively.

occupation.

First we consider the prevalence of repeat migration. As shown in the first column of Table 5, workers who moved in the past 4 years are more likely than average to move again: a worker who moved recently, is 0.32% more likely to move again, compared to an unconditional probability of 2.26%. The increased probability of repeat migration is well-known and has been documented before (DaVanzo (1983), Coen-Pirani (2010)).

In the second column of Table 5, the dependent variable becomes moving and switching occupations and we again find an increased probability of repeat migration. The effect is quantitatively large, as the probability of moving and switching occupations increases by 0.35%, when the unconditional probability is 0.46%. When we consider however, in the third column of Table 5, the probability of moving, but not switching occupation, we find that this positive relationship no longer holds and that in fact recent migrants are now significantly less likely to move and remain in the same occupation.[8] In other words, the increased probability of repeat migration that has been documented is driven exclusively by workers who move and switch occupations.

Switching occupations in the recent past also affects the probability of moving. As shown in the first column of Table 6, workers who have switched occupations in the past 4 years are more likely to move. When considering the probability of moving and switching occupations in the second column, the same effect is there and it is quantitatively large (a 0.23% increase, when the unconditional probability is 0.46%). The same effect is again absent when considering the probability of moving but not moving occupations, as shown in the third column of Table 6.

[8] The effect is quantitatively smaller: the probability of moving, but not switching occupations falls by 0.22%, when the unconditional probability is 1.71%.

	Prob of Moving (Probit)	Prob of Moving & Occup Switch (Probit)	Prob of Moving & No Occup Switch (Probit)
Switched Occup <4 years	0.0026	0.0023	0.0005
	(0.0006)***	(0.0003)***	(0.0006)

Table 6: Effect of Occupational Switching on Moving Prob. Source: 1996 Panel of Survey of Income and Program Participation. 4-month probabilities. Controls include gender, race, education, marital status, firm size, quadratic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.0224, 0.0046 and 0.0172. 236,467, 226,517 and 226,517 observations respectively.

	$\ln(\text{wage})_t$
$\ln(\text{wage})_{t-1}$	0.85
	(0.004)***
Move & Switch Occupations	0.023
	(0.012)**
Occupation Switch without Moving	0.011
	(0.002)***
Move & No Occupation Switch	0.005
	(0.003)*

Table 7: Impact of Occupation Switching and Moving on Wage. Source: 1996 Panel of Survey of Income and Program Participation. 4-month intervals. Controls include gender, race, education, marital status, firm size, quadratic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. 176,013 observations.

In Table 7, we note that moving and switching occupations is associated with a wage increase of approximately 2.3%. On the other hand, moving without switching occupations has a small impact on the wage, while switching occupations without moving leads to a 1.1% wage increase.

It is also worth considering the path of wages before moving. In the first column of Table 8 we see that if a worker is going to move in period t, then his wage falls by about 1% from period $t-2$ to $t-1$, indicating a declining wage path. This suggests that for at least some of the moves we observe, labor market considerations are important in the decision to move.

In the second column of Table 8, we note that wages are declining beforehand only in the case of workers who move and switch occupations, whose wages fall by approximately 2.4%. Workers who move and keep the same occupation do not experience decreasing

	$\ln(\text{wage})_{t-1}$	$\ln(\text{wage})_{t-1}$
Move_t	-0.01	
	$(0.003)^{***}$	
$\text{Move}_t \times$ Occupation Switch_t		-0.024
		$(0.008)^{***}$
$\text{Move}_t \times$ No Occupation Switch_t		-0.002
		(0.004)
$\ln(\text{wage})_{t-2}$	0.843	0.847
	$(0.004)^{***}$	$(0.004)^{***}$

Table 8: Wage Path Before Moving. Source: 1996 Panel of Survey of Income and Program Participation. 4-month intervals. Controls include gender, race, education, marital status, firm size, quadratic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. 158,087 and 146,462 observations.

wages before moving. Thus the results of the first column are driven by workers who move and switch occupations.[9]

Summarizing, more densely populated areas have more occupations. Workers who recently moved to a large city do not immediately earn higher wage; they are however more likely to switch occupations compared to those who moved to a small city. With time in the city this difference disappears and in the cross-section there is no relationship between city size and occupational switching. Workers in larger cities are less likely to move to another location and switch occupations. The patterns associated with moving and switching occupations differ from those associated with moving and remaining in the same occupation (wage declines before move, wage gains upon moving, recent movers more likely to move, recent switchers more likely to move).

3 Baseline Model

Guided by the above facts we develop a model of occupational choice and geographical mobility that accounts for them. In the model presented in this section, the number of occupations in each location is exogenous and workers who decide to move cannot choose their new location. In Section 5, we relax both assumptions.

The basic environment is the following: different cities have a different number of

[9] If we control for occupation switching separately in Table 8, only occupation switching is statistically significant, whereas moving is not. This is consistent with the predictions of the setup of the next section where workers who switch occupations experience wage declining paths beforehand, regardless of whether they move or not; moving does not imply a steeper wage decline, compared to switching occupations without moving.

occupations. Within a city, workers draw their productivity at each occupation. In a frictionless world, workers enter the occupation in which they are the most productive. We, however, introduce the following friction which induces occupational switching: individuals do not know their occupation-specific match, but learn it over time (Jovanovic (1979), Miller (1984), Moscarini (2005)).[10] If workers fail to find a suitable occupation they move to another city by paying a moving cost.[11]

We focus on occupations for two reasons: first and most importantly, the recent literature has emphasized the importance of occupations rather than firms for worker labor market outcomes.[12] The common theme of this literature is that a worker's wage depends on the type of work they do (their occupation), rather than who is employing them. For instance, an accountant's wage reflects how good he is in his accounting tasks, rather than which particular firm is employing him. Second, work by Baum-Snow and Pavan (2012) has found that worker-firm match qualities and search frictions do not differ much across cities of different sizes.

We next describe our setup in detail.

3.1 Economy

Time is continuous. There is a population of workers who are risk neutral and have discount rate $r > 0$.

There is a measure of cities. Each city is characterized by the number of occupations available, $m \in \{1, 2..., M\}$. The distribution of occupations across cities is exogenous and let s_m denote the fraction of cities with m occupations. Within each city, there is a large mass of firms for each occupation.[13]

[10] We follow the recent literature that has argued that occupational mobility is largely due to information frictions (e.g. Groes, Kircher and Manovskii (2010), Papageorgiou (2012)). However, the assumption that workers don't know their productivity is not crucial. The alternative is for workers to know their productivity in all occupations, but the worker's productivity in his current occupation could be changing over time, leading to occupational switches.

[11] This differs from most urban models where mobility is assumed to be costless or very cheap. See for example Rosen (1979) and Roback (1982), as well as more recently Eeckhout (2004) and Van Nieuwerburgh and Weill (2010).

Kennan and Walker (2011) estimate sizable moving costs across states which are increasing with age. In their setup, a worker who moves pays a deterministic cost which depends on age, distance etc., but benefits from the difference in flow payoffs between the origin and destination. The average value of the cost is large, but the gains from the flow payoff differences are also substantial. They estimate their model using data from the NLSY 79 whose respondents are relatively young. See also Hardman and Ioannides (1995) for a discussion of moving costs related to housing.

[12] Kambourov and Manovskii (2009), Antonovics and Golan (2012), Groes, Kircher and Manovskii (2010), Eeckhout and Weng (2010), Alvarez and Shimer (2009, 2011), Papageorgiou (2012).

[13] Alternatively one can assume away firms and assume that workers are engaged in home production

Workers can move from one city to another. A worker leaves his current city either endogenously, or exogenously according to a Poisson process with parameter $\delta > 0$. Search for cities is undirected. Moving from one city to another entails a cost $c > 0$.

While in a city a worker works in only one occupation any time. Moreover, a worker can switch occupations at no cost. Flow output for worker i, in occupation k, in city l at time t is given by:

$$dY_{tl}^{ik} = \alpha_l^{ik} dt + \sigma dW_{tl}^{ik}$$

where dW_{tl}^{ik} is the increment of a Wiener process and $\alpha_l^{ik} \in \{\alpha_G, \alpha_B\}$ is mean output per unit of time and $\sigma > 0$.

Without loss of generality we assume that $\alpha_G > \alpha_B$. Productivities, α_l^{ik}, are independently distributed across occupations, cities and workers. Furthermore, α_l^{ik} is unknown, and let $p_{0l}^{ik} \in (0,1)$ be the worker's prior belief that $\alpha_l^{ik} = \alpha_G$. When he enters a city, the worker draws his prior, p_{0l}^{ik}, for all occupations in that city. Each prior, p_{0l}^{ik}, is drawn independently from a known distribution with support $[0,1]$ and density $g(\cdot)$.

Workers observe their output and obtain information regarding the quality of their match in that specific occupation. Let p_{tl}^{ik} denote the posterior probability that the match of worker i with occupation k is good, i.e. $\alpha_l^{ik} = \alpha_G$. In particular, a worker observes his flow output, dY_{tl}^{ik}, and updates p_{tl}^{ik}, according to (Liptser and Shyryaev (1977)):

$$dp_{tl}^{ik} = p_{tl}^{ik}\left(1 - p_{tl}^{ik}\right)\zeta \frac{dY_{tl}^{ik} - \left(p_{tl}^{ik}\alpha_G + \left(1 - p_{tl}^{ik}\right)\alpha_B\right)dt}{\sigma} \tag{1}$$

where $\zeta = \frac{\alpha_G - \alpha_B}{\sigma}$. The last term on the right hand side is a standard Wiener process with respect to the unconditional probability measure used by the agents. p_{tl}^{ik} is a sufficient statistic of the worker's beliefs regarding α_l^{ik}. To minimize notation, from now on, we drop the t and l subscripts, as well as the i superscript.

The sequence of actions is the following: a worker moves to a city. He observes the number of occupations there, m, and draws his prior p_{0l}^{ik} for each occupation. He then chooses one of the occupations and begins working there, or alternatively he can pay c and move to another city.

in a particular occupation.

3.2 Behavior

Firm competition for the services of workers, ensures that a worker's compensation equals his expected output in the occupation n, he is employed:[14]

$$w(p^n) = \alpha_G p^n + \alpha_B (1 - p^n)$$

Each posterior evolves independently and only when the worker is employed in the corresponding occupation. Therefore the worker's problem is a multi-armed bandit one. The worker values both high current (expected) output, but also information, which allows him to make better decisions in the future. In other words, he may be facing a trade-off between exploration (trying an arm/occupation to figure out the underlying match quality) and exploitation (working in the occupation the gives him the highest wage).

Following Whittle (1980, 1982) and Karatzas (1984), the solution to this problem consists of finding a retirement value for each occupation and then work in the occupation with the highest retirement value. This retirement value serves as an index for each occupation, which corresponds to that occupation's Gittins index (see Gittins and Jones (1974) and Bergemann and Valimaki (2008)). More specifically, the index of each arm (occupation) is the retirement value at which the worker is exactly indifferent between continuing with that arm or retiring. We are able to use the 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)). The advantage of the Gittins index is that it drastically reduces the dimensionality of the problem. Whereas a worker's value depends on his beliefs regarding m arms (occupations), calculating the index of each arm, k, depends only that arm's beliefs (in this case p^k).

We first compute the optimal retirement policy for every occupation, k, with probability p^k of being good and the option of retiring with value W^k. In other words, at every instant, workers can work in this occupation 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, $V^k(p^k, W^k)$, satisfies the following Hamilton-Jacobi-Bellman equation:

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

[14] Alternatively, one can assume that workers sell their output every period to the firms. In that case, the value function of the worker remains the same, since it depends on the expectation of next instant's output. All the implications derived later continue to hold.

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:

$$J = -c + \sum_{m=1}^{M} s_m E_{\mathbf{p}} V(\mathbf{p}_m)$$

where $E_{\mathbf{p}} V(\mathbf{p}_m)$ 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 fraction of cities with m occupations and:

$$\mathbf{p}_m = \left[p^1 \; p^2 ... \; p^m\right] \in \mathbb{R}^m$$

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:

$$V^k\left(\underline{p}\left(W^k\right), W^k\right) = W^k \tag{2}$$

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

The solution to the above differential equation is given by:

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

where:

$$\underline{p}\left(W^k\right) = \frac{(d-1)\left((r+\delta)W^k - \alpha_B - \delta J\right)}{(d+1)(\alpha_G - \alpha_B) - 2\left((r+\delta)W^k - \alpha_B - \delta J\right)} \tag{3}$$

and $d = \sqrt{\frac{8(r+\delta)}{\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2} + 1}$. 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 $W^k = W\left(p^k\right)$, i.e.:

$$W\left(p^k\right) = V^k\left(p^k, W^k\right) \tag{4}$$

For eq. (4) to hold, from eq. (2), it must be the case that:

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

Substituting condition (5) into the threshold condition, equation (3), we obtain:

$$\begin{aligned} p^k &= \frac{(d-1)\left((r+\delta)W\left(p^k\right) - \alpha_B - \delta J\right)}{(d+1)(\alpha_G - \alpha_B) - 2\left((r+\delta)W\left(p^k\right) - \alpha_B - \delta J\right)} \Rightarrow \\ W\left(p^k\right) &= \frac{1}{r+\delta}\frac{(d+1)(\alpha_G - \alpha_B)p^k + \left(2p^k + d - 1\right)(\alpha_B + \delta J)}{2p^k + d - 1} \end{aligned} \tag{6}$$

which is strictly increasing in p^k, leading to the following proposition:

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

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

We also allow workers the option of moving to another city. This option provides known value to the worker, J. In the bandit problem this is equivalent to a "safe arm" that always pays w_j and is valued at J, i.e.:

$$J = \frac{1}{r}w_j$$

Note that since J is the retirement value associated with playing the safe arm, this corresponds to the Gittins index of the safe arm. A worker therefore will 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), we use equation (6) 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 $\underline{p}$, where:*

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

In the online appendix, we derive optimal worker behavior in the case where workers also accumulate occupation-specific human capital.

Summarizing, consider a worker who has just moved to a city. He immediately draws a prior, p_0^k, for each of the m occupations that are available for him to work in. If all m draws are below $\underline{p}$, he immediately pays the moving cost c and starts over in another city. Otherwise, he picks the occupation with the greatest value of the prior and begins work there. If the value of his posterior in that occupation falls below the value of the second best occupation, he immediately switches. A worker leaves his current city endogenously, only when value of the posteriors of all his occupations reach $\underline{p}$.[15] Some workers however may find that one of the occupations they try out is a good match for them, in which case their posterior drifts towards one and their wage increases. These workers leave their match and city only exogenously at rate δ.

3.3 Implications

In our setup, workers in cities with more occupations have more options and therefore we expect them to be on average better matched. This implies that they are also more likely to have higher output. Since firm competition ensures workers are paid their marginal product, workers in cities with more occupations, m, are expected to earn on average higher wages. This is indeed confirmed by the model's calibration results (Section 4).

We next examine our setup's implications regarding geographical mobility. Consider the probability that a worker moves from a location. From Proposition 2, a worker leaves a city when his posterior for all occupations is less than or equal to $\underline{p}$. Consider a worker who has moved to a another city with m occupations. Assume that $d \leq m$ of his draws are above $\underline{p}$. Then the probability he moves endogenously, conditional on d, is given by:

$$\Pr\left(p^1 \text{ reaches } \underline{p}\right) \times \Pr\left(p^2 \text{ reaches } \underline{p}\right) \times ... \times \Pr\left(p^d \text{ reaches } \underline{p}\right)$$

Since $\Pr\left(p^k \text{ reaches } \underline{p}\right) < 1$ for all k with $p_0^k > \underline{p}$, the probability that a worker moves endogenously is decreasing in d.

However d, the number of draws above $\underline{p}$ is increasing in the total draws, m. Thus the probability that a worker moves endogenously is decreasing in m implying that the rate at which workers move out of a city is lower in cities with more occupations m. This 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, the above result immediately implies that cities with more occupations, m, have larger populations.[16] This is consistent with

[15] For some occupations, the drawn prior may be below $\underline{p}$. The optimal strategy for the worker involves ignoring those occupations and never working there.

[16] In fact the flow into larger cities is slightly larger, since the probability that all prior draws are less

the evidence in Figure 1. Moreover, as we saw in Table 4 above, workers in more densely populated areas are indeed less likely to move and switch occupations.

We now turn to the impact of moving on wages. In our framework, workers pay the cost, c and move because they expect a better match in their new location. Their last wage before the move is $w\left(\underline{p}\right)$, whereas in the new location, the worker chooses to work in the occupation with the highest prior, $p_0^k > \underline{p}$. Thus workers who move experience wage increases.

We also examine the path of wages before moving. In our 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 the evidence in Table 8.

We next turn to how the probability of switching occupations is affected by the number of occupations. If workers in cities enjoy a better selection of occupational choices, then we would expect their occupational switching decisions to differ from workers in less populated areas. From Proposition 1, the worker is always employed in the occupation where he has the highest posterior. Following Karlin and Taylor (1981), ignoring δ shocks, the probability of an occupational switch for a worker whose posterior in his current occupation is equal to $p_{(m)}$, is given by:

$$\text{Occ Switch Prob} = \Pr\left(p_{(m)} \text{ reaches } p_{(m-1)} \text{ before } 1\right) = \frac{1 - p_{(m)}}{1 - p_{(m-1)}} \tag{7}$$

where $p_{(m-1)}$ is value of the worker's second highest posterior. Clearly the above probability is decreasing in $p_{(m)}$ and increasing in $p_{(m-1)}$.

One might expect the setup to predict that occupational switching is higher in larger cities. That is not however, necessarily the case: workers in larger cities have higher posteriors in their current occupations $p_{(m)}$. Their second highest posterior, $p_{(m-1)}$, is also increasing in m, the total number of occupations. Therefore, without additional assumptions, the number of occupations has an ambiguous effect on the rate of occupational switching. Put differently, workers in larger cities are both better matched, which tends to decrease their switching probability, but also have better outside options, which increases the probability they switch. We revisit the model's predictions regarding occupational switching in the next section where we calibrate our setup.

than $\underline{p}$, is decreasing in m. This reinforces the result.

4 Calibration

Our model has implications regarding differences across cities in both wages (mean wage and wage inequality), as well as worker reallocation (differences in occupational switching among cities, differences in moving probabilities). Given that other models of agglomeration economies do not have predictions regarding differences in worker reallocation across city size, we use these moments to calibrate our framework and then examine its predictions regarding the wage premium and differences in wage inequality across cities.

We calibrate our setup to white males with a college education.[17] Moreover because our setup does not allow for moving and remaining in the same occupation, we drop workers who move and keep the same occupation. There are two types of locations in our setup: densely populated areas and less densely populated areas. In the data this corresponds to locations with more than 500,000 inhabitants and those with less.

The calibration proceeds in three steps. First we set the number of occupations in each of the two types of locations. Second, we use worker reallocation moments to pin down the key parameters of our setup (δ, c, ζ, p_0 and the share of each location type). Third, we choose α_G and α_B to match the economy mean wage and standard deviation. In what follows we discuss the calibration procedure. Appendix B contains more details.

Our setup is set in continuous time, but we sample the simulated data every 4 months to match the sampling in the SIPP. The discount rate is set to 5% annually (1.64% at the 4 month frequency).

In order to set the number of available occupations in each location (dense vs. non-dense), we use the data from the Occupational Employment Statistics used in Figure 1. The population-weighted average number of occupations in cities with population more than 500,000 is 468.7, while population-weighted average number of occupations in areas with less than 500,000 inhabitants is 186.9 and so the ratio of the two is 2.51.[18,19] Thus even in non-dense areas there is a substantial number of occupations. However it is reasonable to assume that a much smaller subset of these occupations is relevant for

[17] Gould (2007) documents that the premium is larger for workers in white-collar jobs that are typically held by college graduates.

[18] We calculate the number of occupations in areas with less than 500,000 inhabitants as follows: we first calculate the population-weighted number of occupations in metro areas with less than 500,000 inhabitants, which in this case is equal to 249.4. We 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 our sample lives in non-metropolitan areas and 26.33% lives in metro areas with population less than 500,000, we compute the population-weighted number of occupations in non-dense areas to equal 186.9.

[19] Using "broad" occupational groups instead of "detailed" groups (436 "broad" groups instead of 725 "detailed" ones), this ratio becomes 2.10.

α_G	32.22
α_B	6.97
s	54.96%
δ	0.00489
$\frac{\alpha_G-\alpha_B}{\sigma}$	0.1795
p_0	0.0937
c (implied $\underline{p}$)	111 (0.0304)

Table 9: Parameter Values

each worker (especially those with a college degree). In order to calibrate the number of relevant occupations for each worker, we look at the number of occupations a worker tries out in a given time period, which depends on the number of occupations available to him. The number of occupations a worker tries out in a given time period is effectively given by the inverse of the occupational switching probability to new occupations (which is equal to expected tenure in each occupation). In the data the four-month switching probability to *new* occupations is 4.82%. Based on the above observations we set the number of occupations to 12 for dense areas and 5 to non-dense areas. The ratio of occupations is 2.4. Moreover, as described below, the average four-month switching probability to new occupations in the calibrated model is 4.29%.

As shown in Section 2 workers who just moved to a new location, receive higher wages if they moved to a more densely populated area, but the coefficient is not large (first column of Table 1). If we restrict our sample to college-educated white males and focus on dense vs. non-dense locations, the coefficient becomes statistically insignificant (p-value of 0.42). In other words, workers who move to densely populated areas do not receive initially higher wages than their counterparts who move to non-dense areas. If we view this fact through the lens of our setup, it implies that the distribution from which the initial beliefs are drawn, G, has little variance. In our calibration therefore, we set the 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 3): 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 bigger cities offer. Indeed as we will see, our setup replicates this feature of the data.

We next need to specify values for the following 5 parameters: s, δ, c, p_0 and ζ, where s is the share of dense areas in the economy. In order to do that we use moments relating

Moments:	Data	Model
Pop Share in Dense	58.95%	58.93%
Moving Prob Dense	0.50%	0.49%
Moving Prob Non-Dense	0.60%	0.58%
Higher Occup Sw Prob in Dense	0.20%	0.93%
Higher Occup Sw Prob in Dense (recent)	3.34%	2.98%
Mean Wage	$14.20	$14.20
Wage Standard Deviation	$7.31	$7.31

Table 10: Targeted Moments

to the moving probabilities and occupational switching. More specifically, in order to pin down s, δ, c, p_0 and ζ we use the 5 following moments: the population share that lives in a dense area, the coefficient on dense in the occupational switching probability regression, the coefficient on dense in the occupational switching probability regression when conditioning on recent movers only, the four-month probability of moving for workers living in a non-dense area and the same probability for those living in a dense area. We simulate a discrete time approximation of the model presented in Section 3 and match the simulated moments with the ones from the data. The remaining 2 parameters, α_G and α_B are calibrated afterwards to match exactly the mean level of wages and the cross-sectional standard deviation of wages.[20] The full set of parameters is presented in Table 9. The implied cost of moving, c, equals $74,000.

Although a rigorous identification argument is impossible due to the complexity of our framework, we attempt to give an informal argument of how each parameter is identified from the data. The share of dense areas in the economy, s, is pinned down by the population share that lives in a dense area. The four-month moving probabilities for workers in dense and non-dense areas pin down the moving rate, δ and the moving cost, c. Finally the speed of learning, ζ and the level of the initial belief, p_0, are pinned down by the occupational switching probability regressions.[21]

The targeted moments are presented in Table 10. The calibration matches the targeted

[20]None of the other moments depend on the choice of α_G and α_B, so we are able to calibrate them separately. As described in Appendix B, rather than searching over the moving cost, c, the calibration treats $\underline{p}$ as a parameter and afterwards calculates the associated cost, c, for which the retrieved value of $\underline{p}$ is optimal.

[21]The speed at which workers update their beliefs depends on $p(1-p)\zeta$. Changing ζ affects the speed of learning (and the probability of an occupational switch) at all levels of beliefs. p_0 however affects the distribution of beliefs and therefore of the distribution of occupational switching probabilities: for instance, for beliefs that are high (p above 0.5), as p increases the speed of learning decreases, whereas as p goes down the speed of learning increases (before eventually decreasing). The opposite is true when beliefs are below 0.5.

Moments:	Data	Model
Switch Pr to New Occupations	4.82%	4.29%
Initial Wage	$10.92	$9.34

Table 11: Other Moments

	Data	Model
Wage Premium	20.16%	8.47%
Wage Standard Deviation Ratio	21.21%	7.21%

Table 12: Predicted Wage Premium and Greater Wage Inequality in Large Cities

moments well. In the calibrated model recent workers in dense locations are more likely to switch occupations, as in the data, whereas in the cross-section the differences in the occupational switching probabilities are small. Finally, by construction the calibration matches exactly the level and the standard deviation of wages by appropriately choosing α_G and α_B after the other moments have been matched.[22]

Table 11 presents some additional moments. The switching probability to new occupations is equal to 4.29%, close to the observed one (4.82%). As discussed above, we check this moment to evaluate our choice of setting the number of occupations to 12 in dense areas and 5 in non-dense areas. Moreover, the initial wage, which was not targeted in the calibration, is predicted to equal $9.34 a bit below the initial wage of $10.92 observed in the data.

Table 12 presents the predicted wage premium and the ratio of the cross-sectional standard deviation of wages, neither of which were targeted. The calibrated model replicates 42% of the observed wage premium. Moreover, it replicates about a third of the greater wage inequality that has been documented in larger cities. Wage top coding in the data implies that both numbers should be considered upper bounds. It's worth emphasizing that in our setup there is no ex ante worker heterogeneity. Gautier and Teulings (2009) and Eeckhout et al. (2013) introduce models that generate differences in the wage dispersion, but a key assumption in both setups is ex ante heterogenous workers who choose where to work. In the present paper the differences in both the level of wages, as well as wage inequality across city sizes are driven, not by selection, but greater occupational availability.

[22] In the CPS, median hourly wages for male, college graduates, 25 years of age and older are $19.22, assuming weekly hours are equal to 45 (see BLS Release "Usual Weekly Earnings of Wage and Salary Workers: Third Quarter 1996"). In our data however, hourly wages are topcoded at $30. Moreover, our sample also includes workers below 25 years of age.

5 Endogenous Occupation Creation

In this section, we extend the model developed in Section 3 to allow for the number of occupations in each location to be endogenously determined. Larger markets are able to support more occupations. For instance opera singers exist in larger cities, since an opera house is less likely to be profitable in a small town. A larger city caters to more diverse consumer tastes, producing and hiring in a larger variety of services and products. Related to the above, some occupations may reflect to some extent the degree of increased specialization that is possible in larger cities, for instance specialized engineers (see also Baumgardner (1988)).

The basic environment is the following: as before, workers learn about the quality of their occupational match and also decide whether to move or not. We now allow workers to choose their destination city. There is a final good produced by intermediate goods. Each intermediate good requires a specific task or occupation and entails a fixed cost of production. We show that profits are increasing in the size of the city (market), so in equilibrium cities with higher populations support more occupations. More occupations, in turn, attract a larger population as workers benefit from increased occupational availability as in the baseline model, but also increased consumption variety. Increased population density however also causes a negative externality, which prevents cities from becoming unboundedly large. In this setup, both the number of occupations, as well as population are endogenously determined; in equilibrium cities with larger populations have more occupations, consistent with the evidence in Figure 1.

5.1 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 will be 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. Each intermediate good is produced by a different occupation.[23] In each location, workers derive utility from the consumption of the final good given by:

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

[23] See also the specification in Teulings (1995) and Costinot and Vogel (2010).

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 density causes a negative externality to workers (e.g. increased congestion and thus commuting time, higher housing prices due to land scarcity etc.), which is captured by $z(N_t)$, where $\frac{dz(N_t)}{dN_t} > 0$ and $\frac{d^2 z(N_t)}{dN_t^2} > 0$.[24] Flow utility per unit of time is given by:

$$C_t - z(N_t)$$

$z(\cdot)$ can differ across locations.

As before workers work in only one occupation at a time. They can switch occupations at no cost. Worker i, in occupation k, in city l, at time t provides the following flow units of *effective labor*:

$$dY_{tl}^{ik} = \alpha_l^{ik} dt + \sigma dW_{tl}^{ik}$$

where dW_{tl}^{ik} is the increment of a Wiener process and $\alpha_l^{ik} \in \{\alpha_G, \alpha_B\}$. As in the model of Section 3, let $\alpha_G > \alpha_B$ and α_l^{ik}, are independently distributed across occupations, cities and workers. Moreover α_l^{ik} is unknown, and let $p_{0l}^{ik} \in (0,1)$ be the worker's prior belief that $\alpha_l^{ik} = \alpha_G$. 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 we drop the t, l and i sub/superscripts in what follows.

A worker with posterior belief p^k, provides $\alpha_G p^k + \alpha_B(1 - p^k)$ (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:

$$q_k = l_k \tag{8}$$

and there is also an fixed cost of production, f, in terms of the final good. l_k is the total

[24] See for instance Lucas and Rossi-Hansberg (2002) and Eeckhout (2004) who micro-found the negative externality assuming increased commuting time.

labor input in occupation k given by:

$$l_k = \theta_k N \int \left(\alpha_G p^k + \alpha_B \left(1 - p^k\right)\right) h_k \left(p^k\right) dp^k$$

where N is total population in the particular location, θ_k is the fraction of the labor force employed in occupation k and H_k is the distribution of beliefs of these workers who choose to be employed in occupation k.

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

5.2 Behavior

In what follows we consider a symmetric equilibrium, where all producers choose the same price, b ($b_k = b$ for all k) and commit to it.[25]

As before, workers observe the realized units of effective labor they supply in the occupation k where they are employed and update their beliefs regarding α^k following the process described by equation (1). Since the worker's problem is a multi-arm bandit one, as discussed in Section 3.2 the optimal solution is to be employed in the occupation with the highest Gittins index, as described in Proposition 1.

As in Section 3, in equilibrium each worker is employed in the occupation with the highest belief, $p_{(m)}$. Following the same steps as in Section 3.2 we show that a worker moves to another city when the posterior of all his occupations reaches:

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

where J value of a worker about to move to another city:

$$J = -c + \overline{V}$$

where:

$$\overline{V} = \max_l E_{\mathbf{p}} V\left(\mathbf{p}_{m_l}, N_l\right)$$

In other words the worker moves to the city, l, that maximizes his ex ante utility.

The predictions of the baseline setup introduced in Section 3 hold here as well. For instance, the effect of city size on occupational switching continues to be ambiguous, as

[25] Considering dynamic pricing by firms poses significant complications and is beyond the scope of this paper.

demonstrated by equation (7) and the related discussion. The only difference is that the moving probability now also depends on the level of the negative externality, $z(N)$, which could vary across different cities.

Workers demand goods for consumption. In particular, they spend their income (wage income and profits, π_k) on the final good of the city which is produced by intermediate goods. As shown in Appendix C, demand for intermediate good k is given by:

$$q_k = \left(\frac{b_k}{P}\right)^{-\gamma} \left(\frac{W}{P} + fm\right) \tag{9}$$

where:

$$P = \left(\sum_{k=1}^{m} b_k^{1-\gamma}\right)^{\frac{1}{1-\gamma}} \tag{10}$$

is the aggregate price level and:

$$W = \sum_{k=1}^{m} w_k l_k + \sum_{k=1}^{m} \pi_k$$

is total expenditure by city residents.

Each intermediate good producer chooses a price, b_k, given the demand he faces given in (9).[26] Equation (9) pins down the quantity of good k produced, q_k, which in turn pins down the amount of labor required, l_k (see equation (8)). Unlike other models of monopolistic competition, the producer here cannot hire as many workers as he wants at a given wage rate, but instead faces an upward-slopping labor supply curve. More specifically, the workers' occupational choice problem dictates the wage level, w_k, required to attract labor input l_k, which is necessary to produce q_k. Each producer takes this into account when choosing a price, b_k.

We now solve for the intermediate good producer's problem. Taking the first order condition leads to the following price for good k:

$$b_k = \frac{\gamma}{\gamma - 1 + \frac{dw(q_k|w_{-k})}{db_k}} w\left(q_k|w_{-k}\right) \tag{11}$$

The upward-slopping 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 $\frac{dw(q_k|w_{-k})}{db_k} < 0$.

[26] m is assumed to be large enough so that the pricing decision of each producer has a negligible impact on the aggregate price level, P.

Free entry of intermediate goods implies that new goods will be created as long as they sustain non-negative profits. In Appendix C we also show that profits, π, are increasing in city population, N. This immediately leads to the following proposition:

Proposition 3 *Cities with larger populations, N, have more occupations, m.*

The endogenous moving decision analyzed above, as well as the inflow decisions of movers pin down city population, N, in this model. Workers benefit from more occupations because they earn higher wage income due to the increased occupational availability and because they consume a greater variety of products.[27] On the other hand, higher population (which as shown above is required for more occupations) creates increasingly higher disutility, thus limiting the size of cities. If the function capturing this higher disutility, $z(\cdot)$, differs across locations, then in equilibrium there will be cities of different sizes. In this setup the standard equilibrium condition that workers are always indifferent across locations is replaced by the condition that only the workers who move are indifferent.

6 Conclusion

This paper documents a number of facts relating to city size, wages, geographical mobility and occupational switching probabilities. Guided by these facts we develop and calibrate a model where workers in larger cities have more occupations available and as a result form better matches. In our setup, agglomeration economies are not the result of larger cities exogenously having higher productivity. Rather, agglomeration economies are endogenously generated. We calibrate our model using moments relating to geographical mobility and occupational switching. The calibrated model replicates approximately 40% 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 reflected 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 flexible. 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."[28]

[27] See also Lee (2010).
[28] http://www.glassdoor.com/blog/find-jobs-small-cities/

Appendix

A Data Description

In our investigation, we exclude workers in the armed forces. Hourly wages are deflated to real 1996 dollars using the Consumer Price Index. Our measure of population in each metropolitan area is from the 2000 Census. Population in non-metropolitan areas is set to 200,000.[29] 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 or not. The third variable identifies one of 93 MSAs (Metropolitan Statistical Areas) and CMSAs (Consolidated Metropolitan Statistical Areas), as defined by the Office of Management and Budget. We also use the three location variables to identify whether a worker has moved. In our specification, a worker moves when (at least) one of the three location variables change from one wave to the next.

B Calibration Details

The four-month switching probability to new occupations is calculated as follows: the four-month occupational switching probability for white males with a college degree is 7.32%. Not all of these however, are switches to new occupations: 30% of workers return to their original occupation within 4 years.[30] 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 four-month switching probability to *new* occupations is 4.82%.

In order to find values for s, δ, c, p_0 and ζ we simulate of discrete-time version of our setup, where each step is 60 days. More specifically, we exploit the ergodicity of the setup and simulate a single worker for 5,000,000 periods. We 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, we 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 any value of the moving cost, we simulate the

[29] In the SIPP data, the lowest population count of a metro area is 252,000.

[30] Kambourov and Manovskii (2008)

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 3. Moreover the coefficients reported both for the simulation and the data are from a linear probability regression.

The moving cost, c, is found to equal 111. The average four-month wage in the model equals \$14.20, so the annual wage equals \$42.60. Taking into account that in the data the average hourly wage in the data is also \$14.20 and assuming that a worker works for 2000 hours a year, then we translate the moving cost found in our setup to dollars as follows 2000*14.20*111/(14.20*3) = \$74,000.

C Endogenous Occupation Creation Derivations

In this section we derive the equilibrium price and also the relationship between population and profits.

Demand for good k comes from two sources: consumers and producers paying for their fixed cost, f which is in terms of the final good. Solving the producer's problem, implies that the demand for good k by the m producers in that location is given by:[31]

$$\left(\frac{b_k}{P}\right)^{-\gamma} fm$$

Therefore total demand for good k is by equation (9).

Producer's k profits are given by:

$$\pi_k = b_k q_k - w_k l_k - Pf$$

where P is defined equation (10). Substituting in for equation (9), using equation (8) and taking first order conditions leads to equation (11).

[31] The producer's problem consists of choosing goods f_k and is given by:

$$\min_{f_k} \sum_{k=1}^{m} b_k f_k$$

subject to:

$$f \leq \left(\sum_{k=1}^{m} f_k^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}}$$

where f is the fixed cost necessary to begin producing.

Since the price is affected by the wage, through the demand for labor, and using $q_k = l_k$, we have:

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

Using equation (9) and focusing on the symmetric equilibrium where that $b_k = b$ for all k and that all firms hire the same number of workers and make the same profits, we have:

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

where:

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

Moreover we have that:

$$q_k = l_k = \theta\left(w_k|w_{-k} = w\right)NI\left(w_k|w_{-k} = w\right)$$

where:

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

Therefore:

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

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

$$\begin{aligned} \frac{1}{N\frac{d\theta\left(w_k|w_{-k}=w\right)I\left(w_k|w_{-k}=w\right)}{dw_k}} &> 0 \Rightarrow \\ \frac{d\theta\left(w_k|w_{-k} = w\right)I\left(w_k|w_{-k} = w\right)}{dw_k} &> 0 \end{aligned}$$

Given the above and normalizing $w_k = w = 1$, we have:

$$\frac{dw\left(q_k|w_{-k}\right)}{db_k} = -\frac{\gamma}{bN\frac{d\theta I}{dw_k}}\left(\frac{IN + \pi m}{bm} + m^{\frac{1}{1-\gamma}}f\right) \tag{12}$$

Furthermore we have:

$$\pi = (b - 1)\,q - Pf$$

Substituting in for q and W and solving leads to:

$$\pi = \frac{(b-1)\, IN}{m} - m^{\frac{1}{1-\gamma}} fb \tag{13}$$

Substituting in equation (12) for π we have:

$$\frac{dw\left(q_k | w_{-k}\right)}{db_k} = -\frac{\gamma I}{bm \frac{d\theta(w_k|w_{-k}=w) I(w_k|w_{-k}=w)}{dw_k}}$$

which we now substitute into the price equation (11) in order to obtain:

$$b = \frac{\gamma \left(m \frac{d\theta I}{dw_k} + I \right)}{(\gamma - 1)\, m \frac{d\theta I}{dw_k}}$$

Therefore profits (equation (13)) are increasing in N, since b does not depend on N.

References

[1] Alvarez F. and R. Shimer (2009): "Unemployment and Human Capital," Mimeo, University of Chicago, Department of Economics.

[2] Alvarez F. and R. Shimer (2011): "Search and Rest Unemployment," *Econometrica,* 79(1), 75-122.

[3] Andersson, F., S. Burgess and J. I. Lane (2007): "Cities, Matching and the Productivity Gains of Agglomeration," *Journal of Urban Economics*, 61(1), 112-128.

[4] Antonovics K. and L. Golan (2012): "Experimentation and Job Choice," *Journal of Labor Economics*, 30(2), 333-366.

[5] Banks, J. S. and R. K. Sundaram (1994): "Switching Costs and the Gittins Index," *Econometrica*, 62(3), 687-694.

[6] Baum-Snow, N. and R. Pavan (2012): "Understanding the City Size Wage Gap," *The Review of Economic Studies*, 79(1), 88-127.

[7] Baumgardner, J. R. (1988): "The Division of Labor, Local Markets, and Worker Organization," *The Journal of Political Economy*, 96(3), 509-527.

[8] Bergemann, D. and J. Valimaki (2008): "Bandit Problems," in *The New Palgrave Dictionary of Economics,* ed. by S. Durlauf and L. Blume, New York: Macmillan.

[9] Bleakley, H. and J. Lin (2012): "Thick-Market Effects and Churning in the Labor Market: Evidence from U.S. Cities," *Journal of Urban Economics*, 72(2) 87-103.

[10] Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics (www.bls.gov/oes/), date accessed November 21, 2012.

[11] Coen-Pirani, D. (2010): "Understanding Gross Worker Flows across U.S. States," *Journal of Monetary Economics*, 57(7), 769-784.

[12] Costinot, A. and J. Vogel (2010): "Matching and Inequality in the World Economy," *Journal of Political Economy*, 118(4), 747-786.

[13] DaVanzo, J. (1983): "Repeat Migration in the United States: Who Moves Back and Who Moves On?," *The Review of Economics and Statistics*, 65(4), 552-559.

[14] Duranton G. and H. Jayet (2011): "Is the Division of Labor Limited by the Extent of the Market? Evidence from French Cities," *Journal of Urban Economics*, 69(1), 56-71.

[15] Duranton G. and D. Puga (2004): "Micro-foundations of urban agglomeration economies," *Handbook of Regional and Urban Economics*, Volume 4, ed. by J.V. Henderson and J.-F. Thisse, Elsevier, 2063-2117.

[16] Eaton, J. and Z. Eckstein (1997): "Cities and Growth: Theory and Evidence from France and Japan," *Regional Science and Urban Economics*, 27(4-5), 443-474.

[17] Eeckhout, J. (2004): "Gilbrat's Law for (All) Cities," *American Economic Review*, 94(5), 1429-1451.

[18] Eeckhout, J., R. Pinheiro and K. Schmidheiny (2013): "Spatial Sorting," mimeo, University College London, Department of Economics.

[19] Eeckhout, J. and X. Weng (2010): "Assortative Learning," Mimeo, University of Pennsylvania, Department of Economics.

[20] Gautier, P. A. and C. N. Teulings (2009): "Search and the City," *Regional Science and Urban Economics*, 39(3), 251-265.

[21] Gittins, J. C. and D. M. Jones (1974): "A Dynamic Allocation Index for the Sequential Design of Experiments," in *Progress in Statistics*, vol. 1, ed. by J. M. Gani, K. Sarkadi, and I. Vincze, 241-266, Amsterdam: North-Holland.

[22] Glaeser, E. L. and J. D. Gottlieb (2009): "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, 47(4), 983-1028.

[23] Glaeser, E. L., H. D. Kallal, J. A. Scheinkman and A. Shleifer (1992): "Growth in Cities," *The Journal of Political Economy*, 100(6), 1126-1152.

[24] Glaeser, E. L. and D. C. Maré (2001): "Cities and Skills," *Journal of Labor Economics*, 19(2), 316-342.

[25] Gould, E. D. (2007): "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium," *Review of Economic Studies*, 74, 477-506.

[26] Groes, F., P. Kircher and I. Manovskii (2010): "The U-Shapes of Occupational Mobility," Mimeo, University of Pennsylvania, Department of Economics.

[27] Hardman, A. M. and Y. M. Ioannides (1995): "Moving Behavior and the Housing Market," *Regional Science and Urban Economics*, 25(1), 21-39.

[28] Helsley, R. W. and W. C. Strange (1990): "Matching and Agglomeration Economies in a System of Cities," *Regional Science and Urban Economics*, 20(2), 189-212.

[29] Hill, D. H. (1994): "The Relative Empirical Validity of Dependent and Independent Data Collection in a Panel Survey," *Journal of Official Statistics*, 10(4), 359-380.

[30] Jacobs, J. (1969): *The Economy of Cities,* New York: Vintage Books.

[31] Jovanovic, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5) Part 1, 972-990.

[32] Jovanovic, B. and R. Rob (1989): "The Growth and Diffusion of Knowledge," *The Review of Economic Studies*, 56(4), 569-582.

[33] Kambourov, G. and I. Manovskii (2008): "Rising Occupational and Industry Mobility in the United States: 1968-1997," *International Economic Review*, 49(1), 41-79.

[34] Kambourov, G. and I. Manovskii (2009): "Occupational Specificity of Human Capital," *International Economic Review*, 50(1), 63-115.

[35] Karatzas, I. (1984): "Gittins Indices in the Dynamic Allocation Problem for Diffusion Processes," *The Annals of Probability*, 12(1), 173-192.

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

[37] Kennan, J. and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211-251.

[38] Kim, S. (1991): "Heterogeneity of Labor Markets and City Size in an Open Spatial Economy," *Regional Science and Urban Economics*, 21(1), 109-126.

[39] Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3), 483-499.

[40] Lee, S. (2010): "Ability Sorting and Consumer City," *Journal of Urban Economics*, 68(1), 20-33.

[41] Liptser, R. and A. Shyryaev (1977): *Statistics of Random Processes,* Vol. 2 Berlin: Springer-Verlag.

[42] Lucas, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3-42.

[43] Lucas, R. E. and E. Rossi-Hansberg (2002): "On the Internal Structure of Cities," *Econometrica*, 70(4), 1445-1476.

[44] Miller, R. A. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy, 92(6), 1086-1120.*

[45] Moscarini, G. (2005): "Job Matching and the Wage Distribution," *Econometrica*, 73(2), 481-516.

[46] Moscarini, G. and K. Thomsson (2007): "Occupational and Job Mobility in the US," *Scandinavian Journal of Economics*, 109(4), 807-836.

[47] Van Nieuwerburgh, S. and P.-O. Weill (2010): "Why Has House Price Dispersion Gone Up?," *Review of Economic Studies*, 77(4), 1567-1606.

[48] Papageorgiou, T. (2012): "Learning Your Comparative Advantages," Mimeo, Penn State University, Department of Economics.

[49] Roback, J. (1982): "Wages, Rents and the Quality of Life," *Journal of Political Economy*, 90(6), 1257-1278.

[50] Rosen, S. (1979): "Wage-Based Indexes of Urban Quality of Life," in *Current Issues in Urban Economics*, ed. by P. Mieszkowski and M. Straszheim, 74-104, Baltimore and London: Johns Hopkins University Press.

[51] Teulings, C. N. (1995): "The Wage Distribution in a Model of the Assignment of Skills to Jobs," *Journal of Political Economy*, 103(2), 280-315.

[52] Whittle, P. (1980): *Optimization over Time: Programming and Stochastic Control*, New York: Wiley.

[53] Whittle, P. (1982): "Multi-Armed Bandits and the Gittins Index," *Journal of the Royal Statistical Society*, Series B, 42(2), 143-149.