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."

This paper contributes to our understanding of agglomeration economies in three ways. First, using several different data sets, I document a number of facts related to the number of occupations in large and small cities and the relationship between city size, occupational switching patterns, wages, and moving patterns. Second, guided by the findings, I introduce a dynamic model of occupational choice where larger cities have more occupations. This both provides a microfoundation for agglomeration economies and also explains these facts. Third, I calibrate the model to match worker reallocation moments and find that my mechanism accounts for approximately 35% of the observed wage premium and a third of the greater inequality in larger cities.

More specifically, using a comprehensive data set of online vacancies for the US, I find that workers in larger cities have significantly more occupational options than workers in smaller ones: the largest cities have more than 450 occupations, whereas small cities have fewer than 200. This difference is not driven by occupations that would interest few workers.

I next use a worker panel survey and confirm the well-known regularity that workers in more highly populated cities earn higher wages. However, this wage difference is not instantaneous, but instead appears with time in a location. More specifically, when I focus 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.

[1] See, for instance, Jacobs (1969), Lucas (1988), Jovanovic and Rob (1989), Krugman (1991), Glaeser et al. (1992), Eaton and Eckstein (1997), Glaeser (1999). See also Duranton and Puga (2004) and Carlino and Kerr (2015) for literature surveys.

Guided by these findings, I develop a spatial model with geographical mobility and occupational switching. My setup follows the literature arguing that agglomeration improves the expected quality of matches (see the discussion in Duranton and Puga, 2004); in particular it explains agglomeration economies as the result of workers forming better occupational matches. The model's key ingredients are the following: there are more occupations in larger cities; workers match with occupations; the quality of the match is uncertain and learned over time; it is costly to move across cities.

In equilibrium, greater occupational availability allows 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. However, they have more occupational options: this leads to higher occupational mobility for recent movers, consistent with the data, who over time form better matches and obtain higher wages. Overall, occupational mobility is not higher in larger cities: on the one hand, workers have more options in larger cities; on the other hand 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; so now these two effects work in the same direction. In addition, workers who move experience wage declines before moving and wage gains upon moving, consistent with the data.

In order to assess whether my mechanism is quantitatively important, I take my model to the data. One of the key issues that Duranton and Puga (2004) note in their survey of agglomeration economies is that almost all the proposed mechanisms are "observationally equivalent," implying that "empirically identifying and separating these mechanisms becomes very difficult." However the dynamic nature of the mechanism proposed here has a number of additional predictions regarding worker reallocation both within and across cities, as well as how this reallocation interacts with city size, wages, and time in the city that differentiate it from other mechanisms. I thus use these moments to calibrate the model, so as to pick up only the importance of my mechanism. The model matches these moments well. It also matches the magnitude of occupational switching to new occupations, as well as workers' initial wage. I then look at the calibrated model's predictions regarding the wage premium and the greater wage inequality in larger cities: the model replicates approximately 35% of the observed wage premium and a third of the greater inequality in larger cities.

In the baseline setup, some cities exogenously have more occupations and the results do not depend on the reasons behind this fact. In Appendix C I 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 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]

The paper also contributes to an extensive literature on migration (see Greenwood, 1997 and Lucas, 1997 for surveys). I document that migration patterns when coupled with occupational switching differ substantially from those where workers remain in the same occupation. In my setup, migration across metropolitan areas is driven by the desire to find better occupational matches, consistent with the literature that emphasizes the importance of income prospects as a key driver behind the migration decision (Kennan and Walker, 2011).

Finally, this paper contributes to the literature on multi-armed bandit problems by combining the use of Gittins indices (Miller, 1984) with a binary formulation of match qualities (Bolton and Harris, 1999 and Moscarini, 2005). The resulting setup is analytically tractable and delivers closed-form expressions for workers' optimal occupational choice and moving decisions.

The rest of the paper is organized as follows: Section 2 documents the relationship between the number of occupational postings and city size; it also documents a number of facts on wages, moving patterns, and occupational switching patterns in large cities. In Section 3, I introduce my model, which is consistent with these facts, and in Section 4, I calibrate it. Section 5 extends the model to allow for more rapid occupation-specific human capital accumulation in larger cities. Section 6 concludes. The Appendix contains the extension of the model that endogenizes the number of occupations in each location, data description, details of the model simulation and calibration, as well as additional results.

2 Facts

The goal of this section is to investigate the empirical relationships between city size, number of available occupations, occupational switching, and wages. More specifically, I examine the number of available occupations by city size; the city size wage premium and its evolution with time in the city; occupational switching and how it varies with city size; and the patterns associated with moving and switching occupations and how they are different from those associated with moving and remaining in the same occupation.

[2] De la Roca and Puga (2017) note that "the innate ability of surgeons or lawyers in big cities and in smaller places is not that different to start with, it is working in bigger cities and the experience it provides that makes those working there better over time on average." Their results are consistent with the mechanism of the current paper, which argues that better occupational match quality accounts for the observed productivity differences.

[3] The mechanism described in the present paper, can be thought of as the worker counterpart of the mechanism described in Duranton and Puga (2001). In their work, they find that diversified cities offer firms more opportunities to experiment and discover their ideal production process. Similarly here, large cities offer workers more opportunities to discover a good occupational match.

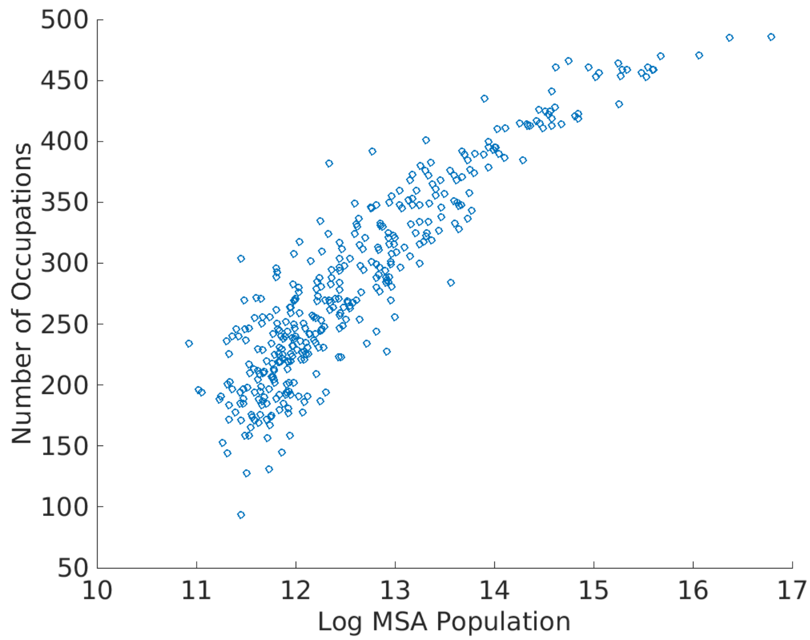

Figure 1: Number of Occupations vs. Log MSA Population - Burning Glass Vacancy Data

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 find that workers in larger cities have more occupations available to work in and this difference 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 Classification (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 Classification) 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 first to use the BG data.

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

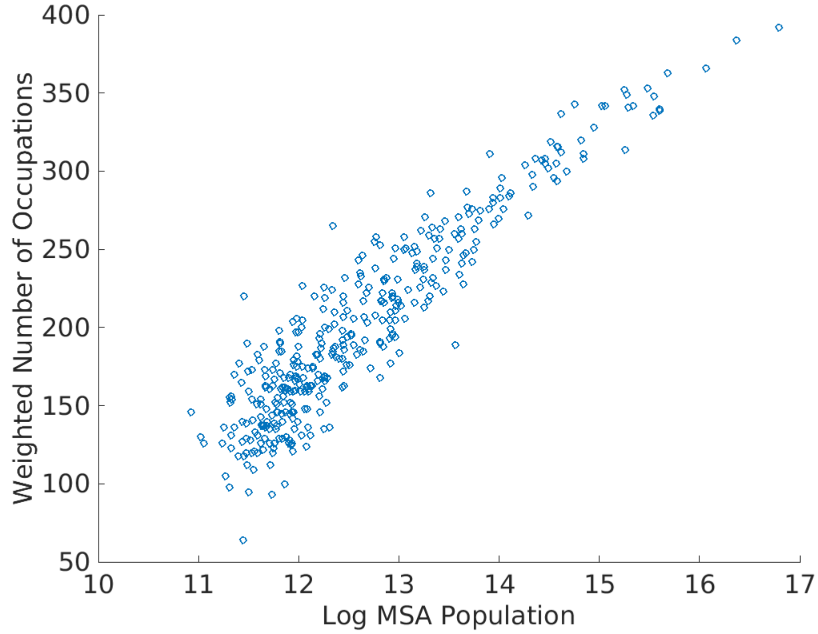

Figure 2: Weighted Number of Occupations vs. Log MSA Population - Burning Glass Vacancy Data

Examples of occupations in the data include actuaries, proofreaders, theatrical and performance makeup artists, manicurists and pedicurists, parking lot attendants, and skin care specialists.

Figure 1 assumes that an occupation is available if there is at least one posting in that occupation. However the same relationship emerges if I consider a stricter definition where either 5, 10, or 50 postings are needed for an occupation to be available.

One may worry that the difference is mostly driven by occupations that would interest few workers. To explore this, I generate weights for each occupation to capture how "popular" it is. More specifically, using data from the Current Population Survey (CPS) from 2003 through 2010, I generate weights for every occupation, using the number of workers who switch into every occupation in large cities.[6] Figure 2 presents the same relationship using the weighted occupational measure. While the overall level falls for both small and large cities, the same relationship remains. It also holds if I use a larger threshold of at least 5, 10, or 50 (weighted) postings.

[6]More specifically, I compute the occupational transition matrix and find the number of workers who switch into every occupation. I restrict this exercise to cities with population greater than 5 million where there are vacancies for almost all occupations. The weight for each occupation is then given by the ratio of the number of workers who switch into the occupation over the average occupational inflow. In other words, if an occupation has the same inflow as the average, the weight is equal to one, whereas occupations into which few workers switch, receive a weight less than one and vice versa. Now each posting is multiplied by the respective occupational weight, so that postings for popular occupations matter more.

The CPS uses the 2002 Census Occupational Classification, while BG reports the data using the 2010 Standard Occupational Classification (SOC) codes. In order to create the weights I use the cross-walk between the two classifications provided by the Census Bureau (https://www.census.gov/people/eeotabulation/data/2010_OccCodeswithCrosswalkfrom2002-2011nov04.xls).

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.

In what follows, the 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, occupation, industry and 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). 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 and how the moving variable is constructed and discusses how both the moving probabilities and the occupational switching rates are consistent with other data sets.[10]

[7]The ACS data contain 294 MSAs. In order to obtain consistent occupational classifications, I use the cross-walk provided by the IPUMS (https://usa.ipums.org/usa/volii/acs_occtooccsoc.shtml). See Ruggles et al. (2015).

[8]Unemployed workers in the UK who are claiming a jobseeker's allowance at their local JobCentre Plus, the Public Employment Service for Great Britain, are by law required to actively look for a job. Indeed the JobCentre Plus' primary goal is to assist workers in finding employment, and as such, it maintains a large database of vacancies. Using data for the period between June 2012 and September 2012 by county, I find the same pattern, with larger counties having a larger number of occupations with posted vacancies. In other words, a worker looking for employment at a JobCentre in a small county in September 2012 had significantly fewer options than his counterpart who was searching at a JobCentre in a larger county.

[9]I am grateful to Rafael Lopes de Melo for providing me with moments from the Brazilian RAIS data.

[10]Restricting the sample to respondents who are still present in the last wave and computing the moments below produces qualitatively and quantitatively very similar results, which suggests that sample attrition bias is not severe.

	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, firm 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)	-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

Table 2: Population Impact on Occupational Switching Probability, Conditional on Not Moving. 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, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors in parentheses are clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.1016 (overall), 0.1830 (recent in second column), and 0.1886 (recent in third column).

Fact 2. *Workers in larger cities earn higher wages and the wage premium increases with time in the city.*

I next 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 highly populated areas are paid significantly higher wages. The magnitude of the coefficient is in line with the results from other data sets.[11] As shown in the first column of Table 1, workers who just moved also receive higher wages if they moved to a highly 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 is consistent with the results in Glaeser and Maré (2001). This pattern suggests that the mechanism that generates these wage differences is relevant mostly *after* a worker arrives in a larger city.

Fact 3. *Among recent movers, workers in larger cities are more likely to switch occupations.*

I now investigate the relationship between occupational mobility and city size. The first column of Table 2 shows that occupational mobility is somewhat lower in larger cities. This finding is consistent

[11]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. (2014). In addition, the signs and magnitudes of the Mincerian controls are also consistent with the prior literature (see the full set of coefficients in Table 1 of the Online Appendix).

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

Table 3: Population Impact of Current City 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, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors in parentheses are clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.0054.

with the results in Bleakley and Lin (2012). However, when focusing on workers who moved to a location in the past 4 years, notice that they are more likely to switch occupations in larger cities, suggesting that time spent in a city is a key factor in occupational mobility that has not been previously considered.[12] The result remains and becomes even stronger when I control for the size of the previous city.[13]

Fact 4. *Workers in large cities have a lower probability of moving and switching occupations.*

When I consider geographical mobility, I find that, as shown in Table 3, the probability of moving and switching occupations is lower for residents of larger cities. It is worth noting that the effect is quantitatively large: each doubling of the population reduces the probability of moving out of the city and switching occupations by 13%.[14]

Fact 5. *Wages fall prior to moving and switching occupations.*

Moreover, as shown in the first column of Table 4, wages decline before moving. Indeed, if a worker is going to move in period t, then his wage falls by about 1% from period $t-2$ to $t-1$. This suggests that for at least some of the moves, labor market considerations are important in the decision to move. The second column of Table 4 indicates that wages are declining beforehand only in the case of workers who move and switch occupations; their wages fall by approximately 2.4%. Workers who move and keep the same occupation do not experience decreasing wages before moving. Thus, the result of the first column is driven by workers who move and switch occupations.[15]

[12] Investigating this further shows that almost all of the effect comes from workers who moved between 1 and 2 years ago. This is also true when using data from the National Longitudinal Survey of Youth (NLSY) 1997. For this group of workers, there is a substantial increase in occupational mobility with city size, which ranges from 11% (cities with log population less than 13) to 29% (cities with log population greater than 15).

[13] Most of the flows are offsetting across occupations, suggesting that most of these switches are driven by idiosyncratic reasons rather than aggregate shocks (see Appendix A for more details). In addition, Table 15 in Appendix A reports the destination occupations that switchers enter by city size.

[14] It is also worth noting that 88% of workers who move and switch occupations go to a metropolitan area as opposed to a non-metropolitan area. This probability is higher than the fraction of workers living in metro areas, which is 79%.

[15] In addition, 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.

	$\ln(\text{wage})_{t-1}$	$\ln(\text{wage})_{t-1}$
Move_t	-0.007	
	(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)
$\ln(\text{wage})_{t-2}$	0.847	0.847
	(0.001)	(0.001)
Number of Obs	146462	146462

Table 4: 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, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors in parentheses are clustered by individual.

Facts	**Evidence**
1. More occupations available in larger cities	Figures 1 and 2
2. Higher wages in larger cities, wage premium increases with time in the city	Table 1
3. Among recent movers, higher occupational switching in larger cities	Table 2
4. Probability of moving and switching occupations decreasing in city size	Table 3
5. Wages fall prior to moving and switching occupations	Table 4

Table 5: Summary of Facts

It is also true that occupational switching in the previous period significantly increases the probability of a move in the following period. Interestingly, past occupational switching has a very large and significant impact on the probability of moving and switching occupations, but no impact on the probability of moving and remaining in the same occupation.[16] This further underscores that the patterns associated with moving and switching occupations are very different from those associated with moving and remaining in the same occupation.

Table 5 summarizes the main facts discussed in this section.

3 Model

Guided by the above facts I develop a model of occupational choice and geographical mobility. The model is based on the intuitive idea that because there is greater occupational availability in large cities, workers there are more productive, since they are able to find a better occupational match. A formal model that delivers this idea and at the same time captures the high rate of occupational mobility (approximately a quarter of all workers switch occupations every year) requires a dynamic formulation of occupational choice. In the model presented in this section, the number of occupations in each location is exogenous.

[16]Results available upon request.

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: different cities have a different 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 occupational switching: individuals do not know their occupation-specific match, but learn it over time (Jovanovic, 1979, Miller, 1984, McCall, 1990, Moscarini, 2005).[17] If workers fail to find a suitable occupation they move to another city by paying a moving cost.[18]

I focus on occupations for two reasons: First, the recent literature has emphasized the importance of occupations rather than firms for worker labor market outcomes.[19] The common theme of this literature is that workers' wages depend 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 size.

I next describe the 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.[20]

Workers can move from one city to another. A worker leaves his current city either endogenously, or he may be forced to move exogenously according to a Poisson process with parameter $\delta > 0$ (as in Hu, 2005, Campbell and Cocco, 2007 and Li and Yao, 2007). When moving, a worker randomly goes to a new

[17] I follow the recent literature that has argued that occupational mobility is largely due to information frictions (e.g. Papageorgiou, 2014 and 2018, Groes et al., 2015, and Pastorino, 2019). However, the assumption that workers do not 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 switching.

The importance of information frictions in trade has been previously explored in Clarida (1993) and Allen (2014).

[18] Kennan and Walker (2011) estimate sizable moving costs across states that are increasing with age. In their setup, a worker who moves pays a deterministic cost that depends on age, distance, etc., but benefits from the difference in flow payoffs between the origin and the 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.

[19] Kambourov and Manovskii (2009a, 2009b), Antonovics and Golan (2012), Alvarez and Shimer (2011), Carrillo-Tudela and Visschers (2014), Papageorgiou (2014), Groes et al. (2015), Silos and Smith (2015), Gervais et al. (2016) and others.

[20] Alternatively one can assume away firms and assume that workers are engaged in home production in a particular occupation.

city.[21,22] Moving from one city to another entails a cost $c > 0$.

While in a city, a worker works in only one occupation at 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

$$dY_{tl}^{ik} = \alpha_l^{ik} dt + \sigma dW_{tl}^{ik}, \tag{1}$$

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

Let $\alpha_G > \alpha_B$. Productivities, α_l^{ik}, are independently distributed across occupations, cities, and workers.[23] 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 $g(\cdot)$. Prior beliefs are rational, i.e. they reflect the true probabilities that $\alpha_l^{ik} = \alpha_G$, and are common knowledge among workers and firms.

Workers and firms observe output and obtain (the same) information regarding the quality of the worker's match in the 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 Shiryaev, 1977)

$$dp_{tl}^{ik} = p_{tl}^{ik}\left(1 - p_{tl}^{ik}\right)\zeta \frac{dY_{tl}^{ik} - \left(p_{tl}^{ik}\alpha_G + \left(1 - p_{tl}^{ik}\right)\alpha_B\right)dt}{\sigma}, \tag{2}$$

where $\zeta = \frac{\alpha_G - \alpha_B}{\sigma}$ is the signal-to-noise ratio. The fraction on the right-hand side is a standard Wiener process with respect to the information set available to the worker. p_{tl}^{ik} is a sufficient statistic of the worker's beliefs regarding α_l^{ik}. Intuitively the change in beliefs as new information arrives, dp_{tl}^{ik}, depends on i) the variance of beliefs, $p_{tl}^{ik}\left(1 - p_{tl}^{ik}\right)$, ii) the signal-to-noise ratio, $\frac{\alpha_G - \alpha_B}{\sigma}$, and iii) the normalized difference between actual and expected output, $\frac{1}{\sigma}\left(dY_{tl}^{ik} - \left(p_{tl}^{ik}\alpha_G + \left(1 - p_{tl}^{ik}\right)\alpha_B\right)dt\right)$. To minimize notation, from now on I 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 and draws his prior p_0^k for each of the

[21] In the model presented in Appendix C, I relax this assumption and workers are allowed to choose which city to go to. However, in equilibrium they are indifferent across all cities, so they also end up choosing randomly as in the present model. This is also true for most urban models going back to Rosen (1979) and Roback (1982).

[22] All results go through if I allow the inflow of workers into a city to depend positively on the number of its occupations.

[23] In the model's calibration the number of occupations in each city will effectively reflect the average number of "new" occupations a worker encounters in each location. See the discussion in Section 4.

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

Firm competition for the services of workers ensures that a worker's compensation equals his expected output in the occupation, n, in which he is employed[24]

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

At any point in time, the worker needs to decide in which occupation he will work. 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 that pays him the highest wage). In addition, the worker needs to decide when to pay the moving cost and move to another city (optimal stopping).

The worker's state space consists of m variables, i.e., his belief for each occupation in his current city; beliefs in past cities are not part of the state space, since there is zero probability that the worker will return to a past city. Solving this problem numerically with more than a handful of occupations is computationally intractable. For instance, in the calibration presented in Section 4, m equals 12 in some cities. Using a 100-point grid for each belief implies that the dimension of the state space is in the order of 100^{12}.

The difficulty of solving multi-armed bandit problems numerically is well-known, but fortunately, these problems become tractable using Gittins indices (see Gittins, 1979 and Bergemann and Valimaki, 2008). Rather than solving the original problem whose state space can be intractably large, the Gittins index approach transforms the problem into m individual problems. The relevant state variables for each one of these new problems comprises only the state variables of that particular arm (occupation). Therefore, 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 all m arms (occupations), calculating the index of each

[24] Alternatively, one can assume that workers sell their realized 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 the next instant's output. All the implications derived later continue to hold.

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

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

[25]See also Silos and Smith (2015) for a recent example of an application of the multi-armed bandit framework to occupational choice.

[26]As shown in Section 5, the model can, however, accommodate one important source of occupational switching costs, namely human capital that is lost when switching occupations. In addition, numerical simulations using parameter estimates from the model calibration discussed in Section 4 and assuming two available occupations show that when there exists a switching cost of up to 4 months of wages, the "error" in the predicted economy average wage is small (less than 4%). Results are available upon request.

where c is the moving cost, $E_{\mathbf{p}}V\left(\mathbf{p}_m\right)$ 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

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

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 $\widetilde{p}\left(W^k\right)$ such that the value matching and the smooth pasting conditions hold:

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

$$V_p^k\left(\widetilde{p}\left(W^k\right), W^k\right) = 0.$$

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

The solution to the above differential equation is given by

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

where

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

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

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

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

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

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

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

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, for each of the m occupations that are available to 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 the value of the posteriors of all his occupations reaches $\underline{p}$.[28] However, some workers may find that one of the occupations they try out is a good match for them, in which case their posterior drifts toward one and their wage increases. These workers leave their match and city only exogenously at rate δ.

3.3 Implications

In this section I explore the model's main predictions and link them back to the facts documented in Section 2 (see Table 5).

- I first examine the setup's implications regarding geographical mobility, as well the relationship between city size and the number of occupations. 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 another city with m occupations. Assume that $d \leq m$ of his draws are above $\underline{p}$. Then the probability that 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 output signals are independent across occupations. 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

[28] 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.

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. If workers in cities enjoy a better selection of occupational choices, then we 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) (Chapter 15.3, Problem A), if we ignore exogenous moving shocks, the probability that a worker whose posterior in his current occupation is equal to $p_{(m)}$ switches occupations at some future date (i.e., the probability that $p_{(m)}$ reaches the value of his second highest posterior before it reaches 1) is given by

$$\Pr\left(p_{(m)} \text{ reaches } p_{(m-1)} \text{ before } 1\right) = \frac{1-p_{(m)}}{1-p_{(m-1)}}, \tag{9}$$

  where $p_{(m-1)}$ is the 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. However, that is not necessarily the case: workers in larger cities have higher posteriors in their current

[29] In fact, the flow into larger cities is slightly larger, since the probability that all prior draws are less than $\underline{p}$ is decreasing in m. This reinforces the result.

In addition, allowing the inflow of workers into a city to depend positively on the number of its occupations strengthens the result.

[30] In addition, workers who move experience wage increases: 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, $p_0^k > \underline{p}$, and therefore enjoys a wage increase. This is consistent with the evidence discussed in footnote 15.

occupations $p_{(m)}$. Their second highest posterior, $p_{(m-1)}$, is also increasing in m, the total number of occupations. Therefore, 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, and have better outside options, which increases the probability that they will switch.[31] Recent movers, however, have not yet formed good matches, so the first effect is muted. As a result, the second effect (more options) dominates, and as shown in the calibration discussed in Section 4, they are more likely to switch occupations (Fact 3).

- Finally, in my 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. In addition, since forming a good match takes time, one expects the wage premium to grow with time in the city, consistent with Fact 2. This is indeed confirmed by the model's calibration results (Section 4).

3.4 Why Do Large Cities Have More Occupations?

The paper's baseline model, presented above, explored the extent to which greater occupational availability accounts for agglomeration economies. This setup takes as given that some cities have more occupations.

In Appendix C, 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.

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. I now allow workers who move to choose their destination city; in equilibrium, they select randomly, as in the baseline model. 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. I show that profits are increasing in the size of the city (goods market), so, in equilibrium, cities with higher populations support more occupations. More occupations, in turn, attract a larger population as workers benefit not only from increased occupational availability, as in the baseline model, but also from increased consumption variety. Increased population, however, also causes a negative externality, which prevents cities from becoming unboundedly large.[32] In this setup, both the

[31] As discussed, above however, both of these effects move in the same direction when it comes to geographical mobility and imply that workers in large cities are unambiguously less likely to move.

[32] It is necessary to introduce some form of negative externality with increased population to prevent all workers who move from choosing to go to large cities. Examples of this "negative externality" are higher housing prices or increased commuting time.

number of occupations and the population are endogenously determined; in equilibrium, cities with larger populations have more occupations, consistent with the evidence in Section 2.

Of course, there could be other mechanisms accounting for why larger cities have more occupations. For instance, Gautier and Teulings (2009) show that when there are increasing returns in the matching technology, large cities have a comparative advantage in producing goods that use scarce worker types, i.e. complex goods, and, as a result, are more heterogeneous in the jobs they offer. Similarly, mechanisms that emphasize the gains from worker specialization also predict more occupations in larger cities.

3.5 Alternative Mechanisms of Agglomeration Economies

Before moving on to the model's calibration, I discuss whether the documented empirical patterns can be explained by other mechanisms. In their survey, Duranton and Puga (2004) classify the different proposed microfoundations of agglomeration economies into three broad categories, those based on: sharing, matching, and learning mechanisms, which I discuss in turn.

In terms of *sharing*, the closest mechanism pertains to workers in large cities specializing in more narrow occupations (see also Baumgardner, 1988, Becker and Murphy, 1992, Duranton and Jayet, 2011, Hsu, 2012, Kok, 2014, and Tian, 2019).[33] Indeed, one may wonder whether narrower specialization is driving the differences in the number of occupations across cities documented in Section 2. To investigate this, I use the BG data and find that if I consider all cities that have fewer than 160,000 residents, there are 485 3-digit occupations available in this group of cities out of a total of 503 possible occupations. By comparison, in the New York metropolitan area, which is the largest MSA, there are 486 occupations available. This holds despite the fewer vacancy postings in the small city group (303,142) compared to the NY MSA (431,937), suggesting that increased occupational availability is not driven by specific occupations that are available in large cities but not in small ones.

In addition, if one groups occupations according to their 2-digit codes instead of their 3-digit codes, most of the difference between large and small cities disappears. In particular, most cities above approximately half a million inhabitants have postings for almost all 2-digit occupations. Again, this is consistent with the previous result: that it is not a specific "type" of occupation that is absent from small cities. Rather within each broad occupational category, a smaller number of occupations are available.[34]

[33] There are also other mechanisms that fall under this category, but are unrelated to the above-mentioned facts, such as the ability to share indivisible goods, like a hockey ring.

[34] It is worth noting that the occupations that are classified under the same 2-digit occupation can still be quite different. For instance, chiropractors, dentists, dietitians and nutritionists, optometrists, pharmacists, physicians and surgeons all fall under the same 2-digit occupation (30).

	Initial	All Years
Full Sample	ln(wage)	ln(wage)
ln(current city pop)	0.016	0.041
	(0.009)	(0.001)
Number of Obs	1261	169536
Singles Only		
ln(current city pop)	0.025	0.039
	(0.011)	(0.002)
Number of Obs	821	80140
Married Workers Only		
ln(current city pop)	-0.003	0.044
	(0.017)	(0.002)
Number of Obs	440	89396

Table 6: Robustness: 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 (full sample only), firm size, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors in parentheses are clustered by individual.

As discussed in the introduction, the mechanism proposed in this paper falls under the *matching* category, as I argue that workers in larger cities are able to form better occupational matches. However, large cities also offer better marriage markets for singles (Gautier et al., 2010), so I repeat my analysis using only married workers; as shown in Tables 6 through 8, the results continue to hold. In addition, given the increasing prevalence of power couples in large cities (Costa and Kahn, 2000), I also show that the results hold for the subsample of singles only as well.[35] Indeed, all the signs and magnitudes are consistent with the baseline, though naturally the standard errors are a bit larger due to the smaller sample.[36]

In terms of *learning* mechanisms, the most relevant ones emphasize the importance of learning and knowledge spillovers in larger cities that lead to more rapid human capital accumulation (see Jacobs, 1969 and Lucas, 1988, as well as the subsequent literature that focuses on agents learning from other agents[37]). Indeed, such a story can explain why the city-size wage premium grows with time in the city. To investigate this, I first check whether large cities also offer more low-skill occupations, where such spillovers are presumably less important; I find that they do.[38] Moreover, in Section 5 that follows the

[35] As an aside, the rise in the importance of power couples is not inconsistent with the basic mechanism of the paper.

[36] The largest difference is the impact of city size on the probability of moving and switching occupations, shown in Table 8. However, it is important to note that singles are much more likely to move and switch occupations than married workers (0.86% vs 0.29%), so the impact of city size is economically large for both groups of workers.

[37] Jovanovic and Rob (1989), Jovanovic and Nyarko (1995), Eaton and Eckstein (1997), Glaeser (1999) and Lucas and Moll (2014).

[38] Indeed the positive relationship between city size and the number of available occupations documented in Figure 1 holds if I consider "low-skill" and "high-skill" occupations separately ("high skill" being 001-354, Management, Professional and Related Occupations and "low-skill" all others). The difference across cities is quantitatively large: a linear regression implies that cities with double the size have approximately 48 more low-skill occupations and 22 more high-skill occupations.

	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)
ln(previous city pop)			-0.0005
			(0.0093)
Number of Obs	63361	1968	1140
Married Workers Only			
ln(current city pop)	-0.0013	0.0201	0.03
	(0.0007)	(0.0111)	(0.0163)
ln(previous city pop)			-0.0228
			(0.0123)
Number of Obs	77481	1392	907

Table 7: Robustness: Population Impact on Occupational Switching Probability, Conditional on Not Moving. 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 (full sample only), firm size, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors in parentheses are clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which for the full sample equals 0.1016 (all years), 0.1830 (recent in second column), and 0.1886 (recent in third column) and, respectively, 0.1348, 0.1895, and 0.1904 for singles and 0.0744, 0.1739, and 0.1863 for married workers.

model's calibration, I extend the model to allow for more rapid human capital accumulation in large cities and examine how the model's main predictions, as well as the calibration results change. I conclude that while more rapid human capital accumulation in larger cities correctly predicts higher growth of wages, it cannot account for the remaining patterns documented above.[39] This is precisely why in the model's calibration presented in the next section, I target worker reallocation moments, so as to pick up the importance of greater occupational availability, rather than other potential mechanisms, such as more rapid human capital accumulation or knowledge spillovers.

[39] The extension allows for occupation-specific human capital. If I consider general human capital instead, one expects that general human capital accumulation should not have an impact on occupational mobility: precisely because general human capital is accumulated in every occupation, it does not affect occupational choice and therefore occupational switching behavior.

Full Sample	Prob of Move & Switch Occup		$\ln(\text{wage})_{t-1}$
ln(current city pop)	-0.0007	$\text{Move}_t \times \text{Occupation Switch}_t$	-0.024
	(0.0002)		(0.008)
Number of Obs	144635	$\text{Move}_t \times \text{No Occupation Switch}_t$	-0.002
			(0.004)
		$\ln(\text{wage})_{t-2}$	0.847
			(0.001)
		Number of Obs	146462
Singles Only			
ln(current city pop)	-0.0013	$\text{Move}_t \times \text{Occupation Switch}_t$	-0.018
	(0.0003)		(0.009)
Number of Obs	65280	$\text{Move}_t \times \text{No Occupation Switch}_t$	0.002
			(0.005)
		$\ln(\text{wage})_{t-2}$	0.822
			(0.005)
		Number of Obs	65762
Married Only			
ln(current city pop)	-0.0002	$\text{Move}_t \times \text{Occupation Switch}_t$	-0.041
	(0.0001)		(0.017)
Number of Obs	79067	$\text{Move}_t \times \text{No Occupation Switch}_t$	-0.007
			(0.005)
		$\ln(\text{wage})_{t-2}$	0.863
			(0.007)
		Number of Obs	80700

Table 8: Robustness: Population Impact of Current City on Probability of Moving and Switching Occupations (Probit) and Wage Path Before Moving. Source: 1996 Panel of Survey of Income and Program Participation. Population data from 2000 Census. 4-month probabilities/intervals. Controls include gender, race, education, marital status (full sample only), firm size, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors in parentheses are clustered by individual. Probit coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.0054 for the full sample, 0.0086 for singles, and 0.0029 for married workers.

4 Calibration

I next investigate the quantitative importance of the proposed mechanism. In particular, to what extent can greater occupational availability account for the observed wage premium and greater wage inequality in larger cities? To address this question, I calibrate the model.

The model has implications regarding differences across cities for both wages (mean wage and wage inequality) and worker reallocation. Given that other models of agglomeration economies do not have predictions regarding differences in worker reallocation across city size, I use these moments to calibrate my framework and then examine its predictions regarding the wage premium and differences in wage inequality across cities.

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.

The second moment pins down the level of available occupations in every city. Although even small cities have a substantial number of occupations, it is reasonable to assume that a much smaller subset of these occupations is relevant for each worker (especially those with a college degree). In order to calibrate the number of relevant occupations for each worker, I use 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 the expected tenure in each occupation). In the data the 4-month switching probability to new occupations is 4.82%, which is my second moment.

In my baseline calibration, I set the number of occupations to 12 for large cities and 5 to smaller cities. This implies a ratio of 2.4, which is close to 2.28, and the predicted 4-month switching probability to new occupations is 4.29%. By targeting the total number of occupations a worker tries out throughout his life, I ensure that the total number of occupations that an average worker can sample is the same in the model and in the data.

Moreover, given that the number of occupations in large and small cities is central to the model's mechanism, in Appendix E I present results using a number of different combinations of these two parameters.

Step 2: In the second step, I jointly retrieve values for the following 5 parameters: s, δ, c, p_0 and ζ.[44] In order to do that, I use moments related to occupational switching and moving probabilities. More specifically, I use the 5 following moments: the population share that lives in a large city, the coefficient on large city in the occupational switching probability regression, the coefficient on large city in the occupational switching probability regression when conditioning on recent movers only, the 4-month probability of moving for workers living in a small city and the same probability for those living in a large city. I simulate the model presented in Section 3 and match the simulated moments with the ones from the data.

Step 3: In the third step, I calibrate the remaining 2 parameters, α_G and α_B, to match exactly the mean level of wages and the residual standard deviation of wages. None of the reallocation moments previously used depend on the choice of α_G and α_B, so I am able to calibrate these two parameters separately.[45] The full set of parameters is presented in Table 9. The implied cost of moving, c, equals

[44] By allowing s, the probability that a worker who moves goes to a large city, to be a parameter to be estimated, I allow for the possibility that large cities are oversampled among movers relative to smaller ones. This is important given the findings reported in footnote 14.

[45] As described in Appendix D, rather than searching over the moving cost, c, the calibration treats $\underline{p}$ as a parameter and afterward calculates the associated cost, c, for which the retrieved value of $\underline{p}$ is optimal.

α_G	28.93
α_B	8.29
s	54.96%
δ	0.00489
$\frac{\alpha_G-\alpha_B}{\sigma}$	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%	0.58%
Higher Occup Sw Prob in Large	0.20%	0.93%
Higher Occup Sw Prob in Large (recent)	3.34%	2.98%
Mean Wage	$14.20	$14.20
Residual Wage Standard Deviation	$5.97	$5.97

Table 10: Targeted Moments. The higher occupational switching probabilities in large cities are computed by regressing occupational switching on a large city dummy and the same set of controls as in Table 2 for the subsample used. Moving probability refers to the 4-month probability of moving of workers living in large and small cities.

$60,667.

Although a rigorous identification argument is impossible due to the complexity of the framework, I provide an informal argument of how each parameter is identified from the data. The probability that a worker who moves goes to a large city, s, is pinned down by the population share that lives in a large city. The 4-month moving probabilities for workers in large and small cities pin down the exogenous moving rate, δ and the moving cost, c. More specifically, δ, affects the moving probability in the same way for small and large cities. c, however, affects largely the moving probability in small cities, as workers there are more likely to move endogenously, essentially pinning down the excess mobility in small cities relative to large. Finally, the signal-to-noise ratio, ζ, and the level of the initial belief, p_0, are pinned down by the occupational switching probability regressions for recent movers and the entire sample, respectively.[46]

The targeted moments are presented in Table 10. The calibration matches the targeted moments well. In the calibrated model, recent workers in large 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.

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, I check this moment to evaluate the choice

[46] The speed at which workers update their beliefs depends on $p(1-p)\zeta$ (see equation (2)). Changing ζ affects the speed of learning (and the probability of an occupational switch) at all levels of beliefs. p_0, however, affects the speed of learning particularly for recent movers, whose beliefs are initially concentrated near that region.

Moments:	**Data**	**Model**
Prob of Switch to New Occupation	4.82%	4.29%
Initial Wage	$10.92	$10.23

Table 11: Other Moments. Switching probability to a new occupation refers to the probability that a worker switches to an occupation in which he has not previously been employed.

	Data	**Model**
Wage Premium	20.16%	6.86%
Wage Standard Deviation Premium	21.21%	7.21%

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

of setting the number of occupations to 12 in large cities and 5 in small cities. Moreover, the initial wage, which was not targeted in the calibration, is predicted to equal $10.23 compared to the initial wage of $10.92 observed in the data.

Table 12 presents the predicted wage premium, as well as the cross-sectional standard deviation premium, neither of which was targeted. The calibrated model replicates approximately 35% of the observed wage premium. Moreover, it replicates about a third of the greater wage inequality that has been documented in larger cities.[47]

In the results presented in Appendix E, I examine the sensitivity of the results with respect to the number of occupations in large and small cities. In particular, Tables 16 through 18 of Appendix E present estimated parameters, as well as the targeted and untargeted moments for various combinations of the number of occupations in large and small cities. The baseline specification generally has a better fit, especially in terms of matching the initial wage, but also in terms of matching the switching probability to new occupations. In all specifications, the model matches a sizable fraction of the observed wage premium, with estimates varying from 23% to 52%. As one might expect, a higher ratio of occupations in large cities relative to small leads to a higher wage premium, though the relationship is not always one-to-one. The model also captures part of the greater wage inequality in large cities, though the fraction explained in the other specifications is lower than that of the baseline estimates.

Sensitivity Analysis

Finally, I consider the importance of the various parameters in obtaining these results. More specifically, I vary different model parameters, one at a time, and then I present the key moments that are affected

[47] See also Baum-Snow and Pavan (2013) for an investigation of how the increase in inequality in larger cities contributed to the overall increase in inequality in the US over three decades. They conclude that agglomeration economies played a key role in the change in wages over that period.

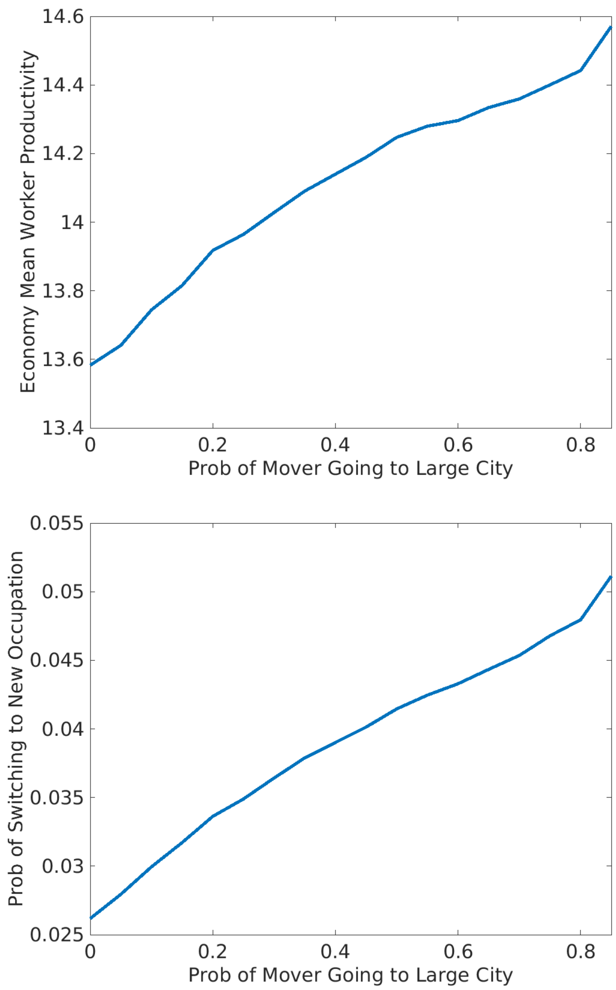

Figure 3: Sensitivity Analysis - Probability of Mover Going to Large City

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.

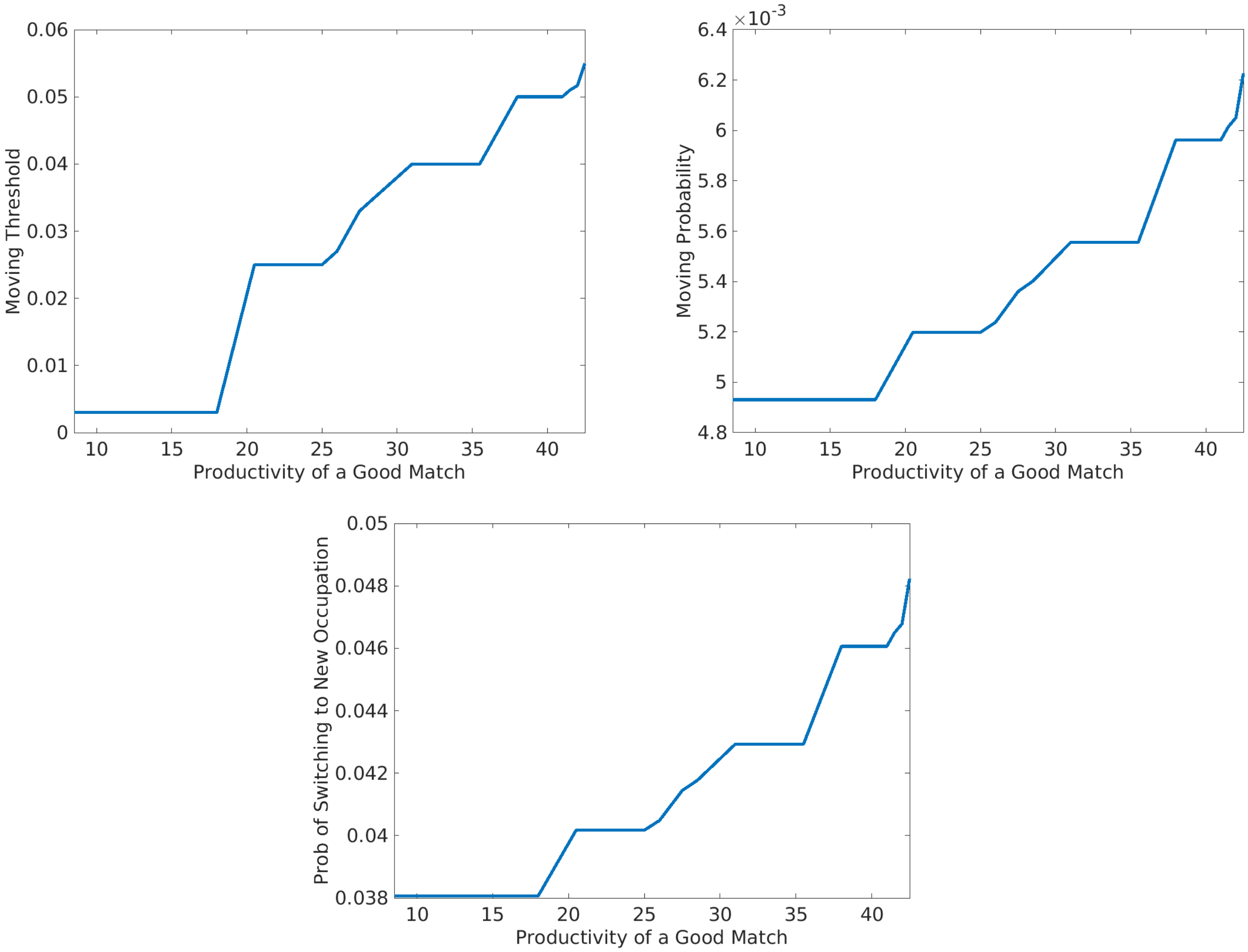

Figure 4: Sensitivity Analysis - Productivity of a Good Match, α_G

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 α_G, I hold constant the signal-to-noise ratio, $\frac{\alpha_G - \alpha_B}{\sigma}$, by appropriately adjusting σ, so as not to conflate different effects. I consider the impact of changing $\frac{\alpha_G - \alpha_B}{\sigma}$ below.

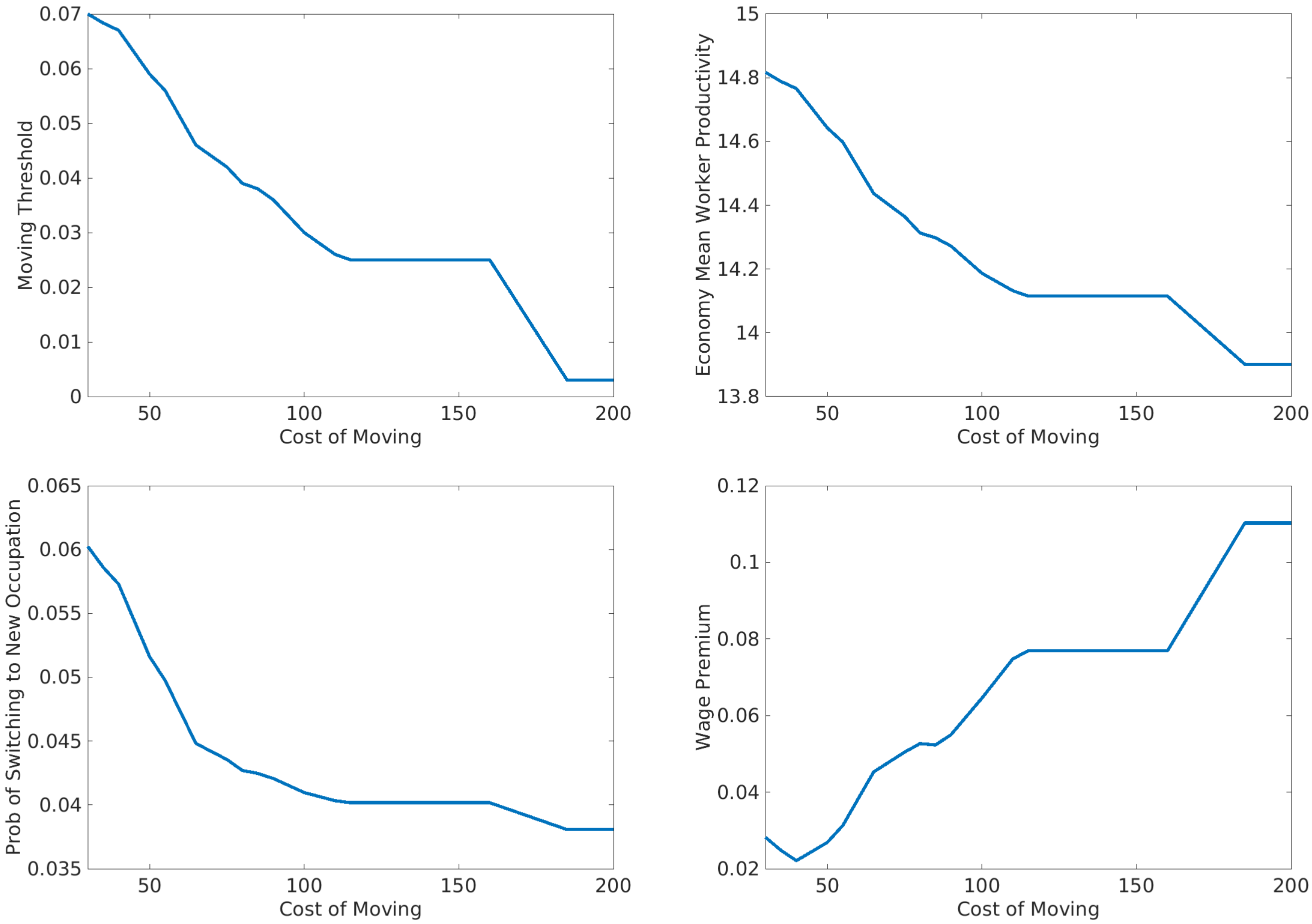

Figure 5: Sensitivity Analysis - Cost of Moving

distribution with mean p_0 and consider various levels of its standard deviation. As shown in Figure 6, when initial belief dispersion increases, the difference 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 difference 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 significant).

Finally, I examine how changing the signal-to-noise ratio, $\frac{\alpha_G - \alpha_B}{\sigma}$, affects workers in this economy. When the signal-to-noise ratio is close to zero, there is almost no switching, as workers learn extremely slowly. As shown in Figure 7, increasing the signal-to-noise ratio leads to an increase in experimentation, and progressively, workers try out more occupations. More interestingly, the wage premium exhibits an inverse U-shape pattern: for very low values of the signal-to-noise ratio, since there is very little

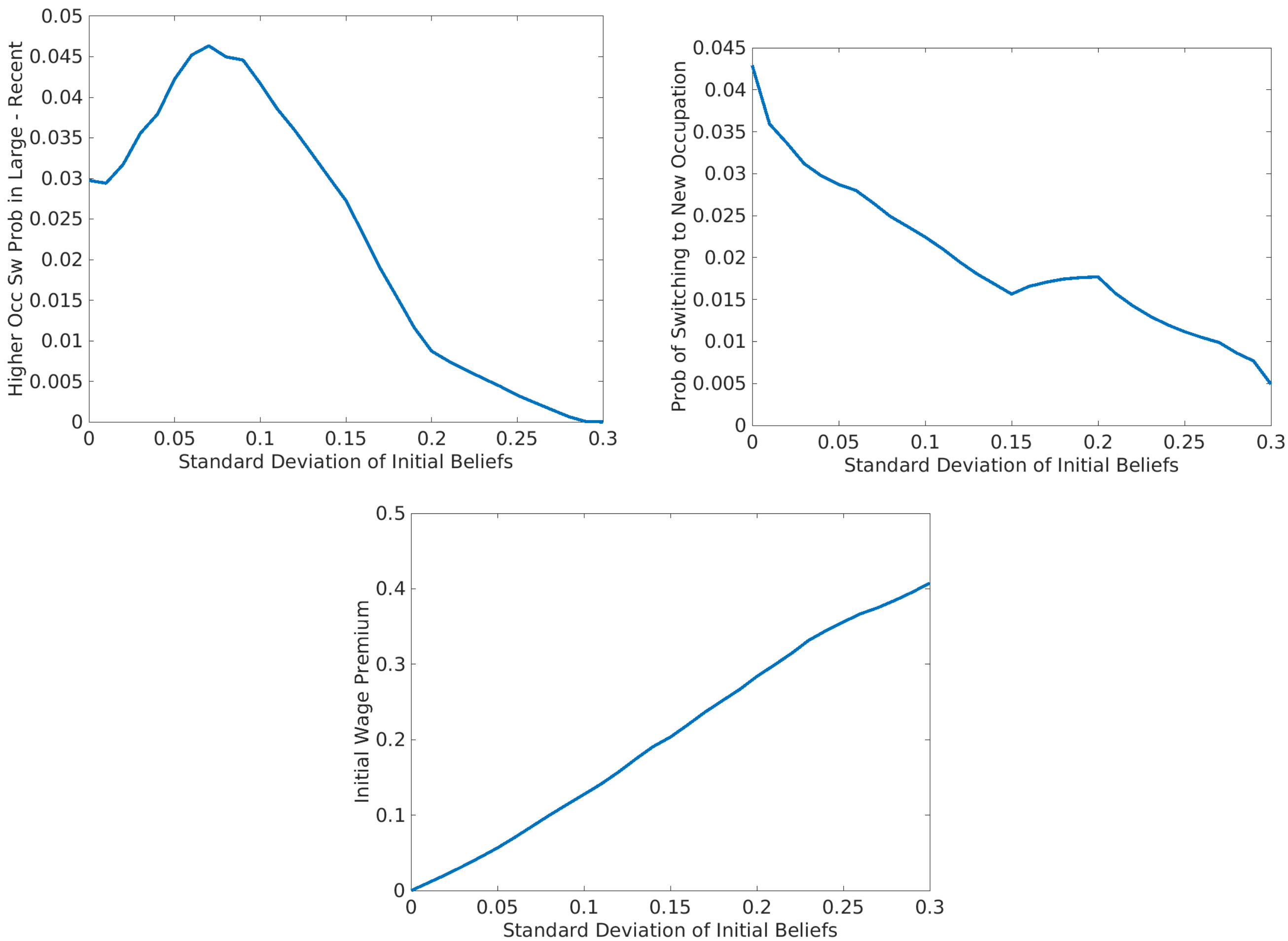

Figure 6: Sensitivity Analysis - Dispersion in Initial Beliefs

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 different 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 benefit of being in a large city with many occupations declines and so does the wage premium.

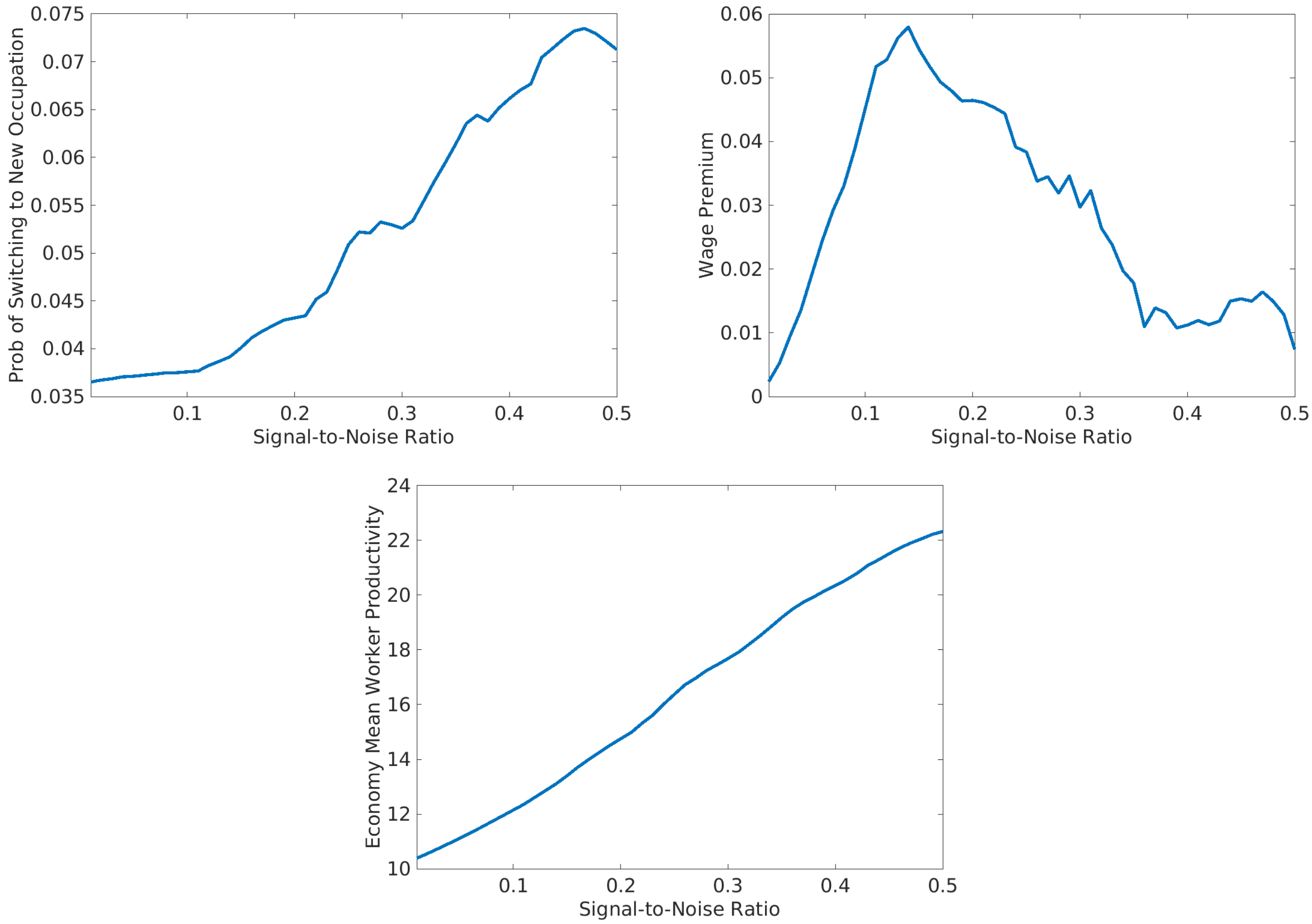

Figure 7: Sensitivity Analysis - Signal-to-Noise Ratio

5 Allowing for Occupation-Specific Human Capital

In this section I extend the model to allow for more rapid occupation-specific human capital accumulation in larger cities in order to understand whether the empirical patterns can be explained by that mechanism instead. In particular, I first consider how the framework's implications change when I allow for occupation-specific human capital; I also examine how they are affected when I change the rate of human capital accumulation so that larger cities offer faster human capital accumulation. I then recalibrate the model while allowing for faster human capital accumulation in larger cities and show that greater occupational choice in larger cities remains quantitatively important.

I introduce occupation-specific human capital by assuming there are two experience levels in each occupation: workers are either experienced or inexperienced. A worker who enters an occupation for the first time is inexperienced. As in Kambourov and Manovskii (2009a), a worker becomes experienced stochastically at rate θ. Output production is higher by a known amount u for all experienced workers.

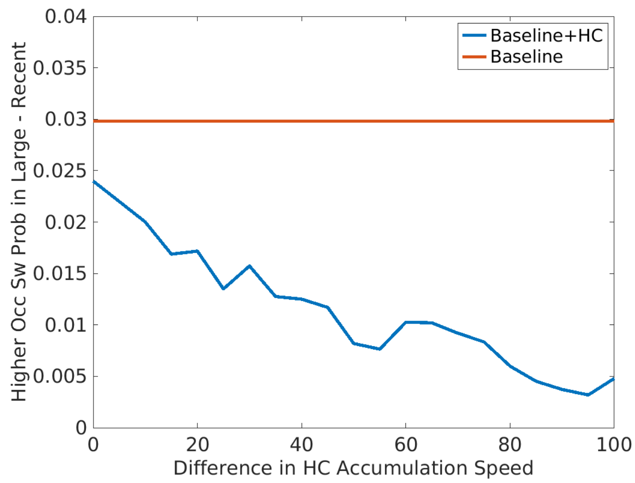

Figure 8: Higher occupational switching probability among recent movers to larger cities, compared to those who moved to smaller ones. Horizontal axis is percent difference in speed of occupation-specific human capital accumulation in large cities versus small.

The model is otherwise identical to that presented in Section 3: some cities offer more occupations than others; workers choose their occupation and are unsure of their occupational match quality; and workers also have the option of moving to a new city by paying a moving cost.

The solution of the expanded model, i.e., the occupational choice and moving decision of the worker, is now substantially more complicated. In particular, the Gittins index of an occupation no longer corresponds to that occupation's belief, but is instead given by expression (6) and, implicitly, by expression (22), both of which can be found in the Online Appendix, which contains the detailed derivation of the solution.

To understand the implications of adding occupation-specific human capital, I use the calibration parameters of my baseline specification and see how the predictions change. In order to specify values for the human capital accumulation parameters, I borrow the estimates of Kambourov and Manovskii (2009b), who find that after 5 years of occupational tenure wages increase between 12% and 20%. Given that initial wages equal \$10.92 and assuming that becoming "experienced" leads to a 20% increase in productivity, which is the upper limit of the Kambourov and Manovskii (2009b) estimates, I set u to 2.184. I allow the rate at which workers acquire human capital, θ, to vary across cities of different sizes while ensuring that the average speed of human capital accumulation in the economy is such that workers become experienced after 5 years on average, as in the Kambourov and Manovskii (2009b) estimates. This allows for learning and knowledge spillovers in larger cities that lead to more rapid human capital accumulation.

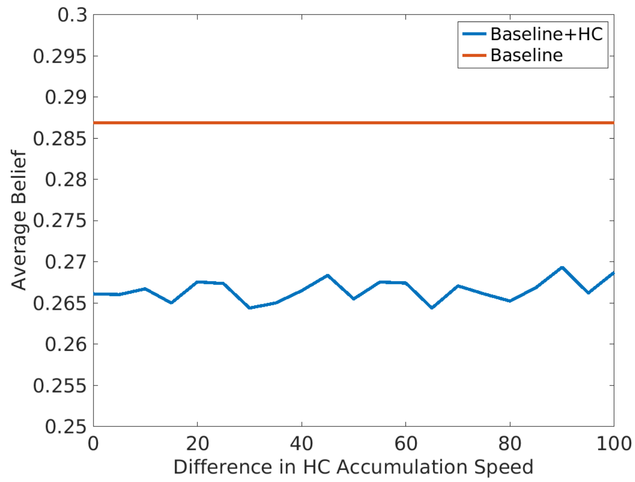

Figure 9: Average belief. Horizontal axis is percent difference in speed of occupation-specific human capital accumulation in large cities versus small.

I first 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 difference in the speed of human capital accumulation across cities of different 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 difference in occupational switching among recent workers, as the importance of "finding a good match" is now diminished. Moreover, as human capital accumulation becomes faster in larger cities, the difference in occupational switching approaches zero: on the one hand, as in the model with no human capital, large cities offer more options, which leads to more occupational switching among recent movers searching for a good match; on the other hand, workers in large cities accumulate human capital faster, which reduces their incentives to switch occupations. Indeed, when the difference in the speed of human capital accumulation is approximately double, the difference in occupational switching among recent workers is very close to zero. Therefore, the implications of these two mechanisms with respect to occupational switching point in opposite directions.

Similarly, when human capital is introduced, the probability of moving out of a small city falls from 0.58% to 0.53%. In particular, the introduction of human capital implies that finding a good match becomes less important, and therefore, there is less of an advantage to working in a large city. As a result, workers in small cities are less willing to move out and also the difference in the moving probabilities across cities is now almost half of what it used to be.

To better understand better the economic mechanism, I next examine how the introduction of human

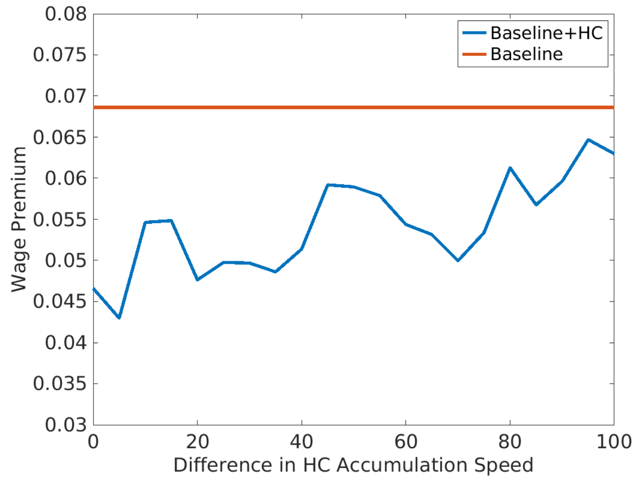

Figure 10: Wage premium between large and small cities. Horizontal axis is percent difference in speed of occupation-specific human capital accumulation in large cities versus small.

capital impacts average match quality. Figure 9 plots the average belief as a function of the difference in the speed of human capital accumulation across cities of different sizes. When occupation-specific human capital is introduced, match quality drops: workers are now more willing to tolerate matches that are potentially "bad", as long as they have accumulated human capital. Average match quality does not change when human capital accumulation becomes faster in larger cities.

I now turn to its implications regarding the wage premium, shown in Figure 10. The introduction of occupation-specific human capital leads to a reduction in the wage differences between large and small cities: since finding a good match becomes less important, there is less of an advantage to working in a large city. Indeed, as shown earlier, the probability of moving away from a small city drops when human capital is introduced. However, as human capital accumulation becomes faster in larger cities, the wage premium increases, as expected, since workers in larger cities are more likely to have accumulated human capital.

The previous results suggest that the implications of human capital accumulation for several moments are often the opposite of what we observe in the data.

I next recalibrate the model to examine the importance of higher occupational availability when occupation-specific human capital is allowed. I assume that the speed of human capital accumulation is 50% higher in large cities and calibrate the importance of human capital using the estimates of Kambourov and Manovskii (2009b) as described above. The results are shown in Table 13. See Appendix D for more details on the calibration of the expanded model.

Given the previous results, it is perhaps not surprising that the model with occupation-specific human

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:			
Prob of Switch to New Occupation	4.82%	4.29%	3.46%
Initial Wage	$10.92	$10.23	$9.05
Wage Premium	20.16%	6.86%	7.35%
Wage Standard Deviation Premium	21.21%	7.21%	6.11%
Parameters:			
# of Occupations in Large and Small		12 and 5	12 and 5
α_G		28.93	28.94
α_B		8.29	6.88
s		54.96%	55.28%
δ		0.00489	0.00483
$\frac{\alpha_G-\alpha_B}{\sigma}$		0.1795	0.153
p_0		0.0937	0.0982
c		91	121.5
u			2.184
θ in Large			0.0772
θ in Small			0.0515

Table 13: Calibration with occupation-specific human capital. The higher occupational switching probabilities in large cities are computed by regressing occupational switching on a large city dummy and the same set of controls as in Table 2 for the subsample used. Moving probability refers to the 4-month probability of moving of workers living in large and small cities. Switching probability to a new occupation refers to the probability a worker switches to an occupation in which he has not previously been employed.

capital captures less well the higher occupational switching rates in large cities among recent movers. Similarly, it has difficulty generating differences in the moving probabilities across cities of different sizes. It can now account for a higher fraction of the wage premium, but less of the greater inequality in larger cities. When I shut down differences across cities in the speed of human capital accumulation, the resulting wage premium is 5.4%, while the standard deviation premium is 4.1%, suggesting that greater occupational availability accounts for most of the wage differences across cities in the calibrated model.[49]

[49]If I shut down occupation-specific human capital completely, the wage premium becomes 6.5% and the standard deviation premium equals 4.9%.

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 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."[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/find-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 difference between flows in and out of every occupation over their sum. The numbers reported are the weighted ratio for the 13 major occupational groups in the 1996 SIPP. This implies that the large majority of occupational flows is offsetting, in both large and small cities, underscoring the importance of idiosyncratic rather than aggregate shocks.

B Proof of Lemma 2

Straightforward derivations lead to:

$$\frac{dW\left(p^k\right)}{dp^k} = \frac{1}{r+\delta}\frac{(d+1)(d-1)(\alpha_G-\alpha_B)}{\left(2p^k+d-1\right)^2} > 0$$

since $r+\delta>0$, $\alpha_G-\alpha_B>0$ and $d=\sqrt{\frac{8(r+\delta)}{\left(\frac{\alpha_G-\alpha_B}{\sigma}\right)^2}+1}>1$.

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 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 $\frac{d^2 z(N_t)}{dN_t^2} > 0$.[55] Flow utility per unit of time is given by

$$C_t - z(N_t)$$

where $z(\cdot)$ 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} be independently distributed across occupations, cities, and workers. Moreover, α_l^{ik} is

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

[55] See, for instance, Lucas and Rossi-Hansberg (2002) and Eeckhout (2004), who microfound the negative externality by assuming increased commuting time.

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, 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

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

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

In what follows, I consider a symmetric equilibrium where all producers choose the same price, b, for their good ($b_k = b$ for all k) and commit to it.[56]

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 (2). Since the worker's problem is a multi-armed bandit one, as discussed in Section 3.2 the optimal solution

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

is to be employed in the occupation with the highest Gittins index, as described in Proposition 1. In the symmetric equilibrium, each worker is employed in the occupation with the highest belief, $p_{(m)}$.

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. Total demand for intermediate good k by both consumers and firms is given by

$$q_k = \left(\frac{b_k}{P}\right)^{-\gamma}\left(\frac{W}{P} + fm\right), \tag{12}$$

where

$$P = \left(\sum_{k=1}^{m} b_k^{1-\gamma}\right)^{\frac{1}{1-\gamma}}, \tag{13}$$

is the local price level and

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

is total expenditure by city residents.[57]

Each intermediate good producer chooses a price, b_k, given the demand he faces given in (12).[58] Equation (12) pins down the quantity of good k produced, q_k, which in turn pins down the amount of labor required, l_k (see equation (10)). 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-sloping 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.

I now solve for the intermediate good producer's problem.

[57] Demand for good k comes from two sources: a) consumers and b) producers paying for their fixed cost, f, which is in terms of the final good. 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.

Solving the producer's problem implies that the demand for good k by the m producers in that location is given by

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

Therefore, total demand for good k, i.e., from both consumers and producers, is given by equation (12).

[58] 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.

Producer's k profits are given by

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

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

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

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

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

Using equation (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

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

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

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

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

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

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

$$\frac{dw\left(q_k|w_{-k}\right)}{db_k} = -\frac{\gamma}{bN\frac{d\theta I}{dw_k}}\left(\frac{IN+\pi m}{bm} + m^{\frac{1}{1-\gamma}}f\right). \tag{15}$$

Furthermore

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

Substituting in for q and W and solving leads to

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

Substituting in equation (15) for π leads to

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

which I now substitute into the price equation (14) 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 (16)) are increasing in N, since b does not depend on N.

This immediately leads to the following proposition:

Proposition 3. *In an economy where all goods are local, cities with larger populations, N, have more occupations, m.*

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

$$d = \sqrt{\frac{8(r+\delta)}{\left(\frac{\alpha_G - \alpha_B}{\sigma}\right)^2} + 1},$$

and J is the 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 demonstrated by equation (9) and the related discussion. The only difference is that the moving probability now also depends on the level of the negative externality, $z(N)$, and also on the number of goods, m.

The endogenous moving decision and the inflow decisions of movers pin down city population, N, in this model. Workers benefit from more occupations because they earn a higher wage income due to the increased occupational availability and because they consume a greater variety of products.[60] 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

[59] In equilibrium, there are no profits.

[60] See also Lee (2010) and Schiff (2015).

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 $\widetilde{x}$ where

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

and

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

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

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

which simplifies to

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

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

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

where $\widetilde{x}$ 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 $\delta \cdot \Delta$. 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: $2000 \times 14.20 \times 91/(14.20 \times 3) = \$60{,}667$.

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, θ must equal $1/(3 \times 5) = 0.067$.

	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
$\frac{\alpha_G-\alpha_B}{\sigma}$		0.1795	0.2302	0.2406	0.1886	0.2347
p_0		0.0937	0.1049	0.1051	0.1003	0.1082
$\underline{p}$		0.0304	0.02	0.0099	0.04	0.02

Table 16: Robustness. The higher occupational switching probabilities in large cities are computed by regressing occupational switching on a large city dummy and the same set of controls as in Table 2 for the subsample used. Moving probability refers to the 4-month probability of moving of workers living in large and small cities. Switching probability to a new occupation refers to the probability a worker switches to an occupation in which he has not previously been employed.

	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:						
Population Share in Large Cities	58.95%	58.96%	59.12%	58.97%	58.99%	58.96%
Moving Probability in Large Cities	0.50%	0.50%	0.44%	0.51%	0.52%	0.50%
Moving Probability in Small Cities	0.60%	0.59%	0.53%	0.61%	0.62%	0.58%
Higher Occup Sw Prob in Large	0.20%	0.37%	0.93%	0.61%	0.57%	0.55%
Higher Occup Sw Prob in Large (recent)	3.34%	1.82%	3.46%	1.71%	2.47%	1.79%
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%	5.41%	2.93%	3.57%	3.81%	4.73%
Initial Wage	$10.92	$8.87	$9.76	$9.98	$9.61	$9.54
Wage Premium	20.16%	6.25%	6.11%	5.09%	5.99%	4.60%
Wage Standard Deviation Premium	21.21%	-0.40%	1.14%	5.76%	3.94%	0.58%
Parameters:						
α_G		23.99	25.07	29.71	25.94	25.33
α_B		6.76	7.75	7.25	7.39	7.74
s		55.06%	54.71%	54.93%	55%	55.07%
δ		0.00499	0.00433	0.00487	0.005	0.00495
$\frac{\alpha_G-\alpha_B}{\sigma}$		0.2287	0.2167	0.1514	0.2007	0.2234
p_0		0.1225	0.1162	0.1214	0.1201	0.1024
$\underline{p}$		0.0404	0.0197	0.0614	0.0405	0.0299

Table 17: Robustness (continued). The higher occupational switching probabilities in large cities are computed by regressing occupational switching on a large city dummy and the same set of controls as in Table 2 for the subsample used. Moving probability refers to the 4-month probability of moving of workers living in large and small cities. Switching probability to a new occupation refers to the probability a worker switches to an occupation in which he has not previously been employed.

	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	0.20%	0.35%	0.48%
Higher Occup Sw Prob in Large (recent)	3.34%	2.57%	1.78%
Mean Wage	$14.20	$14.20	$14.20
Residual Wage Standard Deviation	$5.97	$5.97	$5.97
Other Moments:			
Prob of Switch to New Occupation	4.82%	4.09%	4.55%
Initial Wage	$10.92	$9.09	$9.29
Wage Premium	20.16%	8.38%	5.36%
Wage Standard Deviation Premium	21.21%	1.91%	1.43%
Parameters:			
α_G		23.88	25.14
α_B		6.95	7.10
s		54.87%	55.10%
δ		0.00464	0.00502
$\frac{\alpha_G-\alpha_B}{\sigma}$		0.2301	0.2104
p_0		0.1265	0.1216
$\underline{p}$		0.029	0.0411

Table 18: Robustness (continued). The higher occupational switching probabilities in large cities are computed by regressing occupational switching on a large city dummy and the same set of controls as in Table 2 for the subsample used. Moving probability refers to the 4-month probability of moving of workers living in large and small cities. Switching probability to a new occupation refers to the probability a worker switches to an occupation in which he has not previously been employed.

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 (2008): "Bandit Problems," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Macmillan.

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, 87–103.

Bolton, P. and C. Harris (1999): "Strategic Experimentation," *Econometrica*, 67, 349–374.

Campbell, J. Y. and J. F. Cocco (2007): "How Do House Prices Affect Consumption? Evidence from Micro Data," *Journal of Monetary Economics*, 54, 591–621.

Carlino, G. and W. R. Kerr (2015): "Agglomeration and Innovation," in *Handbook of Regional and Urban Economics*, ed. by J. V. H. Gilles Duranton and W. Strange, Elsevier, vol. 5.

Carrillo-Tudela, C. and L. Visschers (2014): "Unemployment and Endogenous Reallocation over the Business Cycle," Mimeo, University of Essex.

Clarida, R. H. (1993): "Entry, Dumping, and Shakeout," *American Economic Review*, 83, 180–202.

Costa, D. L. and M. E. Kahn (2000): "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990," *Quarterly Journal of Economics*, 115, 1287–1315.

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.*

De la Roca, J. and D. Puga (2017): "Learning by Working in Big Cities," *The Review of Economic Studies*, 84, 106–142.

Deming, D. and L. B. Kahn (2018): "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals," *Journal of Labor Economics*, 36, S337–S369.

Diamond, P. A. (1982): "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 90, 881–894.

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, 56–71.

Duranton, G. and D. Puga (2001): "Nursery Cities: Urban Diversity, Process Innovation and the Life Cycle of Products," *American Economic Review*, 91, 1454–1477.

——— (2004): "Micro-foundations of urban agglomeration economies," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson and J. F. Thisse, Elsevier, vol. 4.

Eaton, J. and Z. Eckstein (1997): "Cities and Growth: Theory and Evidence from France and Japan," *Regional Science and Urban Economics*, 27, 443–474.

Eeckhout, J. (2004): "Gilbrat's Law for (All) Cities," *American Economic Review*, 94, 1429–1451.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014): "Spatial Sorting," *Journal of Political Economy*, 122, 554–620.

Gautier, P. A., M. Svarer, and C. N. Teulings (2010): "Marriage and the City: Search Frictions and Sorting of Singles," *Journal of Urban Economics*, 67, 206–218.

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

Gervais, M., N. Jaimovich, H. E. Siu, and Y. Yedid-Levi (2016): "What Should I Be When I Grow Up? Occupations and Unemployment over the Life Cycle," *Journal of Monetary Economics*, 83, 54–70.

Gittins, J. C. (1979): "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, 41, 148–177.

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

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, 983–1028.

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

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

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

Greenwood, M. J. (1997): "Internal Migration in Developed Countries," in *Handbook of Population and Family Economics*, ed. by M. R. Rosenzweig and O. Stark, Elsevier, vol. 1B.

Groes, F., P. Kircher, and I. Manovskii (2015): "The U-Shapes of Occupational Mobility," *Review of Economic Studies*, 82, 659–692.

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

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

Hershbein, B. and L. B. Kahn (2018): "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings," *American Economic Review*, 108, 1737–1772.

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

Hsu, W.-T. (2012): "Central Place Theory and City Size Distribution," *The Economic Journal*, 122, 903–932.

Hu, X. (2005): "Portfolio Choices for Homeowners," *Journal of Urban Economics*, 58, 114–136.

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

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

Jovanovic, B. and Y. Nyarko (1995): "The Transfer of Human Capital," *Journal of Economic Dynamics and Control*, 19, 1033–1064.

Jovanovic, B. and R. Rob (1989): "The Growth and Diffusion 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 Specificity of Human Capital," *International Economic Review*, 50, 63–115.

Karatzas, I. (1984): "Gittins Indices in the Dynamic Allocation Problem for Diffusion 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 Effect 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 Different 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 Effects 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.

Lucas, R. E. and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122, 483–499.

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

Lucas, R. E. B. (1997): "Internal Migration in Developing Countries," in *Handbook of Population and Family Economics*, ed. by M. R. Rosenzweig and O. Stark, Elsevier, vol. 1B.

McCall, B. P. (1990): "Occupational Matching: A Test of Sorts," *Journal of Political Economy*, 98, 45–69.

Miller, R. A. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120.

Moscarini, G. (2005): "Job Matching and the Wage Distribution," *Econometrica*, 73, 481–516.

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

Papageorgiou, T. (2014): "Learning Your Comparative Advantages," *Review of Economic Studies*, 81, 1263–1295.

——— (2018): "Large Firms and Within Firm Occupational Reallocation," *Journal of Economic Theory*, 174, 184–223.

Pastorino, E. (2019): "Careers in Firms: the Role of Learning and Human Capital," Mimeo, Stanford University, Hoover Institution.

Petrongolo, B. and C. A. Pissarides (2006): "Scale Effects in Markets with Search," *The Economic Journal*, 116, 21–44.

Roback, J. (1982): "Wages, Rents and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

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

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015): "Integrated Public Use Microdata Series: Version 6.0," *University of Minnesota.*

Schiff, N. (2015): "Cities and Product Variety: Evidence from Restaurants," *Journal of Economic Geography*, 15, 1085–1124.

Silos, P. and E. Smith (2015): "Human Capital Portfolios," *Review of Economic Dynamics*, 18, 635–652.

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

Tian, L. (2019): "Division of Labor and the Productivity Advantage of Cities: Theory and Evidence from Brazil," *mimeo, INSEAD.*

Whittle, P. (1982): *Optimization over Time: Programming and Stochastic Control*, New York: Wiley.