Worker Sorting and Agglomeration Economies

Theodore Papageorgiou*

McGill University, Department of Economics†

March 2018

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 and develop a dynamic model of occupational choice that captures them. The calibration of the model suggests that better occupational match quality accounts for approximately 35% of the observed wage premium and part of the greater inequality in larger cities.

Keywords: Occupations, Multi-armed Bandits, Agglomeration Economies, Urban Wage Premium, 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.

†855 Sherbrooke St West, Room 519, Montreal, QC H3A 2T7, Canada. E-mail: *theodore.papageorgiou@mcgill.ca* Web: *http://www.theodorepapageorgiou.com*

1 Introduction

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

This paper contributes to our understanding of agglomeration economies in three ways. First using several different datasets, I document a number of facts relating to the number of occupations in large and small city and the relationship between city size, occupational switching patterns, wages and moving patterns. These facts are not consistent with the standard urban theories. Second, guided by the findings, I develop a dynamic model of occupational choice where larger cities have more occupations, which 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.

Using a comprehensive dataset 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, but instead holds when I weight occupations based on their worker inflows.[2]

I next use a worker panel data to confirm the well-known regularity that workers in more highly populated cities earn higher wages. I find however that this wage difference is not instantaneous, but instead appears with time in a location. More specifically, when focusing on recent movers, workers who moved to a large city receive approximately the same wage as those who moved to a small city. At the same time recent movers to larger cities switch occupations at a higher rate than workers who moved to smaller cities. This difference reverses with time in the city and overall, the occupational switching rate is the same in large and small cities. Moreover, workers in larger cities are less likely to move to another

[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.

[2] In order to check whether there may be additional occupational opportunities beyond those reported in the vacancy data, I use employment data and find that fewer than 5% of workers are employed in an occupation where a local vacancy is not reported. This suggests that workers have few opportunities besides those captured in my vacancy data.

location and switch occupations. I document that the patterns associated with moving and switching occupations are very different from those associated with moving and remaining in the same occupation. The above facts are not consistent with the standard urban theories where workers become immediately more productive upon arriving in larger cities and there is no worker reallocation in equilibrium.[3]

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

In equilibrium, increased options allow workers in larger cities to form better occupational matches compared to workers in smaller cities. Workers who recently moved to a large city do not initially form better matches than workers in smaller cities. As a result they do not receive higher wages. They have however more occupational options: this leads to higher occupational mobility for recent movers, consistent with the data, who over time form better matches and obtain higher wages. Overall, occupational mobility is not higher in larger cities: on one hand workers have more options in larger cities; on the other they are on average better matched. These two effects roughly offset each other. Workers residing in larger cities are however unambiguously less likely to move, both because in equilibrium they are better matched and because they have more options, so now these two effects work in the same direction. 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 calibrate the model using moments relating to differences in occupational switching and geographical mobility across different size cities. The model matches these moments well. It also matches the magnitude of occupational switching to new occupations, as well as the initial wage. 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 the last part of the paper 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

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

diversity. A larger city caters to more diverse consumer tastes, producing and hiring in a larger variety of services and products. Both occupations, as well as population are endogenously determined.

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. 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" (page 2109). The mechanism proposed here however has a number of additional predictions regarding worker reallocation both within, as well as across cities and how this reallocation interacts with city size, wages and time in the city, that separate it from other mechanisms. For that reason, it is precisely these reallocation moments that I use to calibrate the model, so as to pick up only the importance of my mechanism.[4]

This paper contributes to several strands of literature. Bleakley and Lin (2012) document that young workers switch occupations more often in larger cities. This is related to my finding that recent movers in large cities are more likely to switch occupations. They interpret their findings as evidence of increasing returns to scale in the matching function between workers and vacant firms (see also Diamond, 1982, Petrongolo and Pissarides, 2006 and Gautier and Teulings, 2009).[5] The mechanism of the present paper does not rely on the assumption of increasing returns to matching - in fact there are no search frictions in the model. Instead greater occupational availability in larger cities allows workers to form better occupational matches. My empirical finding on the number of occupational opportunities in large cities is related to Duranton and Jayet (2011) who use French employment data and find that scarce occupations are over-represented in large cities.

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 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

[4]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. See also Baumgardner (1988) who argues that large cities allow for a greater degree of specialization.

[5]Hornstein et al. (2012) argue that search frictions account for only a limited fraction of the differences in wages across workers. In the urban context, Baum-Snow and Pavan (2012) find that differences in search frictions contribute little to the observed urban wage premium.

and Maré, 2001).[6,7]

In addition, the paper is related to the literature on greater inequality in larger cities. Gautier and Teulings (2009) and Eeckhout et al. (2014) introduce models that generate differences in the wage dispersion, but a key assumption in both setups is ex ante heterogeneous workers who choose where to work. In my setup there is no ex ante worker heterogeneity. Instead the differences in both the level of wages, as well as wage inequality across city sizes are driven, not by selection, but greater occupational availability.

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 when 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 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, 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 uses vacancy data to document the relationship between the number occupational postings and city size. Section 3 documents a number of facts on wages, moving patterns and occupational switching patterns in large cities. In Section 4 I introduce a model that is consistent with these facts and in Section 5 I calibrate it. Section 6 extends the model by endogenizing the number of occupations in each location. Section 7 concludes.

2 City Size and Number of Occupations

In this section I consider 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

[6] 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.

[7] 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 more opportunities to workers to discover a good occupational match.

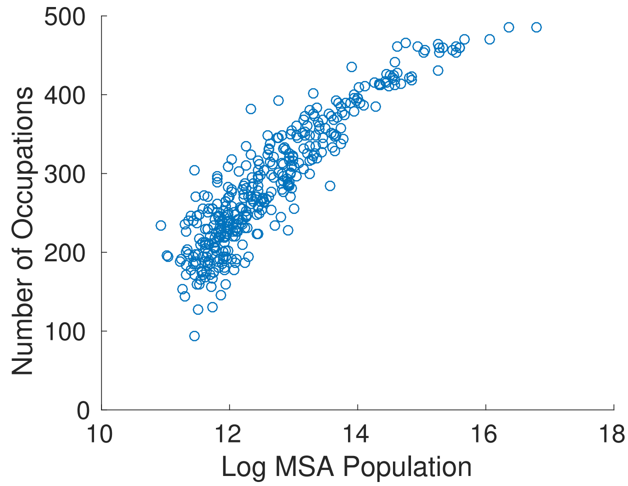

Figure 1: Number of Occupations vs. Log MSA Population - Burning Glass Vacancy Data

that would interest only few workers. Moreover the difference is not driven by specific occupations that are predominantly available in large cities.[8]

I begin by examining whether workers have more occupations available in larger cities using 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 their database covers the near-universe of online job vacancies.[9]

The BG data contains information on the posting's detailed (6-digit) occupation, 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 1st 2016 and April 30th 2016. There are 6,103,537 postings during this period.

Figure 1 plots the number of 3-digit occupations in which there are vacancies in every MSA against its population as reported in the 2010 Census.[10] 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. Examples of occupations in the data include actuaries, proofreaders, theatrical and performance makeup artists, manicurists and pedicurists, parking lot attendants and skin care specialists.

[8]One might expect that the available occupations are endogenous to market characteristics and in Section 6 I introduce a model of endogenous occupation creation. However from the point of view of an individual worker, the distribution of occupations across cities is fixed and taken as given when making decisions.

[9]See also discussion in Hershbein and Kahn (2016) and Deming and Kahn (2017) who were one of the first to use the BG data.

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

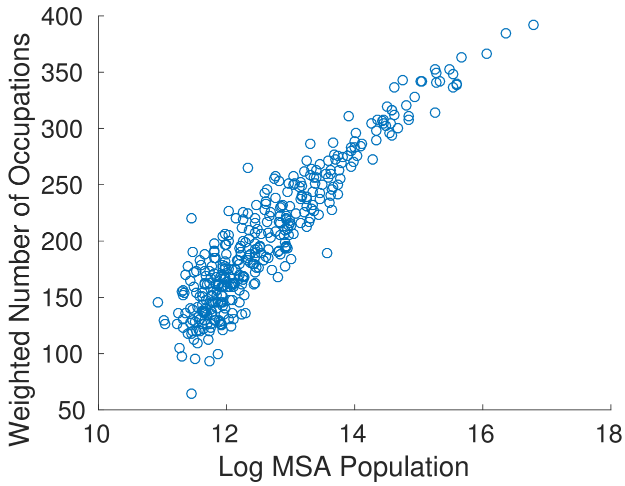

Figure 2: Weighted Number of Occupations vs. Log MSA Population - Burning Glass Vacancy Data

Figure 1 assumes that an occupation is available if there is at least one posting in that occupation. The same relationship however emerges if I consider a stricter definition where either 5, 10 or 50 postings are needed for an occupation to be available.[11]

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.[12] 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.

Similarly, it is not the case that the difference is driven specific occupations that are available in large cities but not in small ones. If I consider all cities that have fewer than 160,000 residents, I find that there are 485 occupations available in this group of cities. By comparison, in the New York metropolitan area, which is the largest MSA, there are 486 occupations available. This holds despite the fewer vacancy

[11]The same relationship also holds if I consider "high-skill" or "low-skill" occupations separately (high skill being 0010-3540, Management, Professional and Related Occupations and low-skill all others).

[12]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 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).

postings in the small city group (303,142) compared to the NY MSA (431,937).

In addition, if one groups occupations according to their 2-digit codes instead, much 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, there is a smaller number of occupations that are available.[13]

The vacancy data used so far is from online postings and while it is comprehensive, it does not include postings that are not also posted online. In order to check whether there may be additional occupational opportunities beyond those reported to 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. It turns out however 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.[14]

Finally, I confirm the same relationship using vacancy postings from the UK.[15] 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 reports 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.[16] The resulting figures can be found in the online appendix.

[13]It is worth noting that within a 2-digit occupation the occupations 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).

[14]The ACS data contains 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). Ruggles et al. (2015).

[15]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 of 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.

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

	Initial	Moved<4 years	Full Sample
	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 clustered by individual. ***, ** and * indicate statistical significance at 1, 5 and 10 percent respectively.

3 City Size, Wages, Occupational Switching and Moving

The goal of this section is to investigate the empirical relationships between: city size, occupational switching and wages. In addition, as agglomeration economies are linked to geographical mobility, I study moving behavior and its relation to occupational switching, wages and city size. More specifically, I examine the city size wage premium and its evolution with time in the city; occupational switching and how it varies with city size; the patterns associated with moving and switching occupation and how they are different from those associated with moving and remaining in the same occupation.

In this section, 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, three-digit occupation, three-digit industry, employer size, as well as the usual demographics, such as gender, age, race, education and marital status. The 1996 panel of the SIPP uses dependent interviewing, which is found to reduce occupational coding error (Hill, 1994). This makes it preferable to use when investigating occupational switching, compared to other panel datasets, such as the National Longitudinal Survey of Youth 1979. Furthermore, the SIPP follows original respondents when they move to a new address, unlike, for instance, the Current Population Survey which is an address-based survey. Appendix A contains more details about the data, how the moving variable is constructed and discusses how both the moving probabilities and the occupational switching rates are consistent with other datasets.[17]

3.1 City Size, Wages, Occupational Switching and Moving

I first examine the evolution of the city size wage premium as a function of time in a city. The last column of Table 1, confirms the well-known empirical regularity that workers in more highly populated areas are

[17]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.

	Full Sample	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 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). ***, ** and * indicate statistical significance at 1, 5 and 10 percent respectively.

paid significantly higher wages. The magnitude of the coefficient is in line with the results from other datasets.[18] 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 implies that the mechanism that generates these wage differences, appears to be relevant mostly *after* a worker arrives in a larger city. Models where agglomeration economies are captured via an increasing function between TFP and city size[19] cannot replicate this fact as they predict that workers moving to larger cities should immediately earn higher wages compared to those moving to less populated locations.

I next explore the patterns of occupational switching and geographical mobility. First, I examine 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 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.[20] The result remains and

[18]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).

[19]Rosen (1979) and Roback (1982).

[20]Investigating this further shows that almost of all the effect comes from workers who moved between 1 and 2 years ago. This is also true when using data from the 1997 National Longitudinal Survey of Youth. 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).

becomes even stronger when I control for the size of the previous city. In addition, workers who arrive from larger cities appear to be less likely to switch occupations, though the result is not statistically significant.

Most of the flows are offsetting across occupations, suggesting that most of these switches are driven by idiosyncratic reasons rather than aggregate shocks.[21] Table 16 in the Appendix reports the destination occupations that 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.[22]

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%.[23] In the setup described in the next section, larger cities offer workers more occupational options who in equilibrium form better occupational matches. As a result, city size has an ambiguous effect on occupational switching (on one hand workers are better matched, on the other they have more options); it predicts however that workers are unambiguously less likely to move from a large city (both because they are better matched and because they have more options).

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

As shown in Table 4, 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.

Moreover, as shown in the first column of Table 5, 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

[21]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 less 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.

[22]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 4, 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.

[23]It is also worth noting that workers who move and switch occupations are more likely to go to a metropolitan area than they are to go to non-metropolitan areas (88%). This probability is higher than the fraction of workers living in metro areas (79%).

	Prob of Moving
	& Switching Occup (Probit)
ln(current city pop)	-0.0007
	$(0.0002)^{***}$
Number of Obs	144653

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 clustered by individual. Coefficients represent marginal effects evaluated at the average value of the 4-month probability, which equals 0.0054. ***, ** and * indicate statistical significance at 1, 5 and 10 percent respectively.

	$\ln(\text{wage})_t$
$\ln(\text{wage})_{t-1}$	0.85
	$(0.004)^{***}$
Move & Switch Occupations	0.023
	$(0.012)^{**}$
Occupation Switch without Moving	0.011
	$(0.002)^{***}$
Move & No Occupation Switch	0.005
	$(0.003)^{*}$
Number of Obs	176013

Table 4: Impact of Occupation Switching and Moving on Wage. Source: 1996 Panel of Survey of Income and Program Participation. 4-month intervals. Controls include gender, race, education, marital status, firm size, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. ***, ** and * indicate statistical significance at 1, 5 and 10 percent respectively.

for at least some of the moves, labor market considerations are important in the decision to move. The second column of Table 5, indicates that wages are declining beforehand only in the case of workers who move and switch occupations, whose wages fall by approximately 2.4%. Workers who move and keep the same occupation do not experience decreasing wages before moving. Thus the results of the first column are driven by workers who move and switch occupations.[24]

In addition, 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 (all results available upon request). This further underscores my earlier assertion that the patterns associated with moving and switching occupations are very different from those associated with moving and remaining in the same occupation.

[24] If I control for occupation switching separately in Table 5, only occupation switching is statistically significant, whereas moving is not. This is consistent with the predictions of the setup of the next section where workers who switch occupations experience wage declining paths beforehand, regardless of whether they move or not; moving does not imply a steeper wage decline, compared to switching occupations without moving.

	$\ln(\text{wage})_{t-1}$	$\ln(\text{wage})_{t-1}$
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 5: 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 clustered by individual. ***, ** and * indicate statistical significance at 1, 5 and 10 percent respectively.

3.2 Robustness

This paper argues that the different patterns related to wage growth and occupational switching in larger cities are linked to larger occupational availability in those cities. One might expect however that this mechanism is less relevant for workers with a low probability of switching occupations, such as dentists or airplane pilots. I therefore repeat the analysis above on the subset of workers who are currently employed in the occupations at the lowest decile of switching probability.[25] Table 6 presents the results for this subset of workers which are remarkably different. While again there is a sizable city wage premium for all workers, this time workers who moved within 4 years also enjoy a wage premium, though there are too few observations in this subsample for the regression to have power. More strikingly, city size no longer affects occupational switching, nor the probability of moving and switching occupations, unlike the results in the full sample. Somewhat surprisingly, for this subset of workers switching occupations without moving is associated with significant wage losses, suggesting that perhaps these workers mostly switch occupation due to exogenous reasons.

3.3 Discussion

In summary, workers in larger cities have significantly more occupational options than workers in smaller ones. Workers who recently moved to a large city do not immediately earn higher wages; they are however more likely to switch occupations compared to those who moved to a small city. With time in the city this

[25] I consider occupations with at least 200 observations.

	Full Sample	Moved<4 years			
			Occ Switch	Prob of Move	
	ln(wage)	ln(wage)	Prob (Probit)	& Sw Occ (Probit)	$\ln(\text{wage})_t$
ln(current city pop)	0.05272	0.0563	0.0002	0.0042	
	$(0.0059)^{***}$	(0.039)	(0.0015)	(0.0058)	
$\ln(\text{wage})_{t-1}$					0.864
					$(0.01)^{***}$
Move & Sw Occ					−0.022
					(0.265)
Occ Sw & No Move					−0.06
					$(0.022)^{***}$
Move & No Occ Sw					0.004
					(0.017)
Number of Obs	9547	212	8403	8589	9986

Table 6: Workers in Occupations with Lowest Switching Probability. Occupational switching probability regression is conditional on not moving. Source: 1996 Panel of Survey of Income and Program Participation. 4-month intervals. Controls for all regressions include gender, race, education, marital status, firm size, quartic in age, 11 industry dummies, 13 occupation dummies. Standard errors clustered by individual. ***, ** and * indicate statistical significance at 1, 5 and 10 percent respectively.

difference disappears and in the cross-section there is no longer a positive relationship between city size and occupational switching. Workers in larger cities are less likely to move to another location and switch occupations. Workers who move and switch occupations experience wage decreases prior to the move. In addition, they enjoy wage gains upon moving. With the exception of the lack of a positive relationship between city size and occupational switching and the size wage premium being small for recent movers, the remaining facts are, to my knowledge, novel.

None of the existing models of agglomeration economies can jointly account for the above facts. As mentioned above, the mechanism that drives these wage differences, appears to be relevant mostly after a worker arrives in a city, so it is not generated by a level effect that makes workers immediately more productive in larger cities. Consistent with this finding, more rapid general human accumulation in larger cities, e.g. due to knowledge spillovers (Lucas, 1988), has often been suggested as a potential mechanism driving agglomeration economies. While more rapid general human accumulation in larger cities can account for the observed higher growth of wages, it cannot account for the remaining patterns documented above. In particular, it is hard to explain why knowledge spillovers would lead to more occupational switching for recent arrivals, but similar overall levels of occupational switching in the cross-section. In fact, it is easy to show that general human capital accumulation, not only can not generate the above occupational switching patterns, but does not have *any* impact on the level of occupational mobility.[26]

[26]Precisely because general human capital is accumulated in every occupation, it doesn't affect the occupational choice

Moreover, if part of the human capital is specific to an occupation, then more rapid accumulation would imply less, not more occupational switching among recent movers in large cities compared to small, which is what we observe in the data.[27] This is precisely why in the model's calibration's presented in Section 5, I target worker reallocation moments, so as to pick up the importance of higher occupational availability, rather than other potential mechanisms, such as more rapid human capital accumulation or knowledge spillovers.

4 Model

Guided by the above facts I develop a model of occupational choice and geographical mobility that accounts for them. The model is based on the intuitive idea that because larger cities have more occupations, workers there are more productive. 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. In Section 6, 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. I introduce however 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).[28] If workers fail to find a suitable occupation they move to another city by paying a moving cost.[29]

---

decision and therefore occupational switching behavior. If some occupations value general human capital more than others, then more rapid accumulation would imply less, not more occupational switching among recent movers in large cities compared to small, which is what we observe in the data.

[27] Kambourov and Manovskii (2009b) find that accumulation of occupation-specific human capital is fast and can lead to a wage increase of 12%-20% within the first five years. In the online appendix to this paper, I extend the model presented in the next section to include occupation-specific human capital.

[28] I follow the recent literature that has argued that occupational mobility is largely due to information frictions (e.g. Pastorino, 2014, Papageorgiou, 2014, Groes, Kircher and Manovskii, 2015). However, the assumption that workers don't know their productivity is not crucial. The alternative is for workers to know their productivity in all occupations, but the worker's productivity in his current occupation could be changing over time, leading to occupational switches.

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

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

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

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.[30] The common theme of this literature is that a worker's wage depends on the type of work they do (their occupation), rather than who is employing them. For instance, an accountant's wage reflects how good he is in his accounting tasks, rather than which particular firm is employing him. Second, work by Baum-Snow and Pavan (2012) has found that worker-firm match qualities and search frictions do not differ much across cities of different size.

I next describe the setup in detail.

4.1 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.[31]

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 or Li and Yao, 2007). When moving a worker randomly goes to a new city.[32,33] Moving from one city to another entails a cost $c > 0$.

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

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

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

Let $\alpha_G > \alpha_B$. Productivities, α_l^{ik}, are independently distributed across occupations, cities and work-

[30] Kambourov and Manovskii (2009a and 2009b), Antonovics and Golan (2012), Eeckhout and Weng (2010), Alvarez and Shimer (2011), Carrillo-Tudela and Visschers (2014), Papageorgiou (2014), Groes, Kircher and Manovskii (2015), Silos and Smith (2015), Gervais, Jaimovich, Siu and Yedid-Levi (2016).

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

[32] In the model presented in Section 6, 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).

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

ers.[34] Furthermore, α_l^{ik} is unknown, and let $p_{0l}^{ik} \in (0,1)$ be the worker's prior belief that $\alpha_l^{ik} = \alpha_G$. When he enters a city, the worker draws his prior, p_{0l}^{ik}, for all occupations in that city. Each prior, p_{0l}^{ik}, is drawn independently from a known distribution with support $[0,1]$ and density $g(\cdot)$.

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

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

where $\zeta = \frac{\alpha_G - \alpha_B}{\sigma}$. The last term on the right hand side is a standard Wiener process with respect to the unconditional probability measure used by the agents. p_{tl}^{ik} is a sufficient statistic of the worker's beliefs regarding α_l^{ik}. 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_{0l}^{ik} for each of the 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.

4.2 Behavior

Firm competition for the services of workers, ensures that a worker's compensation equals his expected output in the occupation n, he is employed[35]

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

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,

[34] 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 discussion in Section 5.

[35] 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 next instant's output. All the implications derived later continue to hold.

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 city. Solving this problem numerically with more than a handful of occupations is computationally intractable. For instance, in the calibration presented in Section 5, m equals 12 in some cities. Using a 20 point grid for each belief, implies the dimension of the state space is in the order of 20^{12}.

The difficulty of solving multi-armed bandit problems numerically is well-known, but fortunately these problems become tractable using Gittins indices (see Gittins and Jones, 1974 and Bergemann and Välimäki, 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 of 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 arm, k, depends only that arm's beliefs (in this case p^k).[36] 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).

More specifically, following Whittle (1980, 1982) and Karatzas (1984), 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 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 being good 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

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

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

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

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

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:

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

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

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

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

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

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

For eq. (5) to hold, from eq. (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

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

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

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

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

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

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 speed of learning varies across occupations, then the Gittins index will be 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", in other words an arm that always pays a constant payoff. Because there is no variation in the safe arm's payoff, there is no belief updating and the worker's state space never changes. Therefore, a safe arm is an absorbing state.

In the context of the setup, because the option of moving also provides a known value, J, this equivalent to a "safe arm" that always pays some payoff w_j, given by

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

Note that since J is trivially the retirement value associated with playing the safe arm, this 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 $\underline{p}$, where*

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

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

Summarizing, consider a worker who has just moved to a city. He immediately draws a prior, p_0^k, for each of the m occupations that are available for him to work in. If all m draws are below $\underline{p}$, he immediately pays the moving cost c and starts over in another city. Otherwise, he picks the occupation with the greatest value of the prior and begins work there. If the value of his posterior in that occupation falls below the value of the second best occupation, he immediately switches. A worker leaves his current city endogenously, only when value of the posteriors of all his occupations reach $\underline{p}$.[37] Some workers however may find that one of the occupations they try out is a good match for them, in which case their posterior drifts towards one and their wage increases. These workers leave their match and city only exogenously at rate δ.

4.3 Implications

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. This is indeed confirmed by the model's calibration results (Section 5).

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

[37]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.

conditional on d, is given by

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

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

However d, the number of draws above $\underline{p}$ is increasing in the total draws, m. Thus the probability that a worker moves endogenously is decreasing in m implying that the rate at which workers move out of a city is lower in cities with more occupations m. This implies that workers stay longer in cities with more occupations, m. Since the flow into a city is the same regardless of the number of occupations, the above result immediately implies that cities with more occupations, m, have larger populations.[38] This is consistent with the evidence in Figure 1. Moreover, as shown in Table 3 above, workers in highly populated areas are indeed less likely to move and switch occupations.

I now turn to the impact of moving on wages. In my framework, workers pay the cost, c and move because they expect a better match in their new location. Their last wage before the move is $w\left(\underline{p}\right)$, whereas in the new location, the worker chooses to work in the occupation with the highest prior, $p_0^k > \underline{p}$. Thus workers who move experience wage increases. This is consistent with the findings in Table 4.

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 the evidence in Table 5. 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 in Section 3, 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), ignoring δ shocks, the probability of an occupational switch for a worker whose posterior in his

[38] 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 for the inflow of workers into a city to depend positively on the number of its occupations strengthens the result.

Model Implications	**Empirical Evidence**
Probability of moving out of city decreasing in city size	Table 3
Larger populations in cities with more occupations	Figures 1 and 2
Wages increase upon moving	Table 4
Wages fall prior to moving	Table 5
Higher occup switching for recent movers in larger cities (C)	Table 2
Higher wages in larger cities (C)	Table 1
Low wage urban premium for recent movers (C)	Table 1
Higher wage dispersion in larger cities (C)	Table 11

Table 7: Summary of Model implications and Corroborating Empirical Evidence. (C) indicates implications that hold for the calibrated model.

current occupation is equal to $p_{(m)}$, is given by

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

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

One might expect the setup to predict that occupational switching is higher in larger cities. That is not however, necessarily the case: workers in larger cities have higher posteriors in their current occupations $p_{(m)}$. Their second highest posterior, $p_{(m-1)}$, is also increasing in m, the total number of occupations. Therefore the number of occupations has an ambiguous effect on the rate of occupational switching. Put differently, workers in larger cities are both better matched, which tends to decrease their switching probability, but also have better outside options, which increases the probability they switch.

Table 7 summarizes the model's implications and also cites the associated empirical evidence. Some of the implications have been derived theoretically in this section, whereas the remaining hold for the calibrated model presented in the next section.

5 Calibration

I next investigate the quantitative importance of the proposed mechanism. In particular, to what extent can it 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 in both wages (mean wage and wage inequality), as well as 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.[39] 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).[40] 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 (δ, c, ζ, p_0 and 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 within-occupation 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 B contains more details.

In order to set the number of available occupations in each location (large vs. small cities), I use two moments. Using the transition matrix weighted number of occupations in each city from the BG data (Figure 2), the population-weighted average number of (weighted) occupations in cities with more than 500,000 inhabitants is 322.6. Similarly the population-weighted average number of (weighted) occupations in cities with fewer than 500,000 inhabitants is 141.6. Therefore the ratio of the number of occupations in large over small cities is 2.28.[41] This ratio is one of the two moments I target.

[39] 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 (2016) find that the college wage premia are higher in larger cities.

[40] As shown in Section 3, in the larger sample, workers who just moved to a new location, receive higher wages if they moved to a highly populated area, but the coefficient is not large (first column of Table 1). This suggests that that static advantages, whereby workers immediately become more productive upon arriving in larger cities, are not important in explaining the wage premium.

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

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

Table 8: Parameter Values

Moments:	**Data**	**Model**
Pop Share in Large	58.95%	58.93%
Moving Prob Large	0.50%	0.49%
Moving Prob Small	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 Within-Occupation Wage St Deviation	\$5.97	\$5.97

Table 9: Targeted Moments

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 expected tenure in each occupation). In the data the four-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 four-month switching probability to new occupations is 4.29%. As shown in Section 4.2, a worker who has tried out an occupation in a previous location would not want to sample it again in his new location. Thus one should not view 12 and 5 as capturing the total number of occupations in a large and small city respectively, but the average number of new occupations a worker encounters in each location. 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 the data.[42]

Moreover, given that the number of occupations in large and small cities is central to the model's

[42] One might expect younger workers to have more new occupations and older workers to have fewer. The calibration here essentially targets the number of occupations tried out by an average worker.

Moments:	**Data**	**Model**
Switch Pr to New Occupations	4.82%	4.29%
Initial Wage	\$10.92	\$10.23

Table 10: Other Moments

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

Table 11: Predicted Wage Premium and Greater Wage Inequality in Large Cities

mechanism, I present results using a number of different combinations of these two parameters.

In the second step, I jointly retrieve values for the following 5 parameters: s, δ, c, p_0 and ζ, where s is the probability that a worker who moves goes to a large city.[43] In order to do that I use moments relating 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 four-month probability of moving for workers living in a small cities and the same probability for those living in a large city. I simulate the model presented in Section 4 and match the simulated moments with the ones from the data.

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 within-occupation 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.[44] The full set of parameters is presented in Table 8. The implied cost of moving, c, equals \$60,667.

Although a rigorous identification argument is impossible due to the complexity of the framework, I attempt to give 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 four-month moving probabilities for workers in large and small cities pin down the moving rate, δ and the moving cost, c. Finally the speed of learning, ζ and the level of the initial belief, p_0, are pinned down by the occupational switching probability regressions for recent movers and the entire sample respectively.[45]

[43] By allowing this probability 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 23.

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

[45] The speed at which workers update their beliefs depends on $p(1-p)\zeta$. Changing ζ affects the speed of learning (and

Targeted Moments:	Data	Baseline				
Pop Share in Large	58.95%	58.93%	0.5895%	58.88%	58.92%	59.06%
Moving Prob Large	0.50%	0.49%	0.50%	0.47%	0.53%	0.47%
Moving Prob Small	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 Within-Occup Wage St Dev	$5.97	$5.97	$5.97	$5.97	$5.97	$5.97
Other Moments:						
Switch Pr to New Occupations	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 Ratio	21.21%	7.21%	3.37%	2.95%	4.45%	3.41%
Parameters:						
# of Occupations in Large		12	11	12	11	15
# of Occupation in Small		5	5	4	6	5
Ratio (# Occ Large over # Occ Small)	2.28 or 2.97	2.4	2.2	3	1.83	3
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 12: Robustness

The targeted moments are presented in Table 9.[46] 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 10 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 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 11 presents the predicted wage premium and the ratio of the cross-sectional standard deviation of wages, neither of which were targeted. The calibrated model replicates approximately 35% of the

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.

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

Targeted Moments:	**Data**					
Pop Share in Large	58.95%	58.96%	59.12%	58.97%	58.99%	58.96%
Moving Prob Large	0.50%	0.50%	0.44%	0.51%	0.52%	0.50%
Moving Prob Small	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 Within-Occup Wage St Dev	\$5.97	\$5.97	\$5.97	\$5.97	\$5.97	\$5.97
Other Moments:						
Switch Pr to New Occupations	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 Ratio	21.21%	-0.40%	1.14%	5.76%	3.94%	0.58%
Parameters:						
# of Occupations in Large		18	9	9	10	14
# of Occupation in Small		6	4	5	5	6
Ratio (# Occ Large over # Occ Small)	2.28 or 2.97	3	2.25	1.8	2	2.33
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 13: Robustness (continued)

observed wage premium. Moreover, it replicates about a third of the greater wage inequality that has been documented in larger cities. It is worth emphasizing that in my setup there is no ex ante worker heterogeneity. Gautier and Teulings (2009) and Eeckhout et al. (2014) introduce models that generate greater inequality in large cities, but in both setups a key assumption is ex ante heterogeneous workers who choose where to work.[47] In the present paper the differences in both the level of wages, as well as wage inequality across city sizes are driven, not by selection, but by greater occupational availability.

I next examine the sensitivity of the results with respect to the number of occupations in large and small cities. Following the discussion above, Tables 12 through 14 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 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

[47]See also Baum-Snow and Pavan (2013) for an investigation of how the increase in the 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.

Targeted Moments:	Data		
Pop Share in Large	58.95%	58.97%	59%
Moving Prob Large	0.50%	0.47%	0.51%
Moving Prob Small	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 Within-Occup Wage St Dev	\$5.97	\$5.97	\$5.97
Other Moments:			
Switch Pr to New Occupations	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 Ratio	21.21%	1.91%	1.43%
Parameters:			
# of Occupations in Large		13	13
# of Occupation in Small		5	6
Ratio (# Occ Large over # Occ Small)	2.28 or 2.97	2.6	2.17
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 14: Robustness (continued)

one might expect, a higher ratio of occupations in large cities relative to small leads to 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 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

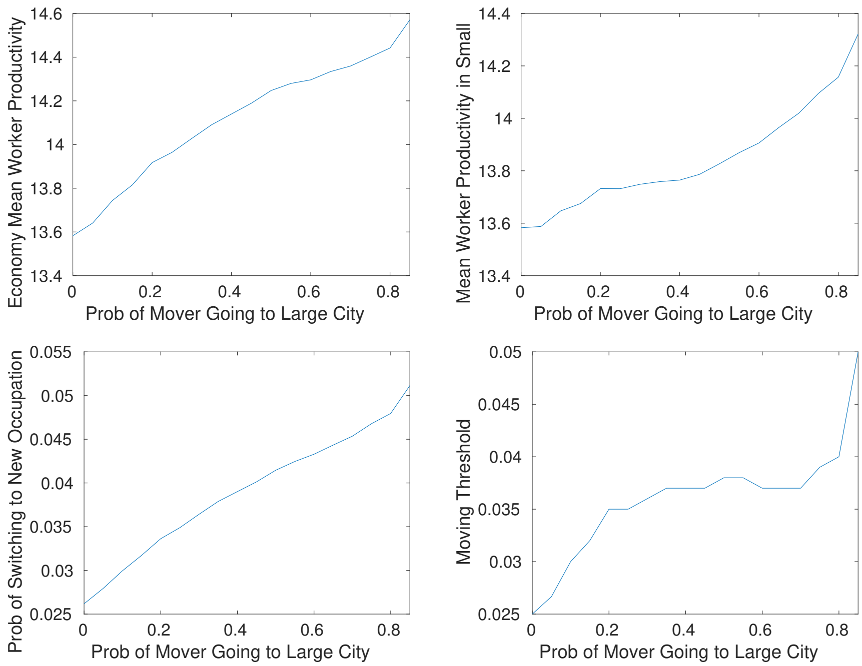

Figure 3: Sensitivity Analysis - Probability of Mover Going to Large City

interestingly however, mean wages and therefore 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 more likely to move. As shown in Figure 3, the moving threshold, $\underline{p}$, is indeed increasing in s. On net 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.

In 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 as 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, as well 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 speed of learning, $\frac{\alpha_G - \alpha_B}{\sigma}$, by appropriately adjusting σ, so as not to conflate different effects. I consider the impact of changing $\frac{\alpha_G - \alpha_B}{\sigma}$ below.

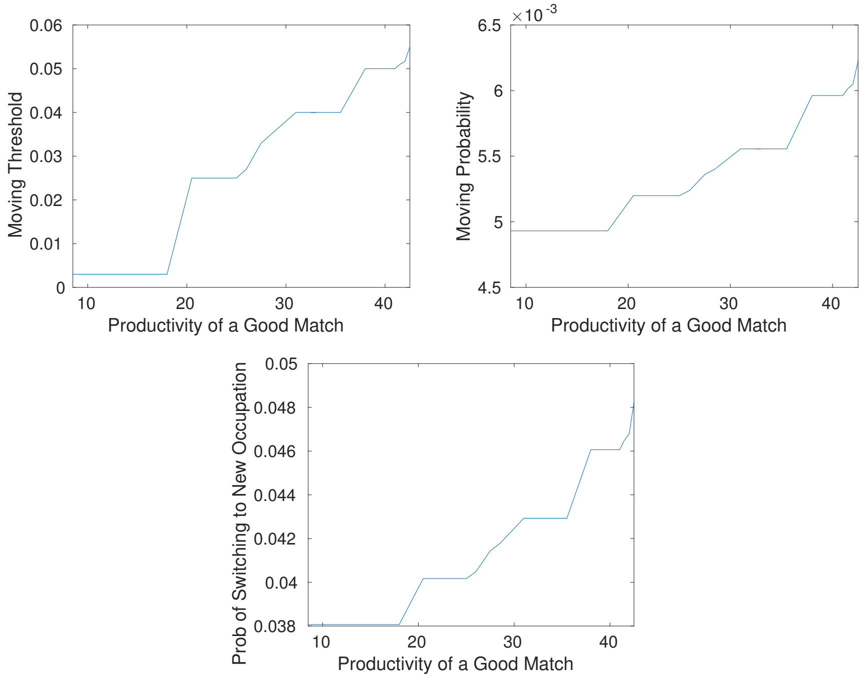

Figure 4: Sensitivity Analysis - Productivity of a Good Match, α_G

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 speed of learning, $\frac{\alpha_G - \alpha_B}{\sigma}$, affects workers in this economy. When the speed of learning is close to zero, there is almost no switching, as workers learn extremely slowly. As shown in Figure 7, increasing the speed of learning 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 speed of learning, since there is very little 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 result sort quickly through their locations' occupations. Both

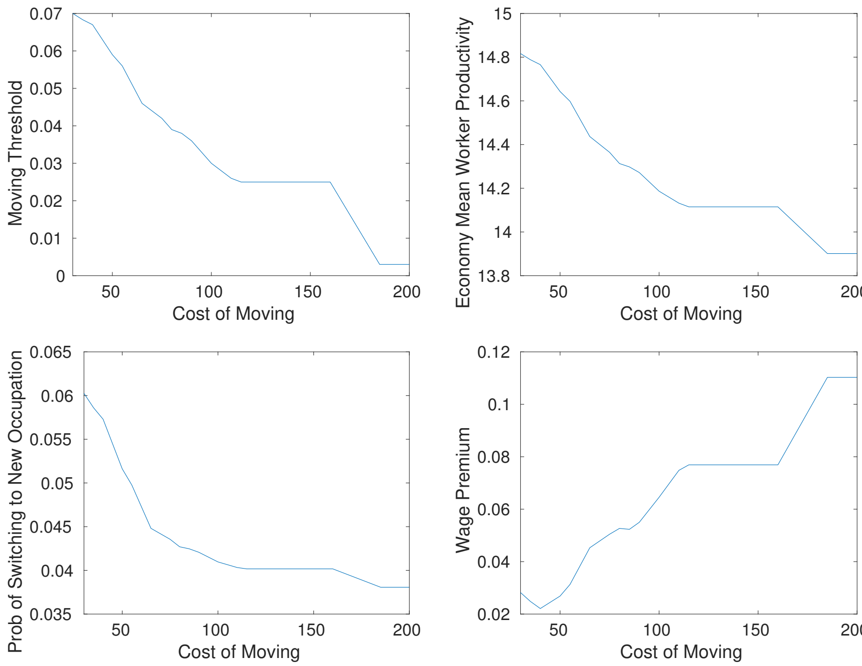

Figure 5: Sensitivity Analysis - Cost of Moving

the switching probability to new occupations, as well as average worker productivity increase, while the benefit of being in a large city with many occupations declines and so does the wage premium.

6 Endogenous Occupation Creation

The previous section explored the extent to which better occupational matching accounts for the observed agglomeration economies. That setup takes as given that some cities have more occupations and the results do not depend on the reasons behind this fact.

In this section, I extend the model developed in Section 4 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. For instance opera singers exist in larger cities, since an opera house is less likely to be profitable in a small town. A larger city caters to more diverse consumer tastes, producing and hiring in a larger variety of services and products. Related to the above, some occupations may reflect to some extent the degree of increased specialization that is possible in larger cities, for instance specialized engineers (see also Baumgardner, 1988 and Kok, 2014). The setup is also related to Hsu's (2012) formalization of central place theory, whereby different goods have different degrees of returns to scale.

The basic environment is the following: as before, workers learn about the quality of their occupational

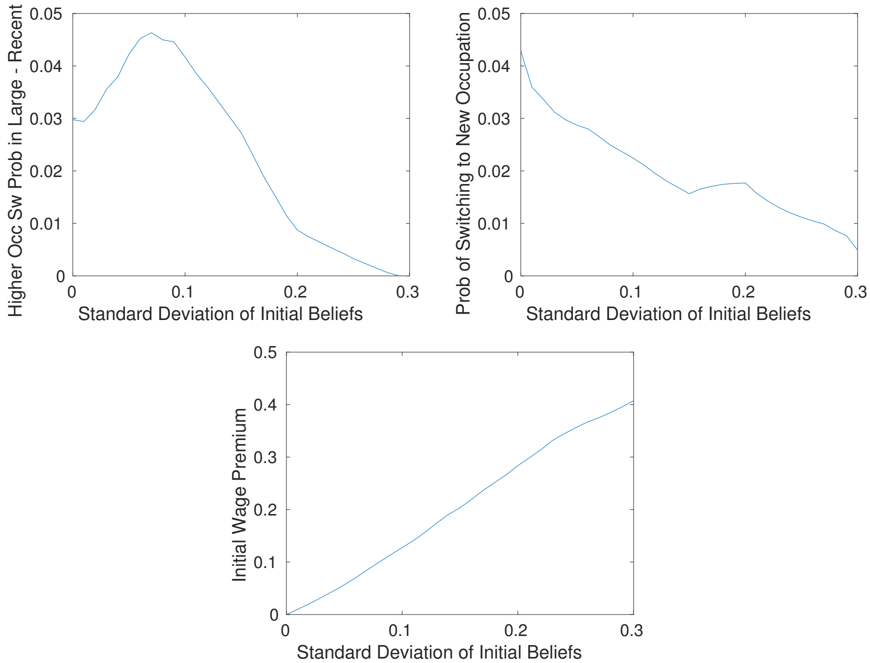

Figure 6: Sensitivity Analysis - Dispersion in Initial Beliefs

match and also decide whether to move or not. I now allow workers to choose their destination city, though in equilibrium they also randomly select their destination, as in the model of Section 4. 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 from increased occupational availability as in the baseline model, but also increased consumption variety. Increased population however also causes a negative externality, which prevents cities from becoming unboundedly large.[49] In this setup, both the number of occupations, as well as population are endogenously determined; in equilibrium cities with larger populations have more occupations, consistent with the evidence in Section 2.

The above setup has a similar flavor to Wolinsky (1983) who also endogenizes the number of producers in a location and shows that under certain conditions, all producers enter a single location and consumers go on shopping trips there. The difference is that here consumers are restricted to shopping from their local producers, so that in equilibrium there are intermediate good producers in many locations, rather than just a single one.[50] In the online appendix I also consider a setup where all goods are tradable across

[49] As in all urban models with agglomeration economies, 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.

[50] If locations in this setup map into MSAs in the data, this assumption make sense. More specifically, Wolinsky's (1983)

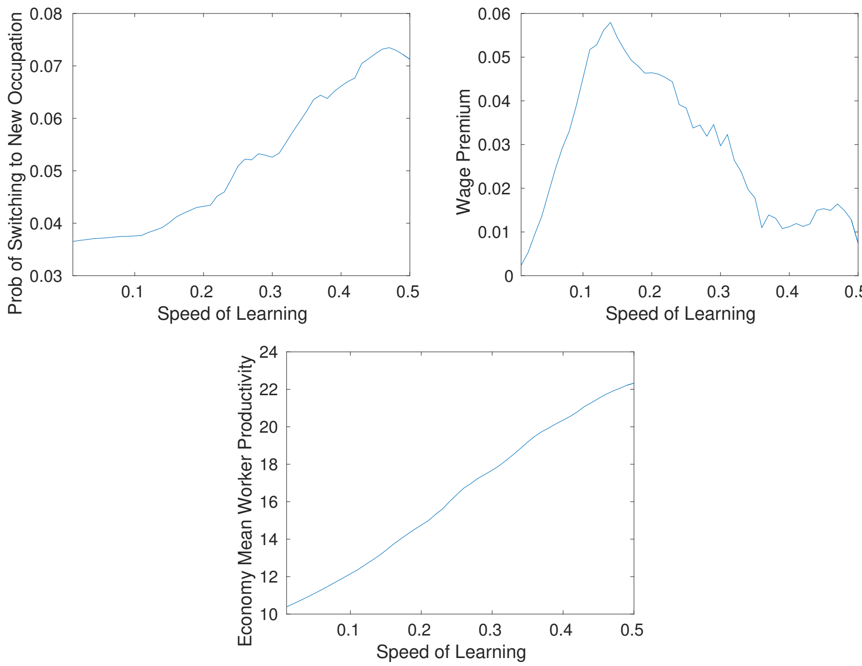

Figure 7: Sensitivity Analysis - Speed of Learning

locations and show that again in equilibrium, cities with larger populations have more occupations.

6.1 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.[51] 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

conditions for a single cluster equilibrium are unlikely to hold here, given the travel costs involved e.g. for a consumer in Baltimore to shop in Austin. However it could be the case that for two MSAs that happen to be located nearby, conditions could be such that only a single cluster emerges for consumers in those two areas.

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

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 $\frac{d^2 z(N_t)}{dN_t^2} > 0$.[52] 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 4, let $\alpha_G > \alpha_B$ and α_l^{ik}, are independently distributed across occupations, cities and workers. Moreover α_l^{ik} is unknown, and let $p_{0l}^{ik} \in (0,1)$ be the worker's prior belief that $\alpha_l^{ik} = \alpha_G$. Priors are drawn independently from a known distribution with support $[0,1]$ and density $g(\cdot)$ when a worker enters a city. To reduce notational congestion 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 an fixed cost of production, f, in terms of the final good. l_k is the total labor input in

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

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.

6.2 Behavior

In what follows I consider a symmetric equilibrium, where all producers choose the same price, b ($b_k = b$ for all k) and commit to it.[53]

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-arm bandit one, as discussed in Section 4.2 the optimal solution 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. As shown in Appendix C, demand for intermediate good k by 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.

Each intermediate good producer chooses a price, b_k, given the demand he faces given in (12).[54]

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

[54] 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.

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-slopping labor supply curve. More specifically, the workers' occupational choice problem dictates the wage level, w_k, required to attract labor input l_k, which is necessary to produce q_k. Each producer takes this into account when choosing a price, b_k.

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

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

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

Free entry of intermediate goods implies that new goods will be created as long as they sustain non-negative profits. In Appendix C I also show that profits, π, are increasing in city population, 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[55]

$$C = \frac{\left(\alpha^G - \alpha^B\right) p^k + \alpha^B}{P\left(m\right)},$$

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

Following the same steps as in Section 4.2, I show that a worker moves to another city when the posterior of all his occupations reaches:

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

[55] In equilibrium there are no profits, π_k. Moreover the wage rate, w_k, which in the symmetric equilibrium is equalized across goods, is normalized to 1.

where

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

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 4 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, $z\left(N\right)$, and also on the number of goods, m.

The endogenous moving decision, as well as the inflow decisions of movers pin down city population, N, in this model. Workers benefit from more occupations because they earn higher wage income due to the increased occupational availability and because they consume a greater variety of products.[56] 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, $z\left(\cdot\right)$, differs across locations, then in equilibrium there will be cities of different sizes. In this setup the standard equilibrium condition that all workers are always indifferent across locations is replaced by the condition that only the workers who move are indifferent.

7 Conclusion

This paper documents a number of facts relating 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 relating to geographical mobility and

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

occupational switching. 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."[57]

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 or not. The third variable identifies one of 93 MSAs (Metropolitan Statistical Areas) and CMSAs (Consolidated Metropolitan Statistical Areas), as defined by the Office of Management and Budget. I also 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 change from one wave to the next.

Table 15 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 datasets (see Moscarini and Thomsson (2007) for estimates from the CPS and Kambourov and Manovskii (2008) for estimates from the PSID). Moreover, 6.78% of the sample moves every year, in line with the estimates from the CPS during the same period (6.72%)[58] 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.[59] Table 16 reports the destination occupations that occupational switchers enter by city size.

[57] http://www.glassdoor.com/blog/find-jobs-small-cities/

[58] 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.

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

Table 15: 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 16: 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.

B 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 four-month switching probability to new occupations is calculated as follows: the four-month occupational switching probability for white males with a college degree is 7.32%. Not all of these however, are switches to new occupations: 30% of workers return to their original occupation within 4 years.[60] This implies an annual rate of "return" switches of approximately 7.5%. In other words, a third of all annual switches are not switches to new occupations. Therefore the four-month switching

http://www.census.gov/hhes/migration/files/cps/p20-531/tab07.txt

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

[60] Kambourov and Manovskii (2008)

probability to *new* occupations is 4.82%.

In my sample I have 7,452 wage observations. In order to calculate the within-occupation 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 within-occupation residual standard deviation is $5.97.

In order to find values for s, δ, c, p_0 and ζ I discretize the setup presented in Section 4 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 (eq. (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 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.

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.

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

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

C Endogenous Occupation Creation Derivations

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

Demand for good k comes from two sources: consumers and producers paying for their fixed cost, f which is in terms of the final good. Solving the producer's problem, implies that the demand for good k by the m producers in that location is given by[61]

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

Therefore total demand for good k is by equation (12).

Producer's k profits are given by

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

[61] 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.

where P is defined equation (13). Substituting in for equation (12), using equation (10) and taking first order conditions leads to equation (14).

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

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

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.

References

[1] Allen T. (2014): "Information Frictions in Trade," *Econometrica,* 82(6), 2041-2083.

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

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

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

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

[6] Baum-Snow, N. and R. Pavan (2013): "Inequality and City Size," *The Review of Economics and Statistics*, 95(5), 1535-1548.

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

[8] Benteley R. (2005): "Publication of JobCentre Plus Vacancy Statistics," Technical Report, Labour Market Trends, June 2005.

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

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

[11] Bolton, P. and C. Harris (1999): "Strategic Experimentation," *Econometrica*, 67(2) 349-374.

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

[13] Campbell, J. Y. and J. F. Cocco (2007): "How Do House Prices Affect Consumption? Evidence from Micro Data," *Journal of Monetary Economics*, 54(3), 591-621.

[14] Carlino G. and W. R. Kerr (2015): "Agglomeration and Innovation," *Handbook of Regional and Urban Economics*, Volumes 5, ed. by G. Duranton, J. V. Henderson and W, C. Strange Elsevier, 349-404.

[15] Carrillo-Tudela, C. and L. Visschers (2014): "Unemployment and Endogenous Reallocation over the Business Cycle," mimeo, University of Essex, Department of Economics

[16] Clarida, R. H. (1992): "Entry, Dumping, and Shakeout," *American Economic Review*, 83(1), 180-202.

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

[18] Davis, D. R. and J. I. Dingel (2016): "A Spatial Knowledge Economy," mimeo, Columbia University, Department of Economics.

[19] De la Roca, J. and D. Puga (2017): "Learning by Working in Big Cities," *Review of Economic Studies*, 84(1), 106-142.

[20] Deming, D. and L. B. Kahn (2017): "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals," *Journal of Labor Economics*, forthcoming.

[21] Diamond, P. (1982): "Aggregate Demand Management in Search Equilibrium," *The Journal of Political Economy*, 90(5), 881-894.

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

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

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

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

[26] Eeckhout, J., R. Pinheiro and K. Schmidheiny (2014): "Spatial Sorting," *Journal of Political Economy*, 122(3), 554-620.

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

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

[29] 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.

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

[31] Glaeser, E. L. (1999): "Learning in Cities," *Journal of Urban Economics*, 46(2), 254-277.

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

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

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

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

[36] Greenwood, M. J. (1997): "Internal Migration in Developed Countries," *Handbook of Population and Family Economics*, Volume 1B, ed. by M. R. Rosenzweig and O. Stark, Elsevier, 647-720.

[37] Groes, F., P. Kircher and I. Manovskii (2015): "The U-Shapes of Occupational Mobility," *Review of Economic Studies,* 82(2), 659-692.

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

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

[40] Hershbein, B. J. and L. B. Kahn (2016): "Do Recessions Accelerate Routine-Based Technological Change? Evidence from Vacancy Postings," Upjohn Institute Working Paper 16-254.

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

[42] Hornstein, A., P. Krusell and G. L. Violante: "Frictional Wage Dispersion in Search Models: A Quantitative Assessment," *American Economic Review*, 101(7), 2873-2898.

[43] Hsu, W.-T. (2012): "Central Place Theory and City Size Distribution," *The Economic Journal*, 122(563), 903-932.

[44] Hu, X. (2005): "Portfolio Choices for Homeowners," *Journal of Urban Economics*, 58(1) 114-136.

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

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

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

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

[49] Kambourov, G. and I. Manovskii (2009a): "Occupational Mobility and Wage Inequality," *Review of Economic Studies*, 76(2), 731-759.

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

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

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

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

[54] Kim, S. (1989): "Labor Specialization and the Extent of the Market," *Journal of Political Economy*, 97(3), 692-705.

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

[56] Kok, S. (2014): "Town and City Jobs: How your Job is Different in Another Location," *Regional Science and Urban Economics*, 49, 58-67.

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

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

[59] Li, W. and R. Yao (2007): "The Life-Cycle Effects of House Price Changes," *Journal of Money, Credit and Banking*, 39(6), 1375-1409.

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

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

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

[63] Lucas, R. E. B. (1997): "Internal Migration in Developing Countries," *Handbook of Population and Family Economics*, Volume 1B, ed. by M. R. Rosenzweig and O. Stark, Elsevier, 721-798.

[64] Machin A. (2003): "The Vacancy Survey: A New Series of National Statistics," Technical Report, Labour Market Trends, July 2003.

[65] McCall, B. P. (1990): "Occupational Matching: A Test of Sorts," *Journal of Political Economy, 98(1), 45-69.*

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

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

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

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

[70] Papageorgiou, T. (2014): "Learning Your Comparative Advantages," *Review of Economic Studies,* 81(3), 1263-1295.

[71] Pastorino, E. (2014): "Careers in Firms: Estimating a Model of Job Assignment, Learning and Human Capital Acquisition," Mimeo, University of Minnesota, Department of Economics.

[72] Petrongolo, B. and C. Pissarides (2006): "Scale Effects in Markets with Search," *The Economic Journal*, 116(1), 21-44.

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

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

[75] Ruggles, S. K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015): Integrated Public Use Microdata Series: Version 6.0 [dataset]. Minneapolis: University of Minnesota.

[76] Schiff, N. (2015): “Cities and Product Variety: Evidence from Restaurants,” *Journal of Economic Geography,* 15(6), 1085-1124.

[77] Silos P. and E. Smith (2015): “Human Capital Portfolios,” *Review of Economic Dynamics,* 18(3), 635-652.

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

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

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

[81] Wolinsky A. (1983): “Retail Trade Concentration due to Consumers’ Imperfect Information,” *The Bell Journal of Economics*, 14(1), 275-282.