

# Understanding the City Size Wage Gap\*

Nathaniel Baum-Snow, Brown University  
Ronni Pavan, University of Rochester

July, 2009

## Abstract

In 2000, wages of full time full year workers were more than 30 percent higher in metropolitan areas of over 1.5 million people than rural areas. The monotonic relationship between wages and city size is robust to controls for age, schooling and labor market experience. In this paper, we decompose the city size wage gap into various components. We propose a labor market search model that incorporates endogenous migration between large, medium and small cities. This model is sufficiently rich to allow for recovery of the underlying ability distributions of workers by city size, arrival rates of job offers by ability and location, and returns to experience by ability and location, when structurally estimated using longitudinal data. Estimates from the structural model facilitate a more complete empirical decomposition of the city size wage gap than is possible using results in existing research. Counterfactual simulations of the model indicate that variation in returns to experience and differences in wage intercepts across location type are the most important mechanisms contributing to the overall city size wage premium. Differences in wage intercepts generate the largest part of the city size wage premium for high school graduates while differences in returns to experience are more important for college graduates. Sorting on unobserved ability within education group and differences in labor market search frictions contribute little to observed city size wage premia.

---

\*The authors gratefully acknowledge financial support for this research from National Science Foundation Award SES - 0720763. Mike Kuklik and Cemal Arbatli provided excellent research assistance. We have benefited from comments in talks at Ca' Foscari University of Venice, Purdue, Syracuse University, Washington University, the University of Pennsylvania, the University of Wisconsin, Yale and at the SED and Econometric Society North American meetings. We thank Emek Basker for generously providing us with ACCRA price index data.

# 1 Introduction

It is widely documented that wages are higher in larger cities. In the 2000 census, average hourly wages of white prime-age men working full-time and full-year were 32 percent higher in metropolitan areas of over 1.5 million people than in rural areas. The relationship between wages and population is monotonically increasing by about 1 percentage point for each additional 100 thousand in population over the full range of metropolitan area size. This monotonic relationship is robust to controls for age, schooling and labor market experience. In addition, it has become considerably steeper since 1980 when large metropolitan areas had wages that were 23 percent greater than rural areas.

In this paper, we investigate the causes of the city size wage gap. In particular, we propose a unified framework for empirically investigating the extent to which selection on latent ability, firm-worker matching, returns to experience and level effects can account for observed differences in wages between cities of different sizes. Our analysis utilizes a model of on-the-job search that incorporates endogenous migration between small, medium and large cities. This model is rich enough to allow for recovery of the underlying ability distributions of workers by location, arrival rates of job offers by ability and location, and returns to experience by ability and location, when structurally estimated by maximum likelihood using longitudinal data. Our estimates facilitate a more complete empirical decomposition of the city size wage gap than is possible using results in existing research. As in Pavan (2009), we employ a non-Gaussian state space approach to construct the likelihood function, allowing for inclusion of an unobserved firm-worker match component of the wage process in the model.

Table 1 summarizes patterns in city size wage premia using census data from 1980, 1990 and 2000. Table 1 presents results from regressing the log hourly wage for full-time full-year white men on indicators for living in metropolitan areas of 250 thousand to 1.5 million and more than 1.5 million. The excluded category includes small metropolitan areas and rural areas. Results in Panel A Specification 1 show that the city size wage premium is monotonic and has been increasing over time. Specification 2 shows that in each year about one-third of the two estimated size premia can be explained with observables.

Counterfactual simulations of our structural model indicate that variation in wage intercepts and returns to experience across location type are the most important mechanisms contributing to the overall city size productivity premium. Both of these mechanisms are important for high school and college graduates throughout the city size distribution,

though differences in wage intercepts across location categories are most important in medium sized cities for high school graduates. However, sorting on unobserved ability within education group and differences in labor market search frictions contribute little to observed city size wage premia. That is, city size wage premia are slightly lower than they would be if search frictions and the distribution of unobserved ability were equal in all locations. These patterns are particularly true for high school graduates.

Our results indicating a lack of importance of matching for generating city size marginal labor productivity premia are consistent with those of Petrongolo & Pissarides (2003) who find no evidence of higher job arrival rates in larger British labor markets despite the higher wage offers in these markets. Furthermore, as discussed in Petrongolo & Pissarides (2001), most empirical evidence on aggregate matching functions indicates that they exhibit constant returns to scale. Our evidence on sorting is consistent with that of Fu & Ross (2008) who also find mild negative sorting of workers on unobserved skill into larger cities.<sup>1</sup> Like Glaeser & Mare (2001), we find positive sorting on observed skill. Our results on the importance of differences in returns to experience for generating city size wage premia are also consistent with those of Gould (2007) and Glaeser & Mare (2001). Our results on level effects generating city size wage premia are consistent with empirical evidence on human capital spillovers by Rauch (1993), Acemoglu & Angrist (2000), Moretti (2004a) and Greenstone, Hornbeck & Moretti (2009).

Roback's (1982) model forms a natural starting point for conceptualizing how wage gaps can persist between cities. Its basic insight is that in order for there to be no incentive for workers to migrate between cities, wages plus the value of amenities minus cost of living must be equalized everywhere for individuals with identical endowments and preferences. Thus, differences in wages adjusted for cost of living differences across cities must solely represent amenity differences for equally productive individuals. Therefore, to the extent that larger cities have more valuable consumer amenities, they should actually exhibit lower wages than small cities for similar workers. Glaeser, Kolko & Saiz (2001) dub this the "Consumer City" phenomenon. Lee (2006) provides evidence that similar workers in the medical professions do indeed earn lower wages in larger cities while Albouy (2009) provides evidence of no strong relationship between city size and quality of life.

A location equilibrium also requires that similar firms have the same profits wherever they are located. Therefore, if nominal wages and rents are higher in cities, productivity

---

<sup>1</sup>Gould (2007) finds positive sorting on unobserved skill but estimates his model on a sample that includes both high school and college graduates.

for firms producing traded goods must also be higher to compensate. Otherwise, firms would move to smaller places in search of cheaper labor and rents. This logic reinforces the intuition from the workers' decision that the city size wage gap implies higher worker productivity in larger cities. Examination of wages that are not deflated for cost of living differences thus directly reveals locations where workers are more productive.

Why are larger cities more productive? Two broad explanations exist. More productive workers may be concentrated in larger cities and/or agglomeration economies may make identical workers more productive in cities<sup>2</sup>. A considerable amount of empirical evidence supports both explanations. In the 2000 census, 41 percent of prime-age white males living in metropolitan areas of over 2.5 million people were college graduates relative to just 20 percent of those living in rural areas, with the fraction monotonically increasing in city size. Glaeser & Maré (2001) argue that sorting on human capital levels accounts for about one-third of the city-size wage gap in the United States. Combes, Duranton & Gobillon (2008) demonstrate using French data that up to half of the wage disparity across French cities can be accounted for by skill differences in their working populations, as captured by individual fixed effects from panel data. Our evidence in Tables 1 and 2 is that these estimates are reasonable.

Several important studies provide evidence supporting the existence of agglomeration economies in cities. Henderson, Kuncoro & Turner (1995) show that firms in several manufacturing industries are more productive when they are located in the same metropolitan area as other firms in the same industry. This phenomenon is known as "localization economies". These authors also provide evidence for cross-industry agglomeration forces, or "urbanization economies", for some new industries. Glaeser et al. (1992) also provide empirical evidence on the existence of "Jacobs" urbanization economies. Ciccone and Hall (1996) find a positive relationship between employment density and productivity at the county level.

While there is fairly conclusive evidence that similar workers are more productive in bigger cities, there is less empirical evidence on the relative importance of the different mechanisms that may generate this productivity difference. Duranton & Puga (2004) review many of the existing micro-founded theories explaining aggregate agglomeration economies. They break up explanations into three broad categories: sharing, matching and learning. Glaeser & Maré (2001) show that wage growth is faster in larger cities and

---

<sup>2</sup>In addition, Manning (2008) proposes monopsonistic labor markets in small cities as a mechanism that could generate a city size wage premium.

that high wages persist for migrants. From this they conclude that larger cities speed human capital accumulation, or that "learning" is important. Moretti (2004) supports this view, providing evidence that human capital spillovers exist from cities to industrial plants located within them. Evidence in support of localized agglomeration economies within industries by Arzhagi and Henderson (2008) and Rosenthal & Strange (2003) for example, indicates that input sharing may also be important. However, other than Petrongolo & Pissarides (2003), there is little direct empirical evidence on the importance of "matching", or differences in search frictions across locations, for generating agglomeration economies. Furthermore, existing evidence in this regard either does not directly measure search frictions or indicates that matching is not more efficient in larger agglomerations. Rosenthal & Strange (2004) exhaustively review the empirical literature on agglomeration economies.

One theme that recurs in much of the empirical work on agglomeration economies is that schooling and city size are complements. Why do highly educated people choose to live in cities? Shapiro (2006) demonstrates that employment growth is higher in better educated cities. He provides evidence that about 60 percent of this employment growth can be attributed to associated productivity enhancements while the remainder is because of improvements in the quality of life associated with skilled cities. Glaeser & Saiz (2003) argue that skilled cities' success comes in part from their ability to better weather economic shocks. Carlino, Chatterjee & Hunt (2007) demonstrate that the patenting rate is increasing in city size, indicating that cities may make higher ability individuals more innovative. Finally, Costa & Kahn (2000) argue that job matching for "power couples" is a force pushing the more highly educated to larger cities.

While convincing evidence exists that larger cities are more productive, existing empirical work on the topic has several limitations. One difficulty for researchers is the likely endogenous sorting on unobserved skill that exists across cities. A common procedure for estimating productivity differences between cities essentially examines the relative wages of migrants in big and small cities. However, there is little reason to believe that migrants are representative of the population as a whole, even when conditioning on observables. An additional limitation exists on the types of mechanisms that can be examined independently. A lack of exogenous variation at many margins has made it difficult to differentiate between various explanations for agglomeration economies using standard regression procedures. This has led many studies to argue for the relative importance of one theory versus another based on descriptions of equilibrium outcomes rather than on evidence from natural experiments. Further, even if exogenous variation can be found, it generally

only occurs on one margin at once, thereby making it difficult to understand potential interactions between different mechanisms.

In this paper, we attempt to fill some of these holes in the literature by specifying and estimating a dynamic model of job search that incorporates many of the elements listed above. The model laid out in Section 4 allows for endogenous migration, unemployment, and job changing decisions. The model is parameterized such that skill differences may imply very different patterns of behavior as a function of underlying parameters and allows for econometric recovery of its deep parameters. The model is sufficiently flexible to allow for separate identification of amenity values, job arrival rates and returns to experience that vary as a function of interactions between underlying unobserved skill of individuals and city size.

Estimated parameters from the model allow for a decomposition of the observed city-size wage gap into four components that all potentially interact: 1) sorting by ability across cities, 2) differences in arrival rates of job offers and job separation rates ("matching") across cities and abilities, 3) wage level effects ("sharing") across cities and abilities and 4) different returns to experience ("learning") across cities and abilities. As part of the exercise, like Kennan & Walker (2009), we also estimate moving cost parameters. Longitudinal data from the work history file of the NLSY with restricted use geocodes allows us to evaluate the relative importance of these explanations for generating city size wage gaps in the United States.

The methodological approach presented below is similar to that used by Gould (2007) to examine the importance of ability sorting in generating the urban wage premium. Gould's estimation of a dynamic model with endogenous migration between urban and rural locations indicates that selective migration of high ability workers is an important force behind the urban productivity premium that gets amplified by steeper experience profiles in urban areas. Our results in this dimension are similar, though our greater sample disaggregation allows us to estimate this selective migration conditional on observed skill. In addition, our paper goes beyond Gould's analysis by incorporating job search in the dynamic model, allowing us to measure whether differences in matching technologies across locations explain part of the wage premia. Our examination of three location size categories rather than two is also important given our observations of striking non-monotonicities in several dimensions with respect to city size. Finally, our accounting for cost of living and amenity differences across locations additionally facilitates a more complete analysis than in existing research as it allows us to more accurately pin down wage level effects across

locations.

We should emphasize that the model specified in this paper is partial equilibrium in nature. That is, firm location is taken as given. Part of the city size productivity gap may come from selection of more productive firms into larger cities. Ellison & Glaeser (1997) document that firms systematically locate in ways that generate industrial agglomerations, though Combes et al. (2009) show that once accounting for agglomeration forces, firm productivity distributions exhibit no statistically significant truncation in larger cities as would be predicted by a firm selection model across markets like Syverson's (2004). Nevertheless, to the extent that some industries are more productive than others, the pattern documented by Ellison & Glaeser (1997) implies that more productive firms may also systematically locate in larger cities. While the framework developed in this paper has little to say about the process that might generate such firm selection, it still allows us to learn much about why cities are more productive. If input costs are higher in cities, it would be difficult for a general equilibrium model to justify the selective location of productive firms producing traded goods to larger cities without their workers also being more productive. Therefore, understanding why workers earn more in larger cities is still informative about why larger cities are more productive.

The next section describes some relevant patterns in the data. Section 3 discusses data construction. Section 4 presents the model. Section 5 discusses how we estimate the model. Section 6 presents the results and various decompositions of the city size wage gap using counterfactual simulations. Finally, Section 7 concludes.

## 2 Empirical Observations

The city size wage premium shows up pervasively in the data. In this section, we present evidence on the existence of the city size wage gap using data from the National Longitudinal Surveys of Youth 1979 (NLSY). We then summarize patterns in wage growth, job transitions, unemployment and human capital as functions of experience and city size. While some of the patterns presented in this section resemble those already documented in the studies cited above, we additionally demonstrate that accounting for differences in cost of living, unemployment spells and job turnover across locations are important inputs to gaining a more complete understanding of patterns in the data. These descriptive results are consistent with agglomeration economies, ability sorting and compensating differentials for higher amenities in large cities all operating simultaneously.

Table 2 Panel A presents estimates of city size wage premia using data from the NLSY. Magnitudes of city size wage premia estimated using the NLSY are very similar to those from the 1990 and 2000 censuses. The wage premium for medium sized cities is estimated at 20 percent while that for large cities is 30 percent, indicating that even though it only includes young adults in 1979, the NLSY is a reasonable data set with which to evaluate reasons for the city size wage premium. Controlling for education and quadratics in age and work experience reduces these coefficients to 0.15 and 0.23 respectively. Controlling for individual fixed effects additionally reduces these coefficients to 0.08 and 0.14. That the inclusion of individual fixed effects generates larger reductions in the city size wage gradient than individual controls may indicate at first glance that positive sorting on unobservable skill is an important component of the city size wage premium. However, the fixed effects results should be interpreted with caution given that the city size coefficients are identified off of movers who are unlikely to be a random sample of the population conditional on controls. Indeed, a simple Roy (1951) model, as developed by Borjas (1987) for example, indicates that movers are predicted to differ on unobservables from the overall population. Evidence presented below indicates that more educated individuals are more mobile. This is an example of why it is fruitful to appeal to the full structural model as we do in the next section.

Table 2 Panel B presents analogous regression results when wages are adjusted for cost of living differences across metropolitan areas.<sup>3</sup> Estimates in Panel B reflect the fact that cost of living in large metropolitan areas is much higher than that in smaller places. All specifications in Panel B exhibit an inverse U shaped pattern. Real wages in medium sized metropolitan areas are the highest, even when controlling for observables and fixed effects. Commensurate with the discussion in the previous section, this evidence is consistent with medium sized cities having the lowest levels of consumer amenities such that individuals are willing to take a 4 percent wage cut to live in large metropolitan areas over medium sized metropolitan areas. While this implied amenity value of large over medium cities is very stable across specifications, estimates of that for small cities are more heterogeneous. Controls for observables reduce the estimated relative amenity value of small places over medium sized cities from 16 to 10 percent while inclusion of individual fixed effects reduces it another 4 percentage points. This inverse U profile of wages adjusted for cost of living seen in Table 2 Panel B persists at most levels of experience. This pattern is evidence that

---

<sup>3</sup>The next section details how we implement this cost of living adjustment.

the relative amenity values of cities do not change much over the life-cycle. Our structural estimates presented in Section 6 indicate that most groups assign medium sized cities the lowest amenity value, followed by large cities then rural areas, results consistent with these reduced form observations.

The results in Table 2 exhibit several features that should invite consideration in any analysis of the city size wage premium. First, adjustment for cost of living is crucial for understanding compensating wage differentials between the largest cities and other locations. Results after this adjustment indicate that using information on workers to understand sources of the city size wage premium requires a model with endogenous migration between at least three size categories. Second, while controlling for observables reduces the premium, doing so does not eliminate it. Therefore, there is significantly more to be learned beyond endogenous sorting on observables about the reasons for city size productivity premia. To this end, Tables 3 to 6 examine various outcomes as a function of labor market experience using the NLSY data.

Table 3 examines wage growth over the first 1, 5, 10 and 15 years of labor market experience. To make the sample as consistent as possible across experience categories, we restrict it to include only those for whom we observe at least 15 years of labor market experience. This represents 80 percent of the NLSY sample described in more detail in the next section. (Only 48 percent of the sample survives to at least 20 years of labor market experience.) The left side of Table 3 shows wages only deflated by the CPI and the right side shows wages deflated over both time and space.

Table 3 Panel A shows that workers in the largest cities saw wage growth of 76 percent on average relative to just 52 percent in the smallest areas over the first 15 years of experience. Panels B and C show that both college and high school graduate subsamples exhibit the same overall pattern of wage growth, though high school graduates have slower growth in all locations. Accounting for cost of living differences does not have much of an effect on wage growth profiles.

As we discuss above, one potential explanation for the city size wage growth premium is that faster turnover generates more efficient firm-worker matches in larger cities. Table 4 describes patterns in job turnover and weeks of unemployment as functions of experience, city size and education. The left side of Table 4 shows that the mean number of jobs held is roughly constant or decreasing in city size at every experience level, especially for the high school sample. College graduates held about 2 fewer jobs by 15 years of experience than did high school graduates and had much a much flatter profile of jobs held with respect to

city size. This is descriptive evidence that job offer arrival rates of employed individuals are unlikely to generate more efficient matching in larger cities. However, the right side of Table 4 shows that the mean number of weeks of unemployment is strongly decreasing in city size for high school graduates while this profile is not monotonic and exhibits much lower levels for the college sample. This pattern indicates that separation rates and/or job offer arrival rates for the unemployed may indeed generate smaller search frictions in larger cities for high school graduates.

A competing potential explanation for the city size wage growth premium is that more skilled individuals systematically migrate to larger cities over the life-cycle. This phenomenon is certainly evident for observable skill. College graduation rates and years of schooling are increasing in city size at all levels of experience. Furthermore, both measures of the human capital gap widen with experience.<sup>4</sup> Table 5 presents evidence of such selective migration. It presents transition matrices between city size categories for the full sample and the high school and college subsamples. Panel A shows that high school graduates are disproportionately located in smaller cities. Panel C shows that under 20 percent of high school graduates move between city sizes during their first 15 years in the labor force. Those that do move out of medium and large cities are more likely to move to small cities and rural areas than medium sized cities. Those who move out of small places are more likely to move to medium sized cities than large cities. In contrast, Panel B shows that college graduates are more mobile and exhibit migration patterns that are more oriented toward larger cities. 36 percent of college graduates entering the labor force in small places move compared to just 13 percent of high school graduates. Of the 24 percent who move out of medium sized cities, more than half migrate to large cities. Of the 26 percent who migrate out of large cities, about two-thirds move to medium sized cities. These differences in migration patterns as a function of education indicate the utility of estimating parameters of the structural model separately by education. Indeed, as we discuss in Section 6, we estimate very different structural parameters for the two groups.<sup>5</sup>

---

<sup>4</sup>Analogous results using contemporaneous education reveals that the human capital growth as a function of experience is more rapid in larger cities, indicating that post labor force entry educational attainment is more rapid in larger cities.

<sup>5</sup>Similar though more pronounced patterns are seen in transition matrices between residential location at age 14 and labor force entry. Among college graduates, 39 percent move to work their first jobs. Their mobility rate out of small locations is more than 20 percentage points higher than out of medium and large locations. Among high school graduates, 18 percent move to take their first jobs with mobility rates out of all three location types with 8 percentage points of each other and non-monotonic in city size.

Table 6 presents decompositions of the mean log wage growth numbers by city size for 15 years of experience reported in Table 3 into four components: within job, between jobs with no unemployment in between, between jobs when individuals experience an unemployment spell in between and unknown. The unknown category consists of wage growth that occurred between jobs sandwiched by a third job for which we have no wage information. Reported values are means across all individuals in each city size-experience cell. Regardless of how wages are deflated and for both subsamples, within job wage growth is increasing in city size whereas between job wage growth is an inverse U in city size. The job to unemployment component of wage growth is small and negative in small and medium sized cities and 0 in large cities. Therefore, the bulk of the faster wage growth rates in larger cities comes from steeper tenure profiles. Table 6 provides no evidence that differences in search frictions are important for generating city size wage growth premia.

Consistent with other studies, the descriptive evidence in Tables 3 to 6 points to systematic differences in observed skill levels as one important driver of city size productivity differences. In addition, attributes of larger cities that contribute to human capital accumulation appear important for generating higher wages in larger cities. However, we find scant evidence that differences in search frictions for the employed generate much of the city size wage premium, though high school graduates do spend less time unemployed in larger cities. It is hard to know, however, the importance of selection on unobserved skill for generating these patterns. Estimation of the model specified in Section 4 therefore facilitates an improved understanding of the mechanisms behind the city size productivity premium.

## 3 Data

### 3.1 NLSY Sample and Data Construction

The primary data set used for the analysis is the National Longitudinal Survey of Youth (NLSY) 1979 restricted use geocoded and work history files. With this data set, we construct information on jobs, unemployment, wages and migration patterns for a sample of young white men ages 14 to 21 on December 31st, 1978 from the time of their entry into the labor force until 2004 or their attrition from the survey. The sample includes 1,758 men from the NLSY79 random sample of 3,003 men. We lose 20 percent of the full sample because they entered the labor force before we observe their initial attachment.

An additional 12 percent of individuals are dropped because they were in the military at some point, never entered the labor market, dropped out of the labor market for at least 4 contiguous years, or had significant missing job history data. The remaining individuals excluded from the sample are nonwhites.

We sample the weekly job history data four times every year for those who become attached to the labor force after January 1st, 1978. Sampled weeks always include the annual interview date (which varies) and the seventh week of the three remaining quarters. Our goal is to sample often enough such that we capture all job and location changes but not so often such that numerical maximization of the likelihood function implied by our structural model is computationally infeasible. Given the number of individuals in the NLSY, quarterly sampling maximizes the number of job and location transitions observed under the constraint that the likelihood function is computable in a reasonable period of time. Wages are only observed on interview dates and in the last observation on each job.<sup>6</sup> In Section 5 we discuss how we deal with missing wage data econometrically. We keep track of the number of weeks in each unemployment spell that occurs in between sampled weeks. Individuals enter the sample when they begin working full-time, providing a convenient initial condition for the likelihood function. We define full-time as working at least 300 hours per quarter if not in school and at least 500 hours per quarter if in school.

We assign individuals to locations based on reported state and county of residence, which is available on interview dates and between interviews during the periods 1978-1982 and 2000-2004 only. We assign most location observations in remaining quarters by assuming that individuals must remain at one location for the duration of each job. We impose that unemployed individuals must remain at the same location as the last job held.<sup>7</sup> Those jobs with multiple reported locations are assigned to the modally reported location. Jobs with multiple modes are assigned the modal location that occurred latest in time.<sup>8</sup> This leaves 5 percent of quarterly observations with no location information. Sixty percent of these observations are for jobs sandwiched between two other jobs at the same location. In these cases, we assume that individuals did not move. For the remaining 2 percent of

---

<sup>6</sup>More precisely, up until 1993 we observe wages on up to 5 jobs per year. After 1993, we only observe wages of the jobs most recently worked prior to interview dates, which occur about every two years.

<sup>7</sup>In the vast majority of cases, we do not observe the location at which individuals are unemployed. The model specified in the next section assumes that individuals cannot move to a new location to go unemployed.

<sup>8</sup>The model specified in the next section imposes that workers must remain at one location throughout each job and that the unemployed remain at their previous work location.

the sample, we impute locations to be that of the first job after the unobserved location spell for which we observe location.<sup>9</sup>

For the purpose of assigning locations into size categories, we use metropolitan area definitions from county agglomerations specified in 1999 but assign them into size categories based on aggregated component county populations in 1980. We select the three size categories used throughout the paper such that the sample is split roughly into thirds.

Evidence in the previous section shows that it is important to allow potential mechanisms behind the city size wage premium to differ by worker skill. As such, we estimate the model specified in the next section separately for those achieving high school graduation only and those with a college education or more. In the high school sample we have 50,665 observations on 675 individuals. In the college sample, we have 42,334 observations on 586 individuals. We observe a wage in about one-quarter of the observations.

### 3.2 Spatial Price Index

Using wages and migration patterns to understand productivity differences across cities requires accounting for cost of living differences across space and time. We denote the exogenously given price of good  $i$  in time/location  $j$  as  $p_i^j$ . We assume that consumers have Cobb-Douglas utility over  $I$  goods, meaning expenditure shares for each good are the same in each location. Our price index measures the relative expenditure required across time and locations to hold utility constant given observed price differences. The expenditure function in every location  $j$  must thus achieve the same level of utility  $U_0$  as that achieved in an arbitrarily chosen base time period and location 0.

$$\sum_i \alpha_i \ln\left(\frac{\alpha_i E^j}{p_i^j}\right) = \ln(U_0) \quad (1)$$

Equating utility in time/locations  $j$  and 0, we obtain the ideal index relating prices in time/location  $j$  to those in 0, capturing the percent increase in expenditure required to keep an individual at the same utility.

$$INDEX_j = \frac{E^j}{E^0} = \exp \sum_i \alpha_i \ln(p_i^j/p_i^0) = \prod_i (p_i^j/p_i^0)^{\alpha_i} \quad (2)$$

---

<sup>9</sup>We re-estimate the model using location of the previous job instead and results are very similar.

This is the index we use to deflate wages across locations and over time.<sup>10</sup>

Building this index requires price data by time and location for different goods and information on expenditure shares. We get prices by location from the American Chamber of Commerce Research Association (ACCRA) data sets from 2000 to 2002. These data report prices in six broad expenditure categories for most metropolitan areas and some rural counties nationwide. When possible, we take data from 2001. For the few regions not sampled in 2001 we take data from either 2000 or 2002. ACCRA reports provide us with price data for 244 metropolitan areas and 179 rural counties.<sup>11</sup> We impute price data for remaining areas as follows. Metropolitan counties are assigned the average prices from other MSAs in the same size category and state when possible. If there are none other of the same size in their state, we impute using data from MSAs of the same size by census division. Price data for rural counties are imputed analogously.

For time series variation in prices, we use regional and metropolitan price index data from the BLS disaggregated into the same six categories used for the ACCRA data. We assign each county to be represented by the most geographically specific index possible in each year. Together, the ACCRA and regional CPI data allow us to calculate the relative price in each expenditure category for location/time period  $j$  relative to the base location/time period. The base time/location we define as the average ACCRA location from 2001 but deflated to be index value 100 in 1999.

Rather than take expenditure shares  $\alpha_i$  directly from the CPI-U, we build expenditure shares for households including white men working full time using data from the biannual Consumer Expenditure Surveys (CEX) starting in 1982. We build shares directly from the CEX in order to best capture preferences of those in our sample and because the weights used for the CPI-U sometimes fluctuate significantly from year to year. We found

---

<sup>10</sup>Albouy (2009) uses a more general methodology to account for cost of living differences across locations. Starting with a framework that includes federal taxes and nonlabor income, Albouy differentiates a generalized spatial indifference condition like that given by (1) to generate a deflated log wage in location  $j$  of  $\ln(w_j) - \frac{1-\delta\tau}{1-\tau} \frac{s_p}{s_w} \ln(p_j)$  where  $\delta$  is a tax deduction parameter,  $\tau$  is the average federal income tax rate,  $s_w$  is the wage share of income,  $s_p$  is the share of income spent on local goods and  $p_j$  is the price of nontraded goods in location  $j$ . As is typical in the literature, of which Carrillo et al. (2009) provides an extensive overview, our index accounts for  $s_p$  only, implying that we may understate real price differences across locations. Because the  $\frac{1-\delta\tau}{(1-\tau)s_w}$  adjustment would have to be different for each example of migration in our data, young workers are likely to face low federal taxes and have  $s_w$  near 1, and to be consistent with the specification of our model presented in the next section, we do not make this additional adjustment. Its exclusion should if anything only lead the structural amenity parameters to be estimated incorrectly.

<sup>11</sup>ACCRA reports prices separately for different counties within some large metropolitan areas. In these cases, we allow our price index to differ accordingly within MSA. Otherwise, we assign the ACCRA reported prices to all counties in a given MSA.

that expenditure shares implied by the CEX are within a few percentage points across education groups and city sizes. As such, we prefer to use the sample from the CEX that best matches our full sample to calculate one set of expenditure weights that we apply to all studied individuals.

## 4 The Model

The model described in this section is specified to be simple enough to be tractably estimated yet sufficiently rich to capture all of the potential explanations for the city size wage and productivity gaps discussed in the introduction. We specify a "finite mixture" model, meaning that we have a finite number of latent agent types by which some parameters of interest are indexed.<sup>12</sup> Our most constraining simplifying assumptions limit the number of these underlying worker types to two and city size categories to three. Though it would be possible to expand the number of both objects, our specification allows for simpler interpretation of estimated parameters and simulations of the estimated model. In addition, our specification facilitates computational tractability.

Individuals derive utility from the sum of a location and type specific amenity and their log wage or unemployment benefit. The different types of locations, characterized by different population size categories, are denoted with subscripts  $j \in \{1, 2, 3\}$  respectively. We denote "ability" levels as  $h_i \in \{h_L, h_H\}$ . These are intended to capture underlying productivity differences between workers either from innate ability or because of different amounts of human capital accumulation prior to entrance into the labor market. We allow the probability that a given worker is of type  $i$  to depend on the location in which he enters the labor market.

The observed log wage depends on the worker's ability, labor market experience in each location type, a firm-specific stochastic component and classical measurement error. The returns to experience and the individual specific intercepts are functions of worker type. The firm-specific stochastic component of the wage  $\varepsilon$  is drawn from a distribution of productivities from which workers sample when they receive a job offer. This distribution  $F_{\varepsilon_j}(\varepsilon)$  differs by location and is taken as exogenous. The unexplained component of the wage, which can be thought of as a measurement error term, is independent across individuals and time and is drawn from the distribution  $F_u(u)$ . Put together, we parameterize

---

<sup>12</sup>Finite mixture models are widely used in the structural estimation literature. Heckman and Singer (1984) and Keane and Wolpin (1997) are two notable examples of studies using this approach.

the wage process of an individual working in location type  $j$  and having experience from location types indexed by  $k$  as follows:

$$\ln w_j(h, x_1, x_2, x_3, \varepsilon, u) = \varepsilon + \beta_0^j(h) + \sum_{k=1}^3 \theta_j \beta_1^k(h) x_k + \beta_2^k \left( \sum_{k=1}^3 x_k \right)^2 + \beta_3 \left( \sum_{k=1}^3 x_k \right)^3 + u. \quad (3)$$

The wage process expressed in Equation (3) captures the extent to which sorting on ability may influence wage growth and levels differently for small and large cities. The unit of time is a quarter. We denote experience at time  $t + 1$  for an individual working at location  $j$  and time  $t$  by  $x'_j = x_j + 1$ , while experience in each other location type remains constant. Individuals accrue experience at the beginning of each working period. We assume that an individual works for 160 periods (40 years) and then retires with a pension equal to the last wage. After retiring he will live for an additional 80 periods (20 years).<sup>13</sup>

We allow the job search technology to differ by city size, ability and employment status. We denote the arrival rates of job offers from the same location to be  $\lambda_j^u(h)$  for unemployed workers and  $\lambda_j(h)$  for employed workers, where  $j$  is the worker's location. The arrival rates of job offers from different locations are  $\lambda_{jj'}^u(h)$  for unemployed workers and  $\lambda_{jj'}(h)$  for employed workers, where  $j'$  is the location of the job offer. We allow job arrival rates for the city of residence and other cities of the same size to differ. For analytical simplicity, we assume that individuals may only receive one job offer each period. Workers who choose to switch jobs at the same location must pay a stochastic switching cost  $v_S$  with zero mean and finite variance. This cost potentially captures differences in non-pecuniary benefits across jobs that might lead workers to accept wage cuts. Exogenous separation rates  $\delta_j(h)$  similarly depend on location and ability. With a job offer at location  $j'$ , individuals have the option to move and pay a one-time cost of  $C_M + v_M$ , where  $v_M$  is a random component with zero mean and finite variance. To keep the model simple and because we only observe the location of unemployment in at most 1 week per year, we assume that all unemployment occurs in the same location as the previous job.

We denote the value of being unemployed at location  $j$  as  $V_j^{UN}$  and the value of holding a job with match quality  $\varepsilon$  at location  $j$  as  $V_j^{WK}(\varepsilon)$ . The state spaces of all value functions that we discuss contain individual specific "ability"  $h$  and experience in all location

---

<sup>13</sup>In order to reduce the computational burden we assume that the unit of time is a quarter for the first 60 quarters while it is a year for the remaining 25 years of working life. The likelihood function is computed only using information on the first 15 years of labor force participation for each individual. More than 75% of observations and 90% of location changes in our data happen during this period.

types  $x_1, x_2, x_3$ . For expositional simplicity we suppress this dependence in the notation. Current utility of an unemployed worker depends additively on the amenity  $\alpha_j(h)$ , the unemployment benefit  $b$  and on an *i.i.d.* preference shock  $v_U$  with zero mean and finite variance which shows up below in (4) and (5). This shock captures the random component of the disutility of work. Current utility of workers is linear in the amenity and their log wage each period net of measurement error.<sup>14</sup> Given these definitions, the deterministic component of the present value of being unemployed and the present value of working are given by the following expressions:

$$\begin{aligned} V_j^{UN} &= \alpha_j(h) + b + \beta^{\frac{1}{3}} V_j^u, \\ V_j^{WK}(\varepsilon) &= \alpha_j(h) + \ln w(\varepsilon) + \beta V_j(\varepsilon). \end{aligned}$$

Individuals receive their flow utility at the beginning of each period. At the end of each period, available options and the value of shocks for the following period are revealed and individuals make job transition and/or migration decisions. Because doing so does not significantly affect computation time, we take advantage of the high frequency of the job history data and index time in months for the unemployed. For this reason, the unemployed agent discounts utility by  $\beta^{\frac{1}{3}}$ , whereas the employed worker discounts by  $\beta$  to represent quarters. The expressions above are of use for clarity of exposition and notational convenience. The key elements of interest are  $V_j^u$  and  $V_j$  which we specify next.

We first consider the environment for an unemployed individual at location  $j$  with  $X < 160$ . At the beginning of each period, the agent observes whether he is faced with one of 5 possible scenarios. With probability  $\left(1 - \lambda_j^u(h) - \sum_{j'=1}^3 \lambda_{jj'}^u(h)\right)$  he does not receive a job offer, with probability  $\lambda_j^u(h)$  he receives a job offer from the same location, and with probability  $\lambda_{jj'}^u(h)$  he receives an offer in location type  $j'$ .<sup>15</sup> The individual decides at the beginning of each period whether to accept a potential job offer or remain unemployed. If he accepts an offer, he pays a cost to move to the relevant new location if necessary.

---

<sup>14</sup>We impose that individuals cannot transfer wealth across periods and unlike Gemici (2008) that households are unitary.

<sup>15</sup>The parameter  $\lambda_{jj'}^u(h)$  represents the probability that an unemployed individual at location  $j$  receives a wage offer in a different city also of size category  $j$  whereas the parameter  $\lambda_j^u(h)$  represents the probability that this individual receives a wage offer in the same city as his last job.

Equation (4) shows the value function for an unemployed agent in location  $j$ :

$$\begin{aligned}
V_j^u &= \left( 1 - \lambda_j^u(h) - \sum_{j'=1}^3 \lambda_{jj'}^u(h) \right) E_{v_U} (V_j^{UN} + v_U) \\
&\quad + \lambda_j^u(h) E_{v_U \varepsilon_j} \max [V_j^{UN} + v_U, V_j^{WK}(\varepsilon)] \\
&\quad + \sum_{j'=1}^3 \lambda_{jj'}^u(h) E_{v_U v_M \varepsilon_{j'}} \max [V_j^{UN} + v_U, V_{j'}^{WK}(\varepsilon) - [C_M + v_M]].
\end{aligned} \tag{4}$$

The first term of Equation (4) represents the case in which the individual receives no job offers. In this case, he has no choice and must remain unemployed for an additional period, receiving utility from the amenity in his location, the unemployment benefit and a utility shock. The second term gives the case in which the individual receives a job offer in his city of residence. Under this scenario, he may choose to accept the job immediately or remain unemployed. The third term states that the unemployed agent will accept a potential job offer in another city of type  $j'$  if the job's option value  $V_{j'}^{WK}(\varepsilon)$  net of the moving cost exceeds that of remaining unemployed. The option value of having a job offer in location  $j'$  is the discounted value of holding the job next period plus the current utility implied by the wage offer. The expectation is taken with respect the distribution of  $\varepsilon$  in location type  $j'$  and the distribution of the random components  $v_U$  and  $v_M$  (expressed as  $E_{v_U v_M \varepsilon_{j'}}$ ).

The value function for a worker at location  $j$  resembles that for an unemployed individual except that it also includes potential exogenous job separations and job switching costs. A worker in location  $j$  faces six potential scenarios: being exogenously separated and going unemployed, not receiving a wage offer, receiving an offer in any of the 3 types of locations and receiving a wage offer in the same city. To simplify the computational intensity of the model and because this assumption has little impact on the results, we assume that a worker decides whether to go unemployed before knowing whether he will receive a wage offer from a different employer. As such, the value to a worker with ability

$h$  at location  $j$  of being employed with firm match  $\varepsilon$  is given by Equation (5):

$$\begin{aligned}
V_j(\varepsilon) &= \delta_j(h) V_j^{UN} + (1 - \delta_j(h)) E_{v_U} \max \{ V_j^{UN} + v_U, \\
&\quad (1 - \lambda_j(h) - \sum_{j'=1}^3 \lambda_{jj'}(h)) V_j^{WK}(\varepsilon) \\
&\quad + \lambda_j(h) E_{v_S \varepsilon'_j} \max [V_j^{WK}(\varepsilon), V_j^{WK}(\varepsilon') - v_S] \\
&\quad + \sum_{j'=1}^3 \lambda_{jj'}(h) E_{v_S v_M \varepsilon'_{j'}} \max [V_j^{WK}(\varepsilon), V_{j'}^{WK}(\varepsilon') - v_S - [C_M + v_M]] \}.
\end{aligned} \tag{5}$$

As is evident in Equation (5), an exogenously separated worker at location  $j$  may only become unemployed in location  $j$ . If the worker is not exogenously separated, he can still choose to go unemployed if  $v_U$  is large enough. If he chooses to keep working, with probability  $(1 - \lambda_j(h) - \sum_{j'} \lambda_{jj'}(h))$  he does not receive a wage offer. In this event, he remains employed in the same job. If he receives an offer, he either accepts it and moves if necessary, or remains at his old job.

To conceptualize how the model works, it is convenient to define a set of reservation functions  $\{\varepsilon^R\}$  that can be thought of as hypothetical firm-worker matches at which agents would be indifferent between two choices conditional on the regime in which a certain choice set is available. We define these functions such that if a new draw for a firm-worker match is  $\varepsilon' > \varepsilon^R$ , then the agent optimally chooses to accept a job offer if available or remain employed if unemployment is the only other option. Specification of and details on construction of these reservation functions are in Appendix A.2.

This model captures each of the components of wage growth discussed in the introduction. The location specific component of utility  $\alpha_j(h)$  captures the difference in amenity values and regulates ability sorting across locations. The job arrival rate parameters  $\lambda_{jj'}(h)$  and  $\lambda_{jj'}^u(h)$  capture the potential importance of matching. The coefficients on experience in Equation (3) capture differences in "learning". The intercepts in the wage process capture fixed productivity differences across locations that may be consequences of nonlabor input sharing and human capital spillovers.

## 5 Estimation

This section outlines how we estimate the parameters of the model detailed in the previous section using maximum likelihood. We then intuitively explain how parameters of the model are identified.

### 5.1 The Likelihood Function

The general form for the contribution to the likelihood of an individual that enters in the market in location  $j$  and is observed for  $T$  periods is given by:

$$L(\theta) = \pi_j f(Y^T|h_L; \theta) + (1 - \pi_j) f(Y^T|h_H; \theta)$$

where  $\pi_j$  is the probability that an individual is of type  $h_L$  given that he enters in location  $j$  and  $\theta$  is the vector of parameters.<sup>16</sup>

Define  $Y_t$  to be the vector of labor market outcomes at time  $t$  which consist of a wage, if any, the location of the worker and the type of labor market transition that the worker has experienced since the previous period. We define  $Y^t = \{Y_1, \dots, Y_t\}$  as the vector of all labor market observations in an individual's job history up to and including period  $t$ . We decompose  $f(Y^T|h; \theta)$  as follows:

$$f(Y^T|h; \theta) = f(Y_1|h; \theta) \prod_{t=2}^T f(Y_t|Y^{t-1}, h; \theta).$$

It is convenient to express the likelihood function in terms of probabilities that one of five types of event occurs. The five events are: finding a job in the same location if unemployed, finding a job in a different location if unemployed, having a job to job transition within the same location, having a job to job transition changing location, and entering unemployment. These probabilities are defined as functions of  $\varepsilon$  and, if relevant,  $\varepsilon'$  but they also depend on the other state variables  $\{h, x_1, x_2, x_3\}$ :

$$\left\{ P_{eu}^j(\varepsilon), P_{ue}^j(\varepsilon), P_{ee}^j(\varepsilon, \varepsilon'), P_{ue}^{jj'}(\varepsilon), P_{ee}^{jj'}(\varepsilon, \varepsilon') \right\}_{j, j'=1}^3$$

As an example, the transition probability between unemployment and employment at lo-

---

<sup>16</sup>The individual index is suppressed for notational simplicity.

cation  $j$  with firm specific component  $\varepsilon$  is as follows.

$$P_{ue}^j(\varepsilon) = \lambda_j^u(h) f_{\varepsilon_j}(\varepsilon) \int 1(\varepsilon > \varepsilon_j^A(v_U)) dF(v_U)$$

The offer  $\varepsilon$  at location  $j$  is received by an unemployed worker with probability  $\lambda_j^u(h) f_{\varepsilon_j}(\varepsilon)$  and is accepted only if it exceeds the reservation wage  $\varepsilon_j^A$ . Appendix A.3 specifies the remaining probabilities.

Computation of  $f(Y_t|Y^{t-1})$  is complicated by the fact that we do not observe the firm match  $\varepsilon$ . While it is not observed directly, we can treat it as a latent variable in a non-Gaussian state space model. That is, we can recover the conditional density of  $\varepsilon$  and then integrate the likelihood function with respect to  $\varepsilon$  given that we know the likelihood contribution for each value of  $\varepsilon$ . Assuming that we know the unconditional distribution of the firm match, we use Bayes' rule to update the conditional distribution of  $\varepsilon$  with new wage information each period. This implies the following updating rule:

$$f(\varepsilon|Y^t) = \frac{f(Y_t|Y^{t-1}, \varepsilon) f(\varepsilon|Y^{t-1})}{f(Y_t|Y^{t-1})} \quad (6)$$

This expression is used extensively to build components of the likelihood function. Appendix A contains a detailed explanation of how we construct the likelihood function.

## 5.2 Identification

The model we specify in the previous section is in the class generally known as finite mixture models. This class of models features a finite number of latent agent types in the economy and a subset of parameters that are indexed by type. By following individuals over time, these type-specific parameters are identified, subject to standard constraints on identification. The distribution of types is nonparametrically identified. Kasahara & Shimotsu (2009) discusses identification of parameters in this class of models.

We cannot nonparametrically identify distributions of the firm specific wage components  $\varepsilon_j$ . This is a standard limitation of structural estimation of search models that occurs because the set of wage offers generated by the left tails of the  $\varepsilon_j$  distributions are not accepted and therefore are not observed. As such, we are required to make assumptions about the forms of the  $F_{\varepsilon_j}(\varepsilon)$  distributions. We assume that these firm-specific components are distributed  $N(0, \sigma_{\varepsilon_j})$ . For the purpose of implementation, the distribution of  $\varepsilon$  is a

grid of 10 elements which is restricted to the range  $(-2\sigma_{\varepsilon_j}, 4\sigma_{\varepsilon_j})$ .

Migration plays a crucial role for the identification of many parameters of the model. If no migration were observed, there would be no way to distinguish between the differences in the composition of the population across locations and the inherent differences that exist between location types. When we observe an individual that moves across location types, the variation in labor market histories within each location type are informative about differences across locations in parameters indexed by location. Parameters indexed by type are identified from the full labor market histories of individuals regardless of their location. Parameters indexed by both type and location are identified from the relative labor market experiences across locations of workers of a given type. Identification of these type and location specific parameters does not require that migration is exogenous, but only that workers' types are constant over time. We leverage the life cycle nature of the model to strengthen separate identification of these different parameters.

Table 10 describes all of the estimated parameters of the model. We partition them into six broad groups: components of the wage in Equation (3), amenities, arrival and separation rates, costs and benefits, type probabilities, and distributional measures. We normalize amenities to 0 in location type 0 as they would not otherwise be separately identified from the wage shifters and returns to experience in location type 0.<sup>17</sup> The one parameter of the model that we do not estimate is the discount factor  $\beta$ . This is standard practice in the structural estimation of search models. Based on estimates from the literature, Gourinchas & Parker (2002) for example, we set the discount factor to 0.95 per year.

## 6 Results

### 6.1 Model Fit

Figures 1 and 2 show graphs of spatially deflated average log hourly wages (in cents) from the data and those predicted by the model as functions of experience and city size. Figure 1 shows results for college graduates while Figure 2 presents results for high school graduates. Figure 1 Panel A exhibits a very good fit for small locations while Panel B

---

<sup>17</sup>As seen in the list of parameters in Blocks A and C of Table 10, we make some assumptions that limit the number of parameters capturing returns to experience and heterogeneity in arrival rates of job offers from one location type to another. Details on our normalizations to job offer arrival rates are in Appendix B.

shows that we have a good fit in medium sized cities up to 6 years of experience after which the model slightly over-predicts wages. Panel C shows that for large cities we also slightly over-predict at over 5 years of experience. Figure 2 shows that the fit for the high school sample is generally even better than that for the college sample.

Despite overprediction in the college sample at higher levels of experience, examination of statistics on transitions between jobs, to and from unemployment, and between locations reveals that the simulated data match the actual data remarkably well in many dimensions. Table 7 Panel A shows actual and predicted job, unemployment and location transitions. The model generates simulated data that imply transition statistics that are at most 0.1 percentage points off from the actual data in both samples. Panel B shows job to job and job to unemployment transitions within location type. Once again, neither simulated statistic by location differs from the actual data by more than 0.2 percentage points. Panel B also shows that the simulated data match observed unemployment duration data remarkably well.

Table 8 presents predicted and actual migration conditional on changing location. At left are location types in period  $t - 1$  while along the top are location types at time  $t$ . Diagonal entries give the fraction of moves between different cities of the same location type. The largest gaps between actual and predicted in the college sample are in transitions from large to small and small to large locations which are under-predicted by 2 percentage points. The high school sample exhibits a slightly less good fit, with the largest gaps being the fractions of small and medium to large transitions, which are both over-predicted by 3 percentage points.

## 6.2 The Role of Ability

Our model is designed to facilitate estimation of parameters indexed by unobserved ability. It is natural to wonder whether unobserved ability in the model captures something beyond what could be measured with a test score and whether a test score would be a superior measure of ability to the one we use. We investigate this question by evaluating the relative predictive powers of the age-adjusted Armed Forces Qualifying Test (AFQT) score available in the NLSY79 data set and our estimated ability measure for wages.

Using data on labor market history  $Y_i^T$  up to each individual  $i$ 's final observation  $T$ , the probability that individual  $i$  is high ability is given by the following expression, which

employs Bayes' Rule.

$$\begin{aligned} \Pr(h = h_H | Y_i^T) &= \frac{\Pr(h = h_H, Y_i^T)}{\Pr(Y_i^T)} \\ &= \frac{\Pr(h = h_H, Y_i^T | h = h_H) \times \Pr(h = h_H)}{\Pr(Y_i^T)} \end{aligned}$$

Each of the three elements in this expression are by-products of the same likelihood function maximization used to estimate parameters of the model.

Table 9 presents strong evidence that our estimated  $Pr(h = h_H)$  is a better predictor of the log wage than is the age-adjusted AFQT score. Each column in Table 9 presents results from regressing log wages on a quadratic in experience and various combinations of the two ability measures and two city size dummy variables. Specification (1) shows that our ability measure has a much higher t-statistic than does AFQT (measured in standard deviations) when predicting log wage. Indeed, the pairwise correlation between unobserved ability and log wages is about 3 times greater than that between AFQT and log wages for the college sample and twice as large for the high school sample. Specifications (2)-(5) in Table 9 show that unobserved ability remains more informative about wages when accounting for city size effects. Coefficients on city size hardly change when AFQT is included as a control in the regressions yet increase when unobserved ability is included as a control. Furthermore, the R-squared markedly increases when estimated ability is included in these regressions. Low correlations between the two measures of ability of 0.16 among college graduates and 0.06 for the high school sample are consistent with these observations.<sup>18</sup>

### 6.3 Parameter Estimates

While we present counterfactual simulation results in the following subsection, many of the lessons from these exercises can be seen qualitatively through examination of parameter estimates. Table 10 presents parameter estimates of the structural model for both samples. As discussed in the identification subsection, we have broken the parameter set into six

---

<sup>18</sup> An additional piece of evidence on the utility of our unobserved ability measure is revealed by examining changes in sorting by ability and city size between age 14 and labor force entry. While the high school sample exhibits little change over this period using either ability measure, the college sample exhibits markedly stronger sorting on unobserved ability to larger locations at labor force entry than at age 14. In contrast, the distributions of AFQT scores across locations in the college sample are very similar at these two points in the life-cycle.

categories.

Category A includes location and ability specific wage level and growth estimates. As expected, high ability types have greater estimated wage intercepts in all cases and higher returns to experience in most cases. Constants of the wage process for both samples and types exhibit an inverse U in city size. However, once cost of living adjustments are removed from wages using average prices by size category, intercepts are monotonically increasing in city size for both samples and types, indicating that level effects are an important component of the city size productivity premium.<sup>19</sup>

Estimated returns to experience are monotonically increasing in city size for low ability types in both samples at half to one additional percentage point per year for larger city size categories while estimated returns to experience for high ability types are non-monotonic. The large differences in returns to experience for some groups indicates that differences in human capital accumulation by location may be an important driver of the city size productivity premium. The value of experience is greater in larger cities for all groups. We also find that wages are more concave in experience in larger cities.

Block B of Table 10 reports estimated amenity parameters. Since amenities for small locations are not separately identified, all estimates are relative to the amenity value of small cities. Results indicate that for the college sample, the highest amenity locations are the rural areas for low ability types and large cities for high ability types, though this second estimate is only significant at the 10 percent level. For the high school sample, both ability groups prefer to live in the smallest location category.

Table 10 Block C reports estimated job offer arrival rates by location and worker type. Interestingly, estimated arrival rates for both types of workers are higher in the high school sample than the college sample. These estimates reflect the fact that workers in the high school sample switch jobs more often. Arrival rates are approximately flat as a function of city size except for high ability college graduates who receive more job offers in larger cities. Arrival rates from unemployed are also very flat across city sizes for all groups except low ability high school graduates who also receive more offers in larger cities. Estimates on arrival rates of offers from different locations reveal that migrating workers are about twice

---

<sup>19</sup>Because the model has no predictions about the exact city of each size to which migrating individuals move, it is impossible for us to exactly undeflate wages back to marginal productivities. Given the equilibrium location of college graduates observed in the data, prices are 6 percent higher in medium sized cities and 18 percent higher in large cities compared to small cities and rural areas. Because within size category they live in smaller cities, these price premia are 3 percent and 19 percent respectively for high school graduates.

as likely to move within city size category than across categories and that receive rates for college graduates are declining in city size. Finally, exogenous separation rates are fairly flat as a function of city size for low types and roughly increasing in city size for high types. Overall, these results on search frictions show that low ability college graduates and high ability high school graduates cycle through different jobs at higher rates than the other groups. Given that worker type is a much more important predictor of differences in search frictions than is location type, these estimates support the descriptive evidence in Tables 4 and 6 that differences in search frictions are not likely to explain much of the city size wage premium. However, these results indicate the potential importance of accounting for variation in the ability distribution across locations so as not to attribute differences in search frictions across locations to location-specific effects.

Block D reveals that the benefit required to convince individuals to go unemployed is higher for high school graduates than for college graduates. This result may reflect high school graduates' higher disutility of working. Additionally, high school graduates are estimated to have higher implied moving costs than college graduates. This result may reflect their higher psychic costs of setting up in a new city. Results in Block E demonstrate that there is not a lot of selection on unobserved worker ability from the point of entry into the labor market. Results in Block F show estimated standard deviations of all distributions in the structural model.

## 6.4 Simulations

Using the parameter estimates from the structural model, Tables 11 and 12 evaluate the importance of potential mechanisms for generating city size wage and productivity premia for the two samples. In these tables, we report counterfactual city size wage premia after independently shutting off each mechanism in the model that may contribute to these premia. For each experiment, we assign the average value of the relevant parameters for each latent type listed at left in Panels B and C across locations and then regress the resulting counterfactually simulated wage on a quadratic in experience and two city size dummy variables. Tables 11 and 12 report the coefficients on the city size dummy variables from these regressions, all of which are statistically significant. Tables A1 and A2 report confidence intervals for these counterfactual premia calculated by simulating the model using parameters sampled from their full estimated joint distributions.

The ultimate goal of these simulations is to evaluate the extent to which marginal

productivity gaps between cities of different sizes change under counterfactual scenarios. Performing such an evaluation is not straightforward because the model only has predictions about city size category of residence, not the particular city of residence within a given category. Therefore, for the channels that generate a significant reduction in city size wage premia, we report the associated percent reduction in marginal product calculated incorporating the equilibrium average price differences reported in Tables 11 and 12 Row 4. This procedure is not perfect because the price differences in Row 4 are not calculated using a counterfactual distribution of people across location within each size category. Indeed, the key assumption for the validity of this exercise is that such equilibrium and counterfactual distributions are the same. Nevertheless, given the magnitudes of our results we believe that they still provide a pretty clear picture of the factors that generate the largest marginal productivity gaps.

Because wages used for estimation are adjusted for cost of living differences, recovery of marginal products by equalizing  $\beta_0^j(h)$  parameters of the wage process requires special consideration. The goal is to generate a counterfactual by equalizing marginal productivities with 0 experience in each location. To achieve this, for the rows entitled "Equalize Wage Constant Across Locs" we equalize  $\beta_0^j(h) + \ln(P^j)$ , where  $\ln(P^j)$  are given in Row 4. We set the population weighted mean to be equal to the population weighted mean of  $\beta_0^j(h)$ . There is no such issue for equalizing returns to experience because wages are already expressed in logs.

Table 11 reports the counterfactual simulation results for college graduates. As a baseline, Panel A Row 1 shows that regressing log wage on a quadratic in experience and two city size dummies using the raw data implies city size wage premia of 0.14 for medium sized cities and 0.09 for large cities. Data simulated from the parameter estimates in Table 10 imply wage premia of 0.15 and 0.12 for the two size categories respectively, very close to the true numbers. Panel A row 3 gives counterfactual city size wage premia in the case where individuals are forced to stay in their location of labor market entry. Restricting mobility actually slightly increases wage premia between rural areas and cities to 0.16 implied by parameter point estimates, as it forces individuals to stay in locations where returns to experience are greater. However, this increase is not statistically significant. Results in Row 2 are used as a benchmark for counterfactuals in Panel B, for which we shut off potential channels other than mobility for the city size wage premia. Results in Row 3 benchmark results in Panel C for which we shut off the same potential channels as in Panel B in addition to mobility.

Equalizing returns to experience and wage level effects by type and location generate approximately equal percentage declines in marginal products for the college sample and account for virtually the entire city size labor productivity gap. Absent differences in returns to experience across locations and maintaining endogenous sorting, the counterfactual city size nominal wage premium would be reduced by one-third to one-half. Absent differences in wage level effects, the counterfactual city size nominal wage premium is estimated to decline by 38% for both medium and large cities.<sup>20</sup> Search frictions and ability sorting at entry into the labor force (conditional on education) have small and statistically insignificant impacts on city size wage premia and marginal productivity premia in the college sample.<sup>21</sup>

Simulation results with the mobility restriction reported in Table 11 Panel C show that mobility is not an important driver of the results in Panel B, though the impact of returns to experience relative to marginal productivity level effects are amplified with this additional restriction. With restricted mobility, equalization of returns to experience across locations and types reduces nominal wage gaps by about 60% with the remaining 40% accounted for by level effects. Other counterfactual changes in wage premia are similarly small and not statistically significant.

Table 12 reports analogous results to those in Table 11 for high school graduates. Though overall patterns are largely in line with those for college graduates, the other two channels also play important roles. Equalizing returns to experience decreases nominal wage premia by 70 percent and 37 percent for medium and large cities respectively with free mobility. Analogous results once restricting mobility are more similar at 63 and 50 percent respectively. Equalizing level effects in the wage equation decreases city size productivity premia by about twice as much with free mobility and at least 40 percent more with restricted mobility.

The resulting total reductions of more than 100% come about because equalizing search parameters or the ability distribution at labor force entry increases wage premia, particularly in medium sized cities. Increases in the premia due to equalization of the initial ability distribution occurs for a similar reason that restricting mobility increases the city size wage premium. This reallocation pushes more low ability individuals into larger loca-

---

<sup>20</sup>Equalizing the  $\beta_0^j(h)$  parameters across locations without adjusting for price differences generates small and insignificant reductions in wage premia for the college sample.

<sup>21</sup>Unlike us, Gould (2007) finds that sorting on unobservables is an important component of the city size wage premium. The reason is that he pools high school and college graduates in his sample. When we use the combined sample, we also find that sorting is an important driver of the premium.

tions in which their returns to experience are higher, meaning higher wages over the rest of their life-cycles. These low ability types are likely to be worse off though because of the relative negative amenity value they place on living in larger cities. Counterfactual premia absent search frictions also indicate that the equilibrium locations of individuals by type result in slightly lower job offer arrival rates and/or higher separation rates than are experienced on average across locations.

## 7 Conclusions

In this paper, we lay out a systematic framework to empirically examine reasons for which larger cities have higher wages and are more productive. Using data from the NLSY, we show that hourly wages are higher and grow faster in bigger cities, workers in larger cities have higher observed skill levels, hold fewer jobs and experience less time unemployed. A decomposition of log wage growth over the first 15 years of experience reveals that within job wage growth generates more of the city size wage gap than between job wage growth. Estimation of the model specified in this paper allows us to sort out the extent to which sorting across locations on ability interacts with sharing, learning and matching to generate city size wage and productivity gaps.

Counterfactual simulations of our structural model indicate that variation in the return to experience and wage level effects across location type are the most important mechanisms contributing to the overall city size wage premium. These mechanisms are important for both high school and college graduates throughout the city size distribution. Differences in wage intercepts across location categories are more important for high school graduates while differences in returns to experience are more important for college graduates. However, sorting on unobserved ability within education group and differences in labor market search frictions independently contribute negatively if at all to observed city size wage premia.

Our results indicate that larger cities foster more rapid human capital accumulation in workers. However, given our data we can do little more than speculate about the mechanisms that might be behind this phenomenon. More rapid human capital accumulation may come from traditional Marshallian externalities in which larger cities contain more "ideas in the air" to be accumulated over time. Alternatively, if technologically intensive industries are more likely to locate in larger cities, technologically advanced capital and labor are complements in production and it takes time for individuals to learn how to use

such technology, steeper returns to experience in larger cities would result. Our result that workers are more productive immediately at moving to a larger city is also consistent with such a mechanism.

Our evidence that level effects are more important for high school graduates than college graduates is striking. This indicates that while those with more human capital learn more quickly in larger agglomerations, there may be something about the technologies used in larger cities that generates larger immediate changes in productivity for less skilled individuals who move there. Given this observation, one interesting topic for future research may be to evaluate the extent to which technological differences across cities of different sizes may generate these level effects.

## A Construction of the Likelihood Function

In this appendix, we present expressions for the contribution of each potential type of event in an individual's job history to the likelihood function. Though we suppress this dependence in the notation, the objects  $A(\cdot)$ ,  $B(\cdot)$ ,  $\varepsilon(\cdot)$ ,  $P(\cdot)$  and  $f(\cdot)$  derived below are functions of type  $h$  and location-specific work experience  $\{x_1, x_2, x_3\}$ .

### A.1 Fundamentals

Wages are not always observed when they should be. To deal with this, we define the functions  $B_{jt}(\cdot)$  and  $A_{jt-1}(\cdot)$ .  $B_{jt}(\cdot)$  gives the distribution of wage information for the final observations covered by each interview while the function  $A_{jt-1}(\cdot)$  gives that for job changes that are reported within an interview cycle. As mentioned in the data section, wages are observed once a year for up to 5 different jobs. Therefore if a worker does not change employer, we have only one wage observation a year for that worker, while if a worker changes employer within a cycle, we may have more than one wage observation. Because the wage is recorded in  $t - 1$  only because the worker has changed job in the previous period, this information must be included in the contribution to the likelihood function of period  $t$  using the function  $A_{jt-1}(\cdot)$ . These functions include the parameter  $p_n$ , the probability of observing a wage.

$$\begin{aligned}
 B_{jt}(\varepsilon) &= \left[ [p_n F_u(u_t)]^{1(w_t \text{ obs})} [1 - p_n]^{1(w_t \text{ not obs})} \right]^{1(int_t \neq int_{t+1})} \\
 A_{jt-1}(\varepsilon) &= \left[ [p_n F_u(u_{t-1})]^{1(w_{t-1} \text{ obs})} [1 - p_n]^{1(w_{t-1} \text{ not obs})} \right]^{1(int_{t-1} = int_t \ \& \ job_{t-1} \neq job_t)}
 \end{aligned}$$

Because we have no interest in the value of  $p_n$  and we take it as exogenous, we can simplify the expressions above by conditioning the likelihood on observing the wages. Therefore, we define these functions to be

$$\begin{aligned}
 B_{jt}(\varepsilon) &= F_u(u_t)^{1(w_t \text{ obs} \ \& \ int_t \neq int_{t+1})} \\
 A_{jt-1}(\varepsilon) &= F_u(u_{t-1})^{1(w_{t-1} \text{ obs} \ \& \ int_{t-1} = int_t \ \& \ job_{t-1} \neq job_t)}
 \end{aligned}$$

instead.

## A.2 Reservation Rules

Regime A occurs when an unemployed agent receives an own-location job offer. Regime B occurs when a worker is choosing whether to go unemployed. Regime C occurs when an unemployed agent receives an offer in another location. Regime D occurs when a worker receives an own-location offer. Regime E occurs when a worker receives an offer in another location. In cases where a new match is drawn and the worker has an existing match quality,  $\varepsilon'$  denotes the new match and  $\varepsilon$  denotes the firm-specific component of the existing job. If the worker is unemployed,  $\varepsilon$  denotes the new match draw.

$$\begin{aligned}
\varepsilon_j^A(v_U) \text{ solves} & : V_j^{UN} + v_U = V_j^{WK}(\varepsilon), \\
\varepsilon_j^B(\varepsilon, v_U) \text{ solves} & : V_j^{UN} + v_U = (1 - \lambda_j(h) - \sum_{j'=1}^3 \lambda_{jj'}(h))V_j^{WK}(\varepsilon) + \\
& + \lambda_j(h) E_{v_S \varepsilon'_j} \max[V_j^{WK}(\varepsilon), V_j^{WK}(\varepsilon') - v_S] + \\
& + \sum_{j'=1}^3 \lambda_{jj'}(h) E_{v_S v_M \varepsilon'_{j'}} \max[V_j^{WK}(\varepsilon), V_{j'}^{WK}(\varepsilon') - v_S - [C_M + v_M]] \\
\varepsilon_{jj'}^C(v_U, v_M) \text{ solves} & : V_j^{UN} + v_U = V_{j'}^{WK}(\varepsilon) - C_M - v_M, \\
\varepsilon_j^D(\varepsilon, v_S) \text{ solves} & : V_j^{WK}(\varepsilon) = V_j^{WK}(\varepsilon') - v_S, \\
\varepsilon_{jj'}^E(\varepsilon, v_S, v_M) \text{ solves} & : V_j^{WK}(\varepsilon) = V_{j'}^{WK}(\varepsilon') - C_M - v_M - v_S.
\end{aligned}$$

## A.3 Transition Probabilities

The probability of exiting unemployment and finding a job with match  $\varepsilon$  in the same location  $j$  is given by:

$$P_{ue}^j(\varepsilon) = \lambda_j^u(h) f_{\varepsilon_j}(\varepsilon) \int 1(\varepsilon > \varepsilon_j^A(v_U)) dF(v_U).$$

The probability of exiting unemployment and finding a job with match  $\varepsilon$  in a different location  $j'$  is:

$$P_{ue}^{jj'}(\varepsilon) = \lambda_{jj'}^u(h) f_{\varepsilon_{j'}}(\varepsilon) \int \int 1(\varepsilon > \varepsilon_{jj'}^C(v_U, v_M)) dF(v_U) dF(v_M)$$

The probability of entering unemployment given that a worker had a job with match  $\varepsilon$

is:

$$P_{eu}^j(\varepsilon) = \delta_j(h) + (1 - \delta_j(h)) \int 1(\varepsilon < \varepsilon_j^B(v_U)) dF(v_U)$$

The probability of changing employer from match  $\varepsilon$  to match  $\varepsilon'$  in the same location is:

$$P_{ee}^j(\varepsilon, \varepsilon') = [1 - P_{eu}^j(\varepsilon)] \times \lambda_j(h) f_{\varepsilon_j}(\varepsilon') \times \int 1(\varepsilon' > \varepsilon_j^D(\varepsilon, v_S)) dF(v_S)$$

The probability of changing employer from match  $\varepsilon$  in location  $j$  to match  $\varepsilon'$  in location  $j'$  is:

$$P_{ee}^{jj'}(\varepsilon, \varepsilon') = [1 - P_{eu}^j(\varepsilon)] \times \lambda_{jj'}(h) f_{\varepsilon_{j'}}(\varepsilon') \times \int \int 1(\varepsilon' > \varepsilon_{jj'}^E(\varepsilon, v_S, v_M)) dF(v_S) dF(v_M)$$

#### A.4 First Period

Because we condition on working in the first period, the contribution to the likelihood of an individual entering in location  $j$  is:

$$L_1^j = \frac{\int B_{j1}(\varepsilon) P_{ue}^j(\varepsilon) d\varepsilon}{\int P_{ue}^j(\varepsilon) d\varepsilon}$$

The resulting posterior distribution of the firm match is:

$$f(\varepsilon|Y^1) = \frac{B_{j1}(\varepsilon) P_{ue}^j(\varepsilon)}{\int B_{j1}(\varepsilon) P_{ue}^j(\varepsilon) d\varepsilon}$$

#### A.5 Unemployment

An individual of ability  $h$  enters unemployment in location  $j$  and has an unemployment spell that lasts  $NW_t$  weeks. The probability of not accepting a job for  $NW_t - 1$  weeks is given by

$$\Pi_2^j(NW_t) = \left( 1 - \int P_{ue}^j(\varepsilon) d\varepsilon - \sum_{j=1}^3 \int P_{ue}^{jj'}(\varepsilon) d\varepsilon \right)^{NW_t-1}$$

After  $NW_t$  weeks, the worker finds a job in location  $j$  or in location  $j'$ . If he finds a job in location  $j$ , the total contribution of this unemployment spell to the likelihood function is:

$$L_{2a}^j = \Pi_2^j(NW_t) \int B_{jt}(\varepsilon) P_{ue}^j(\varepsilon) d\varepsilon$$

The posterior distribution of the match then becomes:

$$f(\varepsilon|Y^t) = \frac{B_{jt}(\varepsilon) P_{ue}^j(\varepsilon)}{\int B_{jt}(\varepsilon) P_{ue}^j(\varepsilon) d\varepsilon}$$

If after  $NW_{it}$  weeks he finds a job in location  $j'$ , the contribution of the unemployment spell to the likelihood function is:

$$L_{2b}^j = \Pi_2^j(NW_t) \int B_{j't}(\varepsilon) P_{ue}^{jj'}(\varepsilon) d\varepsilon$$

The posterior distribution of the match is then:

$$f(\varepsilon|Y^t) = \frac{B_{j't}(\varepsilon) P_{ue}^{jj'}(\varepsilon)}{\int B_{j't}(\varepsilon) P_{ue}^{jj'}(\varepsilon) d\varepsilon}$$

## A.6 Becoming Unemployed

A worker in location  $j$  goes unemployed with probability  $P_{eu}^j(\varepsilon)$  and the density of the observed wage is  $A_{jt-1}(\varepsilon)$ . From the previous period we know  $f(\varepsilon|Y^{t-1})$ . Given this, we can express the contribution of becoming unemployed to the likelihood as:

$$L_3^j = \int A_{jt-1}(\varepsilon) P_{eu}^j(\varepsilon) dF(\varepsilon|Y^{t-1})$$

Given that the firm specific component  $\varepsilon$  is lost, we do not need to update its conditional distribution.

## A.7 Working

If the worker remains with the same employer, the likelihood contribution and the conditional distribution of the firm match after this period can be written as:

$$L_{4a}^j = \int B_{jt}(\varepsilon) \left( 1 - P_{eu}^j(\varepsilon) - \int P_{ee}^j(\varepsilon, \varepsilon') d\varepsilon' - \sum_{j=1}^3 \int P_{ee}^{jj'}(\varepsilon, \varepsilon') d\varepsilon' \right) dF(\varepsilon|Y^{t-1})$$

$$f(\varepsilon|Y^t) = \frac{B_{jt}(\varepsilon) \left( 1 - P_{eu}^j(\varepsilon) - \int P_{ee}^j(\varepsilon, \varepsilon') d\varepsilon' - \sum_{j=1}^3 \int P_{ee}^{jj'}(\varepsilon, \varepsilon') d\varepsilon' \right) f(\varepsilon|Y^{t-1})}{L_{4a}^j}$$

Alternately, the employed worker may move to a different employer in the same type of location.

$$L_{4b}^j = \int \int A_{jt-1}(\varepsilon) B_{jt}(\varepsilon') P_{ee}^j(\varepsilon, \varepsilon') dF(\varepsilon|Y^{t-1}) d\varepsilon'$$

$$f(\varepsilon'|Y^t) = \frac{B_{jt}(\varepsilon') \int A_{jt-1}(\varepsilon) P_{ee}^j(\varepsilon, \varepsilon') dF(\varepsilon|Y^{t-1})}{L_{4b}^j}$$

Note that the inclusion of the function  $A_{jt-1}(\varepsilon)$  captures the fact that we in  $t-1$  we have observed a wage only because the worker has changed job in period  $t$ . Hence this wage information is included in period  $t$  and not in period  $t-1$ .

Finally, the employed worker may move to a different employer in a different type of location.

$$L_{4c}^j = \int \int A_{jt-1}(\varepsilon) B_{j't}(\varepsilon') P_{ee}^{jj'}(\varepsilon, \varepsilon') dF(\varepsilon|Y^{t-1}) d\varepsilon'$$

$$f(\varepsilon'|Y^t, h) = \frac{B_{j't}(\varepsilon') \int A_{jt-1}(\varepsilon) P_{ee}^{jj'}(\varepsilon, \varepsilon') dF(\varepsilon|Y^{t-1})}{L_{4c}^j}$$

## B Normalizations of Job Offer Arrival Rates

In the model there are 48 free parameters that measure arrival rates of job offers of which 36 are probabilities of receiving a wage offer from a different location. Given that changing location is a rare event in the data, these parameters cannot be estimated precisely. Instead

we estimate freely the 12 parameters that describe the probability of receiving a wage offer from the same location and we assume that the remaining probabilities are scaled by the 4 estimated parameters  $\rho_{j'}$  and  $\gamma$ . We define  $\rho_{j'}$  to be a multiplier for arrival rates to a given city  $j'$  and  $\gamma$  to be a parameter that scales the product  $\lambda_j(h)\rho_{j'}$  if the two location sizes are the same but the individual changes location. We use the same scaling factors for unemployed and worker arrival rates. Arrival rates of job offers across locations are thus specified as follows.

$$\begin{aligned}\lambda_{jj'}^u(h) &= \lambda_j^u(h)\rho_{j'} \text{ if } j \neq j' \\ \lambda_{jj'}^u(h) &= \lambda_j^u(h)\rho_{j'}\gamma \text{ if } j = j'\end{aligned}$$

$$\begin{aligned}\lambda_{jj'}(h) &= \lambda_j(h)\rho_{j'} \text{ if } j \neq j' \\ \lambda_{jj'}(h) &= \lambda_j(h)\rho_{j'}\gamma \text{ if } j = j'\end{aligned}$$

These normalizations reduce the number of arrival rate parameters to be estimated from 48 to 16.

## References

- Acemoglu, Daron and Josh Angrist. 2000 "How Large Are the Social Returns to Education? Evidence from Compulsory Schooling Laws," NBER Macroeconomics Annual 9-59.
- Albouy, David. 2008 "Are Big Cities Really Bad Places to Live? Improving Quality of Life Estimates Across Cities," NBER Working Paper #14472
- Arzhagi, Muhammad and J. Vernon Henderson, 2008 "Networking off Madison Avenue," *Review of Economic Studies* 75:4, 1011-1038.
- Basker, Emek 2005 "Selling a Cheaper Mousetrap: Wal-Mart's Effects on Retail Prices" *Journal of Urban Economics* 58:2 203-229
- Baum-Snow, Nathaniel and Ronni Pavan. 2009 "Inequality and City Size" manuscript.
- Borjas, George. "Self-Selection and Earnings of Immigrants." *American Economic Review* 77:4, 531-553.
- Carlino, Gerald, Satyajit Chatterjee and Robert Hunt, 2007. "Urban Density and the Rate of Invention," *Journal of Urban Economics* 61 389-419.
- Carrillo, Paul E., Dirk W. Early and Edgar O. Olsen, 2009. "New Cross-Sectional Price Indices for All Areas of the United States," manuscript.
- Ciccone, Antonio and Giovanni Peri, 2006. "Identifying Human Capital Externalities: Theory with Applications," *Review of Economic Studies* 73 381-412.
- Ciccone, Antonio and Robert Hall, 1996. "Productivity and Density of Economic Activity," *American Economic Review* 86 54-70.
- Combes, Pierre-Philippe, Gilles Duranton and Laurent Gobillon, 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63:2, 723-742.
- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, & Sebastian Roux. 2009. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection" manuscript.
- Costa, Dora and Matthew Kahn, 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990," *Quarterly Journal of Economics* 115 1287-1315.
- Duranton, Gilles and Diego Puga, 2004. "Micro-Foundations of Urban Agglomeration Economies," in *Handbook of Urban and Regional Economics*, J.V. Henderson & J-F Thisse eds. North Holland-Elsevier.
- Ellison, Glenn and Edward Glaeser, 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* 105 889-927.
- Flinn, Christopher, 2002. "Labour Market Structure and Inequality: A Comparison of Italy and the U.S." *Review of Economic Studies* 69 611-645.

- Fu, Shihe and Stephen Ross. 2008 "Wage Premia in Employment Clusters: Agglomeration Economies or Worker Heterogeneity." Working Paper
- Gemici, Ahu. 2008 "Family Migration and Labor Market Outcomes", manuscript
- Glaeser, Edward, Kallal, Hedi, Scheinkman, Jose and Andrei Shleifer, 1992. "Growth in Cities," *Journal of Political Economy* 100 1126-1152.
- Glaeser, Edward, Jed Kolko and Albert Saiz, 2001. "Consumer City," *Journal of Economic Geography* 1 27-50.
- Glaeser, Edward and David Mare, 2001. "Cities and Skills," *Journal of Labor Economics* 19 316-342.
- Glaeser, Edward and Albert Saiz, 2003. "The Rise of the Skilled City," NBER Working Paper #10191.
- Gould, Eric. 2007 "Cities, Workers and Wages: A Structural Analysis of the Urban Wage Premium" *Review of Economic Studies*, 74:2, 477-506.
- Gourinchas, Pierre-Olivier and Jonathan A. Parker. 2002. "Consumption Over the Life-Cycle," *Econometrica*, 70:1, 47-89
- Greenstone, Michael, R. Hornbeck and Enrico Moretti. 2009 "Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants" working paper
- Heckman, James and Burton Singer, 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica* 52 271-320.
- Henderson, J. Vernon, 1997. "Medium Sized Cities," *Regional Science & Urban Economics* 27 583-612.
- Henderson, J. Vernon, Ari Kuncoro & Matthew Turner. "Industrial Development in Cities," 1995. *Journal of Political Economy* 103:5 1067-1090.
- Kasahara, Hiroyuki & Katsumi Shimotsu. 2009 "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices." *Econometrica* 77:1, 135-175.
- Keane, Michael and Kenneth Wolpin, 1994. "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *The Review of Economics and Statistics* 76 648-72.
- Keane, Michael and Kenneth Wolpin, 1997. "The Career Decisions of Young Men," *Journal of Political Economy* 105 473-522.
- Kennan, John and James R. Walker. 2009. "The Effect of Expected Income on Migration Decisions," manuscript
- Lee, Sanghoon. "Ability Sorting and Consumer City," manuscript.
- Moretti, Enrico, 2004a. "Workers' Education, Spillovers, and Productivity: Evidence from

Plant-Level Production Functions," *American Economic Review* 94 656-690.

Moretti, Enrico, 2004b. "Human Capital Externalities in Cities," in *Handbook of Urban and Regional Economics*, J.V. Henderson & J-F Thisse eds. North Holland-Elsevier.

Pavan, Ronni, 2009. "Career Choice and Wage Growth," manuscript.

Peri, Giovanni, 2002. "Young Workers, Learning and Agglomerations," *Journal of Urban Economics* 52 582-607.

Petrongolo, Barbara and Christopher Pissarides. 2001. "Looking Into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature* 39 390-431.

Petrongolo, Barbara and Christopher Pissarides. 2003. "Scale Effects in Markets With Search," *Economic Journal* 116:508 21-44.

Rauch, James, 1993. "Productivity Gains from the Geographic Concentration of Human Capital: Evidence from Cities," *Journal of Urban Economics* 34 380-400.

Roback, Jennifer, 1982. "Wages, Rents, and the Quality of Life," *Journal of Political Economy* 90 1257-1278.

Rosen, Sherwin, 1979. "Wage-Based Indexes of Urban Quality of Life." in Peter Mieszkowski and Mahlon Straszheim (eds.) *Current Issues in Urban Economics*, Baltimore: Johns Hopkins University Press.

Rosenthal, Stuart and William Strange, 2003. "Geography, Industrial Organization and Agglomeration," *Review of Economics and Statistics* 86 438-444.

Rosenthal, Stuart and William Strange, 2004. "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Urban and Regional Economics*, J.V. Henderson & J-F Thisse eds. North Holland-Elsevier.

Roy, A.D.. 1951 "Thoughts on the Distribution of Earnings" *Oxford Economic Papers*, 3, 135-146.

Shapiro, Jesse, 2006. "Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital," *Review of Economics and Statistics* 88 324-335.

Syverson, Chad, 2004. "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy* 112:6, 1181-1222.

Wheeler, Christopher, 2006. "Cities and the Growth of Wages Among Young Workers: Evidence from the NLSY," *Journal of Urban Economics* 60 162-184.

Yankow, Jeffrey, 2006. "Why Do Cities Pay More? An Empirical Examination of Some Competing Theories of the Urban Wage Premium," *Journal of Urban Economics* 60 139-161.

**Table 1: Estimates of City Size Wage Premia from Census Micro Data**

	Spec. 1: No Controls			Spec 2: Individual Controls		
	1980	1990	2000	1980	1990	2000
<b>MSAs 0.25 - 1.5 million</b>	0.141*** (0.0104)	0.175*** (0.00984)	0.185*** (0.00963)	0.0931*** (0.00843)	0.130*** (0.00755)	0.119*** (0.00673)
<b>MSAs &gt; 1.5 million</b>	0.238*** (0.0141)	0.311*** (0.0106)	0.315*** (0.00967)	0.170*** (0.0106)	0.238*** (0.00769)	0.219*** (0.00643)
<b>R-squared</b>	0.029	0.046	0.039	0.251	0.336	0.293

Notes: The sample includes white men 20-64 working at least 40 weeks and at least 35 hours per week from the 5% census micro data samples. Wages of less than \$1 or more than \$300 (1999 dollars) are excluded from the sample. Main entries are coefficients and standard errors from regressions of log hourly wage on indicators for metropolitan area size of residence. Individual controls include age, age squared and indicators for nine levels of education. Standard errors are clustered by county group or PUMA. Sample sizes are 1,269,114 in 1980, 1,330,116 in 1990 and 1,245,865 in 2000.

**Table 2: Estimates of City Size Wage Premia from NLSY Data**

Specification	No Controls	Individual Controls	Individual Controls and Fixed Effects
	1	2	3
<b>Panel A: CPI Deflated</b>			
<b>MSAs 0.25 - 1.5 million</b>	0.20 (0.008)**	0.15 (0.007)**	0.08 (0.012)**
<b>MSAs &gt; 1.5 million</b>	0.30 (0.008)**	0.23 (0.007)**	0.14 (0.013)**
<b>R-squared</b>	0.04	0.30	0.25
<b>Panel B: ACCRA Deflated</b>			
<b>MSAs 0.25 - 1.5 million</b>	0.16 (0.008)**	0.10 (0.006)**	0.06 (0.011)**
<b>MSAs &gt; 1.5 million</b>	0.13 (0.008)**	0.06 (0.007)**	0.02 (0.013)
<b>R-squared</b>	0.01	0.29	0.28

Notes: Each regression uses 1,754 individuals and has 30,367 observations based on quarterly data. Individual control variables include years of education and quadratics in age and experience. The  $R^2$  for Specification 3 is the within  $R^2$ . The sample includes white men in the NLSY79. Complete sample selection rules are explained in the text.

**Table 3: Log Wage Growth As a Function of Labor Force Experience**

Years of Labor Force Experience	CPI Deflated				CPI & ACCRA Deflated			
	1	5	10	15	1	5	10	15
<b>Panel A: Full Sample</b>								
<b>Rural and MSAs &lt; 0.25 million</b>	0.06	0.28	0.36	0.52	0.06	0.30	0.41	0.58
<b>MSAs 0.25 - 1.5 million</b>	0.11	0.36	0.54	0.68	0.11	0.37	0.57	0.72
<b>MSAs &gt; 1.5 million</b>	0.12	0.44	0.65	0.76	0.11	0.43	0.66	0.78
<b>Panel B: College or More</b>								
<b>Rural and MSAs &lt; 0.25 million</b>	0.10	0.34	0.57	0.68	0.10	0.38	0.63	0.75
<b>MSAs 0.25 - 1.5 million</b>	0.13	0.43	0.61	0.89	0.14	0.45	0.64	0.93
<b>MSAs &gt; 1.5 million</b>	0.07	0.44	0.70	0.90	0.07	0.44	0.72	0.93
<b>Panel C: High School Graduates Only</b>								
<b>Rural and MSAs &lt; 0.25 million</b>	0.03	0.27	0.31	0.46	0.03	0.30	0.36	0.52
<b>MSAs 0.25 - 1.5 million</b>	0.09	0.32	0.48	0.53	0.09	0.33	0.51	0.57
<b>MSAs &gt; 1.5 million</b>	0.14	0.41	0.63	0.67	0.13	0.40	0.63	0.68

Notes: Each entry gives the average growth in log wages from labor force entry to the indicated level of experience. Workers are assigned to the location type in which they reside at the indicated level of experience. The sample is the same as that for Table 2 with the additional limitation that we exclude individuals observed for less than 15 years of labor force experience. This limitation reduces the number of individuals in the sample to 1,418 in Panel A, 455 in Panel B and 568 in Panel C.

**Table 4: Job Turnover and Unemployment**

Years of Labor Force Experience	Number of Jobs				Weeks of Unemployment			
	1	5	10	15	1	5	10	15
<b>Panel A: Full Sample</b>								
Rural and MSAs < 0.25 million	1.6	3.5	5.5	6.7	2.9	16.2	27.7	34.3
MSAs 0.25 - 1.5 million	1.6	3.5	5.2	6.6	2.0	12.7	21.2	27.1
MSAs > 1.5 million	1.5	3.2	4.8	5.9	1.7	9.9	15.7	21.0
<b>Panel B: College or More</b>								
Rural and MSAs < 0.25 million	1.5	2.7	4.3	5.0	0.7	3.6	8.3	9.6
MSAs 0.25 - 1.5 million	1.4	2.8	3.9	5.0	0.6	3.9	6.7	9.0
MSAs > 1.5 million	1.4	2.7	4.0	4.9	0.4	4.1	7.5	10.8
<b>Panel C: High School Graduates Only</b>								
Rural and MSAs < 0.25 million	1.7	3.8	6.0	7.2	4.1	21.8	37.4	47.6
MSAs 0.25 - 1.5 million	1.6	3.7	5.6	7.2	2.9	18.2	28.8	36.7
MSAs > 1.5 million	1.6	3.5	5.1	6.4	3.2	16.2	25.0	32.8

Notes: See Table 3 for a description of the sample. Weeks of unemployment are counted from the end of the first full time job.

**Table 5: Migration Patterns by Education Shares**

Location of Labor Force Entry	Location at 15 Years of Labor Force Experience		
	Rural and MSAs < 0.25 million	MSAs 0.25 to 1.5 million	MSAs > 1.5 million
<b>Panel A: Full Sample (Fraction High School)</b>			
Rural and MSAs < 0.25 million	0.78 (0.82)	0.14 (0.65)	0.07 (0.63)
MSAs 0.25 - 1.5 million	0.10 (0.69)	0.82 (0.68)	0.08 (0.37)
MSAs: > 1.5 million	0.08 (0.66)	0.14 (0.54)	0.78 (0.65)
<b>Panel B: College or More</b>			
Rural and MSAs < 0.25 million	0.64	0.24	0.12
MSAs 0.25 - 1.5 million	0.10	0.76	0.14
MSAs: > 1.5 million	0.08	0.18	0.74
<b>Panel C: High School Graduates Only</b>			
Rural and MSAs < 0.25 million	0.87	0.09	0.04
MSAs 0.25 - 1.5 million	0.11	0.87	0.02
MSAs: > 1.5 million	0.10	0.09	0.81

Notes: See Table 3 for a description of the sample. Entries give the fraction of those entering the labor force at the location listed at left residing in the location along the top at 15 years of labor force experience. Numbers in parentheses indicate the fraction of high school only workers in each cell.

**Table 6: Log Wage Growth Decomposition  
0 to 15 Years of Labor Force Experience**

	Within Job	CPI Deflated			Within Job	CPI & ACCRA Deflated			
		Job to Job	Job to Unemp	Un- Known		Job to Job	Job to Unemp	Un- Known	
<b>Panel A: Full Sample</b>									
<b>Rural and MSAs &lt; 0.25 million</b>	0.30	0.25	-0.04	0.01	0.33	0.27	-0.03	0.01	
<b>MSAs 0.25 - 1.5 million</b>	0.44	0.30	-0.06	0.00	0.45	0.31	-0.05	0.00	
<b>MSAs: &gt; 1.5 million</b>	0.50	0.24	0.01	0.00	0.51	0.24	0.02	0.00	
<b>Panel B: College or More</b>									
<b>Rural and MSAs &lt; 0.25 million</b>	0.42	0.30	-0.01	-0.02	0.46	0.32	-0.01	-0.02	
<b>MSAs 0.25 - 1.5 million</b>	0.56	0.34	-0.03	0.01	0.57	0.36	-0.02	0.01	
<b>MSAs: &gt; 1.5 million</b>	0.59	0.26	0.01	0.02	0.60	0.28	0.01	0.02	
<b>Panel C: High School Graduates Only</b>									
<b>Rural and MSAs &lt; 0.25 million</b>	0.30	0.23	-0.07	0.00	0.33	0.25	-0.05	0.00	
<b>MSAs 0.25 - 1.5 million</b>	0.35	0.31	-0.10	-0.02	0.37	0.32	-0.09	-0.01	
<b>MSAs: &gt; 1.5 million</b>	0.44	0.15	0.08	-0.01	0.45	0.15	0.08	0.00	

Notes: Reported numbers decompose average wage growth from 0 to 15 years of experience reported in Table 3. Sums of wage growth components reported in this table may differ from the totals in Table 3 due to rounding. The unknown component of wage growth comes from jobs for which we do not observe a wage.

**Table 7: Actual and Predicted Mobility Rates**

Mobility Rates	College Sample		High School Sample	
	Data	Model	Data	Model
<b>Panel A: Transition Rates</b>				
Job Changes	7.1%	7.2%	11.0%	11.1%
Job to Job Transitions	5.2%	5.2%	6.3%	6.2%
Job to Unemployment Transitions	1.9%	2.0%	4.7%	4.8%
Location Changes	2.1%	2.2%	1.7%	1.7%
<b>Panel B: Transition Rates by Location</b>				
Job to Job in Small	4.1%	4.1%	5.9%	5.8%
Job to Job in Medium	4.5%	4.4%	6.1%	5.9%
Job to Job in Large	4.8%	4.7%	5.6%	5.6%
Job to Unemployment in Small	1.6%	1.8%	4.4%	4.6%
Job to Unemployment in Medium	1.7%	1.9%	4.3%	4.5%
Job to Unemployment in Large	1.5%	1.6%	4.4%	4.6%
Unemployment Duration in Small	5.6	5.4	5.9	5.9
Unemployment Duration in Medium	4.6	5.1	5.5	5.1
Unemployment Duration in Large	6.1	5.6	5.0	4.8

Notes: Each entry above the bottom three rows gives the percent of observations exhibiting the indicated transition in actual data and data simulated using estimated parameters reported in Table 10. The numbers in panel B are calculated using only transitions that did not involve a change of location. The final three rows show averages of actual and simulated data across unemployment spells whose lengths are measured in weeks.

**Table 8: Actual and Predicted Migration Conditional on Moving**

Period t-1 Location	Period t Location: Data			Period t Location: Model		
	Small	Medium	Large	Small	Medium	Large
<b>Panel A: College Sample</b>						
<b>Small</b>	11.2%	8.9%	4.7%	10.6%	7.3%	6.8%
<b>Medium</b>	8.3%	20.0%	10.6%	7.3%	21.5%	10.6%
<b>Large</b>	4.6%	11.1%	20.6%	6.8%	9.9%	19.5%
<b>Panel B: High School Sample</b>						
<b>Small</b>	22.2%	10.6%	5.0%	22.6%	8.7%	7.8%
<b>Medium</b>	11.4%	19.4%	4.8%	9.2%	21.1%	7.6%
<b>Large</b>	5.6%	7.6%	13.3%	5.9%	5.8%	11.3%

Notes: Each entry gives the percent of individuals who change locations that move from the location type listed at left to the location type listed at top. The left half of the table reports these numbers from the data and the right half of the table reports these numbers from data simulated using the parameter estimates reported in Table 10.

**Table 9: Log Wages and AFQT Versus Unobserved Ability**

Specification	(1)	(2)	log wage (3)	(4)	(5)
<b>Panel A: College Sample</b>					
<b>MSAs 0.25 - 1.5 million</b>		0.137 (0.016)	0.120 (0.016)	0.167 (0.014)	0.159 (0.014)
<b>MSAs: &gt; 1.5 million</b>		0.093 (0.017)	0.078 (0.017)	0.162 (0.015)	0.155 (0.015)
<b>AFQT</b>	0.035 (0.005)		0.067 (0.006)		0.028 (0.005)
<b>Pr(h=h<sub>H</sub>)</b>	0.616 (0.013)			0.641 (0.013)	0.631 (0.013)
<b>R-Squared</b>	0.37	0.18	0.19	0.38	0.38
<b>Panel B: High School Sample</b>					
<b>MSAs 0.25 - 1.5 million</b>		0.090 (0.010)	0.084 (0.010)	0.157 (0.010)	0.150 (0.010)
<b>MSAs: &gt; 1.5 million</b>		0.044 (0.011)	0.047 (0.011)	0.087 (0.011)	0.087 (0.011)
<b>AFQT</b>	0.054 (0.004)		0.061 (0.004)		0.051 (0.004)
<b>Pr(h=h<sub>H</sub>)</b>	0.273 (0.010)			0.312 (0.010)	0.304 (0.010)
<b>R-Squared</b>	0.21	0.14	0.16	0.22	0.23

Notes: Each entry is a coefficient or standard error from a regression of log wage on the variables listed at left and a quadratic in experience. Each regression uses all available wage observations from those for whom an AFQT score is observed up to 15 years of experience. Regressions have 7,360 observations in Panel A and 10,075 observations in Panel B. AFQT is adjusted for a quadratic in age and is measured in standard deviation units with standard deviations calculated separately for each sample. Pr(h=h<sub>H</sub>) is the estimated probability that each individual in the data set is high type calculated as explained in the text.

Table 10: Parameter Estimates from the Structural Model

Parameter	Description	College Sample		High School Sample	
		Low Type	High Type	Low Type	High Type
A. Components of Wages					
$\beta_0^1$	Wage Constant for Living in Small Locations	6.69***	7.09***	6.41***	6.67***
$\beta_0^2$	Wage Constant for Living in Medium Locations	6.73***	7.13***	6.51***	6.86***
$\beta_0^3$	Wage Constant for Living in Large Locations	6.66***	7.05***	6.31***	6.79***
$\beta_1^1$	Return to Experience from Work in Small	0.029***	0.068***	0.032***	0.027***
$\beta_1^2$	Return to Experience from Work in Medium	0.040***	0.060***	0.040***	0.053***
$\beta_1^3$	Return to Experience from Work in Large	0.045***	0.068***	0.046***	0.032***
$\theta_2$	Multiplier for Living in Medium Locations		1.46***		1.02***
$\theta_3$	Multiplier for Living in Large Locations		1.64***		1.30***
$\beta_2$	Coefficient On Experience Squared in Small Locations		-0.0033***		-0.0022***
$\phi_2$	Multilier for Medium Locations: $\beta_2^2 = \phi_2\beta_2$		1.18***		1.08***
$\phi_3$	Multiplier for Large Locations: $\beta_2^3 = \phi_3\beta_2$		1.37***		1.21***
$\beta_3$	Coefficient On Experience Cubed		0.00017***		0.00009***
B. Amenities					
$\alpha_2$	Amenity of Medium Locations	-0.19***	0.06	-0.06***	-0.30***
$\alpha_3$	Amenity of Large Locations	-0.04	-0.06*	-0.17***	-0.11***
C. Arrival and Separation Rates					
$\lambda_1$	Job Offer Arrival Rate Within Small Locations	0.12***	0.07***	0.13***	0.25***
$\lambda_2$	Job Offer Arrival Rate Within Medium Locations	0.11***	0.09***	0.15***	0.25***
$\lambda_3$	Job Offer Arrival Rate Within Large Locations	0.11***	0.12***	0.14***	0.24***
$\lambda_1^u$	Job Offer Arrival Rate Within Small from Unemployed	0.14***	0.11***	0.11***	0.19***
$\lambda_2^u$	Job Offer Arrival Rate Within Medium from Unemployed	0.16***	0.11***	0.15***	0.22***
$\lambda_3^u$	Job Offer Arrival Rate Within Large from Unemployed	0.14***	0.11***	0.16***	0.20***
$\rho_1$	Receive Rate in Small: $\lambda_{21} = \rho_1\lambda_2$ and $\lambda_{21}^u = \rho_1\lambda_2^u$		0.10***		0.15***
$\rho_2$	Receive Rate in Medium		0.11***		0.12***
$\rho_3$	Receive Rate in Large		0.11***		0.10***
$\gamma$	Own Location Type Multiplier: $\lambda_{11} = \lambda_1\rho_1\gamma$ and $\lambda_{11}^u = \lambda_1^u\rho_1\gamma$		1.98***		2.30***
$\delta_1$	Separation Rate in Small	0.004**	0.004***	0.014***	0.024***
$\delta_2$	Separation Rate in Medium	0.002**	0.012***	0.013***	0.028***
$\delta_3$	Separation Rate in Large	0.005***	0.014***	0.008***	0.031***
D. Benefits and Costs					
b	Unemployment Benefit		0.41***		1.54***
C	Moving Cost		-0.65*		9.36***
E. Heterogeneity					
$\pi^1$	Probability of Type Given Start in Small Location	0.63***	0.37***	0.56***	0.44***
$\pi^2$	Probability of Type Given Start in Medium Location	0.68***	0.32***	0.75***	0.25***
$\pi^3$	Probability of Type Given Start in Large Location	0.66***	0.34***	0.68***	0.32***
F. Distributions					
$\sigma_1$	Standard Deviation of Match in Small		0.37***		0.36***
$\sigma_2$	Standard Deviation of Match in Medium		0.38***		0.30***
$\sigma_3$	Standard Deviation of Match in Large		0.38***		0.33***
$\sigma_u$	Standard Deviation of Utility Shock for Unemployed		5.71***		3.08***
$\sigma_s$	Standard Deviation of Job Switching Cost Shock		5.13***		2.69***
$\sigma_M$	Standard Deviation of Moving Cost Shock		3.34***		9.05***
$\sigma_w$	Standard Deviation of Wage Measurement Error		0.29***		0.26***

Notes: This table shows the universe of estimated model parameters. \*\*\* indicates significance at the 1% level, \*\* indicates significance at the 5% level and \* indicates significance at the 10% level. The discount rate is calibrated to 0.95.

**Table 11: Counterfactual City Size Wage Premia  
College Sample**

Experiment	Regression Coefficients		Gap With Reference	
	Medium	Large	Medium	Large
<b>Panel A: Premia for Comparison</b>				
1. Baseline Data	0.14	0.09		
2. Model	0.15	0.12		
3. Restricted Mobility	0.16	0.16	0.01	0.05
4. Price Differential	0.06	0.18		
<b>Panel B: Counterfactuals With Free Mobility (Row 2 Reference)</b>				
Equalize Ability Distribution at LF Entry	0.17	0.13	0.02	0.01
Equalize Search Across Locations	0.18	0.12	0.04	0.01
Equalize Return to Experience Across Locs	<b>0.08</b>	<b>-0.03</b>	-0.07	-0.14
Implied Reduction in Marginal Productivity Gap			31%	49%
Equalize Nominal Intercepts Across Locs	<b>0.07</b>	<b>0.01</b>	-0.08	-0.11
Implied Reduction in Marginal Productivity Gap			38%	38%
<b>Panel C: Counterfactuals With Restricted Mobility (Row 3 Reference)</b>				
Equalize Ability Distribution at LF Entry	0.19	0.19	0.03	0.03
Equalize Search Frictions Across Locations	0.14	0.15	-0.02	-0.01
Equalize Returns to Experience Across Locs	<b>0.02</b>	<b>-0.04</b>	-0.13	-0.21
Implied Reduction in Marginal Productivity Gap			61%	61%
Equalize Nominal Intercepts Across Locs	<b>0.06</b>	<b>0.03</b>	-0.10	-0.14
Implied Reduction in Marginal Productivity Gap			46%	40%

Notes: Panel A Row 1 presents average wage premia from the raw data, Row 2 shows premia based on simulated data and Row 3 is based on simulated data for which mobility cost is infinite. Other estimates are based on simulated data using parameter values achieving the listed scenario. For ability distribution equalization, we set probabilities of labor force entry by type across locations equal to their weighted average across locations. For search friction equalization, we set all arrival and separation rates equal to their weighted averages across locations, where the weights are based on composition by type across locations at labor force entry. Equalization of returns to experience across locations is achieved analogously. Imposing all restrictions in Panel C simultaneously generates simulated wage premia of 0 in both location types. Equalization of nominal intercepts across locations is achieved by setting intercepts from the wage process to differ by the average price differences across location type given in Row 4 and renormalized such that the weighted average equals the weighted average of real intercept terms reported in Table 10. All wage premia are adjusted by a quadratic in experience. Bolded entries indicate that both bounds of the 95% confidence interval of the counterfactual reported in Table A1 fall outside of the reference.

**Table 12: Counterfactual City Size Wage Premia  
High School Sample**

Experiment	Regression Coefficients		Gap With Fitted Values	
	Medium	Large	Medium	Large
<b>Panel A: Premia for Comparison</b>				
1. Baseline Data	0.09	0.04		
2. Model	0.09	0.05		
3. Restricted Mobility (Compare to Row 2)	0.11	0.06	0.02	0.01
4. Price Differential	0.03	0.19		
<b>Panel B: Counterfactuals With Free Mobility (Row 2 Reference)</b>				
Equalize Ability Distribution at LF Entry	<b>0.12</b>	0.07	0.04	0.02
Equalize Search Frictions Across Locations	<b>0.17</b>	<b>0.11</b>	0.08	0.06
Equalize Returns to Experience Across Locs	<b>0.01</b>	<b>-0.04</b>	-0.08	-0.09
Implied Reduction in Marginal Productivity Gap			70%	37%
Equalize Nominal Intercepts Across Locs	<b>-0.08</b>	<b>-0.13</b>	-0.17	-0.18
Implied Reduction in Marginal Productivity Gap			147%	77%
<b>Panel C: Counterfactuals With Restricted Mobility (Row 3 Reference)</b>				
Equalize Ability Distribution at LF Entry	<b>0.17</b>	0.08	0.06	0.02
Equalize Search Frictions Across Locations	<b>0.17</b>	0.08	0.06	0.02
Equalize Returns to Experience Across Locs	<b>0.03</b>	<b>-0.06</b>	-0.09	-0.12
Implied Reduction in Marginal Productivity Gap			63%	50%
Equalize Nominal Intercepts Across Locs	<b>-0.05</b>	<b>-0.12</b>	-0.16	-0.18
Implied Reduction in Marginal Productivity Gap			119%	71%

Notes: See the notes to Table 11 for a detailed description of the table. Bolded entries indicate that both bounds of the 95% confidence interval of the counterfactual reported in Table A2 fall outside of the reference.

**Table A1: 95% Confidence Intervals for Counterfactual City Size Wage Premia  
College Sample**

Experiment	Medium Locations		Large Locations	
	Low	High	Low	High
<b>Panel A: Wage Premia for Comparison</b>				
Model	0.08	0.21	0.04	0.18
Restricted Mobility	0.07	0.24	0.09	0.26
<b>Panel B: Counterfactuals With Free Mobility</b>				
Equalize Ability Distribution at LF Entry	0.11	0.21	0.07	0.18
Equalize Search Frictions Across Locations	0.10	0.26	0.04	0.20
Equalize Return to Experience Across Locs	0.00	0.14	-0.11	0.04
Equalize Nominal Intercepts Across Locs	0.00	0.14	-0.08	0.07
<b>Panel C: Counterfactuals With Restricted Mobility</b>				
Equalize Ability Distribution at LF Entry	0.14	0.24	0.14	0.25
Equalize Search Frictions Across Locations	0.06	0.22	0.07	0.23
Equalize Return to Experience Across Locs	-0.07	0.13	-0.13	0.06
Equalize Nominal Intercepts Across Locs	-0.02	0.14	-0.06	0.13

Notes: Entries show 95% confidence intervals for the counterfactual simulations reported in Table 11. These confidence intervals are calculated by sampling from the distribution of parameters reported in Table 10 and simulating the data 500 times.

**Table A2: 95% Confidence Intervals for Counterfactual City Size Wage Premia  
High School Sample**

Experiment	Medium Locations		Large Locations	
	Low	High	Low	High
<b>Panel A: Wage Premia for Comparison</b>				
Model	0.06	0.12	0.02	0.09
Restricted Mobility	0.07	0.15	0.02	0.10
<b>Panel B: Counterfactuals With Free Mobility</b>				
Equalize Ability Distribution at LF Entry	0.10	0.15	0.04	0.10
Equalize Search Frictions Across Locations	0.14	0.20	0.07	0.14
Equalize Return to Experience Across Locs	-0.02	0.04	-0.08	0.00
Equalize Nominal Intercepts Across Locs	-0.10	-0.05	-0.17	-0.10
<b>Panel C: Counterfactuals With Restricted Mobility</b>				
Equalize Ability Distribution at LF Entry	0.14	0.19	0.06	0.11
Equalize Search Frictions Across Locations	0.14	0.21	0.04	0.12
Equalize Return to Experience Across Locs	-0.01	0.07	-0.11	-0.02
Equalize Nominal Intercepts Across Locs	-0.09	-0.01	-0.16	-0.08

Notes: Entries show 95% confidence intervals for the counterfactual simulations reported in Table 12. These confidence intervals are calculated by sampling from the distribution of parameters reported in Table 10 and simulating the data 500 times.

Figure 1: Model Fit , College Sample

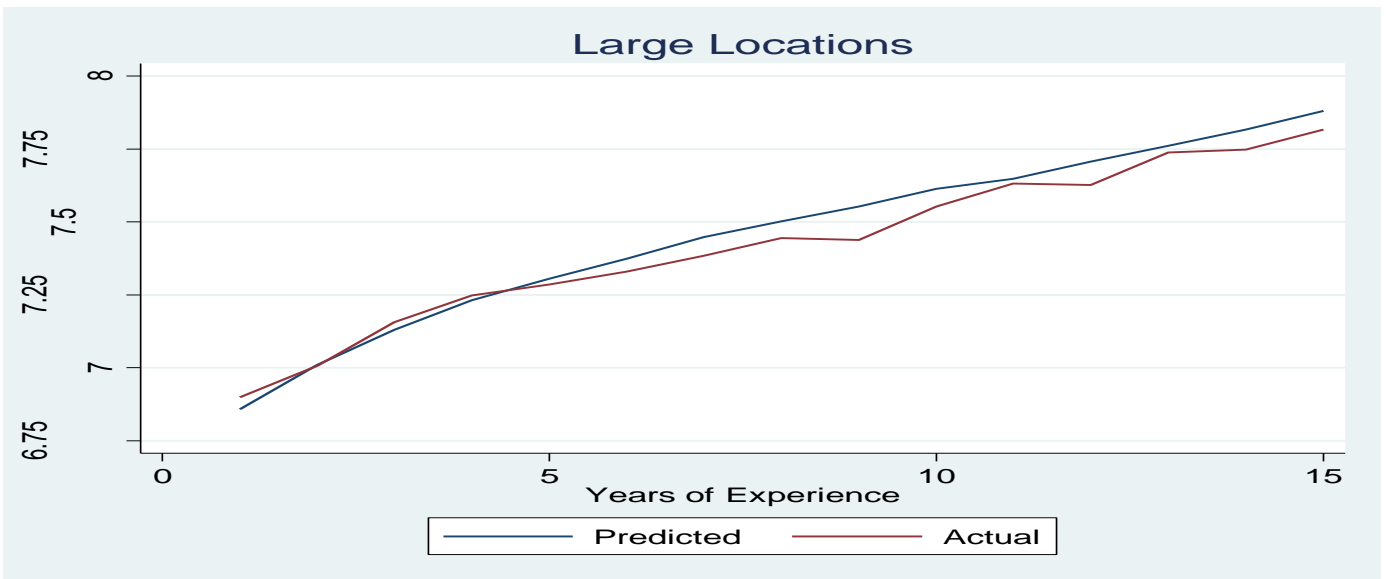
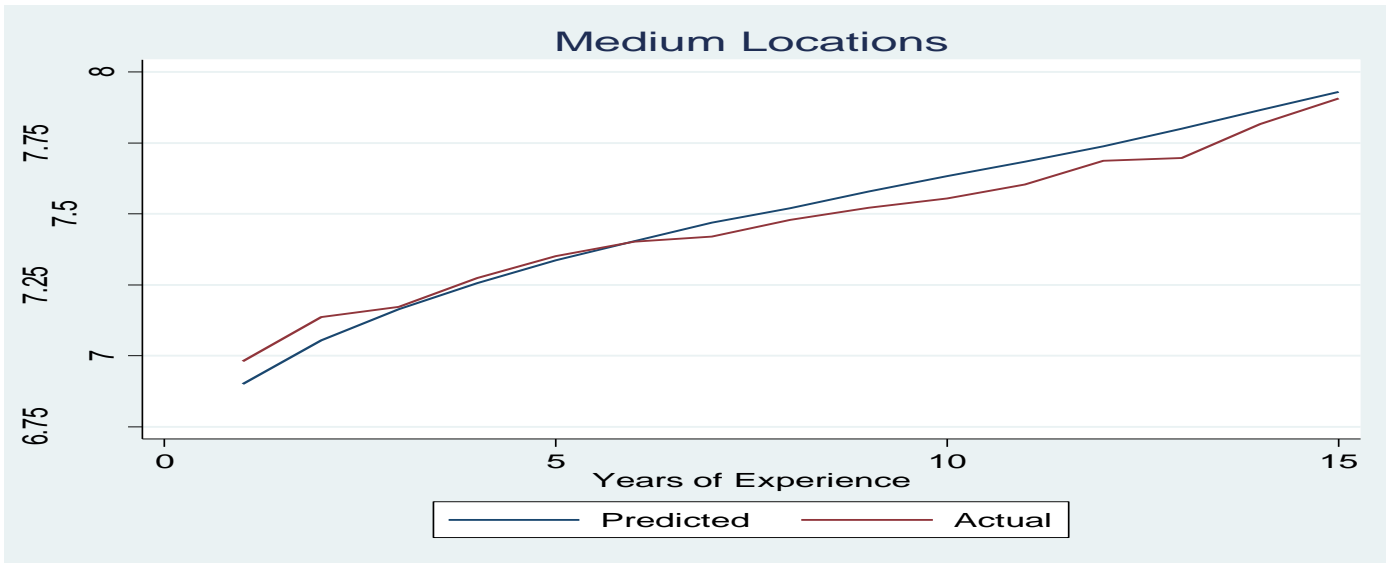
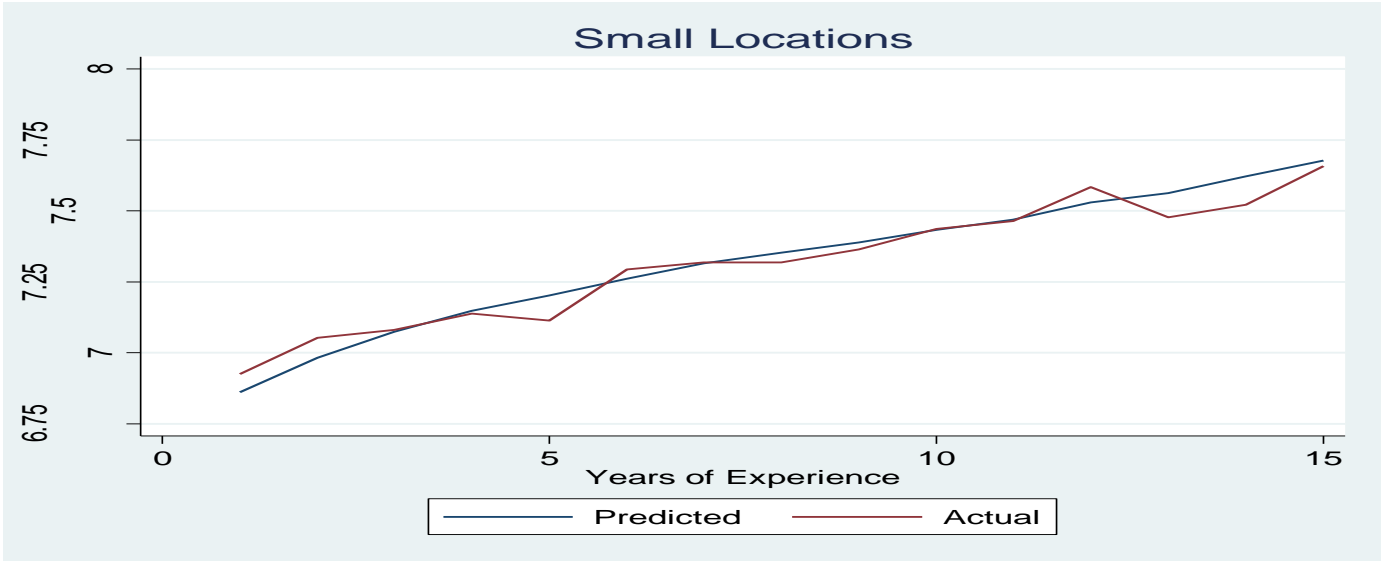


Figure 2: Model Fit , High School Sample

