Understanding the City Size Wage Gap*

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

December, 2008

Abstract

In 2000, wages of full time full year workers were over 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 indicate important roles for sorting on unobserved ability, matching, heterogeneity in returns to experience and the interactions of these factors for generating the city size wage premium. These estimates facilitate a more complete empirical decomposition of the city size wage gap than is possible using results in existing research.

*The authors gratefully acknowledge financial support for this research from the National Science Foundation. Mike Kuklik and Cemal Arbatli provided excellent research assistance. 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 for 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 amenity differences can account for observed differences in wages between cities of different sizes. Our analysis utilizes a model of 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 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. In particular, this paper produces new empirical evidence on the relative importance of various mechanisms by which workers in larger cities are more productive. In addition, given evidence in Baum-Snow & Pavan (2009) that stronger sorting over metropolitan area size has been an important component of increased wage inequality since 1980, our estimates allow us to evaluate the relative importance of reasons behind growing inequality.

Table 1 summarizes patterns in city size wage premia using census data from 1980, 1990 and 2000. Table 1 presents results from regressing 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 group 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 one-quarter to one-third of the two estimated size premia can be explained with observables. Specification 3 shows that including additional controls for local average education levels reduces the gaps further to about one-half of those implied by unconditional means, suggesting that productivity premia can be partially explained through local spillovers.

Parameter estimates indicate that sorting on unobserved ability is an important component of the city size wage premium. Much of the reason this sorting is important is that the profiles of returns to experience and labor market search frictions with respect to city size are different for low and high ability individuals. While low ability individuals are less productive in all types of locations, they hold a clear comparative advantage in small cities and rural areas. Counterfactual experiments reveal that the city size wage premium would be much smaller were returns to experience homogenous across locations and latent types.

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 for equally productive individuals across cities must solely represent amenity differences. 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.

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 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[1]. A considerable amount of empirical evidence supports both explanations. In the 2000 census, 41 percent if 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 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 (2006) 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.

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

[1] In addition, Manning (2008) proposes monopsonistic labor markets in small cities as a mechanism that could generate a city size wage premium.

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 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. However, based on comparisons of within versus between job wage growth, Wheeler (2006) argues that better matching is the primary mechanism by which wage growth is faster in larger cities. Evidence in support of localized agglomeration economies within industries by Arzhagi and Henderson (2006) and Rosenthal & Strange (2003) for example, indicates that input sharing may also be important. 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 (2005) 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 is the potential endogenous sorting that exists across cities on unobserved skill. 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: 1) amenity differences across cities, 2) sorting by ability across cities, 3) differences in arrival rates of job offers ("matching") across cities and abilities, and 4) different returns to experience ("sharing" and "learning") across cities and abilities. 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.

Our results allow us to gain a better understanding of the importance that differences in matching, learning and their interaction with ability between city size categories have for generating productivity differences. These questions are particularly difficult to address with cross-sectional studies because of the inherently dynamic nature of the underlying data generating process. Even given panel data, special care has to be taken to account for the fact that more able workers who are more likely to locate in larger cities also are more likely to receive more wage offers. Structural estimation of a model incorporating both endogenous job search and migration along with a latent ability distribution handles these difficulties.

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 search 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. This paper complements Gould's analysis by incorporating an additional location type, job search, matching and amenities into a similar model. Further, we account for cost of living differences across locations in our data construction.

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 is likely to come from selection of more productive firms into larger cities. Ellison & Glaeser (1997) document that firms systematically locate in ways that generate industrial agglomerations. To the extent that some industries are more productive than others, this pattern 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 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 a 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 the city size wage premium estimated from the NLSY data are very similar to those from the 1990 and 2000 censuses with the wage premium for medium sized cities at 20 percent and that for largest cities at 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 indicates that positive sorting on unobservable skill may be 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. Indeed, evidence presented below indicates that more educated individuals are more mobile. This is an example of why it is fruitful to appeal to a full structural model to evaluate the sources of the city size wage premium.

Table 2 Panel B presents analogous regression results when wages are adjusted for cost of living differences across metropolitan areas.[2] Estimates in Panel B reflect the fact that

[2] The next section details how we implement this cost of living adjustment.

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 5 percent wage cut to live in large metropolitan areas over medium sized metropolitan areas. While the estimated 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 15 to 10 percent while inclusion of individual fixed effects reduces it another 5 percentage points.

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. This indicates 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 city size wage premium, doing so does not eliminate it. Therefore, there is a significant amount to be learned about the reasons for agglomeration economies.

Tables 3 to 6 examine 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 while wages of those entering the labor market are higher in larger cities, this gap increases as a function of experience. At labor force entry, the city size wage premia of medium and large MSAs are 12 and 13 percent respectively. By 15 years of experience, they grow to 34 and 43 percent respectively. The inverse U profile of wages adjusted for cost of living persists at all indicated levels of experienced except 10. This pattern is evidence that the relative amenity values of cities do not change much with experience. Table 3 Panel B shows wage growth rates by experience and city size. The largest cities saw wage growth of 72 percent on average relative to just 49 percent in the smallest areas. Glaeser & Maré (2001) find the same pattern and conclude that the city size wage gap is primarily generated from wage growth differences rather than higher initial wage levels in larger cities. Accounting for cost of living differences does not have much of an effect on wage growth profiles.

As discussed above, one potential explanation for the city size wage growth premium is systematic migration of higher human capital individuals to larger cities over the lifecycle while another is that faster turnover generates more efficient firm-worker matches in larger cities. Table 4 describes patterns in job turnover, unemployment and general human capital accumulation as functions of experience and city size. The left side of Panel A shows that the mean number of jobs held is constant or decreasing in city size at every experience level. The right side of Panel A shows that the mean number of weeks of unemployment is also decreasing in city size. Panel B provides evidence that sorting on human capital levels may explain the patterns in Panels A and B. It shows the fraction who have ever graduated from college and mean years of schooling ever completed by city size. 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. This implies that skilled individuals are systematically migrating to larger cities over the course of their careers.[3]

Table 5 presents evidence of such selective migration. It presents transition matrices between city size categories for high school and college graduates. Panel A 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 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, the structure of the labor market may be such that improvements in match quality is a smaller component of wage growth for low skilled workers than high skilled workers.

Table 6 presents a decomposition of the mean log wage growth numbers by city size for 15 years of experience reported in Table 3 Panel B into four components: within job, between jobs with no unemployment in between, between jobs when individuals experienced 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, within job wage growth is increasing in city size whereas between job wage growth is a fairly flat 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.

[3] Analagous 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.

Consistent with other studies, the descriptive evidence in Tables 5 and 6 points to systematic differences in human capital levels as one important driver of city size productivity differences. Indeed, weeks of unemployment and wage growth profiles match up well to human capital levels across city sizes at 15 years of experience. However, the profile of wage growth due to job to job transitions seen in Table 6 is a fact that has not been recognized before and invites further investigation. The compendium of evidence in Tables 5 and 6 does not indicate that thick labor market search externalities associated with larger cities is enough to generate city size wage gaps. However, it appears that attributes of larger cities that contribute to job specific human capital accumulation are important for generating higher wages in larger cities.

While the descriptive evidence presented here and in other studies points primarily to wage growth over wage level explanations for the city size productivity premium, the mechanism is still far from clear. Evidence presented above indicates that individuals in larger cities match into their primary job more quickly than in smaller cities, potentially allowing for more specific human capital to develop. Existing research does not address the extent to which productivity gaps of larger cities come through better matching or returns to experience, and how this interacts with observed and unobserved skills. 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 computationall 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. 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.

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.[4] Those jobs with multiple reported locations are assigned to the modally reported location. Jobs with multiple modes are assigned the modal location that occured latest in time.[5] 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 the sample, we impute locations to be that of the first job after the unobserved location spell for which we observe location.[6]

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.[7]

We think it is important to allow potential mechanisms behind the city size wage premium to differ by the skill set of workers. Indeed, the city size wage premium is increasing in education. 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

[4] 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.

[5] 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.

[6] We re-estimate the model using location of the previous job instead and results are very similar.

[7] There is a compelling argument to instead split the sample such that rural areas have their own category as in Gould (2007). We plan to reëstimate the model with this categorization as a robustness check.

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.

$$\sum_i \alpha_i \ln(\frac{\alpha_i E^j}{p_i^t}) = ln(U_0) \tag{1}$$

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.

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

This is the index we use to deflate wages across locations and over time.

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.[8] 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.[9] 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 no other MSAs 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 time location/time period j relative to the base location/time period. We define the base time/location as the average ACCRA location from 2001 but deflated to be index value 100 in 1999.[10]

[8] The ACCRA data set commenced being collected in the early 1980s. However, since its coverage and quality increased over time, we choose to use one high quality and complete cross section over trying to use the data set as a panel.

[9] 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.

[10] We choose 1999 as the base year so that 2000 census data does not have to be temporally deflated.

Rather than take expenditure shares α_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 find that the expenditure shares implied by the CEX differ slightly for different education groups and city sizes but these differences have a minimal impact on the resulting price index. As such, we prefer to use the sample from the CEX that best matches our full census sample to calculate one set of expenditure weights that we apply to all individuals in our sample.

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 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. Our most constraining simplifying assumptions limit the number of these underlying worker types to two and city size categories to three.[11] 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 have utility functions that are linear in the sum of a location specific amenity that changes linearly with total work experience $\alpha_{0j} + \alpha_{1j}X$ and their log wage or unemployment benefit b. The different types of locations, characterized by different population size categories, are denoted with subscripts $j \in \{0, 1, 2\}$ respectively. We denote "ability" levels as h_i where $i \in \{0, 1\}$. 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 are functions of worker type. Total labor market experience we express as $X_t = \sum_{k=0}^{2} x_{kt}$. The firm-specific stochastic component of the wage ε 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. Finally, 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

[11] 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.

together, we parameterize the wage process of an individual working in location type j and having experience from location types indexed by k as follows.

$$\ln w_j\left(h, x_0, x_1, x_2, \varepsilon\right) = h + \varepsilon + \beta_0^j + \sum_{k=0}^{2} \beta_{1k}^j\left(h\right) x_k + \beta_2 X^2 + u \tag{3}$$

The wage process expressed in Equation (3) captures the extent to which sorting on ability may influences wage growth and levels differently for small and large cities. 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).

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 $\lambda_j^u\left(h\right)$ for unemployed workers and $\lambda_j\left(h\right)$ for employed workers, where j is the worker's location. The arrival rates of job offers from different locations are $\lambda_{jj'}^u\left(h\right)$ for unemployed workers and $\lambda_{jj'}\left(h\right)$ 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_3 with zero mean and finite variance. 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 $C\left(h\right) + v_2$, where v_2 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 ε at location j as $V_j^{WK}(\varepsilon)$. The state space of all value functions contains the individual specific "ability" h and experience in all location types x_0, x_1, x_2. For expositional simplicity we suppress this dependence in the notation. We assume that the unemployed receive a utility shock v_1 with 0 mean and and finite variance each period.[12] Given these definitions, the present value of being unemployed and working are given by the following expressions.

$$\begin{aligned} V_j^{UN} &= \frac{\alpha_{0j} + \alpha_{1j} X}{3} + b + \beta^{\frac{1}{3}} V_j^u \\ V_j^{WK}\left(\varepsilon\right) &= \alpha_{0j} + \alpha_{1j} X + \ln w\left(\varepsilon\right) + \beta V_j\left(\varepsilon\right) \end{aligned}$$

[12] We do not introduce a utility shock for workers as it would not be separately identified.

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 weekly job history data and index time in months for the unemployed. For this reason, the unemployed agent receives $\frac{1}{3}rd$ the amenity benefit of the worker and discounts utility by $\beta^{\frac{1}{3}}$, whereas the employed worker discounts by β 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 0 with $X < 160$. At the beginning of each period, the agent observes whether he is faced with one of 5 possible scenarios. With probability $\left(1 - \lambda_0^u(h) - \sum_{j=0}^{2} \lambda_{0j}^u(h)\right)$ he does not receive a job offer, with probability $\lambda_0^u(h)$ he receives a job offer from the same location, and with probability $\lambda_{0j}^u(h)$ he receives an offer in location type j.[13] The individual decides at the beginning of each period whether to accept a potential job offer or remain unemployed in location 0. If he accepts an offer, he pays a cost to move to the relevant new location if necessary.

Equation (4) shows the value function for an unemployed agent in location 0.

$$\begin{aligned} V_0^u \quad = \quad & \left(1 - \lambda_0^u(h) - \sum_{j=0}^{2} \lambda_{0j}^u(h)\right) E_{v_1}(V_0^{UN} + v_1) \\ & + \lambda_0^u(h) E_{v_1\varepsilon_0} \max\left[V_0^{UN} + v_1, V_0^{WK}(\varepsilon)\right] \\ & + \sum_{j=0}^{2} \lambda_{0j}^u(h) E_{v_1 v_2 \varepsilon_j} \max\left[V_0^{UN} + v_1, V_j^{WK}(\varepsilon) - \left[C_u(h) + v_2\right]\right] \end{aligned} \tag{4}$$

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 random 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 ε in

[13] The parameter $\lambda_{00}^u(h)$ represents the probability that an unemployed individual at location j receives a wage offer in a different city also of size category 0 whereas the parameter $\lambda_0^u(h)$ represents the probability that this individual receives a wage offer in the same city as his last job.

location type j and the distribution of the random components v_1 and v_2 (expressed as $E_{v_1 v_2 \varepsilon_j}$). The problem for an unemployed individual at location $j' \neq 0$ is symmetric.

The value function for a worker at location 0 resembles that for an unemployed individual except that it also includes potential exogenous job separations and job switching costs. A worker in location 0 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. As such, the value to a worker with ability h at location 0 of being employed with firm match ε is given by Equation (5):

$$
\begin{aligned}
& V_0(\varepsilon) = \delta_0(h) E_{v_1}(V_0^{UN} + v_1) + (1 - \delta_0(h)) \{ \\
& (1 - \lambda_0(h) - \sum_{j=0}^{2} \lambda_{0j}(h)) E_{v_1} \max[V_0^{WK}(\varepsilon), V_0^{UN} + v_1] \\
& + \lambda_0(h) E_{v_1 v_3 \varepsilon_j'} \max\left[V_0^{WK}(\varepsilon), V_0^{UN} + v_1, V_0^{WK}(\varepsilon') - v_3\right] \\
& + \sum_{j=0}^{2} \lambda_{0j}(h) E_{v_1 v_2 v_3 \varepsilon_j'} \max\left[V_0^{WK}(\varepsilon), V_0^{UN} + v_1, V_j^{WK}(\varepsilon') - v_3 - [C(h) + v_2]\right] \}
\end{aligned}
\tag{5}
$$

As is evident in Equation (5), an exogenously separated worker at location 0 may only become unemployed in location 0. If the worker is not exogenously separated, he does not receive a wage offer with probability $\left(1 - \lambda_0(h) - \sum_j \lambda_{0j}(h)\right)$. In this event, he has the option of remaining employed in the same job or going unemployed. If he receives an offer, he either accepts it and moves if necessary, remains at his old job, or goes unemployed in his current location. The value function for the worker at other locations is symmetric to that in Equation (5).

To conceptualize how the model works, it is convenient to define a set of reservation functions $\{\varepsilon_j^A, \varepsilon_{jj'}^B, \varepsilon_j^C, \varepsilon_{jj'}^D\}$ 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. Regime A occurs when an unemployed agent receives an own-location job offer or a worker is choosing whether to go unemployed. Regime B occurs when an unemployed agent receives an offer in another location. Regime C occurs when a worker receives an own-location offer. Regime D occurs when a worker receives an offer in another location. For simplicity, we suppress dependence of the reservation functions on type h and work experience in each location $\{x_0, x_1, x_2\}$.The definitions of these reservation rules are given as follow where $\widetilde{\varepsilon}$ is the match quality of the incumbent job.

$$\varepsilon_0^A(v_1) \text{ solves } V_0^{UN}(x_0,x_1,x_2,v_1) = V_0^{WK}\left(x_0',x_1,x_2,\varepsilon\right)$$
$$\varepsilon_{01}^B(v_1,v_2) \text{ solves } V_0^{UN}(x_0,x_1,x_2,v_1) = V_1^{WK}\left(x_0,x_1',x_2,\varepsilon\right) - C(h) - v_2$$
$$\varepsilon_0^C(\widetilde{\varepsilon},v_1,v_3) \text{ solves}$$
$$\max\left[V_0^{WK}(x_0',x_1,x_2,\widetilde{\varepsilon}), V_0^{UN}(x_0,x_1,x_2,v_1)\right] = V_0^{WK}\left(x_0',x_1,x_2,\varepsilon\right) - v_3$$
$$\varepsilon_{01}^D(\widetilde{\varepsilon},v_1,v_2,v_3) \text{ solves}$$
$$\max\left[V_0^{WK}(x_0',x_1,x_2,\widetilde{\varepsilon}), V_0^{UN}(x_0,x_1,x_2,v_1)\right] = V_1^{WK}\left(x_0,x_1',x_2,\varepsilon\right) - C(h) - v_2 - v_3$$

Reservation rules C and D include the max over the values of working the current job or going unemployed because the shock v_1 might be sufficiently high such that if the individual does not accept the job offer, he would go unemployed rather than stay at his current job. The reservation rules are specified to take account of the fact that every example of migration must result in accepting a new job offer. The shocks v_1, v_2 and v_3 are meant to capture the effects of non-monetary factors that can affect workers' behavior. For example, a worker that accepts a lower paid job because he likes that job better can be captured by a low value of v_3.

This model captures each of the components of wage growth discussed in the introduction. The location specific component of utility $\alpha_{0j}+\alpha_{1j}X$ captures the difference in amenity values. The job arrival rate parameters $\lambda_{jj'}(h)$ and $\lambda_{jj'}^u(h)$ capture the potential importance of matching. The moving costs $C(h)$ regulate ability sorting across locations. Finally, the coefficients on experience in Equation (3) capture differences in "learning" while the intercepts capture differences in "sharing".

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:

$$L(\theta) = \sum_{i=0}^{1} \pi_j^i f\left(Y^T | h_i; \theta\right)$$

where π_j^i is the probability that an individual is of type i given that he enters in location j and θ is the vector of parameters.[14]

Define Y^t as the vector of all labor market observations in this individual's job history up to and including period t. These include location at each point in time, location specific experience and wages and unemployment status. We decompose $f\left(Y^T|h;\theta\right)$ as follows.

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

In the previous section we saw how the job switching and migration behavior of individuals depends on the set of four classes of reservation rules. It is more convenient to express the likelihood function in terms of probabilities that one of five types of event occurs. These probabilites depend on $(h,\varepsilon,\varepsilon',x)$, where x indicates the vector of state variables $\{x_0,x_1,x_2\}$ and ε' is the quality of a new match if offered.

$$\left\{P_{eu}^j\left(h,\varepsilon,x\right), P_{ue}^j\left(h,\varepsilon,x\right), P_{ee}^j\left(h,\varepsilon,\varepsilon',x\right), P_{ue}^{jj'}\left(h,\varepsilon,x\right), P_{ee}^{jj'}\left(h,\varepsilon,\varepsilon',x\right)\right\}_{j,j'=0}^{2}$$

These functions capture the probabilities of each transition and that the new match quality ε is drawn, and facilitate derivation of the likelihood function. For example, the transition probability between unemployment and employment at location j with firm specific component ε is as follows.

$$P_{ue}^j\left(h,\varepsilon,x\right) = \lambda_j^u\left(h\right) f_{\varepsilon_j}(\varepsilon)\int_{-\infty}^{V_j^{WK}-V_j^{UN}} dF(v_1)$$

The offer ε at location j is received by an unemployed worker with probability $\lambda_j^u\left(h\right) f_{\varepsilon_j}(\varepsilon)$ and is accepted only if $V_j^{WK} - V_j^{UN} > 0$. Appendix A specifies the remaining probabilities and specifies in more detail how we build the functions $f\left(Y^1|h;\theta\right)$ and $f\left(Y^t|Y^{t-1},h;\theta\right)$ using the information we have in the data. There are two important considerations here. First, wages are not always observed when they should be. We assume that these missing wage observations happen with an exogenous probability and we condition that probability out of the likelihood function since it is not of interest. Secondly, we have to deal with the fact that the firm-worker match ε is unobserved. The Appendix shows how we use Bayesian updating methods to handle this difficulty.

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

[14] The individual index is suppressed for notational simplicity.

over time, these type-specific parameters are identified, subject to standard constraints on identification. The distribution of types is nonparametrically identified. Kasahara (forthcoming) discusses identification of parameters in this class of models.

We cannot nonparametrically identify distributions of the firm specific wage components ε_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 ε_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 $N(0, \sigma_{\varepsilon_j})$.

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 different 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. In fact, we leverage the life cycle nature of model to strengthen separate identification of these different parameters.

Table 7 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. The wage shifter for high ability workers is recoverable from the fact that the mean ability wage shifter is 0 and the estimated probability that a given worker is low skilled that we call "weight" in the heterogeneity group and is thus not separately identified. We choose to 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.[15]

The one parameter of the model that we do not estimate is the discount factor β. This is standard practice in the structural estimation of search models, as β is not generally separately identified from the unemployment benefit b. Based on estimates from the literature, we set the discount factor to 0.95 per year.

[15] As seen in the lists of parameters in the first two groups in Table 7, we make some assumptions that limit the number of parameters needed to capture heterogeneity in prices of experiences in different locations and arrival rates of job offers from one location type to another. Details on our procedure for doing this are in Appendix B.

6 Results

6.1 Model Fit

Figures 1 and 2 show graphs of actual log hourly wages (in cents) and those predicted by the model as functions of experience for high school graduates. Figure 1 shows these objects using the full data set while Figure 2 shows them broken out by location type. Figure 1 shows that the actual and simulated data lines coincide very closely at all levels of experience. Examining graphs of actual and predicted wages by location in Figure 2 reveals some indication about where our specification could use some improvement. Panel A shows that we underpredict the true wage at 8 to 15 years of experience in small cities and rural areas. Panel B shows quite a good fit in medium sized cities. Panel C shows that for large cities we overpredict at low levels of experience and underpredict at high levels of experience. One explanation for these differences is that returns to experience may be more concave in large locations than in small locations and we do not allow the concavity parameter β_2 to differ by location type. We will correct this weakness in future versions of the paper.

6.2 Parameter Estimates

Table 7 presents parameter estimates of the structural model. As discussed in the identification subsection, we have broken the parameter set into six categories. Our discussion focuses on the results for high school graduates given that model simulations fit the data for this population better than the college population.

Category A is location and ability specific wage level and growth estimates. Note that the initial wage levels for each location type for the high school sample follow the inverse U shaped pattern of wages overall, and essentially summarize the CPI and ACCRA deflated patterns in Table 3 at experience 0 for a more select sample. The easiest set of returns to experience parameters to interpret are the triads $(\beta_{10}, \beta_{11}^0, \beta_{12}^0)$ and $(\beta_{10}^0, \beta_{11}^1, \beta_{12}^2)$. β_{10} denotes the common component of the price of a year of experience in location 0 at other locations. The remaining two elements of the first triad give the price of one year of experience in locations 1 and 2 when working in location 0. The second triad gives prices of a year of experience in each location type for those working in that same location type. First, note that in all cases, type 1 individuals get greater returns to experience than type 0 individuals, as should be expected. More importantly, note that in the college sample returns to experience are more steeply increasing in city size for high latent type individuals than low latent type individuals. In the high school sample, returns to experience are very similar for medium and large cities at 0.27 to 0.29 per year for the low types and 0.42 per year for the high types. The college sample exhibits more of an inverse U shaped pattern for the low types.

Block B of Table 7 reports estimated amenity parameters. Since amenities for small locations are not separately identified, all estimates should be viewed as being relative to the amenity value of small cities. Results for amenities reveal that for the high school sample, the highest amenity locations at labor force entry are the medium sized cities, with rural places fairly close behind. The amenity values of both medium and large places decline with experience, however, such that rural places have the highest amenity value after 6 years of experience. Members of the college sample place the highest amenity value on the largest places and the smallest amenity value on rural places. For this group, the large city amenity value grows over time while the medium sized city amenity value is stable. These results are consistent with Albouy (2008) who demonstrates that it is possible to get a structural spatial equilibrium model to generate equilibria in which large cities have the highest amenities.

Table 7 Block C reports estimated job arrival rates by location and worker type. Interestingly, while low ability workers have higher arrival rates in small cities and rural areas, high ability workers have much flatter arrival rates as a function of city size. Arrival rates of high ability types are slightly greater than those for low ability types in medium and large cities. These statements are largely true for arrival rates from unemployed and for job changes across different cities of the same size as well. In addition, exogenous separation rates, though small everywhere, are larger in rural areas than in cities for the low type only. Overall, these results on search frictions provide some evidence that at least for high ability individuals, matching might happen slightly more efficiently in larger cities. However, the results are mixed and we will have to defer to the simulation results for definitive conclusions.

Block D reveals that the benefit required to get individuals to go unemployed is higher for high school graduates than for college graduates. This result makes since given the greater returns to experience, and associated higher opportunity cost of unemployment, enjoyed by college graduates. Additionally, low ability types are estimated to have higher implied moving costs than high ability types. This result may reflect their higher psychic costs of setting up in a new city.

Results in Block E demonstrate that there is selection on unobserved worker ability from the beginning. Results in Block F show estimated standard deviations of all of the distributions in the structural model. The one that stands out is the standard deviation of shock 2, which is the moving shock. Apparently there are many idiosyncratic reasons why individuals move that are not captured by the model.

6.3 Simulations

Using the parameter estimates from the structural model, we evaluate the importance of potential mechanisms for generating the city size wage premium. Table 8 displays the city size wage premia for the high school sample implied under various counterfactual scenarios. As a baseline, the top row shows that the raw data imply city size wage premia of 0.10

for medium sized cities and 0.03 for large cities.[16] Data simulated from the parameter estimates in Table 7 imply numbers of 0.14 and 0.01 respectively. While the model does overshoot a bit on the medium cities, these numbers are reasonable. We treat the city size real wage premia from the predicted data as a baseline for our analysis of counterfactual scenarios.

We generate simulated data sets shutting down various potential drivers of the city size wage premium. We do this in three dimensions. In one dimension, we equalize various objects across locations. These are the ability distribution at experience 0, search frictions, and returns to experience. The second dimension is equalization across latent types. In particular, we run simulations in which we equalize search frictions and returns to experience. The third dimension is whether we allow mobility across locations or individuals are constrained to work in their initial location forever. We evaluate all counterfactuals with and without mobility.

Most of the counterfactual experiments with no mobility restriction reduce the city size wage premium. The reason is that each experiment reduces at least one incentive for individuals to sort across locations. The only exception is when we equalize the ability distribution across locations. In this case, the city size wage premia increase. The reason is that a bunch of people are forced to start off their careers in locations that would not have been their choice but nevertheless have higher returns to experience. This experience in larger cities is capitalized into larger wages for the rest of their working lives.

The largest reductions in the city size wage premia come from equalizing returns to experience across locations and types. Doing so reduces the gap between wages in rural areas and cities by about 10 percentage points regardless of whether mobility is restricted. The fact that the mobility option does not affect this result indicates that sorting does not interact with returns to experience to generate the city size wage premium. However, heterogeneity is still an important element. Indeed, equalizing returns to experience across locations but allowing them to differ by type generates much smaller reductions in the wage premium of large cities with or without the mobility restriction.

Differences in search frictions across locations also appear to generate some of the city size wage premium. With no mobility restriction, equalizing search frictions across locations reduces the premium by 0.06 for medium sized cities and 0.03 for large cities. However, this reduction is smaller when search frictions are equalized across locations and types. Overall, it appears that heterogeneity in returns to experience is a much more important force generating the city size wage premium than other factors.

[16] For the purpose of generating counterfactuals, we choose not to undo the adjustment to wages for cost of living differences. The reason is that the model has nothing to say about in which metropolitan areas individuals will live within size categories. In the data, average cost of living is about 5 percent higher in medium sized places than small places, and an additional 14 percent higher in the largest metropolitan areas.

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 census and the NLSY, we show that hourly wages are higher in bigger cities, and this is associated with higher levels of human capital and lower levels of unemployment. 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. While interesting, these observations are not sufficient to causally determine the mechanism by which cities are more productive. In particular, estimation of the model specified in this paper allows us to sort out the extent to which sorting across locations on ability interacts with learning and matching to generate city size wage and productivity gaps.

Parameter estimates indicate that sorting on unobserved ability is an important component of the city size wage premium. Much of the reason this sorting is important is that the profiles of returns to experience and labor market search frictions with respect to city size are different for low and high ability individuals. While low ability individuals are less productive in all types of locations, they hold a clear comparative advantage in small cities and rural areas. Counterfactual experiments reveal that the city size wage premium would be much smaller were returns to experience homogenous across locations and latent types.

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.

A.1 Fundamentals

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

$$f\left(\varepsilon|Y^t,h\right)=\frac{f\left(Y^t|Y^{t-1},h,\varepsilon\right)f\left(\varepsilon|Y^{t-1},h\right)}{f\left(Y^t|Y^{t-1},h\right)} \tag{6}$$

This expression is used extensively as we build components of the likelihood function below.

Additionally, 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. We index the function A to period $t-1$ because it contributes to the likelihood function in period t as described below. These functions include the parameter p_n, the probability of observing a wage, which is to be estimated.

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

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

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

instead.

A.2 Transition Probabilities

In cases where a new match is drawn and the worker has an existing match quality, ε' denotes the new match and ε denotes the firm-specific component of the existing job. If the worker is unemployed, ε denotes the new match draw.

$$\begin{aligned}
P^j_{ue}(h,\varepsilon) &= \lambda^u_j(h) f_{\varepsilon_j}(\varepsilon)\int_{-\infty}^{V^{WK}_j(\varepsilon)-V^{UN}_j} dF(v_1) \\
P^{jj'}_{ue}(h,\varepsilon) &= \lambda^u_{jj'}(h) f_{\varepsilon_{j'}}(\varepsilon)\int\int_{-\infty}^{V^{WK}_{j'}(\varepsilon)-C_u(h)-v_2-V^{UN}_j} dF(v_1)dF(v_2) \\
P^j_{eu}(h,\varepsilon) &= \delta_j(h)+(1-\delta_j(h))\int_{V^{WK}_j(\varepsilon)-V^{UN}_j}^{\infty} dF(v_1)\{ \\
&\quad \lambda_j(h)(\int\int\int_{V^{WK}_j(\widetilde{\varepsilon})-v_3-V^{UN}_j}^{\infty} dF(v_1)dF(v_3)dF_{\varepsilon_j}(\widetilde{\varepsilon})) \\
&\quad +\sum_{j'}\lambda_{jj'}(h))(\int\int\int\int_{V^{WK}_{j'}(\widetilde{\varepsilon})-C(h)-v_2-v_3-V^{UN}_j}^{\infty} dF(v_1)dF(v_3)dF(v_2)dF_{\varepsilon_{j'}}(\widetilde{\varepsilon}))\} \\
P^j_{ee}(h,\varepsilon,\varepsilon') &= \lambda_j(h) f_{\varepsilon_j}(\varepsilon')\int_{-\infty}^{V^{WK}_j(\varepsilon')-V^{WK}_j(\varepsilon)} dF(v_3) \\
P^{jj'}_{ee}(h,\varepsilon,\varepsilon') &= \lambda_{jj'}(h) f_{\varepsilon_{j'}}(\varepsilon')\int\int_{-\infty}^{V^{WK}_j(\varepsilon')-v_2-C(h)-V^{WK}_j(\varepsilon)} dF(v_3)dF(v_2)
\end{aligned}$$

A.3 First Period

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

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

The resulting posterior distribution of the firm match is:

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

A.4 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

$$\Pi_2^j(h, NW_t) = \left(1 - \int P_{ue}^j(h,\varepsilon)\, d\varepsilon - \sum_{j=0}^{2} \int P_{ue}^{jj'}(h,\varepsilon)\, d\varepsilon\right)^{NW_t - 1}$$

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:

$$L_{2a}^j = \left(1 - \int P_{ue}^j(h,\varepsilon)\, d\varepsilon - \sum_{j=0}^{2} \int P_{ue}^{jj'}(h,\varepsilon)\, d\varepsilon\right)^{NW_t - 1} \int B_{jt}(h,\varepsilon)\, P_{ue}^j(h,\varepsilon)\, d\varepsilon$$

The posterior distribution of the match then becomes:

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

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

$$L_{2b}^j = \left(1 - \int P_{ue}^j(h,\varepsilon)\, d\varepsilon - \sum_{j=1}^{3} \int P_{ue}^{jj'}(h,\varepsilon)\, d\varepsilon\right)^{NW_t - 1} \int B_{j't}(h,\varepsilon)\, P_{ue}^{jj'}(h,\varepsilon)\, d\varepsilon$$

The posterior distribution of the match is then:

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

A.5 Becoming Unemployed

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

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

A.6 Working

The first case is that he stays with the same employer. The likelihood contribution and the conditional distribution of the firm match can be written as:

$$\begin{aligned}
L_{4a}^j &= \int B_{jt}(h,\varepsilon)\left(1-P_{eu}^{j}(h,\varepsilon)-\int P_{ee}^{j}\left(h,\varepsilon,\varepsilon'\right)d\varepsilon'-\sum_{j=0}^{2}\int P_{ee}^{jj'}\left(h,\varepsilon,\varepsilon'\right)d\varepsilon'\right)dF\left(\varepsilon|Y^{t-1},h\right)\\
f\left(\varepsilon|Y^{t},h\right) &= \frac{B_{jt}(h,\varepsilon)\left(1-P_{eu}^{j}(h,\varepsilon)-\int P_{ee}^{j}(h,\varepsilon,\varepsilon')\,d\varepsilon'-\sum_{j=0}^{2}\int P_{ee}^{jj'}(h,\varepsilon,\varepsilon')\,d\varepsilon'\right)f\left(\varepsilon|Y^{t-1},h\right)}{L_{4a}^j}
\end{aligned}$$

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

$$\begin{aligned}
L_{4b}^j &= \int\int A_{jt-1}(h,\varepsilon)\, B_{jt}\left(h,\varepsilon'\right) P_{ee}^{j}\left(h,\varepsilon,\varepsilon'\right) dF\left(\varepsilon|Y^{t-1},h\right) d\varepsilon'\\
f\left(\varepsilon'|Y^{t},h\right) &= \frac{B_{jt}(h,\varepsilon')\int A_{jt-1}(h,\varepsilon)\, P_{ee}^{j}(h,\varepsilon,\varepsilon')\, dF\left(\varepsilon|Y^{t-1},h\right)}{L_{4b}}
\end{aligned}$$

Note that the inclusion of the function $A_{jt-1}(h,\varepsilon)$ captures the fact that we should have observed the wage of the previous job in the previous period.

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

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

B Parameter Normalizations

In order to limit the set of parameters that we estimate, we impose some constraints on the parameter vector. These constraints are for the set of arrival rates λ and the linear returns to experience β_1 and are described below.

B.1 Job Offer Arrival Rates

Define $\lambda_j^u(h)$ as the arrival rate of job offers for an unemployed individual from the same city given that a worker is of type h and the worker is at location type j. These arrival rates have a structural meaning and are estimated. To derive other arrival rates for unemployed workers, we use the $\lambda_j^u(h)$ parameters and multiply them by one or two additional objects: α_j and γ. Define $\alpha_{j'}$ to be a multiplier for arrival rates to a given city $j\prime$ and γ to be a parameter that scales the product $\lambda_j^u(h)\alpha_{j'}$ if the two location sizes are the same or $j = j\prime$. We use the same scaling factor for the worker arrival rates $\lambda_j^u(h)$. This reduces the number of arrival rate parameters estimated from 24 to 10.

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

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

B.2 Returns to Experience

Recall that the linear portion of return to experience for an individual working in location j and with work history in other locations k is specified in (3) as $\sum_{k=0}^{2} \beta_{1k}^{j}(h)\, x_k$. The goal is to limit the number of parameters β_{1k}^j. We estimate 6 parameters $\beta_{1k}^0(h)$ which are the base returns to experience gained at location k when an agent is working at location 0. For example, $\beta_{11}^0(h_0)$ is the price of one year of a low skilled person's past experience working in location 1 while he is working in location 0. These parameters are multiplied by $\beta_1^1(h)$ or $\beta_1^2(h)$ for current work locations 1 and 2 respectively. Additionally, if the work experience location is the same as the current work location, then we multiply by an additional factor m_j.

$$
\begin{aligned}
\beta_{1k}^j(h) &= \beta_{1k}^0(h)\beta_1^j(h) \text{ if } j > 0 \text{ and } j \neq k \\
\beta_{1k}^j(h) &= \beta_{1k}^0(h)\beta_1^j(h)m_j \text{ if } j > 0 \text{ and } j = k \\
\beta_{10}^0(h) &= \beta_{10}(h)m_0
\end{aligned}
$$

References

David Albouy 2008 “Are Big Cities Really Bad Places to Live? Improving Quality of Life Estimates Across Cities” working paper

Arzhagi, Muhammad and J. Vernon Henderson, 2006 “Networking off Madison Avenue,” manuscript.

Basker, Emek 2005 “Selling a Cheaper Mousetrap: Wal-Mart’s Effects on Retail Prices” *Journal of Urban Economics* 58:2 203-229

N. Baum-Snow and R. Pavan. 2009 “Inequality and City Size” working paper

Carlino, Gerald, Satyajit Chatterjee and Robert Hunt, 2007. “Urban Density and the Rate of Invention,” *Journal of Urban Economics,* 61 389-419.

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, 2006. "Spatial Wage Disparities: Sorting Matters!" 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.

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.

Glaeser, Edward, Kallal, Hedi, Scheinkman, Jose and Andrei Shleifer, 1992. “Growth in Cities,” *Journal of Political Economy* 100 1126-1152.

Gould, Eric. 2007 “Cities, Workers and Wages: A Structural Analysis of the Urban Wage Premium” *Review of Economic Studies*

M. Greenstone, R. Hornbeck and Enrico Moretti. 2008 “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.

Kasahara. forthcoming “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices.” *Econometrica*

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.

Lee, Sanghoon. “Ability Sorting and Consumer City,” manuscript.

Moretti, Enrico, 2004. "Workers’ Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions," *American Economic Review* 94 656-690.

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

Pavan, Ronni, 2006. “Career Choice and Wage Growth,” manuscript.

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

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.

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

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 the City Size Wage Premium from Census Micro Data

	S1: No Controls			S2: Individual Controls			S3: Individual + Local Ed. Controls		
	1980	**1990**	**2000**	**1980**	**1990**	**2000**	**1980**	**1990**	**2000**
MSAs: 250,000 - 1.5 million	0.141***	0.175***	0.185***	0.0931***	0.130***	0.119***	0.0689***	0.0967***	0.0754***
	(0.0104)	(0.00984)	(0.00963)	(0.00843)	(0.00755)	(0.00673)	(0.00855)	(0.00783)	(0.00663)
MSAs: > 1.5 million	0.238***	0.311***	0.315***	0.170***	0.238***	0.219***	0.144***	0.186***	0.155***
	(0.0141)	(0.0106)	(0.00967)	(0.0106)	(0.00769)	(0.00643)	(0.0105)	(0.00872)	(0.00696)
R-squared	0.029	0.046	0.039	0.251	0.336	0.293	0.257	0.345	0.305

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 the control variables indicated at top. Individual controls include age, age squared and indicators for nine levels of education. Local education controls are the fraction of the sample population with each of these nine education levels living in the local county group or PUMA. Standard errors are clustered at the county group or PUMA level. Sample sizes are 1,269,114 in 1980, 1,330,116 in 1990 and 1,245,865 in 2000.

Table 2: Estimates of the City Size Wage Premium from NLSY Data

	No Controls	Individual Controls	Individiual + Cnty Educ. Cntrls	Individual Controls and Fixed Effects	All
Specification	**1**	**2**	**3**	**4**	**5**
			Panel A: CPI Deflated		
MSAs: 250,000 - 1.5 million	0.20	0.15	0.11	0.08	0.07
	(0.008)**	(0.007)**	(0.007)**	(0.012)**	(0.012)**
MSAs: > 1.5 million	0.30	0.23	0.17	0.14	0.12
	(0.008)**	(0.007)**	(0.008)**	(0.013)**	(0.014)**
R-squared	0.04	0.30	0.31	0.25	0.26
			Panel B: ACCRA Deflated		
MSAs: 250,000 - 1.5 million	0.15	0.10	0.08	0.05	0.04
	(0.008)**	(0.006)**	(0.007)**	(0.011)**	(0.012)**
MSAs: > 1.5 million	0.11	0.05	0.02	0.00	-0.01
	(0.008)**	(0.007)**	(0.008)**	(0.013)	(0.013)
R-squared	0.01	0.29	0.29	0.27	0.27

Each regression uses 1754 individuals and has 30,367 observations based on quarterly data. The sample includes white men in the NLSY79. Complete sample selection rules are explained in the text.

Table 3: Wages As a Function of Labor Force Experience

	CPI Deflated					CPI & ACCRA Deflated				
Years of Labor Force Experience	**0**	**1**	**5**	**10**	**15**	**0**	**1**	**5**	**10**	**15**
Panel A: Mean Wages										
Rural and MSAs < 0.25 million	9.00	9.56	12.08	13.69	16.07	9.46	10.06	12.94	15.04	17.75
MSAs: 250,000 - 1.5 million	10.18	11.34	14.59	17.79	21.71	10.44	11.60	14.86	18.34	22.60
MSAs: > 1.5 million	10.37	11.89	16.39	21.09	23.99	9.35	10.70	14.81	19.51	22.62
Panel B: Mean log Wage Growth from Entry into Labor Force										
Rural and MSAs < 0.25 million		0.06	0.28	0.36	0.52		0.06	0.30	0.41	0.58
MSAs: 250,000 - 1.5 million		0.11	0.36	0.54	0.68		0.11	0.37	0.57	0.72
MSAs: > 1.5 million		0.12	0.44	0.65	0.76		0.11	0.43	0.66	0.78

The sample is the same as that for Table 2 with the additional limitation that we must observe individuals to at least 15 years of labor force experience. This limitation reduces the number of individuals in the sample by 20 percent. Numbers in Panel A are in units of 1999 dollars (left panel) and 1999 dollars in the average priced metropolitan area (right panel).

Table 4: Turnover, Unemployment and Human Capital

Years of Labor Force Experience	0	1	5	10	15	0	1	5	10	15
Panel A: Turnover & Unemployment										
	Number of Jobs					Weeks of Unemployment				
Rural and MSAs < 0.25 million	1	1.6	3.5	5.5	6.7	0.0	2.9	16.2	27.7	34.3
MSAs: 250,000 - 1.5 million	1	1.6	3.5	5.2	6.6	0.0	2.0	12.7	21.2	27.1
MSAs: > 1.5 million	1	1.5	3.2	4.8	5.9	0.0	1.7	9.9	15.7	21.0
Panel B: Human Capital										
	Fraction College Graduate					Years of Schooling				
Rural and MSAs < 0.25 million	0.22	0.22	0.21	0.21	0.21	13.20	13.22	13.14	13.13	13.12
MSAs: 250,000 - 1.5 million	0.35	0.35	0.35	0.35	0.35	13.84	13.80	13.86	13.79	13.80
MSAs: > 1.5 million	0.38	0.39	0.39	0.39	0.39	14.17	14.22	14.18	14.29	14.31

See Table 3 for a description of the sample. Panel B uses highest grade ever completed to define schooling variables.

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 250,0000 to 1.5 million	MSAs MSAs: > 1.5 million
Panel A: High School Only			
Rural and MSAs < 0.25 million	0.87	0.09	0.04
MSAs: 250,000 - 1.5 million	0.11	0.87	0.02
MSAs: > 1.5 million	0.10	0.09	0.81
Panel B: College or More			
Rural and MSAs < 0.25 million	0.64	0.24	0.12
MSAs: 250,000 - 1.5 million	0.10	0.76	0.14
MSAs: > 1.5 million	0.08	0.18	0.74

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.

|
|
|

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

	CPI Deflated				CPI & ACCRA Deflated			
	Within Job	**Job to Job**	**Job to Unemp**	**Un-Known**	**Within Job**	**Job to Job**	**Job to Unemp**	**Un-Known**
Rural and MSAs < 0.25 million	0.30	0.24	-0.03	0.01	0.33	0.26	-0.02	0.01
MSAs: 250,000 - 1.5 million	0.44	0.30	-0.06	0.00	0.45	0.31	-0.05	0.00
MSAs: > 1.5 million	0.50	0.24	0.01	0.00	0.51	0.25	0.01	0.00

Reported numbers decompose the wage growth results reported in Table 3. For the purpose of these calculations, we assign individuals only to their locations at 15 years of experience. Sums of wage growth components may differ from Table 3 due to rounding.

Table 7: Parameter Estimates from the Structural Model

Parameter	Description	High School Sample Low Type	High School Sample High Type	College Sample Low Type	College Sample High Type
A. Components of Wages					
β_0^0	Wage Constant for Location 0	6.61		6.77	
β_0^1	Wage Constant for Location 1	6.73		6.83	
β_0^2	Wage Constant for Location 2	6.58		6.90	
h	Wage shifter for ability	-0.16	0.17	-0.29	0.18
β_{10}	Return to Experience from Small when elsewhere	0.017	0.032	0.035	0.026
β_{11}^0	Return to Experience from Medium when wrking in Small	0.022	0.043	0.062	0.057
β_{12}^0	Return to Experience from Large when working in Small	0.022	0.041	0.063	0.078
β_1^1	Multiplier, Medium Locations	1.0	0.7	1.1	1.3
β_1^2	Multiplier, Large Locations	1.6	1.3	0.8	0.9
m_0	Own Location Multiplier in Return to Experience, Loc 0	0.9		1.5	
m_1	Own Location Multiplier in Return to Experience, Loc 1	1.3		1.0	
m_2	Own Location Multiplier in Return to Experience, Loc 2	0.8		0.9	
β_{10}^0	Return to Experience from Small, Working in Small	0.015	0.029	0.052	0.039
β_{11}^1	Return to Experience from Medium, Working in Medium	0.029	0.042	0.067	0.074
β_{12}^2	Return to Experience from Large, Working in Large	0.027	0.042	0.047	0.062
β_2	Coefficient On Experience Squared	-0.00002		-0.00096	
B. Amenities					
α_{01}	Amenity of Medium Sized Places	0.24		0.40	
α_{02}	Amenity of Large Places	-0.40		0.57	
α_{11}	Medium Place Amenity Interacted With Experience	-0.04		-0.01	
α_{12}	Large Place Amenity Interacted With Experience	-0.07		0.05	
C. Arrival and Separation Rates					
λ_0	Job Offer Arrival Rate Within Small Locations	0.23	0.18	0.13	0.05
λ_1	Job Offer Arrival Rate Within Medium Locations	0.17	0.18	0.11	0.06
λ_2	Job Offer Arrival Rate Within Large Locations	0.16	0.17	0.12	0.07
λ^u_0	Job Offer Arrival Rate Within Small from Unemployed	0.26	0.23	0.11	0.10
λ^u_1	Job Offer Arrival Rate Within Medium from Unemployed	0.19	0.21	0.23	0.10
λ^u_2	Job Offer Arrival Rate Within Large from Unemployed	0.17	0.24	0.14	0.18
ρ_0	Receive Rate in Small: $\lambda_{10}=\lambda_1*\alpha_0$ and $\lambda^u_{10}=\lambda^u_1*\alpha_0$	0.03		0.22	
ρ_1	Receive Rate in Medium	0.04		0.20	
ρ_2	Receive Rate in Large	0.07		0.12	
γ	Own Location Multiplier: $\lambda_{00}=\lambda_0*\rho_0*\gamma$ and $\lambda^u_{00}=\lambda^u_0*\rho_0*\gamma$	2.86		2.17	
δ_0	Separation Rate in Small	0.02	0.02	0.02	0.00
δ_1	Separation Rate in Medium	0.02	0.02	0.01	0.01
δ_2	Separation Rate in Large	0.02	0.00	0.01	0.01
D. Costs and Benefits					
b	Unemployment Benefit	2.80		1.28	
C_u	Moving Cost for Unemployed	0.05		2.60	
C	Moving Cost for Employed	4.84	4.49	11.30	2.36
E. Heterogeneity					
weight	Fraction of Workers	0.51	0.49	0.38	0.62
π^0_{low}	Probability Start at Small Location	0.30	0.41	0.31	0.33
π^1_{low}	Probability Start at Medium Location	0.42	0.31	0.42	0.42
π^2_{low}	Probability Start at Large Location	0.28	0.28	0.27	0.24
F. Distributions					
σ_1	Standard Deviation of Match in Small	0.16		0.42	
σ_2	Standard Deviation of Match in Medium	0.12		0.46	
σ_3	Standard Deviation of Match in Large	0.17		0.31	
σ_1	Standard Deviation of Shock 1	4.03		6.04	
σ_2	Standard Deviation of Shock 2	20.74		14.60	
σ_3	Standard Deviation of Shock 3	1.87		3.12	
σ_u	Standard Deviation of Wage Measurement Error	0.40		0.51	

Notes: Listed in this table are estimates of the universe of parameters in the structural model. We also report values of some additional functions of these parameters for convenience. These include several probabilities, type specific wage shifters and returns to experience parameters that are not separately identified. We calibrate the discount rate to 0.95 per year.

Table 8: Counterfactual City Size Wage Premia

High School Sample

Experiment	Regression Coefficients Medium	Regression Coefficients Large	Gap With Fitted Values Medium	Gap With Fitted Values Large
Baseline Data	0.10	0.03		
No Mobility Restriction (simulated data)	0.14	0.01		
Equalize Ability Distribution	0.20	0.03	0.06	0.02
Equalize Search Frictions across Locs	0.08	-0.02	-0.06	-0.03
Equalize Search Frictions across Locs & Types	0.13	0.00	-0.01	-0.01
Equalize Return to Experience across Locs	0.06	-0.02	-0.08	-0.03
Equalize Return to Experience across Locs & Types	0.05	-0.09	-0.09	-0.10
Mobility Restricted	0.11	0.00	-0.03	-0.01
Equalize Ability Distribution	0.17	0.06	0.03	0.05
Equalize Search Frictions across Locs	0.11	0.03	-0.03	0.02
Equalize Search Frictions across Locs & Types	0.11	0.01	-0.03	0.00
Equalize Return to Experience across Locs	0.04	-0.02	-0.10	-0.03
Equalize Return to Experience across Locs & Types	0.03	-0.06	-0.11	-0.07

Notes: Estimates in the "No Mobility Restriction" row are based on fitted values. Other estimates are based on simulated data assuming various potential channels for the city size wage premium are shut off. The mobility restriction is imposed by setting the moving costs very high. Ability distribution equalization means we set the probability each type begins in each location to 1/3. Equalize search frictions across locations means we set the arrival rates equal to their arithmetic average across locations and/or types. Equalize return to experience across locations is analogous.

Figure 1: Model Fit for the High School Sample
Actual and Predicted Log Wage Including All Locations

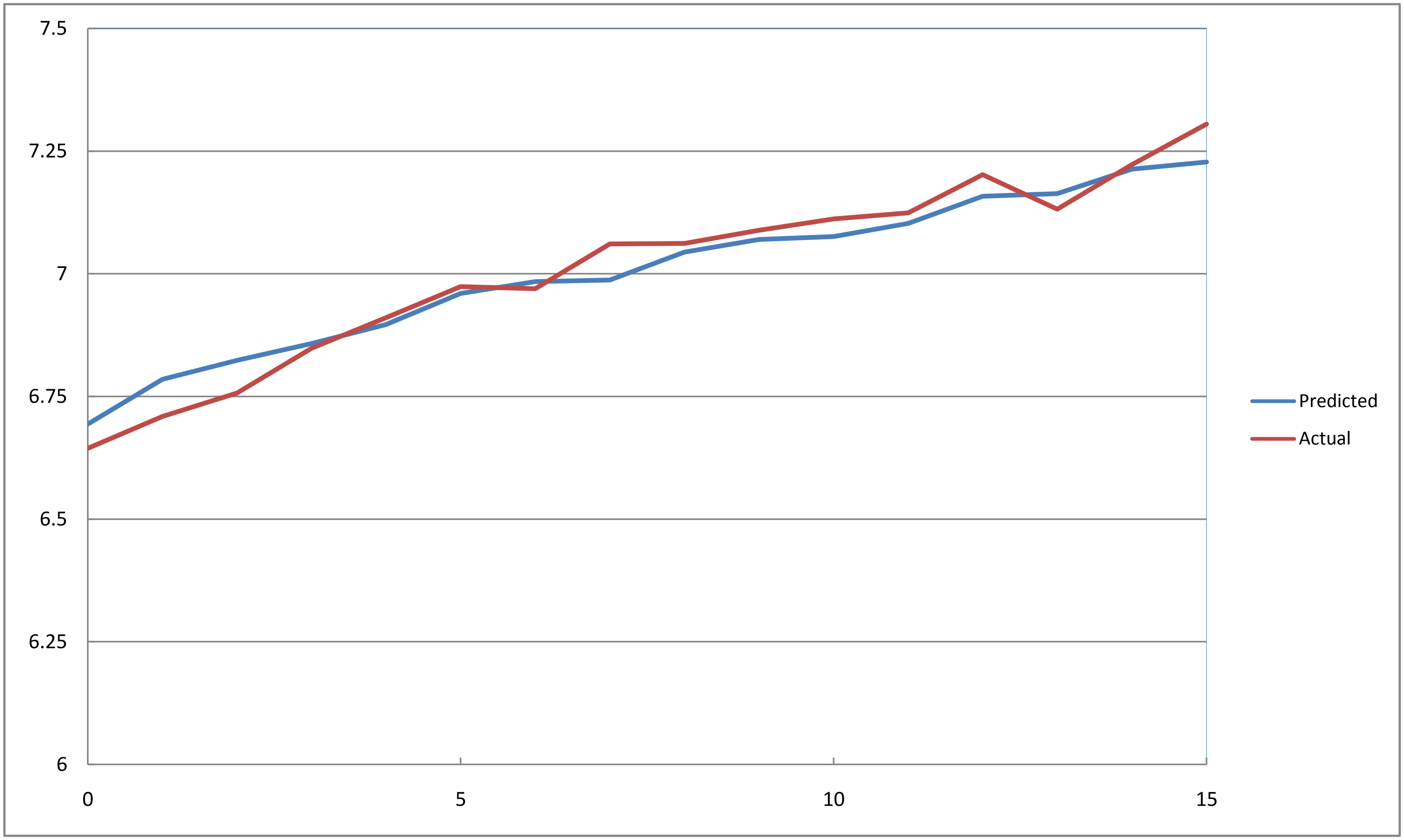
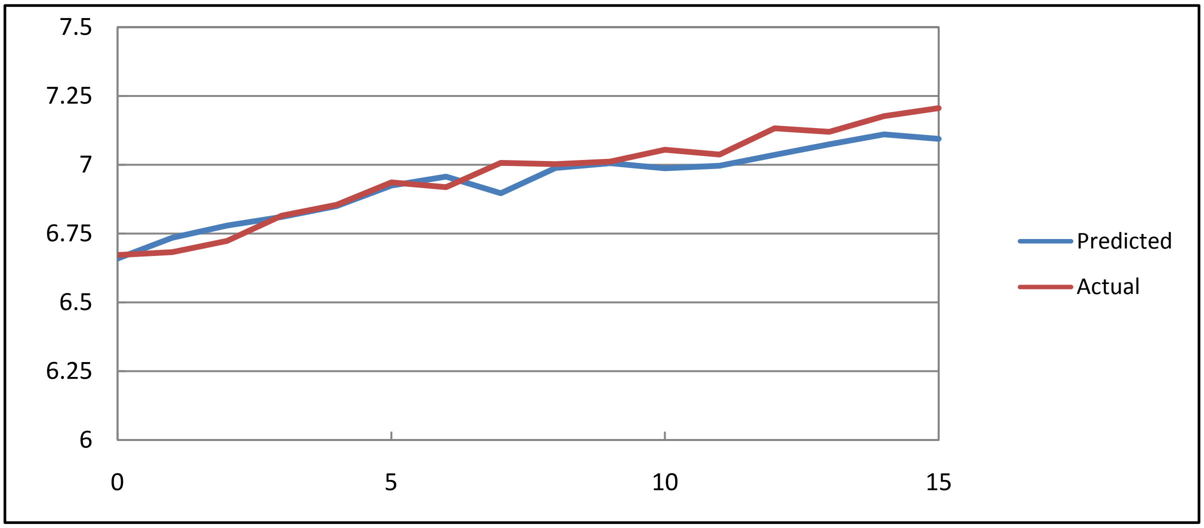
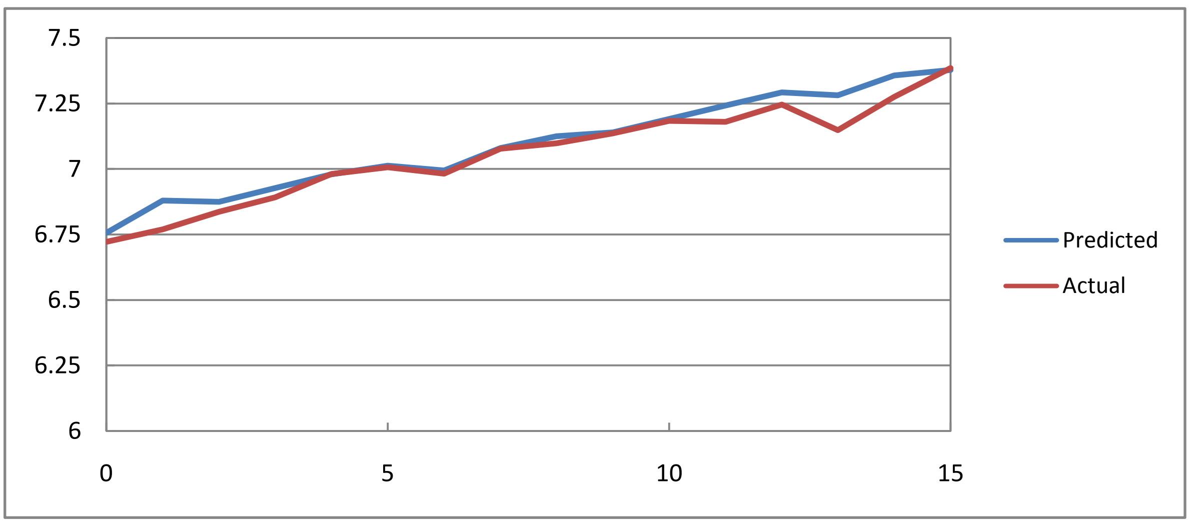


Table 2: Actual and Predicted High School Wages by Location

Panel A: Rural Areas and Small Metropolitan Areas



Panel B: Medium Sized Metropolitan Areas



Panel C: Large Metropolitan Areas