Understanding the City Size Wage Gap*

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

February, 2010
Keywords: Agglomeration, Wage Growth, Urban Wage Premium

Abstract

In 2000, wages of full time full year workers were more than 30 percent higher in metropolitan areas of over 1.5 million people than rural areas. The monotonic relationship between wages and city size is robust to controls for age, schooling and labor market experience. In this paper, we decompose the city size wage gap into various components. We propose an on-the-job search model that incorporates latent ability, search frictions, firm-worker match quality, human capital accumulation and endogenous migration between large, medium and small cities. Counterfactual simulations of the model indicate that variation in returns to experience and differences in wage intercepts across location type are the most important mechanisms contributing to the overall city size wage premium. Steeper returns to experience in larger cities is more important for college graduates while differences in wage intercepts is more important for high school graduates. Sorting on unobserved ability within education group and differences in labor market search frictions and distributions of firm-worker match quality contribute little or slightly negatively to observed city size wage premia in both samples.

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

1 Introduction

It is widely documented that wages are higher in larger cities. In the 2000 census, average hourly wages of white prime-age men working full-time and full-year were 32 percent higher in metropolitan areas (MSAs) of over 1.5 million people than in rural areas. Table 1 presents estimates of city size wage premia using census data from 1980, 1990 and 2000. It shows regressions of the log hourly wage for full-time full-year white men on indicators for living in medium and large population MSAs. The excluded category includes small MSAs and rural areas. Results in Specification 1 show that the city size wage premium is monotonic and has been increasing over time. Indeed, 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 MSA size. Specification 2 shows that in each year about one-third of the two estimated size premia can be explained with observable measures of skill and labor market experience. In addition, the city size wage gap has become considerably steeper since 1980 when large MSAs had wages that were 24 percent greater than rural areas. Because firm productivity is closely related to wages, this is evidence that firms in larger cities are more productive.

In this paper, we investigate the causes of the city size wage and productivity gaps. Our analysis utilizes a model of on-the-job search that incorporates endogenous migration between small, medium and large cities. Estimated parameters from the model allow for a decomposition of the observed city size wage gap into four components that all potentially interact: 1) sorting on unobserved ability across cities, 2) differences in search frictions, unemployment benefits and the distributions of the firm-worker match component of wages across cities and abilities, 3) variation in wage level effects across cities and abilities and 4) variation in returns to experience across cities and abilities. Our use of a finite mixture model as in Keane & Wolpin (1997) allows for recovery of parameters indexed by unobserved ability. As in Pavan (2009), our employment of a non-Gaussian state space approach to construct the likelihood function allows for inclusion of the unobserved firm-worker match component of the wage process in the model.

Counterfactual simulations of the estimated structural model indicate that variation in wage intercepts and returns to experience across location type are the most important components of the overall city size wage premium. For college graduates, each of these channels accounts for roughly half of the wage premia of medium and large cities over small cities and rural areas, though returns to experience are somewhat more important

in the largest cities. For high school graduates these same two channels together also account for the entire wage premia studied. However, level effects are almost three times as important as differences in returns to experience across location type for the wage premium of medium sized cities over small cities and rural areas, while each channel generates roughly equal portions of the large-small city wage premium. Our results provide no evidence that differences in labor market search frictions or firm-worker match quality contribute appreciably to city size wage premia.[1]

While we confirm evidence from elsewhere of positive sorting on observed skill to larger cities, our results indicate that sorting on unobserved ability within education group contributes little to observed city size wage premia. In fact, if anything we find evidence of mild negative sorting on unobserved ability. Among college graduates, high latent ability men are slightly less likely than low ability men to enter the labor force in large cities and they tend to move away from large cities at a faster rate than low ability men over the life-cycle. Among high school graduates, high ability types are less likely to enter the labor force in medium sized cities than small cities. This evidence of mild negative sorting on latent ability into larger cities manifests itself as even greater wage premia for medium and large cities over small cities and rural areas than those observed in the data in a counterfactual environment with no migration and equalized ability distributions across locations.

While results in this paper do not provide a complete taxonomy of all of the mechanisms by which workers in larger cities are more productive, as no single study could, they highlight the relative importance of various classes of explanations for understanding this fact. In particular, our results are consistent with larger cities fostering greater rates of human capital accumulation on the job, or "learning", especially for more highly skilled workers. The importance of level effects in the wage process for generating wage premia of interest could be generated by a host of underlying mechanisms, many of which are discussed in Duranton & Puga (2004). Potential mechanisms include sharing of inputs produced at large efficient scales, sharing risk and taking advantage of greater opportunities for division of labor. Empirical evaluations of micro-founded theories of how larger cities foster more rapid worker learning and sharing can differ by city size we leave to future

[1] Standard urban models typically imply that both the density and level of MSA population may generate agglomeration economies. We follow the convention of most studies of the urban wage premium and use population as our primary agglomeration measure. Ciccone & Hall (1996) is an important study of agglomeration economies that instead uses a density measure. Since MSA population and density are strongly positively correlated, the choice of measure matters little for our purposes.

research. Investigation of why the relative importance of these two forces differ by worker skill would complement such efforts.

Our results indicating a lack of importance of search frictions for generating city size wage productivity premia are of particular interest. Petrongolo & Pissarides (2006) discuss numerous studies finding that the aggregate matching function is constant returns to scale and they themselves find no evidence of higher job offer arrival rates for the unemployed in larger British labor markets. However, the existing literature neither examines the possibility that the distributions of latent firm-worker match components of productivity may be different across labor markets nor allows variation in job offer arrival rates to empirically compete with alternative explanations for higher wages in larger cities. Furthermore, existing evidence that wage growth is more rapid in larger cities after accounting for sorting, in Glaeser & Mare (2001) and Gould (2007) for example, does not distinguish whether this pattern is generated by ascension of steeper job ladders or more rapid human capital accumulation. We also provide new evidence on the nature of sorting on latent ability conditional on observable components of skill and how this sorting contributes to wage premia.[2]

Our results on the importance of differences in returns to experience for generating city size wage premia are consistent with evidence from the existing literature, including Glaeser & Maré (2001) and Gould (2007). However, this is the first study to fully measure the extent to which city size wage premia are generated by differences in wage levels independent of experience and search and matching effects. Such level effects are difficult to determine from panel data on workers because they are only observed when wages change with migration across cities of different sizes, and then only partially. When such examples of migration occur, observed wage changes conflate migration frictions, cost of living and option value differences across locations with differences in wage level effects. Our structural model and careful accounting for cost of living allow us to separately identify each of these components. This cost of living adjustment is crucial. Indeed, our use of three city size categories rather than two, as are used in many other studies, is motivated by the observation that many elements of a model that rationalizes the persistence of a city size productivity premium are not monotonic in city size, including wages adjusted

[2] Gould (2007) finds positive sorting on unobserved skill but estimates his model with only three latent ability levels using a sample that includes both high school and college graduates. His finding that higher latent ability men are more likely to live in larger cities is therefore fully explained by positive sorting on observed education level and says little about sorting on latent ability conditional on education level.

for cost of living. Our results on the role of level effects in city size wage premia are consistent with considerable recent empirical evidence finding positive local spillovers in particular industries, including Rosenthal & Strange (2003), Arzhagi & Henderson (2008) and Greenstone, Hornbeck & Moretti (2009) among others. The key innovation of this paper over the existing literature is that we translate empirical estimates identified off of the structure of the worker's life-cycle problem back to complete and unified decompositions of implied productivity differences across cities of different sizes.

We should emphasize that the model specified in this paper is partial equilibrium in nature. That is, firm location is taken as given. Part of the city size productivity gap may come from selection of more productive firms into larger cities. Ellison & Glaeser (1997) document that firms systematically locate in ways that generate industrial agglomerations, though Combes et al. (2009) argue that once accounting for agglomeration forces, firm productivity distributions exhibit no statistically significant truncation in larger cities as would be predicted by a firm selection model across markets like Syverson's (2004). Nevertheless, to the extent that some industries are more productive than others, the pattern documented by Ellison & Glaeser (1997) implies that more productive firms may also systematically locate in larger cities. While the framework developed in this paper has little to say about the process that might generate such firm selection, it still allows us to learn much about why larger cities are more productive. If input costs are higher in cities, it would be difficult for a general equilibrium model to justify the selective location of productive firms producing traded goods to larger cities without their workers also being more productive. Therefore, understanding why workers earn more in larger cities is informative about why firms in larger cities are more productive.[3]

The next section provides a baseline theoretical framework and presents descriptive evidence that is consistent with the results of the structural estimation exercise. Section 3 discusses data construction. Section 4 presents the model. Section 5 discusses how we estimate the model. Section 6 presents the estimation results and decompositions of the city size wage gap using counterfactual simulations. Finally, Section 7 concludes.

[3] It is also possible that firms are more productive in larger cities because these cities feature natural advantages. However, Ellison & Glaeser (1999) provide evidence that natural advantage is not an important source of agglomeration economies.

2 Empirical Observations

The city size wage premium shows up pervasively in the data. Although we need to appeal to the full structural model to fully understand the relative importance of various channels for generating this premium, in this section we present descriptive evidence that point to the most relevant mechanisms. In particular, we document patterns of transitions between jobs and through unemployment as functions of city size and examine the extent of selective migration. We then examine patterns of wage growth within jobs and over different types of labor market transitions as functions of city size. Putting these quantity and price components together, we perform a statistical decomposition of wage growth over the first 15 years of work experience as a function of city size. The key result of this decomposition is that within job wage growth is the only component of total wage growth that is monotonically increasing in city size on a positive base. While wage growth between jobs is an important component of total wage growth, it has no clear relationship with city size. This evidence is consistent with the estimated model simulation results reported in Section 6 demonstrating that labor market search frictions and matching are not important forces generating the city size wage premium.

A natural starting point for interpreting city size wage premia and conceptualizing how they can persist is the class of models going back to Roback (1982) and Rosen (1979) that emphasize compensating differentials across locations for firms and workers. A general formulation has mobile firms with a constant returns to scale technology whose indirect profit, given by Equation (1), is equalized across locations in long-run equilibrium. In this formulation, labor L, capital K and a composite non-traded good R all enter as factors of production with input prices w_j, r and p_j respectively. Locations are indexed by j and a_j represents a local productive amenity that incorporates agglomeration forces.

$$\overline{\pi} = \max_{L,K,R} \{a_j F(L,K,R) - w_j L - rK - p_j R\} \tag{1}$$

Assuming that the rental rate of capital is the same in every location, total differentiation derives the following equilibrium relationship between any two locations j and j', where ϕ_L is firms' expenditure share on labor and ϕ_R is firms' expenditure share on the non-traded good.

$$\ln(w_j) - \ln(w_{j'}) = \frac{\ln(a_j) - \ln(a_{j'})}{\phi_L} - \frac{\phi_R}{\phi_L}[\ln(p_j) - \ln(p_{j'})] \tag{2}$$

In the simplest one-factor environment, the city size log wage gap for identical workers thus measures the log firm total factor productivity gap. If capital is an additional factor, the log wage gap overstates the productivity gap by $\frac{1}{\phi_L}$. However, if the non-traded good is an additional factor instead, the log wage gap understates total factor productivity differences because local prices are positively correlated with wages.[4]

Table 2 reports city size wage premia in the NLSY data with and without adjustment for cost of living differences across locations. The left side of Table 2 Panel A shows that estimates of city size nominal wage premia using data from the NLSY are very similar to those from the 1990 and 2000 censuses reported in Table 1. The estimated wage premium for medium sized cities is 20 percent while that for large cities is 30 percent, indicating that even though it only includes young adults in 1979, the NLSY is a reasonable data set with which to evaluate reasons for the city size wage gap. 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 reductions in the urban-rural wage gradient, though not the gradient between medium and large cities, may indicate at first glance that positive sorting on unobserved skill is an important component of the urban-rural wage premium. However, the fixed effects results should be interpreted with caution given that they are consistent only if mobility is random conditional on the fixed effect. This assumption would be violated if, for example, a worker moves to a different city because he receives an attractive wage offer there, as would be predicted by our model in Section 4. In general, the fixed effect estimator provides inconsistent estimates if a worker's decision is influenced by a factor that is not perfectly persistent over time and is correlated with wages. Additionally, the fixed effect estimator is unlikely to be identified from a representative sample. Therefore, it would not recover the mean city size wage effect in an environment with heterogeneous effects of city size. Indeed, a simple Roy (1951) model, as developed by Borjas (1987) for example, predicts that movers differ on unobservables from the overall population.

The right side of Table 2 Panel A presents analogous regression results when wages are

[4] In his calibration of a similar model, Albouy (2009) uses $\phi_L = 0.825$ and $\phi_R = 0.025$ based on estimates in the literature, implying that the nominal city size wage gap overstates firm productivity differences by something on the order of 15 percent.

adjusted for cost of living differences across metropolitan areas.[5] These estimates reflect the fact that cost of living in large metropolitan areas is much higher than that in smaller places. In all three of the reported specifications, real wages in medium sized metropolitan areas are the highest. This inverse U profile of wages adjusted for cost of living exists at most levels of experience, evidence that the cost of living adjustment primarily influences measured wage levels rather than measured wage growth.[6]

Table 2 Panels B and C repeat the same analysis using subsamples of college graduates and high school graduates respectively. They show that the same patterns as in Panel A continue to hold conditional on observed skill. Specifications 1 and 2 indicate that college graduates are almost as productive in medium sized cities as they are in large cities while the profile for high school graduates is more monotonic. In addition, we see that the additional measures of observed skill added to the regressions in Column 2 do little to change estimated wage premia conditional on education, providing some indication that education level alone controls well for non-random sorting across locations. Fixed effects estimates produce remarkably similar city size wage premia for the two education groups.

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 workers' decisions. Results after this adjustment indicate that using information on workers to understand sources of the city size wage premium requires a model with endogenous migration between at least three size categories. Second, while controlling for observables reduces wage premia, doing so does not eliminate them. Therefore, mechanisms additional to endogenous sorting on observables must explain observed city size productivity premia. Tables 3 to 5 examine the relative importance of such mechanisms.

One potential explanation for the city size wage premium is that more rapid job turnover generates more efficient firm-worker matches in larger cities. Fewer job separations into unemployment may also generate this phenomenon. To evaluate the potential importance of these mechanisms, the first six columns of Table 3 describe patterns of job turnover

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

[6] In the context of a simple model of compensating differentials, the interpretation of this profile of real wages with respect to city size is that small and large locations have higher consumer amenity values than medium sized cities. The structural model detailed in Section 4 incorporates this idea but adds migration frictions and life-cycle incentives to remain in larger locations, including greater returns to experience and lower search frictions for some groups, that render the compensating differentials interpretation of this empirical pattern incomplete.

and unemployment by city size and education. Columns headed by "Within" indicate transitions within MSAs or rural counties of the indicated size while those headed by "To" indicate transitions that also involve migration to MSAs or rural counties of the indicated size. Such "To" migration may occur between or within size categories but always involves a change of MSA or rural county. Entries are calculated as $\left(\sum_i x_{is}\right) / \left(\sum_i \frac{t_{is}}{T_i}\right)$ and represent estimates of the expected number of transitions or weeks of unemployment experienced if an average individual were to spend his entire first 15 years of work experience living in the indicated city size category. In this ratio, x_{is} is the total quantity of each object given in Table 3 column headers for each man i in location type s over his first T_i years of work experience and t_{is} is the number of years spent working in location type s. To make these results consistent with those to come, we truncate the data at 15 years of work experience and only include those who remain in the sample that long.[7] This represents 80 percent of the NLSY sample, which we describe in more detail in the next section. Our discussion of the results in Table 3 focuses on within location transitions because migration patterns are influenced by cost of living, amenity and option value differences across cities in addition to job offer arrival rates.

The first column of Table 3 shows that while the average number of job to job transitions monotonically increases from 1.7 in small locations to 2.4 in large locations for college graduates, the high school sample exhibits no clear profile. This is evidence that higher job turnover may generate more efficient matching in larger cities, but only for the more skilled. Columns 3 and 4 of Table 3 indicate that the profile of job to unemployment to job transitions is very flat in city size for college graduates and slightly increasing for the high school sample, though selection generates a decreasing profile for the pooled sample. However, college graduates spend less time unemployed in smaller locations while high school graduates spend less time unemployed in larger locations. This could drive part of the city size wage gap for the high school sample, as it is possible evidence of lower search frictions for unemployed individuals in larger cities.

One potential alternative explanation for the city size wage premium is that more skilled individuals disproportionately reside in (and/or migrate to) larger cities. While we are

[7] T_i is not exactly 15 for everybody because the data set does not include all of the required information every quarter. For the purpose of Tables 3, 4 and 5 we only include individuals with a wage observation within two years of 15 years of work experience and truncate the sample at the wage observation closest to 15 years. For cases in which two wage observations are equidistant from 15 years, we use the longer sample.

forced to appeal to the structural model to fully evaluate whether sorting on unobserved skill is important, sorting on observed skill is evident in the data. Table 3 Column 7 indicates that college graduates disproportionately enter the labor force in larger locations. While 32 percent of total labor force entrants are in small locations, only 23 percent of college graduates enter in small locations. In contrast, 36 percent of the high school sample enters in small locations while just 28 percent enters in large locations. By the time they reach 15 years of work experience, this selection only strengthens. Column 8 shows that while there is a general tendency to migrate away from small and large locations to medium locations over the life-cycle, this comes about because college graduates move there from small locations and high school graduates move there from large locations.

Matrices describing location transitions between the time of labor force entry and 15 years of work experience reported in columns 9 to 11 of Table 3 present supporting evidence of such selective migration over the life-cycle. Panel C shows that only 13 percent of high school graduates move between cities of different sizes during their first 15 years of work. Those that do move out of medium and large cities are more likely to move to small cities and rural areas than medium sized cities. Those who move out of small places are more likely to move to medium sized cities than large cities. In contrast, Panel B shows that college graduates are more mobile and exhibit migration patterns that are more oriented toward larger cities. 39 percent of the college graduates entering the labor force in small places move compared to just 12 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 24 percent who migrate out of large cities, about two-thirds move to medium sized cities.[8] This evident sorting on observed skill implies that there might also be sorting on unobserved skill. The structural model handles this possibility.

Table 4 reports the mean log wage growth associated with each type of labor market transition examined in Table 3. Estimates give some indication as to whether returns to experience and firm-worker match quality differ across location type. To calculate these means, we estimate Equation (3) by least squares using data differenced within individual i for each wage observation at time period t. We use the same samples as in Table

[8] Similar though more pronounced patterns are seen in transition matrices between residential location at age 14 and labor force entry. Among college graduates, 33 percent change the sizes of their locations to work their first jobs. Their mobility rate out of small locations is more than 25 percentage points higher than out of medium and large locations. Among high school graduates, 10 percent move to take their first jobs, with mobility rates out of all three location types within 5 percentage points of each other and non-monotonic in city size.

3 excluding non-wage observations and we weight by the inverse of the number of wage observations for each individual. We index each class of labor market transition by location type j.

$$\begin{aligned}\Delta \ln(w_{it}) &= \sum_j \beta_1^j \Delta \exp_{it}^j + \beta_2 \Delta \exp_{it}^2 + \beta_3 \Delta \exp_{it}^3 \\ &\quad + \sum_j (c_1^j JJ_{it}^j + c_2^j JUJ_{it}^j + c_3^j JJm_{it}^j + c_4^j JUJm_{it}^j + c_5^j OTH_{it}^j) + \varepsilon_{it} \quad (3)\end{aligned}$$

This equation includes a cubic in experience, indicators for the four types of job changes for which quantities are reported in Table 3 and OTH_{it}^j, which is an indicator for situations in which a job is skipped because of lack of wage information.[9] The variables JJ_{it}^j, JUJ_{it}^j, JJm_{it}^j, $JUJm_{it}^j$ are indicators for job to job transitions within location, job to job transitions across locations, job to unemployment to job transitions within locations and job to unemployment to job transitions across locations respectively. We estimate this equation separately by skill and for wages that have and have not been adjusted for cost of living.

While the model we specify in Section 4 makes clear that the endogeneity of labor market transitions and migration in (3) makes such a regression potentially problematic, the results in Table 4 describe several important patterns in the wage data. First, the cost of living adjustment hardly influences the estimates when using differenced data. Second, returns to experience is the only element of wage growth that is strongly increasing in city size on a positive base for both subsamples. Third, the fact that average wage growth associated with each job to job transition is not monotonic in city size indicates that firm-worker match quality is not a likely driver of the city size wage premium. Wage responses to transitions through unemployment for the high school sample are increasing in city size but with significant wage losses in small locations. Job to job transitions that involve moving are associated with smaller wage increases in larger cities for both samples and transitions through unemployment that also include migration are associated with larger wage increases in larger cities, but only for college graduates. It should be noted that because we do not typically observe wages at the very beginning of new jobs, wage growth associated with job changes is likely to partially reflect pure experience effects. Estimates

[9] We do not index the quadratic and cubic experience terms by location in order to ease interpretation of the linear term coefficients.

from the structural model do not suffer from this limitation.[10]

To understand how the quantities in Table 3 and the elements of wage growth in Table 4 combine, Table 5 presents decompositions similar to those in Topel and Ward (1992) of log wage growth by city size up to 15 years of experience into four components: within job, between jobs with no unemployment in between, between jobs when individuals experience an unemployment spell in between and unknown. The unknown category consists of wage growth that occurred between jobs sandwiching a third job for which we have no wage information. As in Table 3, each entry is expressed for the average individual if he had spent the totality of his first 15 years of experience in the given location type and is calculated as the ratio of total wage growth from the given component divided by the aggregated fraction of time spent by all sampled individuals in that size location.

Results of this decomposition clearly show that within job growth is the primary driver of the city size wage and productivity premia. For the full sample, within job growth accounted for over 65 percent of the gaps in overall wage growth between medium and small and between large and small cities. Job to job transitions represent the next most important component of the medium-small gap at 34 percent, however, they are a much less important part of the large-small wage growth gap. Instead, transitions through unemployment account for 30 percent of the large-small city size wage growth gap overall. Finally, note that wages at labor force entry are monotonically increasing in city size when not accounting for cost of living differences across locations. These results indicate that steeper returns to experience in larger locations play a central role in the city size wage gap, but that level differences and the propensity to become unemployed may also be important.

Results for the college sample in Panel B and the high school sample in Panel C both exhibit similar patterns, though with different emphases. For the college sample, differences in rates of within job wage growth represent over 80 percent of the gaps between wage growth rates in small versus medium and large cities. For this group, relative wage growth between jobs within locations also appears to be somewhat important for the medium-small gap. In contrast, within job wage growth accounts for around 60 percent of the high school city size wage growth premium. The most important remaining components are job to job within locations for the medium-small gap and job-unemployment-job transitions for

[10] Based on a similar analysis, Wheeler (2006) concludes that wage growth from job transitions is more steeply increasing in city size than that from within jobs. However, that result comes from an estimation equation that does not interact experience with city size and does not account for the number of job transitions experienced by workers relative to the amount of time they spend working.

the large-small gap.

Consistent with other studies, the descriptive evidence in Tables 2 to 5 points to systematic differences in observed skill levels across locations as one important driver of city size productivity differences. In addition, attributes of larger cities that contribute to human capital accumulation appear important for generating higher wages in larger cities for all groups. However, we find less consistent evidence that differences in search frictions and matching generate much of the city size wage premium. It is hard to know the importance of selection on unobserved skill for generating these patterns. In addition, the fact that migration is always associated with job changes makes it difficult to separate out movement up the job ladder with level effects of locations solely through examination of descriptive data. Estimation of the model specified in Section 4 is thus invaluable as it efficiently uses observations about labor force transitions to separate out parameter combinations and it allows for identification of parameters indexed by latent ability.

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. 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.[11] In Section 5 we discuss how we deal with missing wage data econometrically. Individuals enter the sample when they begin working full-time, providing a convenient initial condition for the likelihood function. We define full-time as working at least 300 hours per quarter if not in school and at least 500 hours per quarter if in school.

We assign individuals to locations based on reported state and county of residence, which is available on interview dates and between interviews during the periods 1978-1982 and 2000-2004 only. We assign most location observations in remaining quarters by assuming that individuals must remain at one location for the duration of each job. We impose that unemployed individuals must remain at the same location as the last job held. Those jobs with multiple reported locations are assigned to the modally reported location. Jobs with multiple modes are assigned the modal location that occurred latest in time.[12] This leaves five 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 two percent of the sample, we impute locations to be that of the first job for which we observe location after the unobserved location spell.

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

Evidence in the previous section shows that it is important to allow potential mechanisms behind the city size wage premium to differ by worker skill. As such, we estimate the model 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.

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

[12] In the vast majority of cases, we do not observe the location at which individuals are unemployed. 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.

3.2 Spatial Price Index

When workers make migration decisions, they take into account relative wages and cost of living across locations. Using wages and migration patterns to understand productivity differences across cities thus requires accounting for such cost of living differences across space and time. We adopt the standard approach for calculating a price index that incorporates such a cost of living adjustment. This approach is based on a compensating differentials framework that is the consumer analog to the firm analysis described in the beginning of Section 2.

Equation (4) specifies the indifference relationship for workers with preferences U over the vector of goods x, some of which may differ in price across locations and time periods jointly indexed by j. In order for individuals not to move, indirect utility must be equalized everywhere at some value $\overline{v}$.

$$\overline{v} = \max_{x} \left\{ U(x) + \lambda[w_j - \sum_i p_i^j x_i] \right\} \tag{4}$$

Log-linearizing around a mean location yields the following equilibrium relationship in wages net of cost of living between any two locations j and 0, where s_i is the consumer expenditure share on good x_i.

$$\ln(w_0) = [\ln(w_j) - \sum_i s_i[\ln(p_i^j) - \ln(p_i^0)] \tag{5}$$

We use this formulation to deflate wages by cost of living differences across locations and time periods. The resulting deflator can be expressed as follows.[13]

$$INDEX_j = \prod_i (p_i^j / p_i^0)^{s_i} \tag{6}$$

Building this index requires price data by time and location for different goods and information on expenditure shares. We get prices by location from the American Chamber of Commerce Research Association (ACCRA) data sets from 2000 to 2002. These data report prices in six broad expenditure categories for most metropolitan areas and some rural counties nationwide. When possible, we take data from 2001. For the few regions

[13] The linear approximation for small deviations in prices across locations and periods generalizes to allow for deviations of all sizes if utility is assumed to be Cobb-Douglas.

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.[14] We impute price data for remaining areas as follows. Metropolitan counties are assigned the average prices from other MSAs in the same state and one of five MSA size categories when possible. If there are none other of the same size in their state, we impute using data from MSAs of the same size by census division. Price data for rural counties are imputed analogously.

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

Rather than take expenditure shares s_i directly from the CPI-U, we build expenditure shares for households including white men working full time using data from the biannual Consumer Expenditure Surveys (CEX) starting in 1982. We build shares directly from the CEX in order to best capture preferences of those in our sample and because the weights used for the CPI-U sometimes fluctuate significantly from year to year. We found that expenditure shares implied by the CEX are within a few percentage points across education groups and city sizes. Therefore, we use a sample from the CEX that best matches our full NLSY sample to calculate one set of expenditure weights that we apply to all studied individuals. As with the CPI-U, we allow expenditure shares to evolve over time.

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 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

[14] 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.

indexed.[15] Our most constraining simplifying assumptions limit the number of these underlying worker types to two and city size categories to three. Though it would be possible to expand the number of both objects, our specification allows for simpler interpretation of estimated parameters and simulations of the estimated model.[16] In addition, our specification facilitates computational tractability.

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

Each period, each worker earns a wage which depends on his ability, labor market experience in each location type, a firm-worker specific stochastic component and classical measurement error. The returns to experience and the individual specific intercepts are functions of worker type. The firm-worker specific stochastic component of the wage ε is drawn from the distribution $F^h_{\varepsilon_j}(\varepsilon)$ from which workers sample when they receive a job offer. Because it is impossible to separately identify means of the $F^h_{\varepsilon_j}(\varepsilon)$ distributions and the wage process intercepts $\beta^j_0(h)$ we set these means to 0.[17] The unexplained component of the wage, which can be thought of as measurement error, is drawn from the location-indexed distribution $F_{u_j}(u)$, is mean 0, and is independent across individuals and time. Put 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:

$$\ln w_j\left(h, x_1, x_2, x_3, \varepsilon, u\right) = \varepsilon + \beta^j_0\left(h\right) + \sum_{k=1}^{3} \beta^j_{1,k}\left(h\right) x_k + \beta^j_2 \left(\sum_{k=1}^{3} x_k\right)^2 + \beta_3 \left(\sum_{k=1}^{3} x_k\right)^3 + u \quad (7)$$

[15] 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.

[16] Kennan & Walker (2009) and Coen-Pirani (2009) examine migration between all U.S. states. In order to handle so many locations, they sufficiently simplify their models such that the decompositions like those we perform in Section 6 would not be possible.

[17] One could interpret the $\beta^j_0(h)$ parameters as capturing differences in average match quality across location types. It is perhaps difficult, however, to structurally interpret average match quality as a separate entity from potential sharing mechanisms that influence wage level effects across location types.

After being incorporated into the full model, this specification allows us to estimate separate time and job invariant components of worker wages and returns to experience by city size in a more generalized environment than was possible in the reduced form treatment in Section 2. It additionally allows for estimation of the variances of the distributions of firm-worker specific components, which provide information on the potential importance of matching. Indexing these three objects by latent ability allows the model to handle sorting on unobservables.[18]

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 every other location type remains constant. Individuals accrue experience at the beginning of each working period. Note that we restrict the curvature of the effect of experience to depend only on location and not on the composition of total labor market experience and restrict the cubic term to be the same across worker types and locations. This sensibly reduces the dimensionality of the vector of estimated parameters.[19] We assume that each individual works for 140 quarters and then retires with a pension equal to his last wage. After retiring, he lives for an additional 120 quarters.[20]

We allow the job search technology to differ by city size, ability and employment status. We denote the arrival rates of job offers from the same location to be $\lambda_j^u(h)$ for unemployed workers and $\lambda_j(h)$ for employed workers, where j is the worker's location. The arrival rates of job offers from different locations are $\lambda_{jj'}^u(h)$ for the unemployed and $\lambda_{jj'}(h)$ for 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.[21] Workers who choose to switch jobs at the same location must pay a stochastic switching cost v_S with zero mean and finite variance. This cost potentially captures differences in non-pecuniary benefits across jobs

[18] Based on evidence in the literature that tenure effects are not large, we do not incorporate them into the wage process. In their review of the literature, Altonji & Williams (2005) conclude that log wage growth due to tenure effects is likely only about 0.09 over the first 10 years of experience.

[19] We also impose the that the linear component of experience is restricted such that $\beta_{1,k}^j(h) = \theta_j \beta_{1,k}(h)$. This restriction produces more precise estimates of linear component paramters than is possible with the fully interacted specification. Our estimates of experience effects $\beta_{1,k}^j(h)$ are robust to alternative specifications including $\beta_{1,k}^j(h) = \theta_k \beta_{1,j}(h)$.

[20] In order to reduce the computational burden, we index time to quarters for the first 60 quarters of work after labor force entry and years for the remaining 20 years of working life. The likelihood function is computed only using information on the first 15 years of labor force participation for each individual. More than 75% of observations and 90% of location changes in our data happen during this period.

[21] In order to limit the number of arrival rate parameters that need to be estimated, we impose the parameter restrictions described in Appendix B.

that might lead workers to accept wage cuts. Exogenous separation rates $\delta_j(h)$ similarly depend on location and ability. With a job offer at location j', individuals have the option to move and pay a one-time cost of $C_M(x)+v_M$, that depends on total work experience and has 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 one 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 spaces of all value functions that we discuss additionally include individual specific "ability" h and experience in all location types x_1, x_2, x_3. For expositional simplicity we suppress this dependence in the notation. The current utility of an unemployed worker depends additively on the amenity α_j, the unemployment benefit $b_j(h)$ and on an *i.i.d.* preference shock v_U with zero mean and finite variance which shows up below in (8) and (9). This shock captures the random component of the disutility of work. Given this environment, the deterministic components of the present value of being unemployed and working are given by the following expressions respectively:

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

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

We first consider the environment for an unemployed individual at location j with less than 35 years of work experience. At the beginning of each period, the agent observes whether he is faced with one of five possible scenarios. With probability $\left(1 - \lambda_j^u(h) - \sum_{j'=1}^{3} \lambda_{jj'}^u(h)\right)$ he does not receive a job offer, with probability $\lambda_j^u(h)$ he receives a job offer from the same

location, and with probability $\lambda^u_{jj'}(h)$ he receives an offer in location type j'.[22] The individual decides at the beginning of each period whether to accept a potential job offer or remain unemployed. If he accepts an offer, he pays a cost to move to the relevant new location if necessary.

Equation (8) shows the value function for an unemployed agent in location j:

$$\begin{aligned} V^u_j &= \left(1 - \lambda^u_j(h) - \sum_{j'=1}^{3} \lambda^u_{jj'}(h)\right) E_{v_U}(V^{UN}_j + v_U) \\ &\quad + \lambda^u_j(h)\, E_{v_U \varepsilon_j} \max\left[V^{UN}_j + v_U, V^{WK}_j(\varepsilon)\right] \\ &\quad + \sum_{j'=1}^{3} \lambda^u_{jj'}(h)\, E_{v_U v_M \varepsilon_{j'}} \max\left[V^{UN}_j + v_U, V^{WK}_{j'}(\varepsilon) - [C_M(x) + v_M]\right]. \end{aligned} \tag{8}$$

The first term of Equation (8) 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 $V^{WK}_{j'}(\varepsilon)$ 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 to the distribution of ε in location type j' and the distribution of the random components v_U and v_M (expressed as $E_{v_U v_M \varepsilon_{j'}}$).

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

[22] The parameter $\lambda^u_{jj}(h)$ represents the probability that an unemployed individual at location j receives a wage offer in a different city also of size category j whereas the parameter $\lambda^u_j(h)$ represents the probability that this individual receives a wage offer in the same city as his last job.

will receive a wage offer from a different employer. As such, the value to a worker with ability h at location j of being employed with firm match ε is given by Equation (9):

$$\begin{aligned} V_j(\varepsilon) &= \delta_j(h) V_j^{UN} + (1-\delta_j(h)) E_{v_U} \max\{V_j^{UN} + v_U, \\ &\quad (1-\lambda_j(h) - \sum_{j'=1}^{3} \lambda_{jj'}(h)) V_j^{WK}(\varepsilon) \\ &\quad + \lambda_j(h) E_{v_S \varepsilon'_j} \max\left[V_j^{WK}(\varepsilon), V_j^{WK}(\varepsilon') - v_S\right] \\ &\quad + \sum_{j'=1}^{3} \lambda_{jj'}(h) E_{v_S v_M \varepsilon'_{j'}} \max\left[V_j^{WK}(\varepsilon), V_{j'}^{WK}(\varepsilon') - v_S - [C_M(x) + v_M]\right]\}. \end{aligned} \tag{9}$$

As is evident in Equation (9), an exogenously separated worker at location j may only become unemployed in location j. If the worker is not exogenously separated, he can still choose to go unemployed if v_U is large enough. If he chooses to keep working, with probability $\left(1 - \lambda_j(h) - \sum_{j'} \lambda_{jj'}(h)\right)$ he does not receive a wage offer. In this event, he remains employed in the same job. If he receives an offer, he either accepts it and moves if necessary, or remains at his old job.

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

This model captures each of the components of city size wage premia discussed in the introduction. The job arrival rate parameters for workers $\lambda_{jj'}(h)$ and the distributions of firm-worker matches $F^h_{\varepsilon_j}(\varepsilon)$ capture job turnover and match quality. The job separation rates $\delta_j(h)$, the unemployment benefits $b_j(h)$, the arrival rate parameters for the unemployed $\lambda^u_{jj'}(h)$ along with $F^h_{\varepsilon_j}(\varepsilon)$ capture the propensity to become unemployed and unemployment duration. The coefficients on experience in Equation (7) capture differences in "learning", or any agglomeration effects that increases with experience. The intercepts in the wage process capture fixed productivity differences across locations that may be consequences of nonlabor input sharing and human capital spillovers. Each of the key parameters is indexed by latent ability. Along with the estimated probabilities

that labor force entrants in each location are of low and high ability, we can thus calculate counterfactual wage premia given different ability distributions across location type.

5 Estimation

This section first 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 who enters the labor market in location j and is observed for T periods is given by:

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

where π_j is the probability that an individual is of type h_L given that he enters in location j and θ is the vector of parameters.[23]

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

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

It is convenient to express the likelihood function in terms of probabilities that one of five types of event occurs. These are: finding a job in the same location if unemployed, finding a job in a different location if unemployed, having a job to job transition within the same location, having a job to job transition changing location, and entering unemployment. These probabilities are defined as functions of ε and, if relevant, ε' but they

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

also depend on the other state variables $\{h, x_1, x_2, x_3\}$:

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

As an example, the transition probability between unemployment and employment at location j with firm specific component ε is as follows.

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

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

Computation of $f\left(Y_t|Y^{t-1}\right)$ 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 in 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 new wage information each period. This implies the following updating rule:

$$f\left(\varepsilon|Y^t\right) = \frac{f\left(Y_t|Y^{t-1}, \varepsilon\right) f\left(\varepsilon|Y^{t-1}\right)}{f\left(Y_t|Y^{t-1}\right)} \tag{10}$$

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

5.2 Identification

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

We cannot nonparametrically identify distributions of the firm specific wage compo-

nents 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 $N(0, \sigma_{\varepsilon_j})$. For the purpose of implementation, the distribution of ε is a grid of ten elements which is restricted to the range $(-2\sigma_{\varepsilon_j}, 4\sigma_{\varepsilon_j})$.

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

Table A1 describes all of the estimated parameters of the model. We partition them into six broad groups: components of the wage in Equation (7), arrival and separation rates, amenities, benefits and costs, type probabilities, and distributional measures. We normalize amenities to 0 in location type 0 as they would not otherwise be separately identified from the wage shifters. 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. Based on estimates from the literature, Gourinchas & Parker (2002) for example, we set the discount factor to 0.95 per year.

6 Results

6.1 Model Fit

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

graduates. Figure 1 Panel A exhibits a very good fit for small locations while Panel B shows that we have a good fit in medium sized cities up to 6 years of experience after which the model slightly over-predicts wages. Panel C shows that for large cities we also slightly over-predict at over 5 years of experience. Figure 2 shows that the fit for the high school sample is generally even better than that for the college sample.

Despite overprediction in the college sample at higher levels of experience, examination of statistics on transitions between jobs, to and from unemployment, and between locations reveals that the simulated data match the actual data remarkably well in many dimensions. Table 6 Panel A shows actual and predicted job, unemployment and location transitions. The model generates simulated data that imply transition statistics that are at most 0.1 percentage points off from the actual data in both samples. Panel B shows job to job and job to unemployment transitions within location type. Once again, neither simulated statistic by location differs from the actual data by more than 0.3 percentage points. Panel B also shows that the simulated data match observed unemployment duration data remarkably well.

Table 7 presents predicted and actual migration patterns conditional on changing location. At left are location types in period $t-1$ while along the top are location types at time t. Diagonal entries give the fraction of moves between different cities of the same location type. The largest gaps between actual and predicted in the college sample are in transitions from large to small and small to large locations which are each under-predicted by about two percentage points. The high school sample exhibits a similarly good fit. Its largest actual-predicted gap is for migration from small to large locations, also at minus 2 percentage points.

6.2 The Role of Ability

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

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

employs Bayes' Rule.

$$\Pr\left(h = h_H | Y_i^T\right) = \frac{\Pr\left(Y_i^T | h = h_H\right) \times \Pr\left(h = h_H\right)}{\Pr\left(Y_i^T\right)}$$

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

Table 9 presents strong evidence that our estimated $Pr(h = h_H | Y_i^T)$ is a better predictor of the log wage than is the age-adjusted AFQT score. Each column in Table 9 presents results from regressing log wages on a quadratic in experience and various combinations of the two ability measures and two city size dummy variables. Specification (1) shows that our ability measure has a much higher t-statistic than does AFQT (measured in standard deviations) when predicting log wage. Indeed, the pairwise correlation between unobserved ability and log wages is about three times greater than that between AFQT and log wages for the college sample and twice as large for the high school sample. Specifications (2)-(5) in Table 9 show that unobserved ability remains more informative about wages when accounting for city size effects. Coefficients on city size hardly change when AFQT is included as a control in the regressions yet increase when unobserved ability is included as a control. Furthermore, the R-squared markedly increases when estimated ability is included in these regressions. Low correlations between the two measures of ability of 0.16 among college graduates and 0.06 for the high school sample are consistent with these observations.[24]

6.3 Wage Components

Many parameter estimates are difficult to interpret in isolation because they interact nonlinearly in the model. For this reason, instead of focusing on parameters individually, we primarily describe what groups of parameter estimates imply for patterns in wages as functions of city size.[25]

Figure 3 describes the set of parameters that capture the wage process absent the

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

[25] Table A1 lists individual parameter estimates and indicates their statistical significance.

match component and measurement error terms ε and u. It graphs wage profiles for low and high ability men in each education group as functions of experience setting the match quality ε to 0 and assuming no migration over the life-cycle. These graphs are drawn using parameters in Table A1 Category A.

The estimated level components of wages, plotted as intercepts, are similar across locations for low and high ability college graduates. For low ability high school graduates, however, large locations have the lowest level, followed by small then medium locations. Among high ability high school graduates, small locations have the smallest level with medium and large locations at the same higher level. Low ability members of the high school sample appear to have had an unusually difficult time if they entered the labor force in a large city.

Figure 3 Panels A1 and A2 show that among college graduates, real wages become higher in larger cities at all levels of experience above three years. Furthermore, the returns to experience in larger cities are notably higher than in medium and small locations, especially for high ability individuals. Figure 3 Panels B1 and B2 reveal a somewhat different story for high school graduates. They show that real wages in medium sized cities are highest at all levels of experience for low ability men and most levels of experience for high ability men. While high ability high school graduates have very similar real wages in medium and large cities, low ability high school men earn a very constant 11 percent less in medium sized cities and 18 percent less in small cities beyond 10 years of experience, making nominal wages for this group almost identical in large and small locations. While returns to experience are quite similar across locations for high ability high school graduates, low ability types experience much steeper experience profiles in large locations during the first 10 years of experience, though it comes off of a low base.

In summary, Figure 3 indicates that the city size wage gap has both wage growth and level components for both education groups. However, the growth component appears more important for college graduates than high school graduates.

The other focus of this paper is to understand the role of matching for generating productivity premia across cities of different sizes. As with the parameters of the wage process, it is difficult to interpret the parameters that influence job turnover, job to unemployment transitions and firm-worker match quality in isolation. Instead of attempting to interpret them as such, Figure 4 presents graphs of the mean firm-worker match component ε of the wage process as functions of work experience, city size, education and ability given no mobility across locations. To construct these graphs, we simulate data using all parameters

reported in Table A1 except that we set the moving cost to be infinite. Therefore, the plots in Figure 4 depict the component of wages that comes from ascension of location size specific job ladders.

Figure 4 confirms evidence in Tables 3, 4 and 5 that search and matching are not important components of city size wage premia. The four graphs show few consistent relationships between the match component of wages and city size. Indeed, low ability college graduates is the only group for which ε is monotonically increasing in city size at all levels of experience. Among high ability college graduates, average match quality in small locations exceeds that in large and medium locations at beyond eight years of work experience. Among high school graduates, the ordering of match quality by city size is medium, small, large for low ability and reversed for high ability individuals. Given that low ability individuals earn lower wages on average, that they also have the best labor market matches in large locations serves, if anything, to reduce the city size wage premium.

The patterns in Figures 3 and 4 indicate two incentives for men of each ability level to migrate across location type after labor force entry. Among college graduates, opportunities to improve firm-worker match quality provide incentives for low ability types to move to larger locations at all levels of experience and high ability types to move to small locations at high experience levels. Because relative remaining wage components across locations graphed in Figure 3 do not counteract this force, a larger percentage of low ability than high ability college graduates choose to move to large locations over the course of the life-cycle. A similar pattern exists for high school graduates, though the non-match component of wages does counterbalance the match component across locations for the low ability type. As will be evident in the following subsection, the associated change in worker composition depresses average wages in large locations relative to small locations.

While they are not components of wages, the local amenity value and moving cost do influence migration patterns. Our estimates indicate that medium and large cities have higher amenity values than small locations for both education samples, keeping workers from moving to small locations even if they receive slightly superior wage offers there.[26] We estimate that moving costs are increasing in work experience and have a large stochastic component. Our estimates indicate that about 15 percent of workers have draws from the moving cost distribution that are less than five percent of their wage.

[26] Using a rich static model of compensating differentials, Albouy (2008) also finds that the amenity value of cities is roughly increasing in city size.

6.4 Simulations

Using the parameter estimates from the structural model, Tables 9 and 10 evaluate the importance of potential mechanisms for generating city size wage premia for the two samples. Because each mechanism considered is captured by multiple parameters that interact nonlinearly with remaining parameters of the model, we achieve this by shutting off one or two channels at a time and report implied resulting average city size wage premia. For each experiment, we assign the average value of the relevant parameters for each latent type listed at left in Panels B and C across locations and then regress the resulting counterfactually simulated wage on a cubic in experience and two city size dummy variables. Tables 9 and 10 report the coefficients on the city size dummy variables from these regressions, all of which are statistically significant.

The ultimate goal of these simulations is to evaluate the extent to which observed wage gaps between cities of different sizes change under counterfactual scenarios. Performing such an evaluation is not straightforward because the model only has predictions about city size category of residence, not the particular city of residence within a given category. Therefore, for the channels that generate a significant reduction in city size wage premia, we report the associated percent reduction in nominal wages calculated incorporating the equilibrium average price differences reported in Tables 9 and 10 Row 4. Because of the nonlinearities in the model, these reductions do not necessarily sum to 100 percent. This procedure is not perfect because the price differences in Row 4 are not calculated using a counterfactual distribution of people across location within each size category. Indeed, the key assumption for the validity of this exercise is that such equilibrium and counterfactual distributions are the same. Nevertheless, given the magnitudes of our results we believe that they still provide a very clear picture of the factors that generate the largest wage gaps across locations of different sizes.

Because wages used for estimation are adjusted for cost of living differences, the counterfactual equalizing level effects across locations, involving the $\beta_0^j(h)$ parameters of the wage process, requires special consideration. The goal is to generate a counterfactual by equalizing the level component of observed wages, meaning the non-match component at 0 years of work experience, in each location. To achieve this, for the rows entitled "Equalize Nominal Intercepts Across Locs" we equalize $\beta_0^j(h) + \ln(P^j)$, where $\ln(P^j)$ are given in Row 4. We set the population weighted mean to be equal to the population weighted mean of $\beta_0^j(h)$. There is no such issue for equalizing returns to experience because wages are

already expressed in logs.

Table 9 reports the counterfactual simulation results for college graduates. As a baseline, Panel A Row 1 shows that regressing log wage on a cubic in experience and two city size dummies using the raw data implies city size wage premia of 0.14 for medium sized cities and 0.09 for large cities.[27] Data simulated from the parameter estimates in Table 10 also imply wage premia of 0.14 and 0.09 for the two size categories respectively indicating that the model fits the data very well. Panel A Row 3 gives counterfactual city size wage premia in the case where individuals are forced to stay in their location of labor market entry. Commensurate with the discussion in the previous subsection, restricting mobility actually increases wage premia between rural areas and cities to 0.17, as it forces high ability men to stay in larger locations and keeps low ability men out of these locations. However, this increase is only marginally statistically significant. Results in Row 2 are used as a benchmark for counterfactuals in Panel B, for which we shut off potential channels other than mobility for the city size wage premia. Results in Row 3 benchmark results in Panel C for which we shut off the same potential channels as in Panel B in addition to mobility. 95 percent confidence intervals calculated by simulating the model with draws from the joint parameter distribution in Table A1 are listed below coefficients in Rows 2 and 3.

Differences in returns to experience and wage level effects account for virtually the entire city size nominal wage gap of college graduates. Absent differences in returns to experience across locations and maintaining endogenous sorting, the counterfactual city size nominal wage premium would be reduced by 38 percent for medium locations and 68 percent for large locations. Absent differences in wage level effects, the counterfactual city size nominal wage premium is estimated to decline by about 60 percent for both medium and large cities.[28] Search frictions plus match quality and ability sorting at entry into the labor force have small and statistically insignificant impacts on city size wage premia and marginal productivity premia in the college sample.[29]

It is perhaps more instructive to examine the importance of the same channels on wage

[27] These numbers differ from those in Table 2 because this sample only includes observations up to 15 years of work experience and does not control for age or years of education.

[28] Equalizing the $\beta_0^j(h)$ parameters across locations without adjusting for price differences generates small and insignificant reductions in wage premia for the college sample.

[29] When we break out search and matching into components involving transitions directly between jobs and those involving transitions through unemployment, we find that both channels independently have small impacts on city size wage premia.

gaps without worker mobility, since doing so fully isolates the impact of each channel by city size. Simulation results with the mobility restriction reported in Table 9 Panel C show that mobility is not an important driver of the results in Panel B. With restricted mobility, equalization of returns to experience across locations and types reduces nominal wage gaps by 37 percent for medium locations and 59 percent for large locations while level effects account for 50 percent and 53 percent reductions respectively for the two size categories. Other counterfactual changes in wage premia are small and not statistically significant, though equalizing the ability distribution across locations at labor force entry and restricting mobility does lead to a significant increase in the large city wage premium. This comes about for reasons discussed in the previous subsection.

Table 10 reports analogous results to those in Table 9 for high school graduates. As with college graduates, the estimated model generates predicted city size wage premia perfectly in line with the data of nine percent in medium sized cities and four percent in large cities. While differences in returns to experience and level effects are also the most important channels generating city size wage premia for the high school sample, the other two channels that we consider counteract these two effects somewhat for medium locations. Equalizing returns to experience decreases nominal wage premia by 39 percent and 55 percent for medium and large cities respectively with free mobility. Analogous results given restricted mobility are more similar at 44 percent and 60 percent respectively. Equalizing level effects in the wage equation decreases city size productivity premia by more than twice as much in medium locations with or without free mobility. In large locations, the percentage reductions are similar to those after equalizing returns to experience. Equalization of returns to experience across locations generate smaller reductions in log wages for the high school sample than the college sample. However, equalization of location and ability-specific intercepts across locations leads to similar counterfactual reductions in log wages for the two groups.

The resulting total reductions of more than 100 percent are counteracted somewhat because equalizing search and matching parameters or the ability distribution at labor force entry increase wage premia, particularly in medium sized cities. Increases in the premia due to equalization of the initial ability distribution occurs for the same reason as in with the college sample. Counterfactual premia absent differences in search and matching parameters also indicate that the equilibrium locations of individuals by type result in slightly lower match qualities in medium than small locations. Inspection of Figure 3 reveals that this must be driven by low ability types, which represent about 65

percent of high school graduates.

7 Conclusions

In this paper, we lay out a systematic framework to empirically examine reasons for which larger cities have higher wages and are more productive. Using data from the NLSY, we show that hourly wages are higher and grow faster in bigger cities, and that workers in larger cities have higher observed skill levels. A decomposition of log wage growth over the first 15 years of experience reveals that within job wage growth generates more of the city size wage gap than between job wage growth. Estimation of a model of on-the-job search and endogenous migration between three city size categories allows us to sort out the extent to which sorting across locations on latent ability interacts with level effects, returns to experience, and firm-worker specific wage components to generate city size wage gaps.

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

Our identification of the relative importance of these four broad channels for generating the city size wage premium provides new information about the relevance of certain classes of micro-founded theories over others for generating this phenomenon. We hope that our evidence leads to further investigation of the empirical relevance of various theories that generate wage level and growth effects since these two broad mechanisms are the clear drivers of city size wage, and likely productivity, premia.

A Construction of the Likelihood Function

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

A.1 Fundamentals

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

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

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

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

instead.

A.2 Reservation Rules

Regime A occurs when an unemployed agent receives an own-location job offer. Regime B occurs when a worker is choosing whether to go unemployed. Regime C occurs when an unemployed agent receives an offer in another location. Regime D occurs when a worker receives an own-location offer. Regime E occurs when a worker receives an offer in another location. In cases where a new match is drawn and the worker has an existing match quality, 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.

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

A.3 Transition Probabilities

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

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

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

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

The probability of entering unemployment given that a worker had a job with match ε

is:

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

The probability of changing employer from match ε to match ε' in the same location is:

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

The probability of changing employer from match ε in location j to match ε' in location j' is:

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

A.4 First Period

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

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

The resulting posterior distribution of the firm match is:

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

A.5 Unemployment

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

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

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

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

The posterior distribution of the match then becomes:

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

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

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

The posterior distribution of the match is then:

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

A.6 Becoming Unemployed

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

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

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

A.7 Working

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

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

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

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

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

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

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

B Normalizations of Job Offer Arrival Rates

In the model there are 48 free parameters that measure arrival rates of job offers of which 36 are probabilities of receiving a wage offer from a different location. Given that changing location is a rare event in the data, these parameters cannot be estimated precisely. Instead we estimate freely the 12 parameters that describe the probability of receiving a wage offer from the same location and we assume that the remaining probabilities are scaled by the

4 estimated parameters $\rho_{j'}$ and γ. We define $\rho_{j'}$ to be a multiplier for arrival rates to a given city j' and γ to be a parameter that scales the product $\lambda_j(h)\rho_{j'}$ if the two location sizes are the same but the individual changes location. We use the same scaling factors for unemployed and worker arrival rates. Arrival rates of job offers across locations are thus specified as follows.

$$\begin{aligned}
\lambda^u_{jj'}(h) &= \lambda^u_j(h)\rho_{j'} \text{ if } j \neq j' \\
\lambda^u_{jj'}(h) &= \lambda^u_j(h)\rho_{j'}\gamma \text{ if } j = j'
\end{aligned}$$

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

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

References

Albouy, David. 2009. "Are Big Cities Really Bad Places to Live? Improving Quality of Life Estimates Across Cities," NBER Working Paper #14472

Albouy, David. 2009. "What Are Cities Worth? Land Rents, Local Productivity and the Capitalization of Amenity Values," NBER Working Paper #14981

Altonji, Joseph & Nicolas Williams. 2005. "Do Wages Rise With Job Seniority? A Reassessment," *Industrial and Labor Relations Review* 58:3, 370-397.

Arzhagi, Muhammad and J. Vernon Henderson, 2008 "Networking off Madison Avenue," *Review of Economic Studies* 75:4, 1011-1038.

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

Borjas, George. "Self-Selection and Earnings of Immigrants." *American Economic Review* 77:4, 531-553.

Ciccone, Antonio and Robert Hall, 1996. "Productivity and Density of Economic Activity," *American Economic Review* 86 54-70.

Coen-Pirani, Daniele. 2009. "Understanding Gross Worker Flows Across U.S. States," manuscript.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, & Sebastian Roux. 2009. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection" manuscript.

Duranton, Gilles and Diego Puga, 2004. "Micro-Foundations of Urban Agglomeration Economies," in *Handbook of Urban and Regional Economics Vol. 4*, 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.

Ellison, Glenn and Edward Glaeser, 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review* 89 311-316.

Glaeser, Edward and David Mare, 2001. "Cities and Skills," *Journal of Labor Economics* 19 316-342.

Gould, Eric. 2007 "Cities, Workers and Wages: A Structural Analysis of the Urban Wage Premium" *Review of Economic Studies*, 74:2, 477-506.

Gourinchas, Pierre-Olivier and Jonathan A. Parker. 2002. "Consumption Over the Life-Cycle," *Econometrica*, 70:1, 47-89

Greenstone, Michael, R. Hornbeck and Enrico Moretti. (forthcoming) "Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants" *Journal of Political Economy*

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.

Kasahara, Hiroyuki & Katsumi Shimotsu. 2009 "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices." *Econometrica* 77:1, 135-175.
Keane, Michael and Kenneth Wolpin, 1997. "The Career Decisions of Young Men," *Journal of Political Economy* 105 473-522.

Kennan, John and James R. Walker. 2009. "The Effect of Expected Income on Migration Decisions," manuscript

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

Petrongolo, Barbara and Christopher Pissarides. 2006. "Scale Effects in Markets With Search," *Economic Journal* 116:508 21-44.

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.

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

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

Topel, R.H. and M.P. Ward, 1992. "Job Mobility and the Careers of Young Men" *Quarterly Journal of Economics* 108, 439-479.

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

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

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

	Spec. 1: No Controls			Spec. 2: Individual Controls		
	1980	**1990**	**2000**	**1980**	**1990**	**2000**
MSAs 0.25 - 1.5 million	0.14***	0.18***	0.19***	0.09***	0.13***	0.12***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MSAs > 1.5 million	0.24***	0.31***	0.32***	0.17***	0.24***	0.22***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R-squared	0.03	0.05	0.04	0.25	0.34	0.29

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

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

	No Controls	Individual Controls	Individual Controls and Fixed Effects	No Controls	Individual Controls	Individual Controls and Fixed Effects
	Temporally Deflated Only			Spatially and Temporally Deflated		
Specification	**1**	**2**	**3**	**1**	**2**	**3**
	Panel A: Full Sample					
Medium	0.20***	0.15***	0.08***	0.16***	0.10***	0.06***
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Large	0.30***	0.23***	0.14***	0.13***	0.06**	0.02
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
R-squared	0.04	0.30	0.61	0.01	0.29	0.60
	Panel B: College or More					
Medium	0.24***	0.23***	0.08***	0.18***	0.17***	0.06*
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)
Large	0.28***	0.29***	0.12***	0.11**	0.11**	0.02
	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.03)
R-squared	0.03	0.26	0.61	0.01	0.27	0.61
	Panel C: High School Graduates Only					
Medium	0.13***	0.13***	0.08***	0.10***	0.10***	0.07**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Large	0.23***	0.23***	0.18***	0.04	0.05	0.04
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)
R-squared	0.03	0.18	0.54	0.01	0.19	0.53

Notes: Each regression in Panel A uses data on 1,754 white men and has 30,367 observations based on quarterly data. Panel B has 8,899 observations on 583 individuals. Panel C has 12,646 observations on 674 individuals. Individual controls are years of education and quadratics in age and experience. Standard errors are clustered by location. Complete sample selection rules are explained in the text.

Table 3: Attributes at 15 Years of Work Experience as a Function of Location

	Job-Job Changes		Job-Unemployment-Job Changes				Frac. in Location		Migration Since LF Entry		
	Within	To	Within	Length	To	Length	At Entry	At 15 Yrs	To S	To M	To L
Location	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Panel A: Full Sample											
Small	2.7	0.7	1.8	22.9	0.4	4.7	0.32	0.30	0.78	0.15	0.07
Medium	3.0	0.6	1.7	19.1	0.3	3.4	0.36	0.39	0.10	0.83	0.08
Large	3.0	0.4	1.6	16.0	0.3	2.2	0.32	0.30	0.08	0.13	0.79
Panel B: College or More											
Small	1.7	0.9	0.7	6.0	0.4	2.3	0.23	0.21	0.61	0.26	0.13
Medium	2.2	0.7	0.8	6.2	0.3	3.4	0.40	0.42	0.10	0.76	0.14
Large	2.4	0.6	0.7	7.2	0.3	2.1	0.37	0.37	0.09	0.15	0.76
Panel C: High School Graduates Only											
Small	3.1	0.6	2.1	31.2	0.4	4.6	0.36	0.37	0.88	0.08	0.04
Medium	3.3	0.4	2.2	27.1	0.3	3.0	0.36	0.37	0.10	0.88	0.01
Large	3.0	0.3	2.3	26.7	0.2	2.4	0.28	0.25	0.08	0.08	0.84

Notes: The sample includes all individuals used for the regressions in Table 2 except those who we do not observe for at least 15 years of work experience. Columns marked "Within" report numbers for job changes within location whereas columns marked "To" report job changes across locations to locations of the indicated size. Each entry is calculated as the total amount of the quantity indicated in the column header for the sample indicated in the panel header divided by the sum of the fraction of time spent by everybody in the sample in the location category given in the row header. Therefore, each entry is the amount of each quantity experienced by the average individual over the first 15 years of work experience if he were to live in the indicated location for the full time. "Length" refers to total length of all unemployment spells. "LF" stands for labor force. The full sample includes 1,425 men, including 466 in the college sample and 566 in the high school sample.

Table 4: ΔLog Wage Regressions

	Temporally Deflated Only			Spatially and Temporally Deflated		
	All	**College**	**HS**	**All**	**College**	**HS**
ΔExperience	0.031**	0.046**	0.016	0.031**	0.046**	0.017
in Small	(0.012)	(0.019)	(0.023)	(0.012)	(0.019)	(0.023)
ΔExperience	0.056***	0.062***	0.046***	0.054***	0.060***	0.045***
in Medium	(0.010)	(0.018)	(0.013)	(0.010)	(0.018)	(0.013)
ΔExperience	0.059***	0.064***	0.055***	0.057***	0.062***	0.054***
in Large	(0.011)	(0.023)	(0.012)	(0.010)	(0.022)	(0.012)
ΔExp^2	0.001	0.000	0.003	0.002	0.001	0.004
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
ΔExp^3	-0.000*	-0.000	-0.000*	-0.000**	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Job to Job	0.066***	0.081*	0.078***	0.066***	0.082*	0.079***
in Small	(0.018)	(0.043)	(0.028)	(0.018)	(0.043)	(0.028)
Job to Job	0.097***	0.126***	0.094***	0.096***	0.126***	0.093***
in Medium	(0.010)	(0.024)	(0.016)	(0.010)	(0.024)	(0.016)
Job to Job	0.079***	0.083***	0.078***	0.078***	0.085***	0.077***
in Large	(0.012)	(0.022)	(0.015)	(0.012)	(0.023)	(0.015)
Job-Un-Job	-0.017	0.043	-0.050**	-0.015	0.043	-0.047*
in Small	(0.017)	(0.092)	(0.025)	(0.017)	(0.092)	(0.025)
Job-Un-Job	-0.027	-0.022	-0.028	-0.028	-0.022	-0.029
in Medium	(0.019)	(0.047)	(0.026)	(0.019)	(0.046)	(0.026)
Job-Un-Job	0.021	0.026	0.025	0.021	0.028	0.023
in Large	(0.015)	(0.054)	(0.020)	(0.015)	(0.054)	(0.019)
Job-Job+Move	0.099***	0.212***	0.032	0.126***	0.245***	0.053
to Small	(0.032)	(0.055)	(0.057)	(0.032)	(0.057)	(0.058)
Job-Job+Move	0.116***	0.118***	0.019	0.133***	0.144***	0.027
to Medium	(0.036)	(0.044)	(0.058)	(0.032)	(0.043)	(0.058)
Job-Job+Move	0.085**	0.113***	-0.134	0.055*	0.107***	-0.204**
to Large	(0.034)	(0.041)	(0.101)	(0.032)	(0.040)	(0.100)
Job-Un-Job +	-0.033	-0.013	0.000	0.004	0.019	0.029
Move to Small	(0.042)	(0.101)	(0.058)	(0.041)	(0.100)	(0.057)
Job-Un-Job +	-0.072*	-0.007	-0.089	-0.028	0.028	-0.025
Move to Medium	(0.044)	(0.095)	(0.063)	(0.044)	(0.091)	(0.062)
Job-Un-Job +	0.161***	0.184**	0.234***	0.109*	0.143**	0.158*
Move to Large	(0.055)	(0.074)	(0.071)	(0.058)	(0.060)	(0.087)
Unobs. Job	-0.105***	-0.180**	-0.140**	-0.100**	-0.163**	-0.142**
to Small	(0.038)	(0.082)	(0.068)	(0.040)	(0.081)	(0.069)
Unobs. Job	-0.086	-0.035	-0.164*	-0.085	-0.052	-0.152*
to Medium	(0.074)	(0.197)	(0.087)	(0.073)	(0.191)	(0.086)
Unobs. Job	-0.077	-0.014	-0.107	-0.085	-0.038	-0.093
to Large	(0.080)	(0.186)	(0.099)	(0.081)	(0.186)	(0.100)
Observations	21481	6276	9020	21481	6276	9020
R-squared	0.019	0.028	0.020	0.020	0.030	0.020

Notes: Each column is a separate regression of the change in log wage on functions of experience or labor market transitions listed at left. The regression specification is given by Equation (3) in the text. The 185 cases of gaps between wage observations exceeding 10 quarters are excluded.

Table 5: Log Wage Growth Decomposition
0 to 15 Years of Work Experience

	Wage at LF Entry		Within	Job to Job		Job to Unemp. to Job		Un-	Total
	COL Adj.	Unadjusted	Job	Within	Between	Within	Between	Known	Growth
Panel A: Full Sample									
Small	2.09	2.04	0.28	0.18	0.06	-0.03	-0.02	-0.01	0.47
Medium	2.13	2.12	0.44	0.25	0.06	-0.05	-0.02	0.00	0.67
Frac of Diff With Small			0.74	0.34	-0.01	-0.06	-0.01	0.01	
Large	2.02	2.13	0.51	0.22	0.04	0.01	0.03	-0.01	0.80
Frac of Diff With Small			0.66	0.11	-0.06	0.14	0.16	0.00	
Panel B: College or More									
Small	2.27	2.22	0.39	0.11	0.19	0.00	-0.02	-0.02	0.65
Medium	2.26	2.24	0.56	0.25	0.08	-0.03	0.00	-0.01	0.87
Frac of Diff With Small			0.81	0.64	-0.49	-0.11	0.11	0.05	
Large	2.24	2.34	0.61	0.20	0.07	0.00	0.04	0.00	0.91
Frac of Diff With Small			0.83	0.34	-0.47	0.01	0.22	0.07	
Panel C: High School Graduates Only									
Small	2.00	1.96	0.27	0.24	0.02	-0.08	0.00	-0.02	0.42
Medium	2.04	2.01	0.34	0.28	0.01	-0.07	-0.03	-0.02	0.53
Frac of Diff With Small			0.67	0.38	-0.02	0.13	-0.20	0.05	
Large	1.88	1.99	0.44	0.22	-0.03	0.04	0.05	-0.01	0.70
Frac of Diff With Small			0.59	-0.10	-0.16	0.43	0.18	0.06	

Notes: Reported numbers decompose average temporally deflated wage growth by location from 0 to 15 years of experience. As in Table 3, each entry is expressed for the average individual as if he were to spend the totality of his first 15 years of experience in the given location type. The unknown component of wage growth comes from jobs for which we do not observe a wage. Analogous elements of cost of living adjusted wage growth are very similar to those reported in this table. "COL Adjusted" in the first column refers to wages at labor force entry that have been spatially and temporally deflated. "Unadjusted" in the second column refers to wages that have only been temporally deflated.

Table 6: Actual and Predicted Mobility Rates

Mobility Rates	College Sample Data	College Sample Model	High School Sample Data	High School Sample Model
Panel A: Transition Rates				
Job Changes	7.1%	7.1%	11.0%	11.0%
Job to Job Transitions	5.2%	5.1%	6.3%	6.2%
Job to Unemployment Transitions	1.9%	2.0%	4.7%	4.8%
Location Changes	1.8%	1.8%	1.4%	1.4%
Panel B: Transition Rates by Location				
Job to Job in Small	4.1%	4.0%	5.9%	5.8%
Job to Job in Medium	4.5%	4.3%	6.1%	5.8%
Job to Job in Large	4.8%	4.8%	5.6%	5.6%
Job to Unemployment in Small	1.6%	1.7%	4.4%	4.6%
Job to Unemployment in Medium	1.7%	1.8%	4.3%	4.4%
Job to Unemployment in Large	1.5%	1.6%	4.4%	4.6%
Unemployment Duration in Small	5.6	5.7	5.9	5.8
Unemployment Duration in Medium	4.6	4.7	5.5	5.5
Unemployment Duration in Large	6.1	5.9	5.0	5.0

Notes: Each entry above the bottom three rows gives the percent of observations exhibiting the indicated transition in actual data and data simulated using estimated parameters reported in Table A1. The numbers in panel B are calculated using only transitions that did not involve a change of location The final three rows show averages of actual and simulated data across unemployment spells whose lengths are measured in weeks.

Table 7: Actual and Predicted Migration Conditional on Moving

Period t-1	Period t Location: Data			Period t Location: Model		
Location	**Small**	**Medium**	**Large**	**Small**	**Medium**	**Large**
Panel A: College Sample						
Small	13.0%	10.0%	5.4%	11.6%	9.7%	7.2%
Medium	10.7%	16.7%	13.2%	10.6%	18.1%	11.7%
Large	5.6%	12.9%	12.5%	8.0%	10.6%	12.6%
Panel B: High School Sample						
Small	26.2%	14.1%	6.9%	25.7%	12.4%	8.9%
Medium	13.3%	12.1%	9.3%	13.7%	13.6%	7.5%
Large	6.3%	6.1%	5.9%	7.4%	5.7%	5.0%

Notes: Each entry gives the percent of individuals who change locations that move from the location type listed at left to the location type listed at top. The left half of the table reports these numbers from the data and the right half of the table reports these numbers from data simulated using the parameter estimates reported in Table A1.

Table 8: Log Wages and AFQT Versus Unobserved Ability

			log wage		
Specification	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
			Panel A: College Sample		
Medium		0.17	0.15	0.14	0.13
		(0.02)	(0.02)	(0.01)	(0.01)
Large		0.12	0.10	0.24	0.23
		(0.02)	(0.02)	(0.01)	(0.01)
AFQT	0.04		0.07		0.03
	(0.01)		(0.01)		(0.01)
Pr(h=h_H)	0.60			0.65	0.64
	(0.01)			(0.01)	(0.01)
R-Squared	0.41	0.24	0.25	0.43	0.43
			Panel B: High School Sample		
Medium		0.10	0.10	0.15	0.15
		(0.01)	(0.01)	(0.01)	(0.01)
Large		0.06	0.06	0.09	0.09
		(0.01)	(0.01)	(0.01)	(0.01)
AFQT	0.05		0.06		0.05
	(0.00)		(0.00)		(0.00)
Pr(h=h_H)	0.21			0.31	0.30
	(0.01)			(0.01)	(0.01)
R-Squared	0.22	0.18	0.19	0.23	0.24

Notes: Each entry is a coefficient or standard error from a regression of log wage on the variables listed at left, experience and experience squared. Each regression uses all available wage observations from those for whom an AFQT score is observed. Regressions have 8,673 observations in Panel A and 12,215 observations in Panel B. AFQT is not observed for all individuals and is measured in standard deviation units with standard deviations calculated separately for each sample. Pr(h=h_H) is the estimated probability that each individual in the data set is high type calculated as explained in the text.

Table 9: Counterfactual City Size Wage Premia
College Sample

Experiment	Regression Coefficients Medium	Regression Coefficients Large	Gap With Reference Medium	Gap With Reference Large
Panel A: Premia for Comparison				
1. Baseline Data	0.14	0.09		
2. Model	0.14	0.09		
	[0.07, 0.22]	[0.02, 0.17]		
3. Restricted Mobility (Compare to Row 2)	0.17	0.17	0.03	0.08
	[0.09, 0.27]	[0.08, 0.26]		
4. Price Differential	0.06	0.18		
Panel B: Counterfactuals With Free Mobility (Row 2 Reference)				
Equalize Ability Distribution at LF Entry	0.13	0.14	-0.01	0.05
Equalize Search & Matching Across Locations	0.15	0.15	0.02	0.05
Equalize Return to Experience Across Locs	**0.06**	**-0.09**	-0.08	-0.18
Implied Reduction in Marginal Productivity Gap			38%	68%
Equalize Nominal Intercepts Across Locs	**0.02**	**-0.07**	-0.11	-0.16
Implied Reduction in Marginal Productivity Gap			58%	60%
Panel C: Counterfactuals With Restricted Mobility (Row 3 Reference)				
Equalize Ability Distribution at LF Entry	0.15	0.24	-0.02	0.07
Equalize Search & Matching Across Locations	0.19	0.16	0.02	-0.02
Equalize Returns to Experience Across Locs	**0.08**	**-0.03**	-0.09	-0.21
Implied Reduction in Marginal Productivity Gap			37%	59%
Equalize Nominal Intercepts Across Locs	**0.05**	**-0.01**	-0.12	-0.19
Implied Reduction in Marginal Productivity Gap			50%	53%

Notes: Panel A Row 1 presents average wage premia from the raw data, Row 2 shows premia based on simulated data and Row 3 is based on simulated data for which mobility cost is infinite. Other estimates are based on simulated data using parameter values achieving the listed scenario. For ability distribution equalization, we set probabilities of labor force entry by type across locations equal to their weighted average across locations. For search friction equalization, we set all arrival and separation rates equal to their weighted averages across locations, where the weights are based on composition by type across locations at labor force entry. Equalization of returns to experience across locations is achieved analogously. Imposing all restrictions in Panel C simultaneously generates simulated wage premia of 0 in both location types.

Equalization of nominal intercepts across locations is achieved by setting intercepts from the wage process to differ by the average price differences across location type given in Row 4 and renormalized such that the weighted average equals the weighted average of real intercept terms reported in Table A1. Bolded entries indicate that counterfactual estimates lie outside the indicated 95% confidence interval for the reference. Italics indicate lying outside the 90% confidence interval instead.

Table 10: 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
Panel A: Premia for Comparison				
1. Baseline Data	0.09	0.04		
2. Model	0.09	0.04		
	[0.06, 0.13]	[0.00, 0.08]		
3. Restricted Mobility (Compare to Row 2)	0.10	0.03	0.01	-0.01
	[0.06, 0.15]	[-0.02, 0.07]		
4. Price Differential	0.03	0.19		
Panel B: Counterfactuals With Free Mobility (Row 2 Reference)				
Equalize Ability Distribution at LF Entry	0.12	0.05	0.02	0.01
Equalize Search & Matching Across Locations	**0.13**	*0.08*	0.04	0.04
Equalize Returns to Experience Across Locs	**0.05**	**-0.08**	-0.05	-0.13
Implied Reduction in Marginal Productivity Gap			39%	55%
Equalize Nominal Intercepts Across Locs	**-0.05**	**-0.09**	-0.14	-0.14
Implied Reduction in Marginal Productivity Gap			118%	59%
Panel C: Counterfactuals With Restricted Mobility (Row 3 Reference)				
Equalize Ability Distribution at LF Entry	**0.15**	0.06	0.05	0.02
Equalize Search & Matching Across Locations	*0.13*	0.05	0.03	0.02
Equalize Returns to Experience Across Locs	**0.05**	**-0.10**	-0.06	-0.13
Implied Reduction in Marginal Productivity Gap			44%	60%
Equalize Nominal Intercepts Across Locs	**-0.04**	**-0.11**	-0.14	-0.14
Implied Reduction in Marginal Productivity Gap			111%	63%

Notes: See the notes to Table 9 for a detailed description of the table.

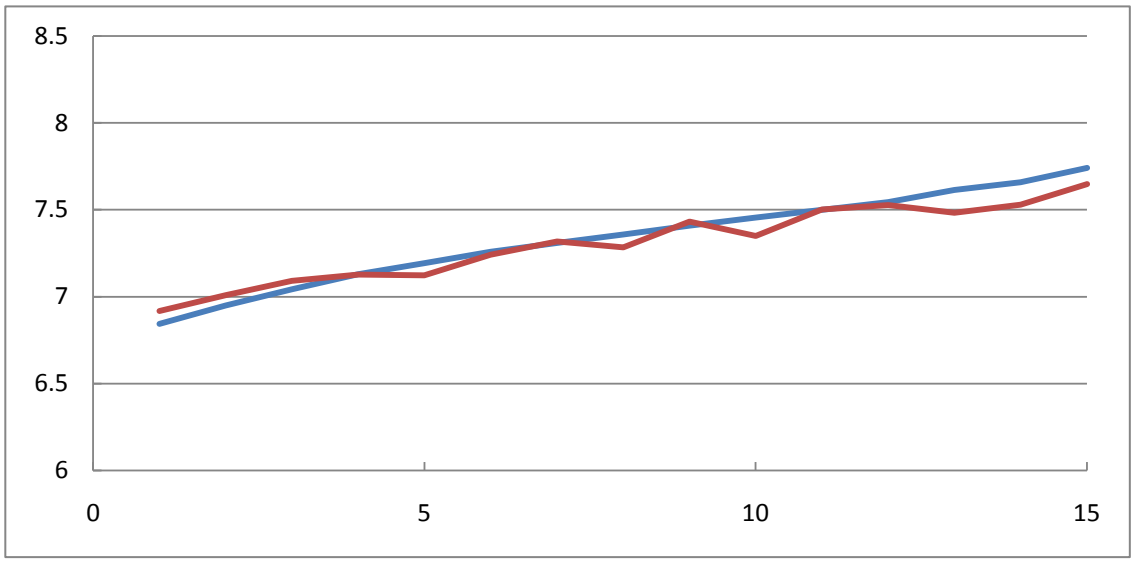
Table A1: Parameter Estimates from the Structural Model

Parameter	Description	College Sample Low Type	College Sample High Type	High School Sample Low Type	High School Sample High Type
A. Components of Wages					
β_0^1	Wage Constant for Living in Small Locations	6.66***	7.02***	6.41***	6.62***
β_0^2	Wage Constant for Living in Medium Locations	6.70***	7.11***	6.52***	6.71***
β_0^3	Wage Constant for Living in Large Locations	6.66***	7.04***	6.26***	6.70***
β_1^1	Return to Experience from Work in Small	0.052***	0.087***	0.061***	0.054***
β_1^2	Return to Experience from Work in Medium	0.058***	0.079***	0.074***	0.071***
β_1^3	Return to Experience from Work in Large	0.069***	0.100***	0.062***	0.051***
θ_2	Multiplier for Living in Medium Locations	1.36***		0.96***	
θ_3	Multiplier for Living in Large Locations	1.35***		1.58***	
β_2	Coefficient On Experience Squared in Small Locations	-0.0056***		-0.0054***	
ϕ_2	Multilier for Medium Locations: $\beta_2^2 = \phi_2\beta_2$	1.16***		1.05***	
ϕ_3	Multiplier for Large Locations: $\beta_2^3 = \phi_3\beta_2$	1.21***		1.27***	
β_3	Coefficient On Experience Cubed	0.00024***		0.00020***	
B. Arrival and Separation Rates					
λ_1	Job Offer Arrival Rate Within Small Locations	0.11***	0.06***	0.10***	0.22***
λ_2	Job Offer Arrival Rate Within Medium Locations	0.12***	0.07***	0.13***	0.26***
λ_3	Job Offer Arrival Rate Within Large Locations	0.11***	0.11***	0.12***	0.24***
λ^u_1	Job Offer Arrival Rate Within Small from Unemployed	0.14***	0.08***	0.13***	0.18***
λ^u_2	Job Offer Arrival Rate Within Medium from Unemployed	0.19***	0.11***	0.18***	0.17***
λ^u_3	Job Offer Arrival Rate Within Large from Unemployed	0.14***	0.09***	0.18***	0.21***
ρ_1	Receive Rate in Small: $\lambda_{21}=\rho_1\lambda_2$ and $\lambda^u_{21}=\rho_1\lambda^u_2$	0.39***		0.20***	
ρ_2	Receive Rate in Medium	0.29***		0.09***	
ρ_3	Receive Rate in Large	0.18***		0.05***	
γ	Own Location Type Multiplier: $\lambda_{11}=\lambda_1\rho_1\gamma$ and $\lambda^u_{11}=\lambda^u_1\rho_1\gamma$	1.34***		2.31***	
δ_1	Separation Rate in Small	0.003	0.006***	0.009***	0.009***
δ_2	Separation Rate in Medium	0.008***	0.008***	0.007***	0.030***
δ_3	Separation Rate in Large	0.004***	0.016***	0.002	0.006***
C. Amenities					
α_2	Amenity of Medium Locations	0.18***		0.06***	
α_3	Amenity of Large Locations	0.20***		0.06***	
D. Benefits and Costs					
$\ln(b_1)$	Unemployment benefit in Small	-0.73***	0.01	0.34***	0.89***
$\ln(b_2)$	Unemployment benefit in Medium	-1.36***	-0.95***	0.39***	0.76***
$\ln(b_3)$	Unemployment benefit in Large	-0.39***	-1.25***	0.31***	0.73***
C	Moving Cost - Constant	14.42***		7.43***	
C	Moving Cost - Linear term	0.95***		3.99***	
C	Moving Cost - Quadratic Term	0.04***		-0.20***	
E. Heterogeneity					
π^1	Probability of Type Given Start in Small Location	0.54***	0.46***	0.53***	0.47***
π^2	Probability of Type Given Start in Medium Location	0.49***	0.51***	0.71***	0.29***
π^3	Probability of Type Given Start in Large Location	0.66***	0.34***	0.65***	0.35***
F. Distributions					
σ_1	Standard Deviation of Match in Small	0.36***	0.34***	0.34***	0.38***
σ_2	Standard Deviation of Match in Medium	0.39***	0.31***	0.26***	0.50***
σ_3	Standard Deviation of Match in Large	0.38***	0.36***	0.32***	0.31***
σ_U	Standard Deviation of Utility Shock for Unemployed	8.82***		6.76***	
σ_S	Standard Deviation of Job Switching Cost Shock	8.58***		3.63***	
σ_M	Standard Deviation of Moving Cost Shock	22.05***		23.94***	
σ_{u1}	Standard Deviation of Wage Measurement Error - Small	0.30***		0.27***	
σ_{u2}	Standard Deviation of Wage Measurement Error - Medium	0.29***		0.22***	
σ_{u3}	Standard Deviation of Wage Measurement Error- Large	0.29***		0.28***	

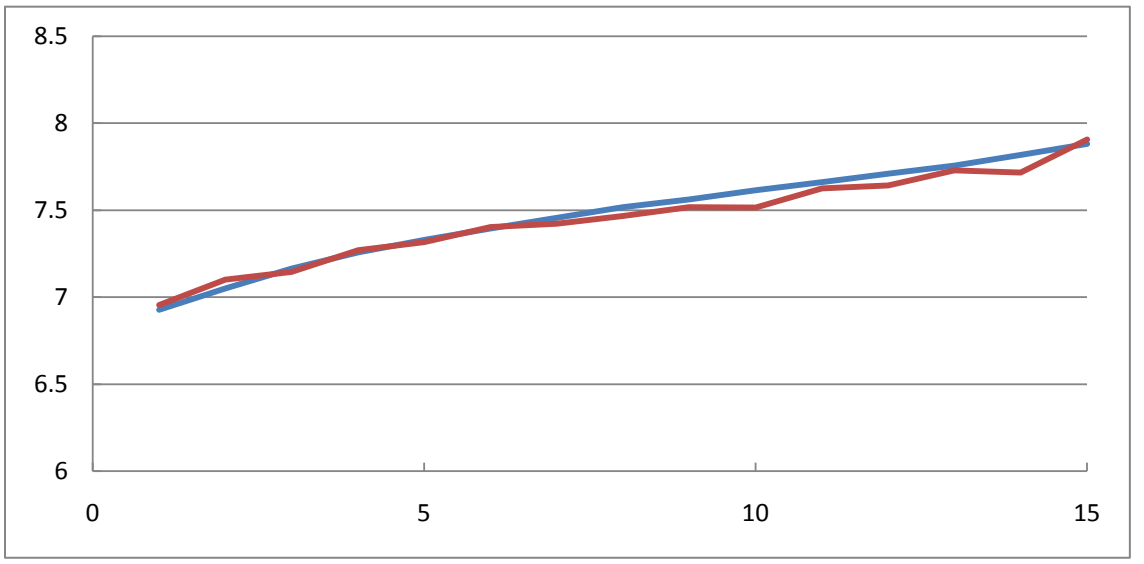
Notes: This table shows the universe of estimated model parameters. *** indicates significance at the 1% level, ** indicates significance at the 5% level and * indicates significance at the 10% level. The discount rate is calibrated to 0.95.

Figure 1: Actual and Predicted Wages by Experience: College Sample

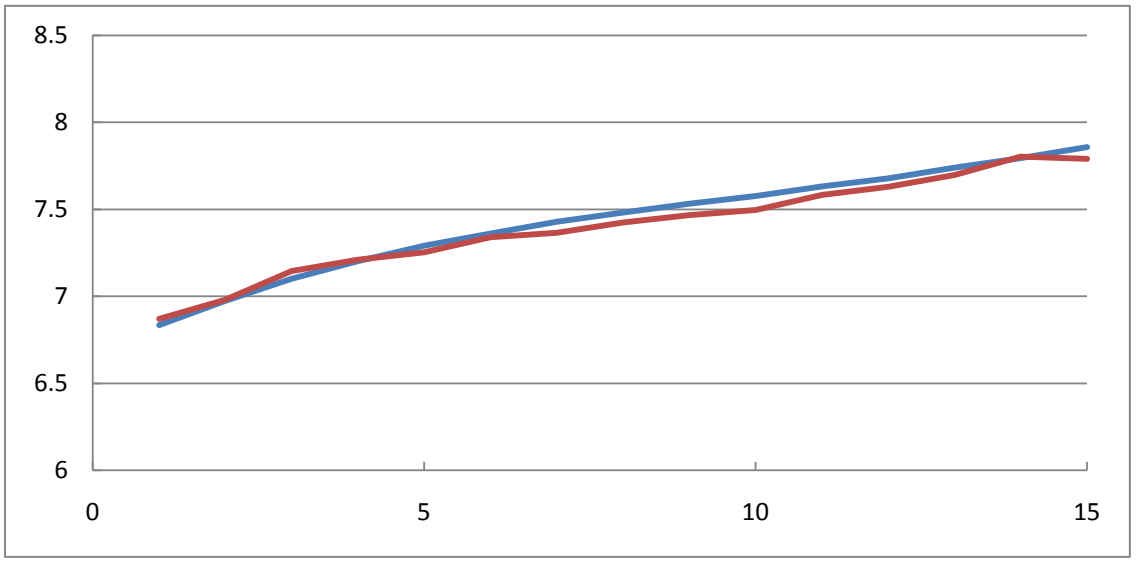
Panel A: Small Locations



Panel B: Medium Locations



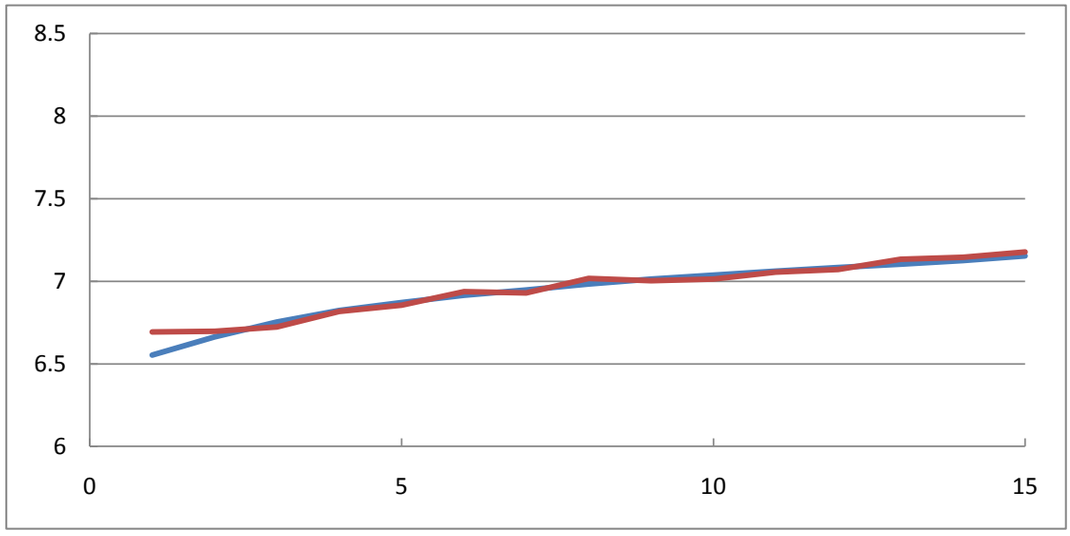
Panel C: Large Locations



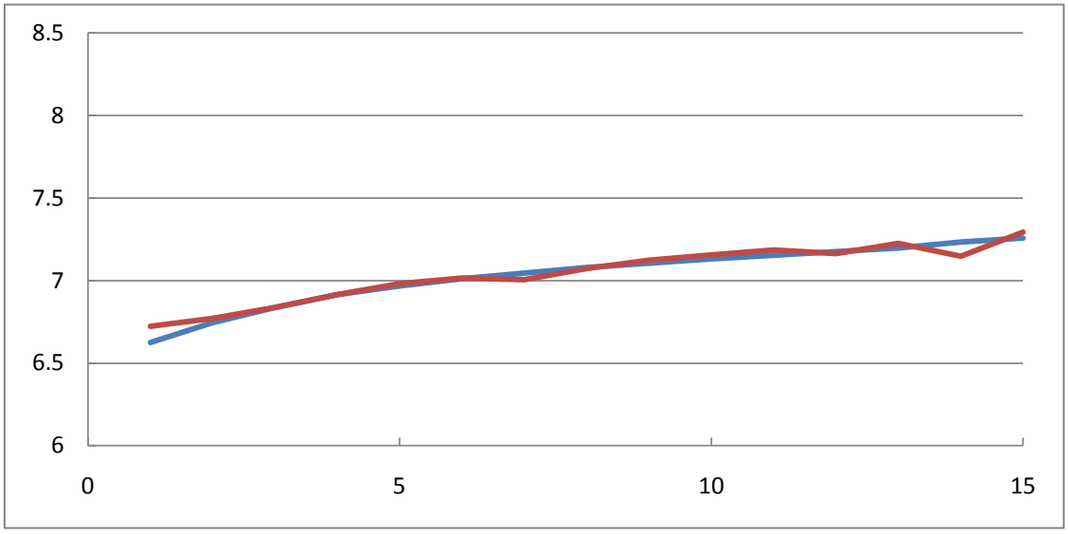
Notes: Each plot shows mean actual and predicted wages by work experience. Means are calculated separately within each year of work experience. Actual mean wages are drawn in red and mean wages based on model simulations are drawn in blue. Table A1 lists the parameter values used for simulations.

Figure 2: Actual and Predicted Wages by Experience: High School Sample

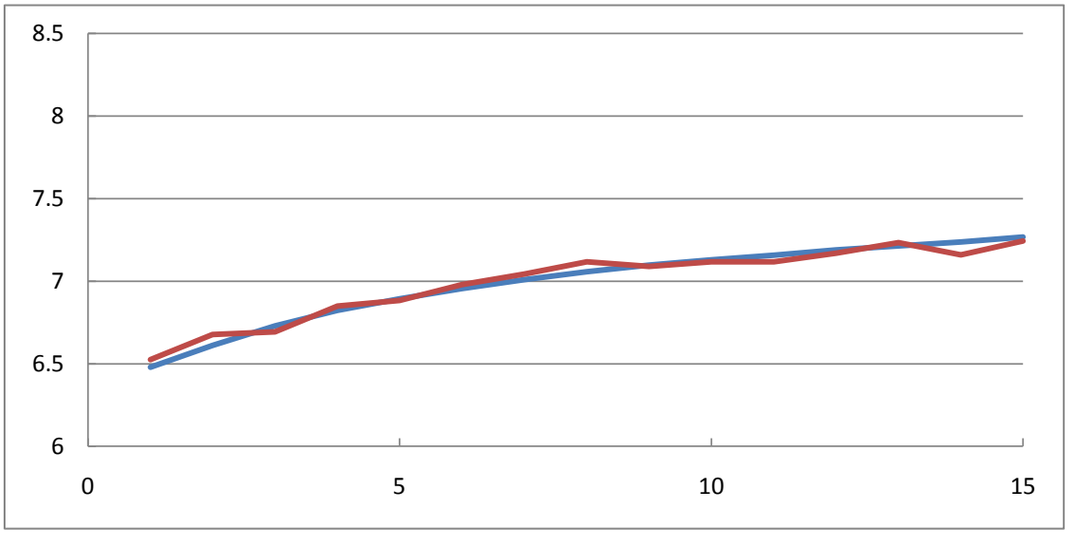
Panel A: Small Locations



Panel B: Medium Locations



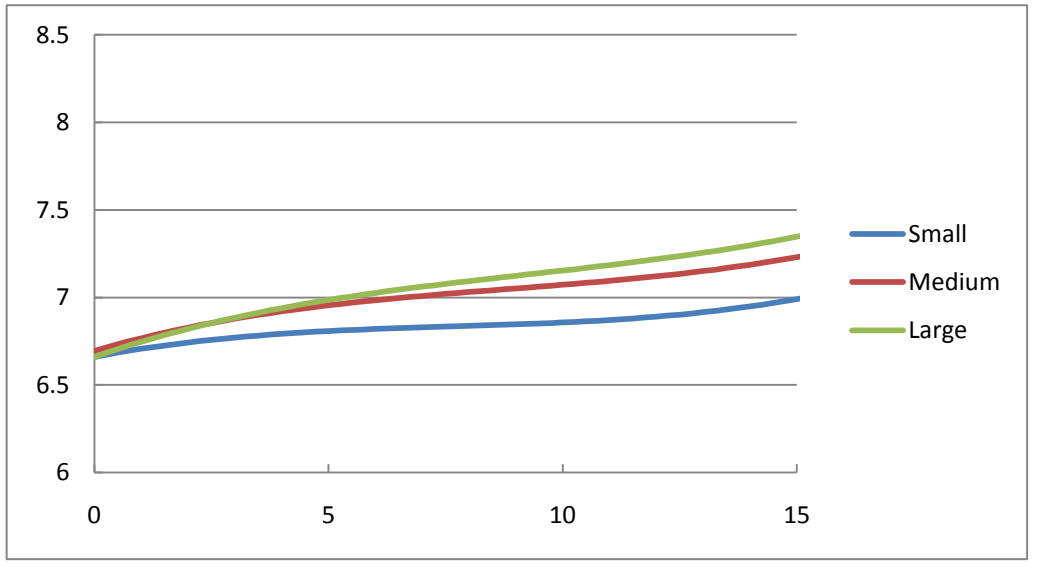
Panel C: Large Locations



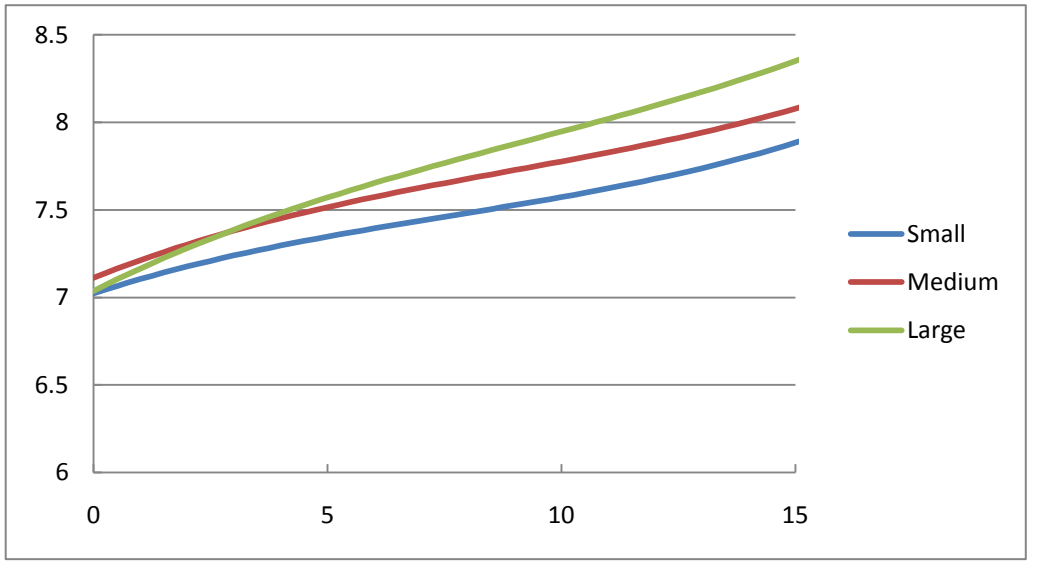
Notes: See the notes to Figure 1 for a description of the plots.

Figure 3: Experience Profiles Implied by Parameter Estimates

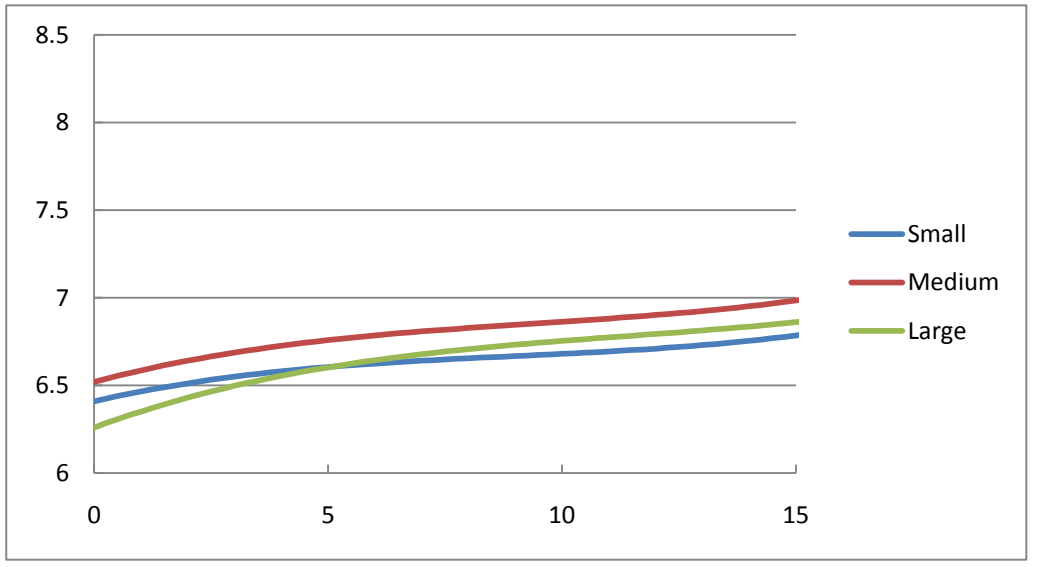
Panel A1: College Graduates, Low Ability



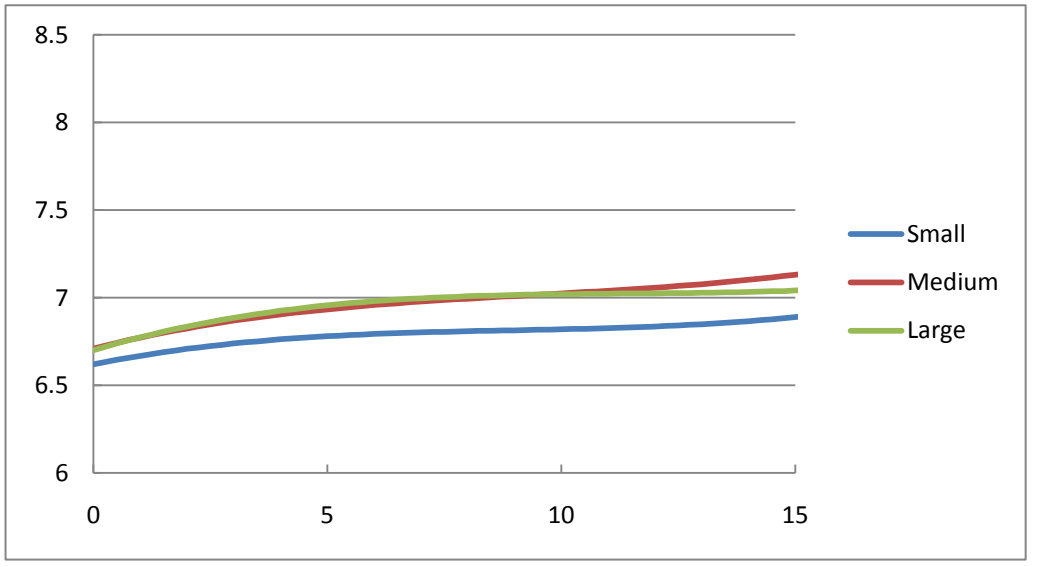
Panel A2: College Graduates, High Ability



Panel B1: High School Graduates, Low Ability

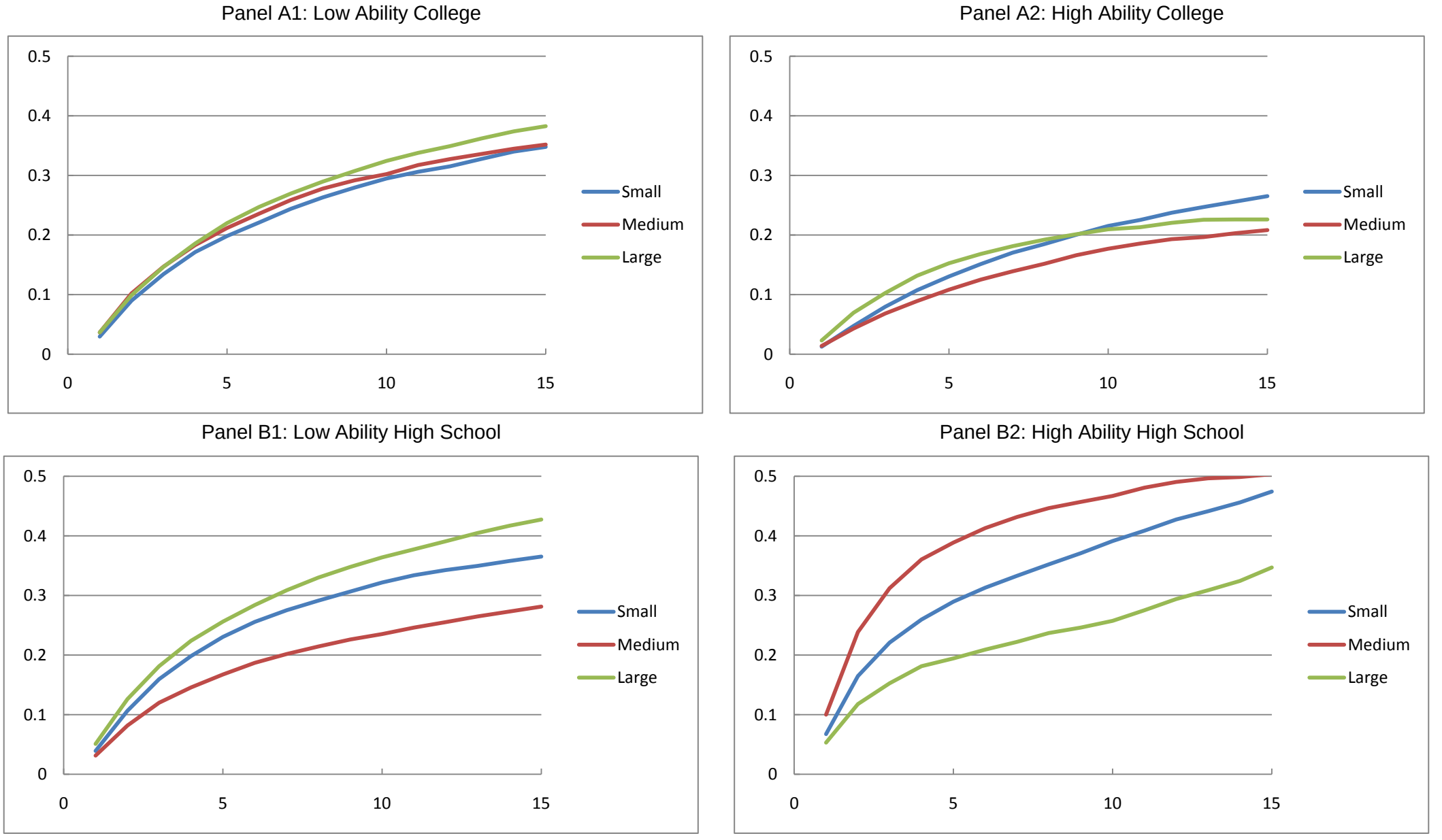


Panel B2: High School Graduates, High Ability



Notes: Each panel graphs real wages excluding the firm-worker match and measurement error components as functions of years of experience by location type, education and ability. These are plots of Equation (7) in the text but restricting ε and u to 0 and assuming no mobility across locations. The parameters used to graph these functions are found in Table A1.

Figure 4: Firm-Worker Match Component of the Wage



Notes: Each panel graphs the mean of the match component ε of the real wage assuming no mobility across locations. Data for the plots are constructed by simulating the model using parameters in Table A1 except that the mobility costs are set to be infinite.