Review of Economic Studies (2012) **79**, 88–127 doi: 10.1093/restud/rdr022
© The Author 2011. Published by Oxford University Press on behalf of The Review of Economic Studies Limited.
Advance access publication 19 August 2011

Understanding the City Size Wage Gap

NATHANIEL BAUM-SNOW
Brown University and NBER
and
RONNI PAVAN
University of Rochester

First version received February 2010*; final version accepted February* 2011 (*Eds.*)

In this paper, we decompose city size wage premia into various components. We base these decompositions on an estimated 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 observed city size wage premia. Variation in returns to experience is more important for generating wage premia between large and small locations, while differences in wage intercepts are more important for generating wage premia between medium and small locations. Sorting on unobserved ability within education group and differences in labour market search frictions and distributions of firm-worker match quality contribute little to observed city size wage premia. These conclusions hold for separate samples of high school and college graduates.

Key words: Agglomeration, Wage Growth, Urban Wage Premium

JEL Codes: J24, J31, J61, J64, R12, R23

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% higher in metropolitan areas (MSAs) of over 1.5 million people than in rural areas and MSAs of less than 250 thousand people. Indeed, the relationship between wages and population is monotonically increasing by about one percentage point for each additional 100 thousand in population over the full range of MSA size. In addition, the city size wage gap has become considerably steeper since the 1980s when large MSAs had wages that were 24% greater than rural areas and small MSAs. Because firm productivity is closely related to wages, these wage premia are evidence that workers and 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

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

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 and Wolpin (1997) allows for recovery of parameters indexed by unobserved worker type. As in Pavan (2011), 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, differences in wage intercepts account for about two-thirds of the wage premium between medium-sized MSAs and smaller areas, with most of the rest accounted for by differences in returns to experience across these locations. For high school graduates, wage intercept differences explain almost the entire medium–small city size wage premium. However, differences in returns to experience are more important for explaining wage premia of the largest MSAs over small MSAs and rural areas. Experience effects explain 57% of this gap for college graduates and 78% of this gap for high school graduates. We find that differences across locations in job offer arrival rates and dispersion of firm-worker match quality distributions do not contribute appreciably to observed city size wage premia.[1]

While we confirm evidence from elsewhere (Combes, Duranton and Gobillon 2008) 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 some evidence of negative sorting on unobserved ability. Among high school graduates, higher latent ability men are less likely to enter the labour force in large cities. Among college graduates, higher ability men tend to move away from large cities at a faster rate than lower ability men over the life-cycle. This evidence of mild negative sorting on latent ability into larger cities manifests itself as even greater wage premia for 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 and Puga (2004). Potential mechanisms include sharing of inputs produced at large efficient scales, sharing risk, and taking advantage of greater opportunities for division of labour. We leave to future research the empirical evaluation of specific micro-founded theories of how larger cities foster more rapid worker learning and sharing. Investigation of why the relative importance of these two forces differ by city size would complement such efforts.

Our results indicating a lack of importance of search frictions for generating city size wage premia are of particular interest. Petrongolo and 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 labour

1. Standard urban models typically imply that both the density and the 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 and Hall (1996) is an important study that uses a density measure instead. Since MSA population and density are strongly positively correlated, the choice of measure matters little for our purposes.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

markets. However, the existing literature neither examines the possibility that the distributions of latent firm-worker match components of productivity may be different across labour 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, e.g. Glaeser and Maré (2001) and Gould (2007), does not distinguish whether this pattern is generated by ascension of steeper job ladders or by 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 and 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. 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 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 and Strange (2003), Arzaghi and Henderson (2008), and Greenstone, Hornbeck and Moretti (2010). 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.

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

2. EMPIRICAL OBSERVATIONS

2.1. *Conceptual environment*

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, labour L, capital K, and a composite non-traded good R all enter as factors of production with input prices w_j,

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.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

r, and p_j, respectively. Locations are indexed by j and a_j represents a local productive amenity that incorporates agglomeration forces.

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

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 labour and ϕ_R is firms' expenditure share on the non-traded good.

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

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 could understate total factor productivity differences because local prices are positively correlated with wages.[3]

Though wages unadjusted for cost of living differences across locations are most informative about workers' marginal productivities, analyses of workers' location decisions would be incomplete without accounting for such differences. When workers consider migration, they evaluate both the relative wage and the relative cost of living across locations. To construct an appropriate cost of living index, we adopt the standard approach using a compensating differentials framework, i.e, the consumer analogue to the firm analysis described above.

Equation (3) specifies the indifference relationship for identical workers with preferences U over the vector of goods z, 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 $\overline{v}$.

$$\overline{v} = \max_{z} \left\{ U(z) + \lambda[w_j - \sum_g p_g^j z_g] \right\}. \quad (3)$$

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_g is the consumer expenditure share on good g.

$$\ln(w_0) = \ln(w_j) - \sum_g s_g[\ln(p_g^j) - \ln(p_g^0)]. \quad (4)$$

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.[4]

$$\text{INDEX}_j = \prod_g (p_g^j / p_g^0)^{s_g}. \quad (5)$$

Appendix A details the process and data sources we use to construct this index.

3. In his calibration of a similar model, Albouy (2010) uses $\phi_L = 0{\cdot}85$ and $\phi_R = 0{\cdot}025$ based on estimates in the literature, implying that the nominal city size wage gap overstates firm productivity differences by about 15%.

4. The linear approximation for small deviations in prices across locations and periods generalizes to allow for deviations of all sizes if utility is of the Cobb–Douglas form, which Davis and Ortalo-Magné (2011) argue accurately captures preferences over housing and other goods.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

2.2. *Data*

The primary data set used for this analysis is the National Longitudinal Survey of Youth 1979 (NLSY79) restricted use geocoded and work history files. Using this data set, we construct information on jobs, unemployment, wages, and migration patterns for a sample of white men aged 14–21 years on 31 December, 1978 from the time of their entry into the labour force until 15 years of work experience, year 2004, or their attrition from the survey, whichever comes first. We sample the weekly job history data four times every year for those who become attached to the labour force after 1 January, 1978 and observe wages in about one-quarter of the observations. Appendix A details how we construct the data including our sample selection rules.

2.3. *Descriptive patterns*

Table 1 reports city size wage premia with and without adjustment for cost of living differences across locations. The estimated nominal wage premium for medium-sized cities over smaller areas is 0·19, while that for large cities is 0·29. Controlling for education and a cubic in work experience reduces these coefficients to 0·14 and 0·22, respectively. Controlling for individual fixed effects additionally reduces these coefficients to 0·07 and 0·15.

TABLE 1
Estimates of city size wage premia

	No controls	Individual controls	Individual controls and fixed effects	No controls	Individual controls	Individual control and fixed effects
1980's MSA	Temporally deflated only			Spatially and temporally deflated		
population	1	2	3	1	2	3
			Panel A: full sample			
0·25–1·5	0·19***	0·14***	0·07***	0·14***	0·09***	0·05***
million	(0·03)	(0·03)	(0·02)	(0·03)	(0·02)	(0·02)
> 1·5 million	0·29***	0·22***	0·15***	0·11***	0·05*	0·03
	(0·03)	(0·03)	(0·02)	(0·03)	(0·03)	(0·02)
R-squared	0·04	0·26	0·60	0·01	0·24	0·59
			Panel B: College or more			
0·25–1·5	0·21***	0·20***	0·07**	0·15***	0·14***	0·05*
million	(0·05)	(0·04)	(0·03)	(0·04)	(0·04)	(0·03)
> 1·5 Million	0·27***	0·27***	0·12***	0·09*	0·09*	0·01
	(0·05)	(0·05)	(0·04)	(0·05)	(0·05)	(0·040)
R-squared	0·03	0·18	0·60	0·01	0·18	0·61
			Panel C: high school graduates only			
0·25–1·5	0·12***	0·12***	0·06*	0·09***	0·09***	0·05*
million	(0·03)	(0·03)	(0·03)	(0·03)	(0·03)	(0·03)
> 1·5 million	0·22***	0·22***	0·18***	0·03	0·04	0·05
	(0·03)	(0·03)	(0·04)	(0·04)	(0·03)	(0·04)
R-squared	0·03	0·14	0·52	0·01	0·14	0·52

Notes: Each regression in Panel A uses data on 1754 white men and has 25,363 observations based on quarterly data. Panel B has 7,555 observations on 583 individuals. Panel C has 10,436 observations on 674 individuals. Individual controls are four educational dummies and cubic polynomials in work experience. We only include observations from the first 15 years of work experience. Standard errors are clustered by location. Complete sample selection rules are explained in Appendix A. *** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

That the inclusion of individual fixed effects generates reductions in the urban-rural wage gradient, though not in 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 3. In general, the fixed-effect estimator provides inconsistent estimates if a worker's decision is influenced by an unobserved 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 effects of city size that are heterogeneous in worker characteristics and/or location characteristics. Indeed, a simple Roy (1951) model as developed by Borjas (1987) predicts that movers differ on unobservables from the overall population.

The right side of Table 1, Panel A presents analogous regression results using wages adjusted for the cost of living differences across MSAs. These estimates reflect the fact that cost of living in large MSAs is higher than that in smaller places. In all three of the reported specifications, real wages in medium-sized MSAs 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.[5] In Table 1, Panels B and C repeat the same analysis as in Panel A separately for samples of college and high school graduates. They indicate the same general patterns, though as should be expected, inclusion of controls for education and work experience have much smaller effects on estimated city size wage premia for these more selected samples.[6]

The results in Table 1 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 2–4 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 2 describe patterns of job turnover and unemployment by city size and education. Columns headed by "Within" indicate transitions within MSAs

5. 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 3 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.

6. Because others including Combes *et al*. (2010) have found that accounting for the potential endogeneity of city size using IV methods leads to only small reductions in estimated city size wage and productivity premia, our primary analysis does not attempt to account for potentially endogenous city size. Indeed using our data, the ordinary least square elasticity of wages with respect to 1980's population, controlling for education and a cubic in experience, is 0·055, while the analogous IV elasticity is 0·087. Using wages adjusted for cost of living, these estimated elasticities are 0·017 and 0·015, respectively. Following Ciccone and Hall (1996), we use 1870's population, population density, and distance to coast as instruments.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE 2
Attributes at 15 years of work experience as functions of location

	Job–job changes		Job-unemployment-job changes				Fraction in location	
Location	Within	To	Within	Length	To	Length	At Entry	At 15 Yrs
Size	1	2	3	4	5	6	7	8
				Panel A: full sample				
Small	2·7	0·7	1·8	22·9	0·4	4·7	0·32	0·30
Medium	3·0	0·6	1·7	19·1	0·3	3·4	0·36	0·39
Large	3·0	0·4	1·6	16·0	0·3	2·2	0·32	0·30
				Panel B: college or more				
Small	1·7	0·9	0·7	6·0	0·4	2·3	0·23	0·21
Medium	2·2	0·7	0·8	6·2	0·3	3·4	0·40	0·42
Large	2·4	0·6	0·7	7·2	0·3	2·1	0·37	0·37
				Panel C: high school graduates only				
Small	3·1	0·6	2·1	31·2	0·4	4·6	0·36	0·37
Medium	3·3	0·4	2·2	27·1	0·3	3·0	0·36	0·37
Large	3·0	0·3	2·3	26·7	0·2	2·4	0·28	0·25

Notes: The sample includes all individuals used for the regressions in Table 1 except those who we do not observe for at least 15 years of work experience. Columns marked "Within" report numbers of 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 labour force. Bootstrapped standard errors with samples clustered by individual reveal that the only statistically significant medium–small and large–small differences in Columns 1–6 are in Panel/Columns A2, A5, B1, and C2. College graduates are significantly more likely to be located in large locations at LF entry and 15 years of experience. The full sample includes 1425 men, including 466 in the college sample and 566 in the high school sample.

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 $\left(\sum_i y_{ij}\right)/\left(\sum_i \frac{t_{ij}}{T_i}\right)$ 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 experience living in the indicated city size category. In this ratio, y_{ij} is the total quantity of each object given in Table 2 column headers for each man i in location type j over his first T_i years of work experience and t_{ij} is the number of years spent working in location type j.[7]

Results reported in Table 2 reveal few patterns in labour market transitions that could plausibly lead to the prevalence of city size wage premia reported in Table 1. Within location, the only statistically significant large–small and medium–small gaps are for job to job transitions for

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 2–4, we only include individuals with a wage observation within 2 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.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE 3
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)
ΔExperience 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)
ΔExp3	−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)
Unobservable 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)
Unobservable 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)
Unobservable 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	21,481	6,276	9,020	21,481	6,276	9,020
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 labour market transitions listed at left. The regression specification is given by equation (6) in the text with standard errors clustered by location. The 185 cases of gaps between wage observations exceeding ten quarters are excluded. *** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

college graduates, monotonically increasing from 1·7 in small locations to 2·4 in large locations. To determine statistical significance, we block bootstrap standard errors using individual level clusters (unreported). For the pooled sample, workers experience significantly fewer of both types of labour market transitions involving migration to larger locations. However, these patterns are difficult to interpret given that many variables change simultaneously with residential relocation.

In Columns 7 and 8 of Table 2, we examine the extent of sorting on observed education and how it changes over the life-cycle. We see that college graduates disproportionately enter the labour force in larger locations. By 15 years of work experience, this sorting has strengthened slightly with a tendency of both education groups to move to medium-sized locations.

Table 3 reports the mean log wage growth associated with each type of labour market transition examined in Table 2. 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 (6) 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 2 excluding non-wage observations and we weight by the inverse of the number of wage observations for each individual. We index each class of labour market transition by location type j.

$$\Delta \ln(w_{it}) = \sum_j \beta_1^j \Delta \exp_{it}^j + \beta_2 \Delta \exp_{it}^2 + \beta_3 \Delta \exp_{it}^3 + \sum_j (c_1^j JJ_{it}^j + c_2^j JUJ_{it}^j + c_3^j JJm_{it}^j + c_4^j JUJm_{it}^j + c_5^j OTH_{it}^j) + \varepsilon_{it}. \quad (6)$$

This equation includes a cubic in experience, indicators for the four types of job changes for which quantities are reported in Table 2 and OTH_{it}^j, which is an indicator for situations in which a job is skipped due to a lack of wage information.[8] The variables JJ_{it}^j, JJm_{it}^j, JUJ_{it}^j, and $JUJm_{it}^j$ 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 3 makes clear that the endogeneity of labour market transitions and migration in equation (6) makes such a regression potentially problematic, the results in Table 3 still 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 monotonically 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 likely an important driver of the city size wage premium.[9]

To understand how the quantities in Table 2 and the elements of wage growth in Table 3 combine, Table 4 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, between jobs including an unemployment spell, and unknown. The

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

9. Based on a similar analysis, Wheeler (2006) concludes that wage growth from job transitions is more steeply increasing in city size than that occurring within jobs. However, his analysis relies on 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.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE 4
Log wage growth decomposition 0–15 years of work experience

	Wage at LF entry		Within	Job to job		Job to Unemployment to job			Total
	COL Adjusted	Unadjusted	Job	Within	Between	Within	Between	Unknown	Growth
	1	2	3	4	5	6	7	8	9
Panel A: full sample									
Small	2·14	2·09	0·28	0·18	0·06	−0·03	−0·02	−0·01	0·47
Medium	2·22	2·20	0·44	0·25	0·06	−0·05	−0·02	0·00	0·67
Fraction of difference with small			0·74	0·34	−0·01	−0·06	−0·01	0·01	
Large	2·11	2·22	0·51	0·22	0·04	0·01	0·03	−0·01	0·80
Fraction of difference with small			0·66	0·11	−0·06	0·14	0·16	0·00	
Panel B: College or More									
Small	2·34	2·28	0·39	0·11	0·19	0·00	−0·02	−0·02	0·65
Medium	2·40	2·39	0·56	0·25	0·08	−0·03	0·00	−0·01	0·87
Fraction of difference with small			0·81	0·64	−0·49	−0·11	0·11	0·05	
Large	2·30	2·42	0·61	0·20	0·07	0·00	0·04	0·00	0·91
Fraction of difference with small			0·83	0·34	−0·47	0·01	0·22	0·07	
Panel C: High School Graduates Only									
Small	2·04	1·99	0·27	0·24	0·02	−0·08	0·00	−0·02	0·42
Medium	2·12	2·07	0·34	0·28	0·01	−0·07	−0·03	−0·02	0·53
Fraction of difference with small			0·67	0·38	−0·02	0·13	−0·20	0·05	
Large	1·97	2·08	0·44	0·22	−0·03	0·04	0·05	−0·01	0·70
Fraction of difference 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 2, 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 labour force entry that have been spatially and temporally deflated. "Unadjusted" in the second column refers to wages that have only been temporally deflated. Bootstrapped standard errors reveal that the only significant medium–small differences are Panel/Columns A2, A3, A9, B3, and B9. Significant large–small differences are in Panel/Columns A2, A3, A7, A9, B2, B3, B5, B9, C2, C3, C6, and C9.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

unknown category consists of wage growth that occurred between jobs sandwiching a third job for which we have no wage information. As in Table 2, each entry is expressed for the average individual as if he were to complete his entire first 15 years of work experience in a given location type.

Results of this decomposition clearly show that within-job growth is the primary dynamic driver of the city size wage and productivity premia. For the full sample, within-job growth accounted for 74% of the gap in overall wage growth between medium and small cities and 66% of the gap between large and small cities, with these differences statistically significant. Consistent with evidence in Table 3, this is the only source of wage growth that is monotonic in city size on a positive base. In addition, note that wages at labour force entry are monotonically increasing in city size when not accounting for cost of living differences across locations. Results for the college sample in Panel B and the high school sample in Panel C both exhibit similar patterns, though within-job wage growth is a more important component of overall wage growth for the college sample than for the high school sample.

Consistent with other studies, the descriptive evidence in Tables 1–4 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 the next section is thus invaluable as it efficiently uses observations about labour force transitions to separate out parameter combinations and it allows for identification of parameters indexed by latent ability.

3. 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 indexed.[10] We limit the number of these underlying worker types and city size categories to three.[11]

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 agent types as $h_a \in \{h_A, h_B, h_C\}$. These are intended to capture underlying productivity differences or comparative advantages between workers either from innate ability or because of different amounts of human capital accumulation prior to entrance into the labour market. We allow the probability that a given worker is of each type to depend on the location in which he enters the labour market.

Each period, each worker earns a wage which depends on his type, work experience in each location type, a firm-worker–specific stochastic component and classical measurement

10. 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.

11. Kennan and Walker (2011) and Coen-Pirani (2010) 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 5 would not be possible.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

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 mean 0 distribution $F^h_{\varepsilon_j}(\varepsilon)$ from which workers sample independently when they receive a job offer. 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:

$$\ln w_j(h, x_1, x_2, x_3, \varepsilon, u) = \varepsilon + \beta_0^j(h) + \sum_{k=1}^{3} \beta_{1,k}^j(h) x_k + \beta_2^j \left(\sum_{k=1}^{3} x_k \right)^2 + \beta_3 \left(\sum_{k=1}^{3} x_k \right)^3 + u. \quad (7)$$

After being incorporated into the full model, this specification allows us to estimate separate time and job invariant level 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 wage components, which provide information on the potential importance of matching. Indexing these three objects by latent ability allows the model to handle potential sorting on unobservables.[12]

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 labour 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. 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.[13]

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 $\lambda^u_j(h)$ 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 $\lambda^u_{jj'}(h)$ 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. We assume that individuals may only receive one job offer each period.[14] Workers who choose to switch jobs at the same location must pay a stochastic switching cost υ_S with zero mean and finite variance. This cost potentially captures differences in non-pecuniary benefits across jobs that might lead workers to accept wage cuts. Exogenous separation rates $\delta_j(h)$ similarly depend on location and ability. With a job offer at location j', individuals have the option to move and pay a one-time cost of $C_M(x) + \upsilon_M$ that depends on total work experience and has a random component with zero mean and finite variance. 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.

12. 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 and Williams (2005) conclude that log wage growth due to tenure effects is likely only about 0·09 over the first 10 years of experience.

13. In order to reduce the computational burden, we index time to quarters for the first 60 quarters of work after labour force entry and years for the remaining 20 years of working life.

14. In order to limit the number of arrival rate parameters that need to be estimated, we impose the parameter restrictions described in Appendix C.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

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 $V_j^{\text{WK}}(\varepsilon)$. The state spaces of all value functions that we discuss additionally include the individual type h and experience in all location sizes x_1, x_2, x_3. For expositional simplicity, we suppress this dependence in the notation. The current utility of unemployed workers depends additively on the local amenity value $\alpha_j(h)$ and unemployment benefit b_j. The current utility of workers depends similarly on this same amenity value and the log wage specified in equation (7). This implies that workers may not save or borrow.[15] Given this environment, the deterministic components of the present value of being unemployed and working are given by the following expressions, respectively:

$$V_j^{\text{UN}} = \frac{\alpha_j(h)}{3} + \ln b_j + \beta^{\frac{1}{3}} V_j^u,$$

$$V_j^{\text{WK}}(\varepsilon) = \alpha_j(h) + \ln w_j(\cdot) + \beta V_j(\varepsilon).$$

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 $\lambda_{jj'}^u(h)$ he receives an offer in location type j'.[16] The individual decides at the beginning of each period whether to accept a potential job offer or to 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:

$$\begin{aligned} V_j^u = &\left(1-\lambda_j^u(h) - \sum_{j'=1}^3 \lambda_{jj'}^u(h)\right) E_{\upsilon_U}\left(V_j^{\text{UN}} + \upsilon_U\right) \\ &+ \lambda_j^u(h)\, E_{\upsilon_U \varepsilon_j} \max\left[V_j^{\text{UN}} + \upsilon_U, V_j^{\text{WK}}(\varepsilon)\right] \\ &+ \sum_{j'=1}^3 \lambda_{jj'}^u(h) E_{\upsilon_U \upsilon_M \varepsilon_{j'}} \max\left[V_j^{\text{UN}} + \upsilon_U, V_{j'}^{\text{WK}}(\varepsilon) - [C_M(x) + \upsilon_M]\right]. \end{aligned} \tag{8}$$

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 an i.i.d. shock υ_U to the

15. Allowing for saving would greatly complicate the model to the point of intractability. Prohibiting saving is not unreasonable given that we estimate the model only using data from the first 15 years of working lives, when wages are growing and individuals are less likely to save.

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

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

option value of being unemployed. This shock has zero mean and finite variance and captures the random component of the disutility of work. The second term gives the case in which the individual receives a job offer in his city of residence. Under this scenario, he may choose to accept the job immediately or remain unemployed. The third term states that the unemployed agent will accept a potential job offer in another city of type j' if the job's option value $V_{j'}^{\mathrm{WK}}(\varepsilon)$ net of the moving cost exceeds that of remaining unemployed. The option value of having a job offer in location j' is the discounted value of holding the job next period plus the current utility implied by the wage offer. The expectation is taken with respect to the distribution of ε in location type j' and the distribution of the random components υ_U and υ_M (expressed as $E_{\upsilon_U \upsilon_M \varepsilon_{j'}}$).

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 a wage offer in the same city, and receiving an offer from elsewhere in one of the three types of locations. To simplify the computational intensity of the model and because this assumption has little impact on the results, we assume that a worker decides whether to go unemployed before knowing whether he will receive a wage offer from a different employer. As such, the value to a worker with ability h at location j of being employed with firm match ε is given by equation (9):

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

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 υ_U is large enough. If he chooses to keep working, with probability $(1-\lambda_j(h) - \sum_{j'} \lambda_{jj'}(h))$ he does not receive a wage offer. In this event, he remains employed in the same job. If he receives an offer, he either accepts it and moves if necessary or remains at his old job.

4. 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 and how the model captures various potential mechanisms behind city size wage premia.

4.1. *The likelihood function*

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

$$L(\theta) = \pi_j^A f(Y^T | h_A; \theta) + \pi_j^B f(Y^T | h_B; \theta) + (1 - \pi_j^A - \pi_j^B) f(Y^T | h_C; \theta),$$

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

where π_j^a is the probability that an individual is of type h_a given that he enters in location j and θ is the vector of parameters.[17]

Define Y_t to be the vector of labour market outcomes at time t which consist of a wage, if observed, the location of the worker and the type of labour market transition that the worker has experienced since the previous period.[18] We define $Y^t = \{Y_1, \ldots, Y_t\}$ as the vector of all labour market observations in an individual's job history up to and including period t. We decompose $f(Y^T|h;\theta)$ as follows:

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

It is convenient to express the likelihood function in terms of probabilities that one of five types of event occurs. These are being unemployed for a certain number of months and then finding a job in the same location, being unemployed for a certain number of months and then finding a job in a different location, 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 also depend on the other state variables $\{h, x_1, x_2, x_3\}$:[19]

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

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

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

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 match ε_j^A, which is the match value that would make this unemployed worker indifferent between remaining unemployed and accepting a job offer in the same location. Appendix B.2 specifies the full set of reservation rules and Appendix B.3 specifies the remaining event probabilities.

Computation of $f(Y_t|Y^{t-1})$ 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. We use Bayes' rule to update the conditional distributions of ε with new wage information each period. This implies the following updating rule:

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

This expression is used extensively to build components of the likelihood function. We also assume that the utility shock, job switching cost, moving cost, and wage measurement error distributions are independent and normal with mean zero and standard deviations to be estimated. Appendix B contains a complete detailed explanation of how we construct the likelihood function.

17. The individual index is suppressed for notational simplicity.

18. Wages are only observed on interview dates and immediately prior to job changes. We specify in Appendix B.1 how we handle this non-randomness in observing the wage in our construction of the likelihood function.

19. Due to data limitations, we assume that all unemployment spells occur in the same location as the previous job.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

4.2. *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 non-parametrically identified. Kasahara and Shimotsu (2009) discusses identification of parameters in this class of models.

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 labour 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 labour market histories of individuals regardless of their location. Parameters indexed by both type and location are identified from the relative labour 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 sets of parameters.

Transitions through unemployment are useful for identification of variances of the match distributions $F^h_{\varepsilon_j}(\varepsilon)$, and consequently all other parameters in the model as well. Unemployed workers draw their firm-specific wage components from unconditional distributions that do not depend on their match histories. Observing the different labour market outcomes of workers that have been unemployed in two different locations allows us to distinguish between the relative contributions of location and worker type and potentially allows for observation of several draws from the same unconditional distribution. A standard limitation in the structural estimation of search models is that we cannot non-parametrically identify these match distributions. This 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. This is why we assume normality for the match quality distributions. For the purpose of implementation, the distributions of ε each are on a grid of ten elements which is restricted to the range $(-2\sigma_{\varepsilon_j}, 4\sigma_{\varepsilon_j})$.

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 attempt to index as many parameters as possible by type with the few deviations coming from our discoveries that small sample sizes preclude achievement of precise estimates. For this reason, we also restrict the sets of parameters describing returns to experience and job arrival and separation rates.[20] We normalize amenities to zero 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, *e.g.* Gourinchas and Parker (2002), we set the discount factor to 0·95 per year.

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

20. We impose that the linear component of experience is restricted such that $\beta^j_{1,k}(h) = \theta_j \beta_{1,k}(h)$. This restriction produces more precise estimates of linear component parameters that is possible with the fully interacted specification. Our estimates of experience effects $\beta^j_{1,k}(h)$ are robust to alternative specifications including $\beta^j_{1,k}(h) = \theta_k \beta_{1,j}(h)$. Appendix C details the set of parameters that describe job offer arrival rates.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

matches $F^h_{\varepsilon_j}(\varepsilon)$ capture job turnover and match quality. The job separation rates $\delta_j(h)$, the unemployment benefits b_j, 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 increase with experience. The intercepts in the wage process capture fixed productivity differences across locations that may be consequences of non-labour input sharing and human capital spillovers. Each of the key parameters is indexed by latent ability. Along with the estimated probabilities that labour force entrants in each location are of each worker type, we can thus calculate counterfactual wage premia given different ability distributions across location type.

While the parameters of our model are all identified in an econometric sense, we emphasize that our model only partially identifies the relevant underlying theoretical mechanisms at play. For example, it is impossible to distinguish whether the level-effect parameters $\beta^j_0(h)$ capture differences in the unconditional means of match quality distributions or one of many different possible "sharing" mechanisms. Similarly, while the returns to experience parameters may represent rates of human capital accumulation, they may also represent the speed at which more efficient within-firm optimization of worker task assignment takes place. For these reasons, we prefer to regard the underlying mechanisms that we can identify broadly as level effects and experience effects, respectively. However, it is clear that we can separately identify these two classes of explanations for city size wage premia from each other and from unobserved worker heterogeneity that is fixed over time and mechanisms that govern transitions between jobs and through unemployment.

5. RESULTS

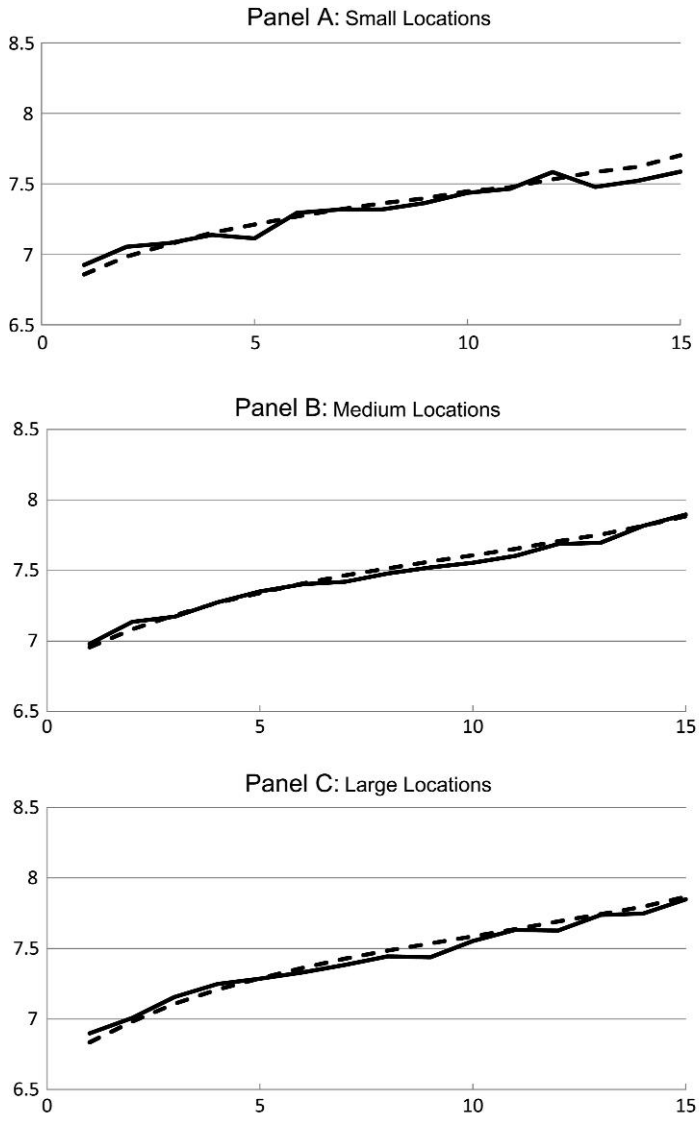
5.1. *Model fit*

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 for the college and high school samples, respectively. Both figures show that the model generally fits the data very well in this dimension. The only minor exceptions are for college graduates in medium and large cities at between 5 and 13 years of experience, where the model slightly over-predicts wages.

Examination of statistics on transitions between jobs, to and from unemployment, and between locations reveals that the simulated data closely match the actual data in these other dimensions as well. In Table 5, Panel A shows actual and predicted job, unemployment, and location transitions. The model generates simulated data that imply transition statistics that are at most 0·2 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 6 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. Among 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 gaps are for migration from small to large locations, also at minus two percentage points, and from medium to large at plus two percentage points.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017



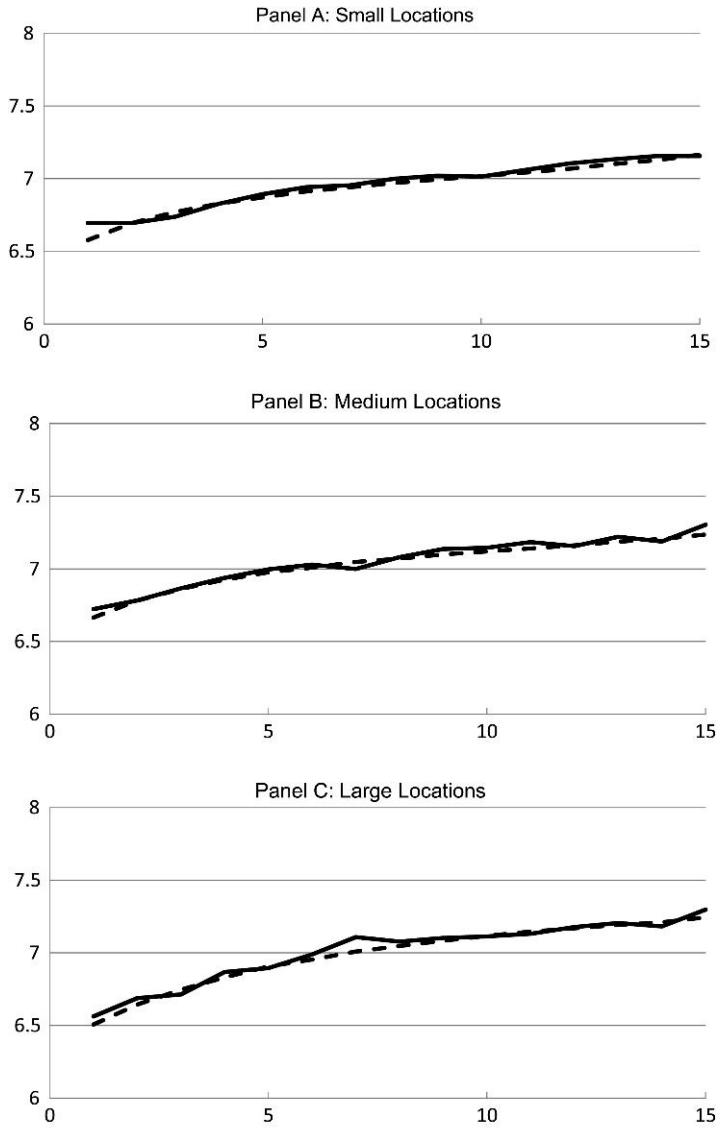
Notes: Each plot shows mean actual and predicted log wages in cents by work experience. Means are calculated separately within each year of work experience. Solid lines are actual mean wages and dashed lines are mean wages based on model simulations. Table A1 lists the parameter values used for simulations.

FIGURE 1
Actual and predicted wages by experience: college sample

5.2. *Wage components*

Rather than discuss individual parameter estimates, we find it more instructive to instead describe what various groups of parameter estimates imply for patterns of wages as functions of unobserved worker type and city size. Indeed, non-linear interactions in the model render many parameters difficult to interpret in isolation. One lesson that comes out of this examination is that

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017



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

FIGURE 2
Actual and predicted wages by experience: high school sample

workers' unobserved heterogeneity is rather complicated. While certain worker types are more productive than others in all environments, others have comparative advantages in one location over others. Therefore, it would be incomplete to equate worker type with level differences in worker productivity as is often assumed in regression models.

Figure 3 describes the set of parameters that capture the wage process absent the match component and measurement error terms ε and u. It graphs wage profiles for low, medium, 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 plots are drawn using parameters

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE 5
Actual and predicted mobility rates

	College sample		High school sample	
Mobility rates	Data	Model	Data	Model
Panel A: transition rates				
Job changes	7·1%	7·2%	11·0%	11·2%
Job to job transitions	5·2%	5·2%	6·3%	6·3%
Job to unemployment transitions	1·9%	2·0%	4·7%	4·9%
Location changes	1·8%	1·8%	1·4%	1·4%
Panel B: transition rates by location				
Job to job in small	4·1%	4·2%	5·9%	5·9%
Job to job in medium	4·5%	4·4%	6·1%	6·1%
Job to job in large	4·8%	4·9%	5·6%	5·8%
Job to unemployment in small	1·5%	1·8%	4·4%	4·5%
Job to unemployment in medium	1·7%	1·7%	4·3%	4·5%
Job to unemployment in large	1·5%	1·7%	4·4%	4·6%
Unemployment duration in small	5·6	5·5	5·9	5·8
Unemployment duration in medium	4·6	4·8	5·5	5·5
Unemployment duration in large	6·0	5·7	5·0	5·1

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.

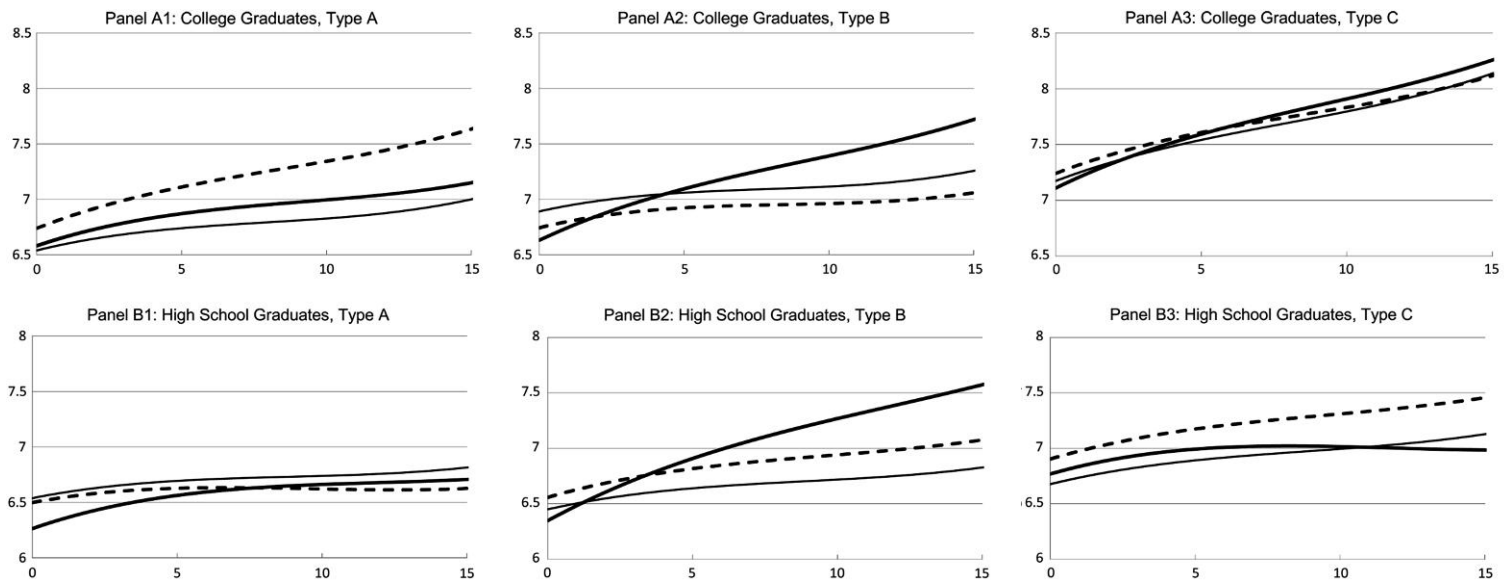
TABLE 6
Actual and predicted migration conditional on moving

Period t-1	Period t location: data			Period t location: model		
Location	Small	Medium	Large	Small	Medium	Large
Panel A: college sample						
Small	13·2%	9·9%	5·4%	12·4%	9·1%	7·1%
Medium	10·6%	16·8%	13·2%	10·4%	17·4%	12·4%
Large	5·6%	13·0%	12·4%	7·9%	11·1%	12·3%
Panel B: high school sample						
Small	26·0%	14·1%	6·9%	25·6%	12·6%	8·7%
Medium	13·3%	12·1%	9·3%	14·1%	13·4%	7·3%
Large	6·3%	6·1%	5·9%	7·7%	6·0%	4·7%

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.

in Table A1 Category A and thus represent real wages as viewed by workers and not marginal products of labour. Because prices are higher in larger cities, marginal products of labour are greater in larger cities for both samples at all ability levels. The price gaps reported in Row 4 of Panels A and B of Table 7 justify this statement.

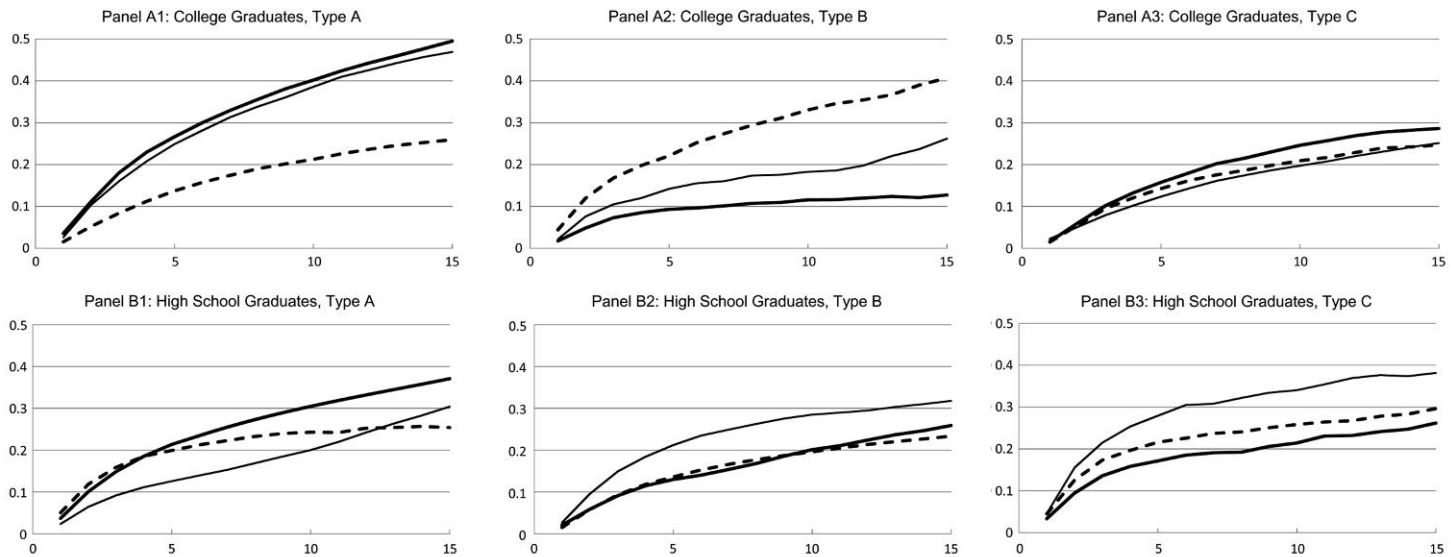
Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017



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. Thin lines are for small sized locations, dashed lines are for medium locations and thick solid lines are for large locations. The parameters used to graph these functions are found in Table A1.

FIGURE 3
Experience profiles implied by parameter estimates

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017



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.

FIGURE 4
Firm-worker match component of the wage

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

Results in Figure 3 highlight the existence of both absolute and comparative advantages across worker type. Among college graduates, Type C workers have clear absolute advantages over workers of the other two types in both their level effects and returns to experience in all locations. However, Type A has a comparative advantage over Type B in medium locations, while Type B is more productive than Type A in small and large locations at most experience levels. Among high school graduates, Type C is more productive than the other two types at most experience levels, though some of this advantage over Type B erodes over the life cycle because of Type B's steeper return to experience in the largest locations. Type A high school graduates have a slight advantage over Type B in small locations only. Overall, 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 generating differences between large and small locations while the level component seems most important for understanding medium–small gaps. Figure 3 also shows that differences between real wage level and growth effects provide few strong incentives for more able individuals to sort into larger locations.

Another focus of this paper is to understand the role of matching for generating productivity premia across cities of different sizes. To get a handle on how the search and matching parameters combine to generate potential wage gaps for each worker type, 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 worker type 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 2–4 that search, and matching are not likely important drivers of city size wage premia as it shows few monotonic relationships between the match component of wages and city size. The two panels of Figure 4 with monotonicities are more than counteracted by stronger non-monotonicities in other panels. Indeed, the only education-type combination that has monotonicities in arrival rates and match quality standard deviations is Type C college graduates, though Type A college graduates and Types B and C high school graduates have estimated job offer arrival rates for the unemployed that are monotonic in city size. We revisit the role of job offer arrival rates for generating wage gaps in the following subsection.

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 generally have higher amenity values than small locations, a force keeping workers from moving to small locations even if they receive slightly superior wage offers there. The only exceptions are Type A college graduates (who dislike medium locations the most) and Type B high school graduates (who dislike large locations the most).[21] We estimate that moving costs are increasing in work experience are larger for high school graduates and have a large stochastic component. At 15 years of experience, our estimates indicate that 24% of high school and 29% of college workers have draws from the moving cost distribution that are less than their monthly wage.

5.3. *Simulations*

Using the parameter estimates from the structural model, Table 7 evaluates the importance of potential mechanisms for generating city size wage premia in the two samples. We achieve this goal by shutting off each channel that could generate wage differences one at a time and report implied resulting average city size wage premia and the resulting reductions in worker marginal

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

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

productivities. For each experiment, we assign the average value of the relevant parameters for each latent type listed at left across locations and then regress the resulting counterfactually simulated wage on a cubic in experience and two city size dummy variables.

TABLE 7
Counterfactual city size wage premia given restricted mobility

	Regression coefficients		Gap with reference	
Experiment	Medium	Large	Medium	Large
Panel A: college sample				
1. Baseline data	0·14	0·09		
2. Model	0·14	0·09		
	[0·07, 0·21]	[0·02, 0·17]		
3. Restricted mobility (compare to Row 2)	0·15	0·13	0·01	0·04
	[0·07, 0·24]	[0·04, 0·23]		
4. Price differential	0·06	0·18		
Equalize ability distribution at LF entry	0·17	0·14	0·02	0·01
Implied reduction in marginal productivity gap			−10%	−3%
Equalize search & matching across locations	0·14	0·11	0·00	−0·03
Implied reduction in marginal productivity gap			2%	8%
Equalize returns to experience across locations	0·06**	−0·04**	−0·08	−0·17
Implied reduction in marginal productivity gap			39%	57%
Equalize nominal intercepts across locations	0·01**	0·02**	−0·13	−0·11
Implied reduction in marginal productivity gap			65%	37%
Panel B: high school sample				
1. Baseline data	0·09	0·04		
2. Model	0·09	0·04		
	[0·05, 0·19]	[−0·05, 0·29]		
3. Restricted mobility (compare to Row 2)	0·10	0·04	0·00	0·00
	[0·04, 0·19]	[−0·02, 0·29]		
4. Price differential	0·03	0·18		
Equalize ability distribution at LF entry	0·08	0·10	−0·02	0·07
Implied reduction in marginal productivity gap			16%	−29%
Equalize search & matching across locations	0·13	0·04	0·04	0·00
Implied reduction in marginal productivity gap			−31%	−1%
Equalize returns to experience across locations	0·06	−0·14**	−0·04	−0·17
Implied reduction in marginal productivity gap			29%	78%
Equalize nominal intercepts across locations	−0·02**	−0·04**	−0·12	−0·07
Implied reduction in marginal productivity gap			97%	34%

Notes: In Panels A and B, 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 labour force entry by type across locations equal to their weighted average across locations. For search and matching equalization, we set all arrival and separation rates plus match distribution standard deviations equal to their weighted averages across locations, where the weights are based on composition by type across locations at labour force entry. Equalization of returns to experience across locations is achieved analogously. Imposing all restrictions 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. ** indicate that counterfactual estimates lie outside the indicated 95% confidence interval for the reference. * indicate lying outside the 90% confidence interval instead.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

As a baseline, Rows 1 and 2 of each panel show that simulated data using estimated parameter values exhibit wage premia that are within 0·01 of those seen in the actual data. Row 3 gives counterfactual city size wage premia in the case where individuals are forced to stay in their location of labour market entry. Restricting mobility slightly increases college graduate wage premia between rural areas and cities from 0·09 to 0·13, as it forces high-ability men to stay in larger locations and keeps low-ability men out of these locations. However, this increase is not statistically significant. Because they are less mobile, restricting mobility hardly changes high school city size wage premia. Ninety-five percent confidence intervals calculated by simulating the model with draws from the joint parameter distribution in Table A1 are listed below the coefficients in Rows 2 and 3.

Results in Row 3 benchmark the counterfactual wage premia presented next to each experiment in the lower part of each panel. For evaluating sources of city size wage premia, we only consider counterfactual simulations that do not allow migration because equalization of parameters across city size categories can lead to large changes in sorting on unobserved ability that feed back into wages. Because workers never actually sort as much as in such counterfactual environments that permit migration, the associated wages are not useful for understanding sources of city size wage premia observed in the data. Changes in wage premia absent our mobility restriction combine the price effects that are our primary interest with the effects of changing worker quantities across locations. We discuss such counterfactual environments that allow for migration below in our discussion of Table 8. The entries marked with double asterisk (**) and single asterisk (*) in the lower part of each panel indicate wage premia that are statistically different from those reported in Row 3 at the 5% and 10% levels, respectively.

One difficulty in performing this exercise is that the model only has predictions about city size category of residence, not the particular city of residence within a given category. However, calculation of counterfactual marginal productivity (or nominal wage) gaps requires an adjustment for price differences across locations. To perform such calculations, we use the equilibrium average price differences observed in the data for each sample, which are reported in Row 4. This procedure is not perfect because the price differences 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.[22] 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 and so persistent price gaps across locations difference out.

Both panels of Table 7 show that differences in returns to experience and wage-level effects account for virtually the entire city size nominal wage gaps for both samples. Absent differences in returns to experience across locations and maintaining endogenous sorting, college

22. Price differences across locations are different for the two samples because they are calculated using locations in the raw data. Conditional on living in medium-sized cities, high school graduates lived in cheaper places than college graduates on average.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE 8
Counterfactual city size wage premia and migration outcomes given free mobility

Experiment		Implied % Reduction	Net migration out of location Type A	Type B	Type C
Panel A: college sample					
Model	Medium		0·001%	−0·065%	−0·101%
	Large		−0·022%	0·024%	0·042%
Equalize ability distribution	Medium	−7·2%	−0·004%	−0·103%	−0·082%
at LF entry	Large	−4·9%	−0·015%	0·017%	0·037%
Equalize search & matching	Medium	−3·9%	−0·270%	0·559%	0·014%
across locations	Large	6·8%	0·177%	−0·721%	−0·121%
Equalize returns to experience	Medium	−2·6%	0·378%	−0·540%	−0·334%
across locations	Large	29·3%**	−0·260%	0·542%	0·216%
Equalize nominal intercepts	Medium	63·1%**	0·202%	−0·139%	−0·004%
across locations	Large	36·3%**	0·098%	0·027%	0·076%
Panel B: high school sample					
Model	Medium		−0·013%	−0·031%	−0·086%
	Large		0·011%	0·041%	0·042%
Equalize ability distribution	Medium	10·3%	0·145%	−0·137%	−0·023%
at LF entry	Large	−16·2%	−0·106%	0·155%	−0·007%
Equalize search & matching	Medium	−0·3%	−0·413%	−0·011%	−0·043%
across locations	Large	1·6%	0·215%	0·070%	0·122%
Equalize returns to experience	Medium	28·4%*	−0·164%	−0·097%	0·080%
across locations	Large	65·6%**	−0·011%	0·187%	−0·255%
Equalize nominal intercepts	Medium	111·1%**	0·004%	0·045%	0·023%
across locations	Large	39·4%**	−0·059%	0·036%	0·179%

Notes: Results are based on the same type of simulations as those performed for Table 7 except that we allow agents to freely migrate across locations subject to estimated migration costs. Column 1 shows the implied percent reduction in medium–small and large–small location marginal productivity gaps using the same procedure as is used for Table 7. The remaining columns indicate the fraction of workers of each type who migrate out minus the fraction who migrate in to and from locations of each listed type between 0 and 15 years of experience using simulated data in the listed counterfactual. ** In the the first column of results indicate that counterfactual estimates lie outside the indicated 95% confidence interval for the reference. * Indicate lying outside the 90% confidence interval instead.

graduates' nominal wage premia would be reduced by 39% for medium locations and 57% for large locations while high school graduates' premia would be reduced by 29% and 78%, respectively. Absent differences in wage-level effects, college graduates' nominal wage premia are estimated to decline by 65% for medium and 37% for large cities while high school graduates' premia are reduced by 97% and 34%, respectively.

Search frictions plus match quality and ability sorting at entry into the labour force have small and statistically insignificant impacts on city size wage premia and marginal productivity premia in both samples. For college graduates neither mechanism accounts for more than 10% of either medium–small or large–small city size wage gaps, with (statistically insignificant) point estimates showing slight negative sorting into both medium and large cities at labour force entry. Point estimates show somewhat stronger negative sorting of high school graduates into large cities. Equalization of search and matching parameters across locations actually increases the counterfactual medium–small city size wage gap by 31% for the high school sample but has almost no effect on the large–small city size wage premium. Both the sorting on unobservables and search and matching mechanisms thus somewhat counteract the contributions of the sharing

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

and learning mechanisms to observed city size wage premia.[23] Analogous simulations conducted on simulated data calculated with alternative parameters estimated using data deflated with other cost of living indexes are reported in Table A2 and yield similar conclusions.[24]

Table 8 explores counterfactual reductions in marginal productivity gaps and net migration by type associated with each of the experiments explored in Table 7 without the migration restriction. Note that our estimated parameters indicate that high-ability college workers migrate out of large locations into medium locations over the life cycle—more direct evidence of mild negative sorting than that described in Table 7. With migration permitted, counterfactual reductions in marginal productivity gaps show similar patterns as in Table 7 with the notable exception that returns to experience appear less important for college graduates. Inspection of migration results reveals that this is because Types B and C depart large locations and enter medium locations much more rapidly than they otherwise would have, induced by the fact that premia in returns to experience are no longer reasons to stay in the large locations. This is consistent with evidence in Panels A2 and A3 of Figure 3 showing relatively high returns to experience for these two worker types in the largest locations. However, these migratory flows are so much stronger than those seen in equilibrium that it is possible that general equilibrium effects would affect the results in Table 8.

Using an alternative method for generating counterfactual wages, we confirm evidence from Tables 7 and 8 about the relative importance of level effects and returns to experience. With the simulated job and migration histories derived using the estimated parameters, we recalculate wages after setting nominal level effects or returns to experience equal across locations. This exercise differs from those in Tables 7 and 8 only in that we maintain the equilibrium distribution of quantities rather than allowing agents' behaviour to endogenously respond to the price changes we impose. For the college sample, equalizing nominal level effects in this environment results in reductions in nominal medium–small and large–small wage premia by 82% and 41%, respectively while for the high school sample, the associated reductions are 72% and 32%. Equalizing returns to experience in this environment results in reductions in nominal medium–small and large–small wage premia by 50% and 52%, respectively in the college sample and 13% and 80%, respectively in the high school sample.

5.4. *The role of ability*

It is natural to wonder whether the unobserved ability that we estimate 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. Understanding the role of unobserved ability also allows us to pinpoint potential pitfalls of more traditional regression–based methods that are often used to try to control for unobserved ability. We investigate these questions by evaluating the relative predictive powers of cohort-specific percentiles of the Armed Forces Qualifying Test (AFQT) score available in the NLSY79 data set and our estimated ability measure for wages.

23. 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 also have small independent impacts on city size wage premia.

24. The first alternative cost of living deflator, as in Albouy (2010), assumes that consumers only care about housing and other goods and uses census micro data from 2000 rather than American Chamber of Commerce Research Association (ACCRA) data to calculate one cross section of the house price index. See Appendix A.3 for details. The second alternative deflator is the simple consumer price index (CPI-U) applied equally to all locations. In addition, we tried estimating the model using a price index that incorporates the result in Handbury and Weinstein (2010) that grocery prices are lower in larger cities. Use of this price index also has little effect on the simulation results (unreported).

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

Using data on labour market history Y_i^T up to each individual i's final observation T, the probability that individual i is of ability h_a is given by the following expression, which employs Bayes' rule.

$$\Pr(h = h_a|Y_i^T) = \frac{\Pr(Y_i^T|h = h_a) \times \Pr(h = h_a)}{\Pr(Y_i^T)}.$$

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

Not surprisingly, Table 9 presents evidence that our estimated $\Pr(h = h_B|Y_i^T)$ and $\Pr(h = h_C|Y_i^T)$ variables are better predictors of log wages than is the AFQT score. Each column in Table 9 presents results from regressing log wages on a cubic in experience and various

TABLE 9
Log wages and AFQT vs. unobserved ability

	Log wage						
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: college sample							
Medium		0·14***	0·12***	0·14***	0·13***	0·05*	0·06
		(0·04)	(0·04)	(0·03)	(0·03)	(0·03)	
Large		0·09*	0·08	0·12***	0·12***	0·01	−0·05
		(0·05)	(0·05)	(0·04)	(0·04)	(0·04)	
AFQT	0·16***		0·39***		0·13**		
	(0·06)		(0·07)		(0·06)		
Pr(h = h_B)	−0·02			−0·03	−0·03		
	(0·04)			(0·04)	(0·04)		
Pr(h = h_C)	0·67***			0·68***	0·67***		
	(0·03)			(0·03)	(0·03)		
Individual FE	No	No	No	No	No	Yes	Yes
Simulated data	No	No	No	No	No	No	Yes
R-squared	0·40	0·18	0·19	0·41	0·41	0·61	0·58
Panel B: high school sample							
Medium		0·09***	0·08***	0·08***	0·08***	0·05*	0·03
		(0·03)	(0·03)	(0·02)	(0·02)	(0·03)	
Large		0·04	0·05	0·06**	0·06**	0·05	0·00
		(0·03)	(0·03)	(0·03)	(0·03)	(0·04)	
AFQT	0·18***		0·25***		0·17***		
	(0·04)		(0·05)		(0·04)		
Pr(h = h_B)	0·17***			0·17***	0·16***		
	(0·03)			(0·03)	(0·03)		
Pr(h = h_C)	0·47***			0·48***	0·47***		
	(0·03)			(0·03)	(0·03)		
Individual FE	No	No	No	No	No	Yes	Yes
Simulated data	No	No	No	No	No	No	Yes
R-Squared	0·26	0·14	0·16	0·26	0·27	0·52	0·41

Notes: Each entry is a coefficient or standard error from a regression of log wage on the variables listed at left and a cubic in experience. Each regression uses all available wage observations from those for whom an AFQT score is observed. Regressions have 7,360 observations in Panel A and 10,075 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 and in a way that allows for the distribution to differ by individual age of taking the test. Pr(h = h_a) is the estimated probability that each individual in the data set is type a calculated as explained in the text. *** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

combinations of the two ability measures and two city size dummy variables. Specification (1) shows that our estimated Type C probabilities alone have much higher t-statistics than does AFQT (measured in standard deviations) when predicting log wage in both samples. Indeed, the pairwise correlations between probability Type C and log wage at labour force entry is more than five times greater than that between AFQT and log wage in both samples. In addition, the estimated Type A probability is negatively correlated with AFQT, mother's education, and parental income, while our estimated Type C ability probability is positively correlated with these three observable proxies for ability in both samples. 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 after type probabilities are already controlled for. However, coefficients on AFQT drop considerably when the type probabilities are included as additional controls. Furthermore, the R-squared markedly increases when estimated ability is included in these regressions. Three of four city size coefficients are slightly greater when ability controls are included, recapitulating evidence from the counterfactual simulations of mild negative ability sorting.

5.5. *Evaluating fixed-effects estimates*

Our model allows for heterogeneous treatment effects of city size on worker marginal products by incorporating various dimensions of unobserved heterogeneity and endogenous migration in a dynamic environment. The fixed-effect regression estimator would consistently measure an average effect of city size on worker wages for any arbitrarily high amount of heterogeneity in wage intercepts and exogenous migration conditional on these intercepts. Even with such exogenous migration, however, this average would be difficult to interpret as it would potentially incorporate underlying treatment effects that are heterogeneous across both individuals and work experience with varying weights. Indeed, De la Roca and Puga (2011) argue that it is appropriate to control for full work location histories to account for such dynamic treatment effect heterogeneity. Because our model necessarily allows for less *ex ante* unobserved heterogeneity in intercepts than the fixed-effect estimator, fixed-effects estimates would also be informative in an environment with considerable heterogeneity in wage intercepts but exogenous migration. We evaluate the validity of fixed-effects estimates by comparing those using simulated data to those using actual data. If simulated wage gaps are similar to actual wage gaps after accounting for fixed effects, the model must be capturing at least as many of the relevant components of the data-generating process as the fixed-effects estimator. Comparison with results of the counterfactual exercises in Tables 7 and 8 that explicitly eliminate sorting on unobservables then reveals whether the model includes additional important aspects of this process that are not accounted for by the fixed-effects estimator.

Columns 6 and 7 of Table 9 report fixed-effects estimates of city size wage premia using actual and simulated data, respectively. Estimates from the simulated data exhibit more accentuated inverse U relationships between wages and city size within individual for both samples than those from the actual data, although the two sets of coefficients are not statistically different from each other. Interpreted in the context of a model that only has heterogeneity in intercepts, these estimates would imply even more positive sorting into large locations than those using the actual data. However, as seen in Tables 7 and 8 (and Columns 2–5 of Table 9), there in fact exists mild negative sorting into large locations, a feature that the fixed-effects estimates produced from the model itself completely misses. This bias may reflect the presence of endogenous migration. Alternatively, because mobility is more prevalent early in the life cycle, fixed-effect estimates weight these periods more heavily. Given the evidence in Figure 3 that treatment effects of city size grow over the life cycle for most workers, fixed-effect estimates are thus likely to understate

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

average treatment effects of city size over the first 15 years of experience, thereby making sorting on unobserved ability look more prevalent than it really is even if migration is exogenous.

6. CONCLUSIONS

In this paper, we lay out a systematic framework to empirically examine reasons for which workers in 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 returns to experience and wage-level effects 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 generating wage gaps between medium and small cities while differences in returns to experience are more important for generating large–small city size wage gaps. However, sorting on unobserved ability within education group and differences in labour 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 provides new information about the relevance of certain classes of micro-founded theories over others for generating the city size wage premium. 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.

APPENDIX

A. Data

A.1. *Sample*

The sample includes 1758 men from the NLSY79 random sample of 3003 men. We lose 20% of the full sample because they entered the labour force before we observe their initial attachment. An additional 12% of individuals are dropped because they were in the military at some point, never entered the labour market, dropped out of the labour market for at least four contiguous years or had significant missing job history data. The remaining individuals excluded from the sample are non-whites. We only use the first 15 years of work time experienced by each individual in the sample.

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. Until 1993, we observe wages on up to five jobs per year. After 1993, we only observe wages of the jobs most recently worked prior to interview dates, which occur about every 2 years. Individuals enter the sample when they begin working full-time or at least 300 hours per quarter if not in school and at least 500 hours per quarter if in school. The resulting high school sample has 37,393 observations on 675 individuals and the college sample has 32,009 observations on 586 individuals.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

A.2. *Location assignment*

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. Because we rarely observe whether workers migrate after going unemployed, 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. This leaves 5% of quarterly observations with no location information. Sixty percent of these observations are for jobs sandwiched between two other jobs at the same location. In these cases, we assume that individuals did not move. For the remaining 2% 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.

A.3. *Cost of living*

Section 2.1 details theory behind construction of the cost of living index. As is discussed there, our cost of living index requires price data by time and location for different goods and information on expenditure shares. Our general approach is to build one accurate cross section of location and commodity-specific prices in various ways and then temporally deflate using regional consumer price index series. The cross sectional cost of living index that forms a basis for the price index used in our primary analysis utilizes the ACCRA data sets from 2000 to 2002, which were also used in Basker (2005) .These data report prices in six broad expenditure categories for most MSAs and some rural counties nationwide. When possible, we take data from 2001. For the few regions not sampled in 2001, we take data from either 2000 or 2002. ACCRA reports provide us with price data for 244 metropolitan areas and 179 rural counties. 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. 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.

We also estimate the model using price indexes constructed using an alternative version of this cross sectional price index as a robustness check. For this index, we use the 2000 census micro data to construct the housing component instead of taking it from the ACCRA data. Following Albouy (2010), we first regress log home price or log rent on MSA fixed effects and indicators for ranges of the number of units in structure fully interacted with the number of rooms in the unit, the year of construction and the number of bedrooms as separate sets of terms in separate regressions, weighting each unit by the census provided household weight. We then reweight by implied value and run the same regression pooling across rental and owner-occupied units with the same controls fully interacted with tenure but with common MSA fixed effects. These fixed effects form our alternative housing component of the price index. To get an index for the non-housing component of prices, we use the ACCRA data cross section described above to predict the fraction of the cross-sectional variation in non-housing price that varies with the housing price, just as is done in Albouy (2010). We also report results for which we perform no cross-sectional cost of living adjustment.

For time series variation in prices, we use regional and metropolitan price index data from the Bureau of Labor Statistics disaggregated into the same categories for which we build cross-sectional price differences. We assign each county to be represented by the most geographically specific index possible in each year. The Bureau of Labor Statistics (BLS) regional price indexes either apply to specific MSAs or to MSA size categories 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_g 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 NLSY79 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.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

City size wage gap decomposition results using our two alternative price indexes are similar to those using our primary cost of living deflator reported in Table 7. They are reported in Table A2.

B. 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 $\{x_1, x_2, x_3\}$.

B.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, and f_{u_j} which is the probability density function of the location-specific measurement error.

$$B_{jt}(\varepsilon) = \left[[p_n f_{u_j}(u_t)]^{1(w_t \text{ obs})}[1-p_n]^{1(w_t \text{ not obs})}\right]^{1(\text{int}_t \neq \text{int}_{t+1})}$$

$$A_{jt-1}(\varepsilon) = [[p_n f_{u_j}(u_{t-1})]^{1(w_{t-1} \text{ obs})}[1-p_n]^{1(w_{t-1} \text{ not obs})}]^{1(\text{int}_{t-1}=\text{int}_t \ \& \ \text{job}_{t-1}\neq\text{job}_t)}.$$

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

$$B_{jt}(\varepsilon) = f_{u_j}(u_t)^{1(w_t \text{ obs } \& \ \text{int}_t \neq \text{int}_{t+1})}$$

$$A_{jt-1}(\varepsilon) = f_{u_j}(u_{t-1})^{1(w_{t-1} \text{ obs } \& \ \text{int}_{t-1}=\text{int}_t \ \& \ \text{job}_{t-1}\neq\text{job}_t)}$$

instead.

B.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.

$$\varepsilon_j^A(v_U) \text{ solves : } V_j^{\text{UN}} + v_U = V_j^{\text{WK}}(\varepsilon),$$

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

$$\varepsilon_{jj'}^C(v_U, v_M) \text{ solves : } V_j^{\text{UN}} + v_U = V_{j'}^{\text{WK}}(\varepsilon) - C_M - v_M,$$

$$\varepsilon_j^D(\varepsilon, v_S) \text{ solves : } V_j^{\text{WK}}(\varepsilon) = V_j^{\text{WK}}(\varepsilon') - v_S,$$

$$\varepsilon_{jj'}^E(\varepsilon, v_S, v_M) \text{ solves : } V_j^{\text{WK}}(\varepsilon) = V_{j'}^{\text{WK}}(\varepsilon') - C_M - v_M - v_S.$$

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

B.3. *Transition probabilities*

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

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

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

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

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

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

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

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

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

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

B.4. *First period*

A worker is included in the sample from his first full-time employment spell. As a result, the contribution to the likelihood function of the first observation must condition on the fact that the worker has accepted a job. The contribution to the likelihood of an individual entering in location j is therefore:

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

The resulting posterior distribution of the firm match is

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

B.5. *Unemployment*

Consider an individual that 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

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

After NW_t weeks, the worker accepts a job in either location j or in location j'. If he accepts a job in location j, the total contribution to the likelihood function is given by

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

The posterior distribution of the match is

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

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

If he instead accepts a job in location j', the contribution of the unemployment spell to the likelihood function is

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

The posterior distribution of the match is then

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

B.6. *Becoming unemployed*

A worker in location j goes unemployed with probability $P_{eu}^{j}(\varepsilon)$. If this happens we may observe some new information about the wage associated with the last job. This new information is characterized by the density $A_{jt-1}(\varepsilon)$. From the previous period, we have calculated the conditional distribution of the match $f(\varepsilon|Y^{t-1})$. This density takes into account all the information revealed up to time $t-1$. Given this, we can express the contribution of becoming employed to the likelihood as follows:

$$L_{3}^{j} = \int A_{jt-1}(\varepsilon) P_{eu}^{j}(\varepsilon) dF(\varepsilon|Y^{t-1}).$$

Because the firm-specific component ε will not affect the future observations, we do not need to update its conditional distribution.

B.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 follows:

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

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

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

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

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

Note again the inclusion of the function $A_{jt-1}(\varepsilon)$. This captures the fact that in $t-1$ we may have observed a wage as a result of the job change. Because we take into account of this job mobility decision in the contribution of time t, this wage information must also be included in period t and not in period $t-1$.

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

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

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

C. Normalizations of job offer arrival rates

In the model, there are 72 free parameters that measure arrival rates of job offers of which 54 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 the 18 parameters that describe the probability of receiving a wage offer from

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE A1
Parameter estimates from the structural model

Parameter	Description	College sample			High school sample		
		Type A	Type B	Type C	Type A	Type B	Type C
A. Components of wages							
β_0^1	Wage constant for living in small locations	6·54***	6·89***	7·17***	6·54***	6·45***	6·68***
		(0·03)	(0·04)	(0·03)	(0·01)	(0·02)	(0·02)
β_0^2	Wage constant for living in medium locations	6·74***	6·74***	7·24***	6·50***	6·56***	6·90***
		(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)
β_0^3	Wage constant for living in large locations	6·58***	6·63***	7·11***	6·26***	6·35***	6·77***
		(0·02)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)
β_1^1	Return to experience from work in small	0·065***	0·058***	0·098***	0·052***	0·059***	0·064***
		(0·003)	(0·004)	(0·003)	(0·001)	(0·001)	(0·001)
β_1^2	Return to experience from work in medium	0·084***	0·052***	0·083***	0·049***	0·076***	0·078***
		(0·001)	(0·002)	(0·002)	(0·002)	(0·001)	(0·002)
β_1^3	Return to experience from work in large	0·068***	0·095***	0·098***	0·053***	0·084***	0·044***
		(0·001)	(0·002)	(0·002)	(0·001)	(0·001)	(0·001)
θ_2	Multiplier for living in medium locations		1·21***			0·99***	
			(0·11)			(0·09)	
θ_3	Multiplier for living in large locations		1·29***			1·68***	
			(0·08)			(0·11)	
β_2^1	Coefficient on experience squared in small locations		-0·006***			-0·005***	
			(0·0002)			(0·0002)	
β_2^2	Coefficient on experience squared in medium locations		-0·007***			-0·006***	
			(0·0002)			(0·0001)	
β_2^3	Coefficient on experience squared in large locations		-0·007***			-0·007***	
			(0·0001)			(0·0002)	
β_3	Coefficient on experience cubed		0·0003***			0·0002***	
			(0·00002)			(0·00002)	
B· Arrival and separation rates							
λ_1	Job offer arrival rate within small locations	0·10***	0·17***	0·03***	0·24***	0·08***	0·20***
		(0·007)	(0·016)	(0·004)	(0·010)	(0·005)	(0·011)
λ_2	Job offer arrival rate within medium locations	0·07***	0·19***	0·08***	0·23***	0·11***	0·25***
		(0·005)	(0·013)	(0·007)	(0·013)	(0·005)	(0·011)

(*continued*)

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE A1
Continued

Parameter	Description	College sample			High school sample		
		Type A	Type B	Type C	Type A	TypeB	Type C
λ_3	Job offer arrival rate within large locations	0·10***	0·18***	0·09***	0·14***	0·15***	0·24***
		(0·006)	(0·009)	(0·007)	(0·007)	(0·011)	(0·014)
λ_1^u	Job offer arrival rate within small from (unemployed)	0·11***	0·16***	0·05***	0·21***	0·10***	0·15***
		(0·008)	(0·019)	(0·010)	(0·008)	(0·005)	(0·009)
λ_2^u	Job offer arrival rate within medium from (unemployed)	0·11***	0·20***	0·11***	0·15***	0·19***	0·21***
		(0·007)	(0·012)	(0·012)	(0·008)	(0·008)	(0·001)
λ_3^u	Job offer arrival rate within large from (unemployed)	0·17***	0·11***	0·11***	0·16***	0·25***	0·21***
		(0·019)	(0·012)	(0·014)	(0·008)	(0·018)	(0·009)
ρ_1	Receive rate from small: $\lambda_{21} = \rho_1 \lambda_2$		0·29***			0·17***	
			(0·04)			(0·02)	
ρ_2	Receive rate from medium		0·15***			0·06***	
			(0·02)			(0·01)	
ρ_3	Receive rate from large		0·09***			0·03***	
			(0·01)			(0·01)	
ρ_1^u	Receive rate from small (unemployed): $\lambda_{21}^u = \rho_1^u \quad \lambda_2^\rho$		0·25***			0·17***	
			(0·06)			(0·03)	
ρ_2^u	Receive rate from medium (unemployed)		0·25***			0·09***	
			(0·05)			(0·01)	
ρ_3^u	Receive rate from large (unemployed)		0·18***			0·05***	
			(0·03)			(0·01)	
γ	Own location type multiplier: $\lambda_{11} = \lambda_1 \rho_1 \gamma$ and $\lambda_{11}^u = \lambda_1^u \rho_1^u \gamma$		1·22***			1·18***	
			(0·11)			(0·11)	
ψ_M	Receive rate (type multiplier): $\lambda_{21}^M = \psi_M^u \lambda_{21}$	1	0·91***	2·39***	0·47***	1	0·69***
			(0·07)	(0·19)	(0·04)		(0·07)
ψ_M^u	Receive rate (type multiplier—unemployed): $\lambda_{21}^M = \psi_M^u \lambda_{21}$	1	0·63***	1·29***	0·97***	1	0·96***
			(0·05)	(0·12)	(0·02)		(0·04)
δ_1	Separation rate in small	0·003	0·013***	0·003*	0·001	0·015***	0·035*
		(0·002)	(0·005)	(0·002)	(0·002)	(0·002)	(0·003)
δ_2	Separation rate in medium	0·004***	0·010***	0·007***	0·051***	0·008***	0·030***
		(0·001)	(0·004)	(0·002)	(0·004)	(0·002)	(0·003)

(*continued*)

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE A1
Continued

Parameter	Description	College sample			High school sample		
		Type A	Type B	Type C	Type A	Type B	Type C
δ_3	Separation rate in large	0·001	0·036***	0·011***	0·003	0·000	0·021***
		(0·001)	(0·003)	(0·002)	(0·002)	(0·003)	(0·003)
C. Amenities							
α_2	Amenity of medium locations	-0·22***	0·34***	0·33***	0·96***	0·20***	0·04
		(0·05)	(0·06)	(0·05)	(0·07)	(0·06)	(0·08)
α_3	Amenity of large locations	0·20***	0·85***	0·45***	0·85***	-0·56***	0·65***
		(0·04)	(0·05)	(0·04)	(0·08)	(0·09)	(0·09)
D· Benefits and costs							
$\ln(b_1)$	Unemployment benefit in small		-0·48***			0·25***	
			(0·09)			(0·04)	
$\ln(b_2)$	Unemployment benefit in medium		-1·03***			-0·01	
			(0·09)			(0·03)	
$\ln(b_3)$	Unemployment benefit in large		-1·52***			0·00	
			(0·10)			(0·03)	
C_M	Moving cost - constant		3·34			-1·14	
			(4·37)			(3·64)	
C_M	Moving cost - linear term		1·18**			1·38**	
			(0·49)			(0·54)	
C_M	Moving cost - quadratic term		0·01			0·07***	
			(0·01)			(0·02)	
E· Heterogeneity							
π^1	Probability of type given start in small location		0·19***	0·32***	0·31***		0·19***
			(0·04)	(0·04)	(0·04)		(0·03)
π^2	Probability of type given start in medium location		0·35***	0·21***	0·22***		0·49***
			(0·05)	(0·04)	(0·03)		(0·04)
π^3	Probability of type given start in large location		0·23***	0·34***	0·18***		0·27***
			(0·03)	(0·04)	(0·03)		(0·03)

(*continued*)

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE A1
Continued

Parameter	Description	College sample			High school sample		
		Type A	Type B	Type C	Type A	Type B	Type C
F· Distributions							
σ_1	Standard deviation of match in small	0·39***	0·37***	0·27***	0·27***	0·38***	0·45***
		(0·01)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)
σ_2	Standard deviation of match in medium	0·28***	0·45***	0·30***	0·40***	0·23***	0·37***
		(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)
σ_3	Standard deviation of match in large	0·41***	0·32***	0·35***	0·29***	0·22***	0·32***
		(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)
σ_U	Standard deviation of utility shock for unemployed		9·15***			5·58***	
			(1·25)			(0·42)	
σ_S	Standard deviation of job switching cost shock		7·48***			3·68***	
			(0·61)			(0·20)	
σ_M	Standard deviation of moving cost shock		15·24***			32·75***	
			(3·79)			(5·52)	
σ_{u1}	Standard deviation of wage measurement error- small		0·30***			0·27***	
			(0·002)			(0·001)	
σ_{u2}	Standard deviation of wage measurement error- medium		0·29***			0·21***	
			(0·001)			(0·001)	
σ_{u3}	Standard deviation of wage measurement error-large		0·28***			0·28***	
			(0·001)			(0·001)	

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.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

TABLE A2
Counterfactual reductions in city size marginal productivity gaps given restricted mobility using two alternative price indexes

Experiment	College sample		High school sample	
	Medium	Large	Medium	Large
Panel A: housing/non-housing index built using census micro data				
Real wage gaps	0·08	0·07	0·01	0·02
Price differential	0·12	0·20	0·10	0·20
Restrict mobility	−6%	−13%	0%	0%
Equalize ability distribution at LF entry	0%	−8%	3%	−33%
Equalize search & matching across locations	4%	11%	−28%	3%
Equalize returns to experience across locations	35%	55%	29%	73%
Equalize nominal intercepts across locations	63%	39%	110%	50%
Panel B: no spatial deflation				
Wage gaps	0·20	0·27	0·12	0·22
Restrict mobility	−3%	−13%	−4%	−2%
Equalize ability distribution at LF entry	6%	4%	41%	−22%
Equalize search & matching across locations	4%	6%	−41%	−5%
Equalize returns to experience across locations	45%	57%	45%	84%

Notes: Entries are analogous to selected entries in Table 7 except that the counterfactual simulations have been run on a model estimated using data with wages deflated by two alternative price indexes. For Panel A, we deflate wages using a price index that has only housing and non-housing components. The housing component is constructed with census micro data only and the non-housing component is constructed as a linear combination of the housing component using one ACCRA data cross section to determine the correlation between the two. See Appendix A for further details. For Panel B, we use data that has no spatial deflation and has only been deflated temporally by the CPI-U.

the same location and we assume that the remaining probabilities are scaled by the 11 estimated parameters $\rho_{j'}, \rho^u_{j'}$, γ, $\psi(h)$ and $\psi^u(h)$. 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 within location size. The multipliers $\psi(h)$ and $\psi^u(h)$ allow the received rates to parsimoniously differ by worker type and are normalized to one for the type A. Arrival rates of job offers across locations are thus specified as follows:

$$\lambda^u_{jj'}(h) = \lambda^u_j(h)\rho^u_{j'}\psi^u(h) \text{ if } j \neq j',$$

$$\lambda^u_{jj'}(h) = \lambda^u_j(h)\rho^u_{j'}\psi^u(h)\gamma \text{ if } j = j',$$

$$\lambda_{jj'}(h) = \lambda_j(h)\rho_{j'}\psi(h) \text{ if } j \neq j',$$

$$\lambda_{jj'}(h) = \lambda_j(h)\rho_{j'}\psi(h)\gamma \text{ if } j = j'.$$

These normalizations reduce the number of arrival rate parameters to be estimated from 72 to 29.

Acknowledgment. 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. The paper has benefited greatly from three anonomous referees' comments and those 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 at Madison, Yale and at the SED, the North American Econometric Society, NARSC, HLUM, RCEF, CRES Applied Microeconomics, and IEB Urban meetings. We thank Emek Basker for generously providing us with ACCRA price index data.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

ALBOUY, D. (2008), "Are Big Cities Really Bad Places to Live? Improving Quality-of-Life Estimates Across Cities" (NBER Working Paper No. 14472).
ALBOUY, D. (2010), "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values" (NBER Working Paper No. 14981).
ALTONJI, J. and WILLIAMS, N. (2005), "Do Wages Rise With Job Seniority? A Reassessment", *Industrial and Labor Relations Review*, **58**, 370–397.
ARZAGHI, M. and HENDERSON, J. V. (2008), "Networking off Madison Avenue", *Review of Economic Studies*, **75**, 1011–1038.
BASKER, E. (2005), "Selling a Cheaper Mousetrap: Wal-Mart's Effects on Retail Prices", *Journal of Urban Economics*, **58**, 203–229.
BORJAS, G. (1987), "Self-Selection and the Earnings of Immigrants", *American Economic Review*, **77**, 531–553.
CICCONE, A. and HALL, R. (1996), "Productivity and the Density of Economic Activity", *American Economic Review*, **86**, 54–70.
COEN-PIRANI, D. (2010), "Understanding Gross Worker Flows Across U.S. States", *Journal of Monetary Economics*, **57**, 769–784.
COMBES, P., DURANTON, G. and GOBILLON, L. (2008), "Spatial Wage Disparities: Sorting Matters!", *Journal of Urban Economics*, **63**, 723–742.
COMBES, P.-P., DURANTON, G., GOBILLON, L. and ROUX, S. (2010), "Estimating Agglomeration Economies with History, Geology and Worker Effects", in Glaeser, E. L. (ed.) *Agglomeration Economics* (The University of Chicago Press) 15–65.
DAVIS, M. and ORTALO-MAGNÉ, F. (2011), "Household Expenditures, Wages, Rents", *Review of Economic Dynamics*, **14**, 248–261.
DE LA ROCA, J. and PUGA, D. (2011), "Learning by Working in Dense Cities", manuscript.
DURANTON, G. and PUGA, D. (2004), "Micro-Foundations of Urban Agglomeration Economies", in Henderson, J. V. and Thisse, J. F. (eds) *Handbook of Urban and Regional Economics, Vol. 4* (North Holland: Elsevier) 2063–2117.
GLAESER, E. and MARÉ, D. (2001), "Cities and Skills", *Journal of Labor Economics*, **19**, 316–342.
GOULD, E. (2007), "Cities, Workers and Wages: A Structural Analysis of the Urban Wage Premium", *Review of Economic Studies*, **74**, 477–506.
GOURINCHAS, P. and PARKER, J. A. (2002), "Consumption Over the Life-Cycle", *Econometrica*, **70**, 47–89.
GREENSTONE, M., HORNBECK, R. and MORETTI, E. (2010), "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings", *Journal of Political Economy*, **118**, 536–598.
HANDBURY, J. H. and WEINSTEIN, D. E. (2010), "Is New Economic Geography Right? Evidence from Price Data", manuscript.
HECKMAN, J. and SINGER, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, **52**, 271–320.
KASAHARA, H. and SHIMOTSU, K. (2009), "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices", *Econometrica*, **77**, 135–175.
KEANE, M. and WOLPIN, K. (1997), "The Career Decisions of Young Men", *Journal of Political Economy*, **105**, 473–522.
KENNAN, J. and WALKER, J. R. (2011), "The Effect of Expected Income on Individual Migration Decisions", *Econometrica*, **79**, 211–251.
PAVAN, R. (2011), "Career Choice and Wage Growth", *Journal of Labor Economics*, **29**, 549–587.
PETRONGOLO, B. and PISSARIDES, C. (2006), "Scale Effects in Markets With Search", *Economic Journal*, **116**, 21–44.
ROBACK, J. (1982), "Wages, Rents, and the Quality of Life", *Journal of Political Economy*, **90**, 1257–1278.
ROSEN, S. (1979), "Wage-Based Indexes of Urban Quality of Life", in Peter, M. and Mahlon, S. (eds) *Current Issues in Urban Economics* (Baltimore: Johns Hopkins University Press).
ROSENTHAL, S. and STRANGE, W. (2003), "Geography, Industrial Organization, and Agglomeration", *Review of Economics and Statistics*, **85**, 377–393.
ROY, A. D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, **3**, 135–146.
TOPEL, R. and WARD, M. (1992), "Job Mobility and the Careers of Young Men", *Quarterly Journal of Economics*, **107**, 439–479.
WHEELER, C. (2006), "Cities and the Growth of Wages Among Young Workers: Evidence from the NLSY", *Journal of Urban Economics*, **60**, 162–184.

Downloaded from http://restud.oxfordjournals.org/ at Brown University on January 18, 2017