Better educated workers form many more long-distance job matches, and they move more quickly following local shocks. I argue this is a consequence of larger dispersion in wage offers, independent of geography. In a frictional market, this generates larger surpluses for workers in new matches, which can better justify the cost of moving - should the offer originate from far away. The market is then “thinner” but better integrated spatially. I motivate my hypothesis with evidence on mobility patterns and subjective moving costs, and I test it using wage returns to local and long-distance matches over the jobs ladder.

1 Introduction

Geographical mobility is known to be crucial to the adjustment of local labor markets (Blanchard and Katz, 1992; Amior and Manning, 2018). But there is severe inequity in the incidence of mobility. Better educated Americans make many more long-distance moves, especially near the beginning of their careers: one year after leaving school, 9 percent of postgraduate degree holders move state annually, compared to just 2 percent of high school dropouts (Figure 1). At the same time, the low educated move more sluggishly in response to local employment shocks (Topel, 1986; Bound and Holzer, 2000; Wozniak, 2010; Notowidigdo, 2011), contributing to substantial persistence in local jobless rates (Amior and Manning, 2018). This is of particular concern today, given the severe contraction of local manufacturing industries: see Autor, Dorn and Hanson (2013), Acemoglu et al. (2016) and Charles, Hurst and Notowidigdo (2016).

Figure 1 is based on Current Population Survey (CPS) waves since 1999. Cross-state mobility has declined since the 1980s (Molloy, Smith and Wozniak, 2011), but this decline was fairly uniform across education groups. Appendix A.2 shows the large education differentials have persisted since at least the 1960s.
This figure reports the fraction of individuals living in a different state 12 months previously in Current Population Survey (CPS) March waves between 1999 and 2018, taken from the IPUMS database (Flood et al., 2018). Data points represent three-year moving averages (separately by education) over the support of potential labor market experience. I exclude individuals with less than one year of experience at the beginning of the 12 month window, those living abroad one year previously, and those with imputed migration status: Kaplan and Schulhofer-Wohl (2012) show there are inconsistencies in the imputation procedure in non-response cases. The non-response rate for migration status is 14 percent in my sample, and this varies little with education. See Appendix A.1 for further details.

The evidence suggests the effect of education is causal (Malamud and Wozniak, 2012; Machin, Salvanes and Pelkonen, 2012), but what explains it? One view is that better educated workers face larger local differentials in expected utility, a consequence perhaps of local variation in the returns to human capital (e.g. Costa and Kahn, 2000; Wheeler, 2001; Dahl, 2002; Lkhagvasuren, 2014; Davis and Dingel, 2019). But one might reasonably argue the reverse, that it is the low educated who face larger differentials: after all, they suffer larger local shocks to wages and employment (Hoynes, 2000; Gregg, Machin and Manning, 2004), and the impact of these shocks persists much longer (Amior and Manning, 2018). For these reasons and others, many have concluded that the low educated face prohibitive migration costs, whether due to financial constraints, lack of information or home attachment (Greenwood, 1973; Topel, 1986; Bound and Holzer, 2000; Wozniak, 2010; Moretti, 2011; Kennan, 2015; Balgova, 2018; Caldwell and Danieli, 2018). However, it has proven difficult to identify exactly which costs might be responsible. Indeed, migration costs are typically estimated as a residual, conditional on the assumptions of the particular model (e.g. Kennan and Walker, 2011).

In this paper, I offer an alternative explanation: high-educated mobility is a consequence of larger dispersion in wage offers, independent of geography. In a frictional market, this generates larger job surpluses for workers in new matches (that is, larger job valuations in excess of their reservations), particularly at the beginning of their careers. Crucially, these surpluses will more frequently justify the otherwise prohibitive cost of moving, even if the}

Better educated individuals may also face larger geographical differentials in expected utility for other reasons. Diamond (2016) finds they have stronger preferences over local utilities. And Notowidigdo (2011) argues the low educated are better insured against local shocks by compensating transfer payments and housing costs.
offer distribution is identical everywhere. They will also amplify the migratory response to local employment shocks, as they place more workers on the margin of moving. In this way, the labor market becomes “thinner” (match quality varies considerably) but better integrated spatially.

Theoretically, larger offer dispersion may be motivated intuitively by a notion of specialized skills - or supermodularity between workers’ abilities and job attributes (such as task complexity or firm quality), in the spirit of Sattinger (1975). Based on workers’ job transitions, Gottfries and Teulings (2016) find that offer dispersion is indeed increasing in education. Lise, Meghir and Robin (2016) attribute this to larger worker-job complementarities. Also, Deming and Kahn (2018) identify large variation in firms’ skill requirements (from vacancy postings) within metro areas and occupations, which are positively correlated with firm performance.

To take a practical example, a good job may plausibly motivate a young engineer to move between two similar cities (despite the cost), but not his older counterpart (who already has a good match) or somebody who cuts hair for a living (for whom a good match yields little reward). A frictional market for specialized jobs is crucial to this hypothesis: otherwise, the young engineer would wait for an equivalent local offer. This reflects evidence on job transitions from Nimczik (2018), who shows that better educated workers participate in labor markets which are more expansive spatially but more specialized in particular industries.

My hypothesis builds on early insights from Schwartz (1976) and Wildasin (2000), who discuss how specialized markets can motivate mobility - though to my knowledge, it is absent from the recent literature. My contribution is to explore the theoretical implications using a simple jobs ladder model - and to evaluate the job surplus hypothesis empirically against competing explanations. The idea is also closely related to Van Ommeren, Rietveld and Nijkamp (1997) and Manning (2003b), who argue that search frictions and the associated job surplus can help explain why workers accept offers with long commutes. I apply this insight to the long-standing debate over education differentials in geographical mobility.

I motivate my hypothesis with three descriptive facts. First, the education differentials in Figure 1 are entirely driven by individuals who report moving for the sake of a specific job, rather than to “look for work” or for non-work reasons: this affirms the central role of long-distance job matching. Second, evidence on self-reported willingness to move (from the Panel Study of Income Dynamics) suggests that moving costs are remarkably similar across education groups. And third, as is well known, net migratory flows between states are small (relative to gross flows); but using the American Community Survey, I also show they are not increasing in education, even within detailed occupation-defined labor markets. This casts doubt on the importance of differential local returns in driving the mobility gap. In light of this evidence, the purpose of this paper is to offer an alternative (and non-geographical) explanation for the mobility gap, based on market frictions and the job surplus.

I set out a model of migration embedded in a jobs ladder, building on Schmutz and Sidibé (forthcoming). Workers draw random offers at a finite rate from an exogenous wage offer
distribution, independent of geography. They search both on and off the job, which gives rise to a jobs ladder - following the logic of Burdett (1978). Job offers may arise locally or elsewhere, as in Jackman and Savouri (1992), Molho (2001), Lutgen and Van der Linden (2015) and Llull and Miller (2018). In the latter case, a random moving cost is drawn; and the worker accepts the offer (and moves) if the associated surplus can justify the cost.

In this environment, larger within-area offer dispersion (relative to moving costs) expands the rate of long-distance matching - and especially for workers with initially lower quality matches. Larger offer dispersion also amplifies the migratory response to local changes in offer rates. These results depend crucially on search frictions and job separations: in their absence, there is no job surplus or geographical mobility, and workers only match locally.

Turning to the data, high-educated job flows are indeed consistent with large offer dispersion relative to moving costs. Better educated workers form many more cross-state (relative to within-state) matches. Also, education differentials in mobility are smaller for workers with lower match quality, as proxied by initial unemployment or low job tenure - and this can partly account for similar effects over the support of experience (Figure 1). But using job flows alone, I am unable to disentangle the effects of offer dispersion and moving costs.

To address this problem, I exploit evidence from wage transitions in the Survey of Income and Program Participation. First, I identify differences in within-area offer dispersion and job surplus using the wage returns to within-state job matches. These returns are strongly increasing in education (see also Bartel and Borjas, 1981; Mincer, 1986); and as the model predicts, the education differentials are larger among workers with initially low match quality - as proxied by either tenure or experience. Quantitatively, I show that mean education differentials in local job surplus can account for the bulk of the observed gap in the rate of long-distance matching - under certain distributional assumptions.

Second, I show the wage returns to cross-state matching are disproportionately large (relative to within-state) for better educated workers. Under my model’s assumptions, this implies that workers’ realized moving costs (conditional on moving) are steeply increasing in education - by a compensating differentials argument. This is a natural consequence of larger offer dispersion: better educated workers are more likely to select into migration because of large job surplus and despite steep moving costs. Crucially, the large wage returns to cross-state matches (among college graduates) cannot be explained by differential local human capital premia - even for detailed occupation-defined tasks.

To summarize, I claim there is strong evidence that within-area offer dispersion and job surplus drive the mobility gap - as well as good theoretical reasons to expect it. My hypothesis also offers an intuition for the differential migratory responses to local shocks (and the local persistence of low educated joblessness), without resorting to unobserved differences in moving costs. Though I focus on education differentials, this paper offers new insights for understanding geographical immobility more generally. And though I apply these insights to internal mobility, it also offers a rationale for the surprising degree of positive selection among
international migrants (see e.g. Grogger and Hanson, 2011).

My focus is on worker behavior, so I have chosen to take the offer distribution and contact rate as given. But in a previous iteration of the paper (Amior, 2015), I show that these can amplify any first order effect of job surplus. Intuitively, high educated workers share their large surplus with the firms which recruit them, according to the wage-setting process. And in an effort to realize this surplus, both parties will spend heavily on long-distance search: e.g. workers by investing in long-distance networks, and firms by funding fly-outs of potential hires. These activities raise the long-distance contact rate, and thereby reinforce my hypothesis that job surplus drives the mobility gap.

In the following section, I present my three new descriptive facts. Section 3 sets out the jobs ladder model and derives the key results, Section 4 tests its predictions for local and cross-state job matching rates, and Sections 5 and 6 estimate the wage returns to local and cross-state matches respectively. I also include Online Appendices with a range of empirical exercises, sensitivity tests and theoretical derivations.

2 Motivating facts

2.1 The mobility gap is driven by workers moving for a specific job

Long-distance job matching is integral to my hypothesis, and this approach is supported by evidence on subjective reasons for moving. Figure 2 shows the mobility gap (between education groups) is entirely driven by individuals who report moving for reasons related to a specific job - whether due to a new job, job transfer, or shorter commute. The effect is strongest for the young, though I show in Appendix B that it is not driven by former students returning home.

Individuals very rarely move speculatively to “look for work”: only 5 percent of cross-state migrants report this motivation, compared to the 47 percent who move for a specific job. This is unsurprising given the associated risks (Molho, 1986). Interestingly, these speculative movers are disproportionately low-educated. Amior (2015) attributes this to different levels of investment in long-distance search and recruitment (see also Balgova, 2018), itself a consequence of offer dispersion and job surplus; and I return to this point below. There is also a mild negative education gradient in the residual “non-job” migration, which consists primarily of family and housing-related motivations.

Appendix A.3 offers a more detailed breakdown of reported reasons for both cross-state and cross-county mobility, together with associated education gradients. Interestingly, the (positive) education slope is proportionally steeper for cross-state than cross-county moves, and my model offers a rationale (job surplus should matter more for costlier longer-distance moves: see Section 3.6). There may be concern that household dependents simply report the motivations of

---

3See Amior (2015) for evidence that firms invest more heavily in the recruitment of better educated workers.
the household breadwinners, but Appendix A.4 shows that restricting the sample to top earners in each household makes no difference to the results.

2.2 **Subjective moving costs are unrelated to education**

The cost of moving is typically estimated as a residual, conditional on the assumptions of the particular model. But here, I offer more direct estimates of moving costs, based on a unique set of hypothetical questions on willingness to move in the Panel Study of Income Dynamics (PSID) in the 1970s. These estimates suggest that moving costs are remarkably similar across education groups. Though this is not contemporary data, mobility differentials between education groups in the 1970s are similar to today (see Appendix A.2). And in recent work, Kosar, Ransom and van der Klaauw (2019) have found similar results using a 2018 survey.\(^4\)

In 1969-72 and 1979-80, employed household heads\(^5\) were asked: “Would you be willing to move to another community if you could earn more money there?” And in 1969-72, those who answered affirmatively were also asked: “How much would a job have to pay for you to be willing to move?” These questions speak to the cost of moving conditional on receiving a job offer, so they exclude the costs of long-distance job search. The first question identifies the set of “marginal residents” who would be willing to accept a good long-distance match (should one materialize). And the second allows me to impute the distribution of moving costs among

---

\(^4\)They elicit subjective preferences using a one-off module in the New York Fed’s Survey of Consumer Expectations: respondents are asked to choose between hypothetical migration options. Though education differentials are not their focus, their Table 5 suggests that non-pecuniary moving costs and the value of living close to family vary little with education.

\(^5\)Household heads are always male, unless there is no husband (or cohabiting partner) present or the husband is too ill to respond to the survey.
Figure 3: Share who “would move” and “might move” for better job

The first panel reports the share of employed household heads who report being willing to move for work. This is based on responses to: “Would you be willing to move to another community if you could earn more money there?” The second panel reports the share of employed heads who both (i) answer affirmatively to the question “Do you think you might move in the next couple of years?” and (ii) report job-related reasons in answer to the question “Why might you move?” The sample is restricted to employed household heads with 1-30 years of experience in the years 1969-72 and 1979-80, when both questions were asked. The full sample consists of 16,947 observations.

In the first panel of Figure 3, I plot the share of respondents who report being “willing” to move. 55 percent answer yes. Though this share is declining somewhat in experience, there is remarkably little variation across education groups. Of course, these subjective responses are only useful if the low-educated do not disproportionately overstate their willingness to move. And it turns out they are indeed realistic about their meager migration prospects. The PSID asks: “Do you think you might move in the next couple of years?” and “Why might you move?” Based on this data, the second panel of Figure 3 plots the share of respondents who claim they “might” move for work. The results here reflect the familiar education-experience mobility patterns from Figures 1 and 2 above. The contrast with the first panel is striking: the fact that low-educated workers expect low mobility is apparently unrelated to their hypothetical “willingness” to move.

I now study the distribution of costs among marginal residents, which I impute as the difference between the log reservation wage (for accepting a long distance offer) and the log current wage. This difference may be interpreted as the annuitized value of a hypothetical fixed moving cost (see Section 6.1 below) or the disutility of living away from home (see the extension in Section 3.6). One might alternatively study a dollar (rather than log) difference. The appropriate formulation would depend on the model for utility, but the log difference is more conservative: the imputed high-educated costs will be smaller (in relative terms), thus working

---

6I have grouped undergraduate and postgraduate degree holders together because of small samples. Between 1970 and 1980, the PSID also asked unemployed individuals: “Would you be willing to move to another community if you could get a good job there?” 77 percent agree; and as with the employed, the fraction answering yes varies little with education.

724.3 percent of college graduates who claimed they “might” move residence for job reasons actually did so (citing those same reasons) in the subsequent year, and the number is very similar (25.2) for non-graduates. This suggests there is little systematic difference by education in the accuracy of these subjective expectations.
This figure plots kernel distributions of the imputed (annuitized) costs of migration. Conditional on expressing willingness to move, employed household heads (in the PSID, 1969-72) answer the question: “How much would a job have to pay for you to be willing to move?” Imputed costs are computed as the difference between the log of this reservation wage and the worker’s current wage (or more specifically, the worker’s average wage over the previous 12 months), where wages are measured in hourly terms. I drop all observations with top-coded reservation wages. Importantly, these are unbalanced: 31 percent of graduate observations are top-coded, compared to just 6 percent for non-graduates. To address concerns about selection, I additionally exclude the top 25 percent of the (remaining) non-graduate reservation wage observations. Beyond this, I also drop the top and bottom 2 percent of the remaining imputed costs within each education group. I restrict the sample to individuals with 1-30 years of experience. The sample consists of 398 college graduates and 2,744 non-graduates. Plots are based on Epanechnikov kernel functions, with a bandwidth of 0.07. Scatterplots 4.2, 4.3 in favor of the costs hypothesis.

The distribution of these imputed costs is implicitly truncated: only those who are “willing to move” (i.e. with sufficiently low costs) report a long-distance reservation. Conveniently though, the fraction who do so varies little with education (Figure 3), so selection should not be a problem for group comparisons: for each education group, I observe the (approximately) bottom 50 percent of costs.

In Figure 4, I plot kernel densities of imputed costs separately for college graduates and non-graduates. The two plots look remarkably similar, with mean costs of 0.35 and 0.38 respectively - relative to workers’ current wages. The individual heterogeneity is substantial, as stressed by Kennan and Walker (2011). And reassuringly, Appendix C shows these imputed costs do have significant predictive power for future migration decisions. See Section 6.4 below for comparisons with existing estimates of migration costs.

2.3 The mobility gap is not driven by net migratory flows

It is commonly argued that the mobility gap is driven by local variation in the returns to human capital. If so, net migratory flows of better educated workers between areas should be disproportionately large (relative to gross flows). But I find no evidence of this - both on aggregate (as has previously been documented) and even for net flows within detailed occupation-defined labor markets.
### Table 1: Net cross-state migration rates by education

<table>
<thead>
<tr>
<th>Basic Flows within 2-digit occ markets</th>
<th>Flows within occ markets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross mig rate (%)</td>
<td>Net mig rate (%)</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>HS dropout</td>
<td>2.51</td>
</tr>
<tr>
<td>HS graduate</td>
<td>2.71</td>
</tr>
<tr>
<td>Some college</td>
<td>3.04</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>3.71</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>3.91</td>
</tr>
</tbody>
</table>

This table reports annual gross and net cross-state migration rates within education groups. The net rate is computed as $\frac{1}{n} \sum_j |n_{in}^j - n_{out}^j|$, where $n$ is the total sample of individuals, $n_{in}^j$ is the number of in-migrants to state $j$, and $n_{out}^j$ are the out-migrants. The first three columns report the overall migratory flows by education group, and the final six report flows within 2-digit and 3-digit occupation-defined labor markets. For each education group, these are constructed by weighting occupation-specific migration rates (gross and net) by occupational employment shares. Migrants are defined as individuals who lived in a different state 12 months previously. The sample consists of individuals with 2 to 30 years of potential experience (at the end of the 12 month window) in the ACS between 2000 and 2009, and this is further restricted to the employed in columns 4-9. Employment status and occupation are recorded at time of survey. Occupational codes are based on the census 2000 scheme.

I estimate the cross-state net migration rate as $\frac{1}{n} \sum_s |n_{in}^s - n_{out}^s|$, where $n$ is the total sample of individuals, $n_{in}^s$ is the number of in-migrants to state $s$, and $n_{out}^s$ are the out-migrants. Dividing by 2 ensures that migrants are not double-counted. Notice the gross migration rate is equal to $\frac{1}{n} \sum_s n_{in}^s$ or equivalently $\frac{1}{n} \sum_s n_{out}^s$. I base my estimates on American Community Survey (ACS) samples between 2000 and 2009. Migrants are defined as people who lived in a different state 12 months previously.

Table 1 reports gross and net migration rates separately by education. As is well known, net flows (column 2) are dwarfed by gross flows (column 1). It is less well known that this is especially true of better educated workers. Though gross migration is steeply increasing in education, net flows are remarkably flat; so the ratio of net to gross migration is decreasing in education: from 0.14 for dropouts to 0.08 for postgraduates (column 3). Thus, high-educated migration has a much weaker “directional” component: see Folger and Nam (1967), Schwartz (1971) and Lkhagvasuren (2014).

This does not entirely rule out the possibility that local returns to human capital are driving the mobility gap - if these returns are tied to particular task specializations. This can be tested by studying net and gross flows within detailed occupation-defined labor markets: for each education category, I estimate the within-occupation net migration rate as $\sum_o \gamma_o \frac{1}{2n-o} \sum_s |n_{in}^{os} - n_{out}^{os}|$, where $\gamma_o$ is the fraction of individuals employed in occupation group $o$, and $n_{in}^{os}$ and $n_{out}^{os}$ are the in/out-migrants to/from state $s$ employed in occupation $o$. I restrict the sample to individuals

---

8 Over 2000-9, the ACS offers a consistent occupation classification (based on the census 2000 scheme) and also substantial samples - important for a detailed occupation decomposition. There are 221,000 cross-state migrants in my ACS sample, compared to 18,000 in the CPS in the same years. I take ACS data from IPUMS (Ruggles et al., 2017).

9 See Shryock (1959); Schwartz (1971); Jackman and Savouri (1992); Coen-Pirani (2010); Monras (2015a).
employed at the time of survey, when occupations are recorded.\footnote{Note this is immediately after the twelve month period in which migration occurs. This is the appropriate time to measure occupation for this exercise: an individual’s ex post occupation is a good indicator of the job market in which they were originally searching.} I report estimates separately using 98 2-digit occupations and 466 3-digit occupations. The education gradient in net migration (columns 5-6) remains remarkably flat, even using the 3-digit definition. And again, the net-gross ratios are strongly decreasing in education (columns 7-8).

Appendix D reproduces these migration rates and wage differentials separately for different experience groups, and the key results are preserved. And in the same appendix, I reject the possibility that gross mobility differentials are merely driven by churn, i.e. software engineers moving to California and then returning home. Specifically, I study stocks of migrants (i.e. individuals living outside their birth state): this excludes return migrants by construction. Though net imbalances of migrant stocks (between state pairs) are increasing in education, the net-gross ratio is remarkably flat - and again, even within occupations.

This all suggests the mobility gap cannot be explained by better educated workers converging on particular states, even within detailed occupation-defined markets. For example, there may be many software engineers flocking to California (relative to hairdressers), but there are also many moving in the opposite direction. Of course, workers may be sorting geographically on unobserved ability, even within detailed occupations. For example, the best engineers may be taking high paid positions in California, and the mediocre ones moving elsewhere. However, I show in Section 6 that the wage returns to long-distance mobility cannot be explained by differential local occupation premia, even within education groups.

This is not to say California does not offer particular productive advantages to software engineers, in the spirit of Dahl (2002). But in the context of spatial arbitrage with inelastic housing (as in Roback, 1982) or some other congestion externality, any such skill-location complementarities may be manifested more in the geographical distribution of population stocks than population flows.

3 Jobs ladder model of migration

3.1 Overview

I set the model in continuous time. The model and its parameters are defined for an individual worker \(i\); but to ease the notation, I suppress the subscript \(i\) until I set out the empirical specification. Workers are distributed across \(J\) areas denoted by subscript \(j\). In deriving the main results, I assume these areas are identical - at least in steady-state. Of course, there is substantial local variation in incomes in practice, but this theoretically should be offset by corresponding variation in housing costs in spatial equilibrium. One can certainly model local invariance in worker values as an outcome of spatial arbitrage (as in Roback, 1982); but to ease the exposti-
tion, I have chosen to take local wage offers and prices as given. Still, I do briefly consider the implications of endogenous wage-setting in Section 3.6 below, together with a range of other theoretical extensions.

Unemployed workers receive real income \( b \) (deflated by local prices), and employed workers are paid a (real) wage equal to:

\[
w_j(\epsilon, X) = \gamma_j' X + \sigma^\epsilon(X) \epsilon
\]

where \( X \) is a vector of individual characteristics defining human capital, and \( \epsilon \) is an idiosyncratic term representing job match quality. In (1), the returns to human capital, \( \gamma_j' \), are permitted to vary locally. But, the evidence on net flows (Section 2.3) suggests this variation does not contribute to the mobility gap between education groups. My approach is to rule out differential local returns in the main exposition (i.e. impose \( \gamma_j' = \gamma \)), but I consider its implications in an extension below and when estimating the wage returns.

The importance of match quality depends on the offer dispersion \( \sigma^\epsilon \), which is permitted to vary with an individual’s human capital \( X \). My central claim is that \( \sigma^\epsilon \) is increasing in education. This may be motivated intuitively by a notion of specialized skills. Or, in the spirit of Sattinger (1975), it may be characterized as supermodularity between human capital \( X \) and job attributes (such as task complexity or firm quality) represented by \( \epsilon \). To ease notation, I suppress the \( X \) argument of \( \sigma^\epsilon \) in the analysis below.

Both employed and unemployed workers in area \( j \) draw local job offers at a finite exogenous rate \( \lambda_j \), and they draw offers from other areas \( k \neq j \) at rate \( \pi \lambda_k \). One might suppose \( \pi < 1 \) if long-distance job search is more costly. For example, many job offers arrive through personal networks (Granovetter, 1995), and networks are weaker at longer distances. I take \( \pi \) as given; but as I point out below, there are good reasons to believe it may vary with education. Again, I derive the main results for identical offer rates (i.e. \( \lambda_j = \lambda \)), though I consider local deviations from this environment below.

Job offers are characterized by match quality draws \( \epsilon \) from an exogenous distribution, \( F^\epsilon \). Both \( F^\epsilon \) and its density \( f^\epsilon \) are continuous and differentiable over their support, and the hazard rate \( \frac{f^\epsilon(\epsilon)}{1 - F^\epsilon(\epsilon)} \) is monotonically increasing in \( \epsilon \). If an offer happens to arrive from outside area \( j \), workers also draw a fixed moving cost equal to:

\[
m = \sigma^\mu(X) \mu
\]

which is payable on acceptance of the offer. The parameter \( \mu \sim F^\mu \) is stochastic and bounded below by zero, where both \( F^\mu \) and its density \( f^\mu \) are continuous and differentiable. I also assume the elasticity of the density \( f^\mu \), i.e. \( \mu \frac{f^\mu(\mu)}{f^\mu(1)} \), is monotonically decreasing in \( \mu \). This is a stronger condition than log concavity (or a monotone hazard), but it is satisfied by standard dis-
tributions with finite lower supports.\footnote{E.g. the two-parameter Weibull, two-parameter gamma, exponential, chi-squared, log normal and uniform. See Andersen (1996) and Nocke and Yeaple (2008).} $\sigma^\mu$ determines the size of moving costs and is permitted to vary with human capital $X$; though as before, I suppress the $X$ argument below.

Workers condition their acceptance of long-distance offers on their $\epsilon$ and $\mu$ draws. Implicitly, the zero lower bound on $\mu$ rules out “non-job” motivations for moving (Figure 2 suggests these do not contribute to the mobility gap). But in an extension below, I show how one can generate equivalent results in a model with heterogeneous local preferences instead of fixed moving costs.

Workers can exit jobs in two ways: either through a quit (if they receive a more attractive offer) or involuntary separation (to unemployment). The latter arrive randomly at rate $\delta$. On separation, workers remain (at least initially) in their current area of residence. I rule out the possibility of moving without a job in hand: i.e. from unemployment in one area to unemployment in another. In practice, the evidence in Figure 2 suggests that such speculative migration is rare; but I briefly consider this possibility in an extension below.

### 3.2 Workers’ value and job matching rates

Given the offer rate is independent of employment status, the reservation wage for unemployed workers is simply the out-of-work income $b$ (Burdett and Mortensen, 1998). Conditional on human capital $X$, the reservation match quality which generates a wage offer equal to $b$ is:

$$\epsilon_{Rj} = \frac{b - \gamma_j X}{\sigma^\epsilon} \quad (3)$$

The unemployment value in area $j$ is therefore equal to $V_j(\epsilon_{R})$, where $V_j(\epsilon)$ is the value of a local job with match quality $\epsilon$:

$$rV_j(\epsilon) = \gamma_j X + \sigma^\epsilon \epsilon + \delta \left[ V_j(\epsilon_{R}) - V_j(\epsilon) \right] + \lambda_j \int_{\epsilon}^{\infty} \left[ V_j(\epsilon') - V_j(\epsilon) \right] dF^\epsilon(\epsilon') \quad (4)$$

$$+ \pi \sum_{k \neq j} \lambda_k \int_{0}^{\infty} \left[ \int_{-\infty}^{\infty} \max \left\{ V_k(\epsilon') - V_j(\epsilon) - \sigma^\mu \mu, 0 \right\} dF^\epsilon(\epsilon') \right] dF^\mu(\mu)$$

To ease notation, I have suppressed $X$ as an argument in $V_j$. Workers discount utility at rate $r$. The first term, $\gamma_j X + \sigma^\epsilon \epsilon$, is the flow utility. The second term accounts for the asset loss associated with job separations, which arrive at rate $\delta$. The final two terms describe the value of local and long-distance search respectively. Workers accept any local offer yielding $\epsilon'$ (distributed $F^\epsilon$) exceeding $\epsilon$, where $\epsilon$ is the initial match quality. And conditional on initial quality $\epsilon$ and a cost draw $\mu$, workers accept an offer $\epsilon'$ outside $j$ if the worker’s job surplus $V_k(\epsilon') - V_j(\epsilon)$ exceeds the moving cost $\sigma^\mu \mu$.

Imposing that local returns and offer rates are the same everywhere (i.e. $\gamma_j = \gamma$ and $\lambda_j = \lambda$),
\[ V_j(\epsilon) \text{ collapses to:} \]

\[ rV(\epsilon) = \gamma X + \sigma^\epsilon \epsilon + \delta [V(\epsilon_R) - V(\epsilon)] + \lambda \int_\epsilon^\infty [V_j(\epsilon') - V_j(\epsilon)] dF^\epsilon(\epsilon') \]

\[ + \pi \lambda \int_0^\infty \left[ \int_\epsilon^\infty \max \{V(\epsilon') - V(\epsilon) - \sigma^\mu \mu, 0\} dF^\epsilon(\epsilon') \right] dF^\mu(\mu) \]

where \( \tilde{\pi} = \pi (J - 1) \). Conditional on initial match quality \( \epsilon \), workers form local matches at rate:

\[ \rho_L(\epsilon) = \lambda [1 - F^\epsilon(\epsilon)] \]

and cross-area matches at rate:

\[ \rho_C(\epsilon) = \tilde{\pi} \lambda \int_\epsilon^\infty \left[ F^\mu \left( \frac{V(\epsilon') - V(\epsilon)}{\sigma^\mu} \right) \right] dF^\epsilon(\epsilon') \]

where \( \rho_C(\epsilon) \) depends on the size of the job surplus, \( V(\epsilon') - V(\epsilon) \).

### 3.3 The crucial role of search frictions and job separations

In the absence of spatial disparities in local values, search frictions (i.e. a finite offer rate \( \lambda \)) are crucial to generating a job surplus and motivating geographical mobility. To see this more clearly, it is useful to rewrite the job surplus in (7) as:

\[ V(\epsilon + z) - V(\epsilon) = \int_\epsilon^{\epsilon + z} V'(x) dx = \sigma^\epsilon \Omega(z, \epsilon) \]

where

\[ \Omega(z|\epsilon) = \int_\epsilon^{\epsilon + z} \frac{1}{r + \delta + \rho_L(x) + \rho_C(x)} dx \]

is the surplus in \( \epsilon \) units accruing to a match which raises match quality by \( z \), for a worker with initial match quality \( \epsilon \); and \( \sigma^\epsilon \Omega(z|\epsilon) \) is the surplus in wage units. Using (9), the cross-area matching rate in (7) can then be expressed as:

\[ \rho_C(\epsilon) = \tilde{\pi} \lambda \int_\epsilon^\infty \left[ F^\mu \left( \frac{\sigma^\epsilon}{\sigma^\mu} \int_\epsilon^{\epsilon'} \frac{1}{r + \delta + \rho_L(x) + \rho_C(x)} dx \right) \right] dF^\epsilon(\epsilon') \]

Search frictions matter for two reasons. First, for given initial match quality \( \epsilon \), the offer rate \( \lambda \) affects the rate at which surplus is discounted - via the \( \rho_L \) term in the denominator of (10). Workers will only accept long-distance offers (at large expense) if they do not expect an acceptable local offer to arrive soon. A smaller \( \lambda \) expands the (discounted) value of long-distance matches and increases their frequency, relative to local matches.

Second, as I show in Appendix H, the offer rate affects the equilibrium distribution of match quality \( \epsilon \) across workers. For an infinite \( \lambda \) (or zero separation rate \( \delta \)), all workers will benefit from the maximum match quality in equilibrium; so there will be no job surplus and no cross-
area matching. But as $\lambda$ declines relative to $\delta$ (i.e. the market becomes “thinner”), workers increasingly find themselves at lower “rungs” of the jobs ladder (i.e. lower $\varepsilon$); so job matches will yield larger surpluses - which facilitate greater geographical mobility.

Similarly, a positive separation rate $\delta$ is crucial to reducing the equilibrium $\varepsilon$s and generating job surplus. But also, $\delta$ cannot be too large: otherwise, workers would discount the surplus in new matches very heavily (see (10)). Intuitively, workers will not move for long-distance matches which they expect to terminate soon.

### 3.4 Impact of $\sigma^\varepsilon$ and $\sigma^\mu$ on job matching rates

I now consider the impact of offer dispersion $\sigma^\varepsilon$ and the size of moving costs $\sigma^\mu$ on local and cross-area matching rates, across the support of match quality $\varepsilon$.

**Proposition 1.** Given a worker’s initial match quality $\varepsilon$, the local matching rate $\rho_L(\varepsilon)$ is independent of the offer dispersion $\sigma^\varepsilon$. But the cross-area matching rate $\rho_C(\varepsilon)$ is increasing in $\sigma^\varepsilon$ and decreasing in the size of moving costs $\sigma^\mu$.

The independence of local job matching follows from (6). Intuitively, local job transitions are costless, so a strictly positive job surplus is not necessary for the acceptance of an offer. So, larger offer dispersion $\sigma^\varepsilon$ (and the larger associated surplus) will have no effect on $\rho_L(\varepsilon)$.

In contrast, cross-area matching is costly, so job surplus does matter. Surpluses are increasing in offer dispersion $\sigma^\varepsilon$, but they justify fewer moves if $\sigma^\mu$ is larger: this is clear from (10). Alternatively, a larger offer dispersion $\sigma^\varepsilon$ can be interpreted as “thinning” the market; and an equivalent argument to Section 3.3 then applies.

It is important to place some caveats on the local matching result. First, Proposition 1 speaks only to the matching rate for given $\varepsilon$: mean matching rates will also depend on the distribution of the $\varepsilon$s, which in turn depends on the reservation $\varepsilon_R$ and the separation rate $\delta$ (these may themselves be sensitive to offer dispersion: see Section 3.6). Second, local matching is clearly not entirely costless in practice; in which case, $\rho_L(\varepsilon)$ may increase (somewhat) in $\sigma^\varepsilon$ for given $\varepsilon$. But any effect will be relatively small\(^{12}\) (compared to cross-state matching), and the mean local rate may depend more on the $\varepsilon$ distribution - as I discuss below.

**Proposition 2.** For sufficiently large moving costs (i.e. sufficiently large $\sigma^\mu$), the positive effect of offer dispersion $\sigma^\varepsilon$ on the cross-area matching rate $\rho_C(\varepsilon)$ is decreasing in a worker’s initial match quality $\varepsilon$. Similarly, the negative effect of $\sigma^\mu$ on $\rho_C(\varepsilon)$ is decreasing in $\varepsilon$.

Intuitively, workers with larger initial match quality $\varepsilon$ have fewer rungs of the jobs ladder left to climb. Consequently, offer dispersion $\sigma^\varepsilon$ will matter less for their surpluses in future matches and, therefore, for their cross-area matching rate. Similarly, workers with larger $\varepsilon$ will be less sensitive to changes in moving costs $\sigma^\mu$: they are unlikely to move either way.

\(^{12}\)Section 3.6 describes how the impact of $\sigma^\varepsilon$ is increasing in migration distance (because of larger costs), and the same principle applies here.
To see this more formally, consider the derivative of (7) with respect to \( \log \sigma^e \):

\[
\frac{d\rho_C(\varepsilon)}{d\log \frac{\sigma^e}{\sigma^\mu}} = \tilde{\pi} \lambda \int_{\varepsilon}^{\infty} \left\{ \left[ \frac{d\log \Omega(\varepsilon' - \varepsilon|\varepsilon)}{d\log \frac{\sigma^e}{\sigma^\mu}} + 1 \right] \frac{\sigma^e}{\sigma^\mu} \Omega(\varepsilon' - \varepsilon|\varepsilon) f^\mu \left( \frac{\sigma^e}{\sigma^\mu} \Omega(\varepsilon' - \varepsilon|\varepsilon) \right) \right\} dF^e(\varepsilon')
\]

(11)

This expression is positive, consistent with Proposition 1. And in Appendix I.1, I show further that it is decreasing in initial match quality \( \varepsilon \). This effect is driven by the \( \Omega(\varepsilon' - \varepsilon|\varepsilon) \) term (representing the job surplus), which is decreasing in \( \varepsilon \).

In principle, the vagaries of the moving cost distribution \( F^\mu \) may upset the result, especially if there are many workers on the margin of moving. A sufficient condition for Proposition 2 is that \( \mu f^\mu(\mu) \) is unambiguously increasing in \( \mu \); or equivalently, the elasticity of the density \( f^\mu \) globally exceeds -1. Alternatively, as I show in Appendix I.1, Proposition 2 must also hold for sufficiently large \( \sigma^\mu \) - given my assumption (above) that the elasticity of the density is decreasing. Intuitively, both the increasing \( \mu f^\mu(\mu) \) and large \( \sigma^\mu \) conditions ensure that a substantial fraction of moving cost draws are “large”.

3.5 Response to local shocks

Until now, I have focused on the implications of within-area offer dispersion \( \sigma^e \) for steady-state mobility. But as I now show, it can also help account for known differences in the migratory responses to local shocks. The model in this paper is not entirely suitable for assessing local labor market adjustment, as I take the firm-side of the economy (i.e. the offer distribution and arrival rate) as given. Nevertheless, the model can assess the initial migratory response, at the moment the shock arrives; and in a more complete model (with endogenous labor market conditions), this should speak to the speed of local adjustment.

For simplicity, I study the response around a steady-state where offer rates \( \lambda_j = \lambda \) and worker values \( V_j = V \) are spatially invariant. Given the exogeneity of the offer rates, this ensures that the population of workers is evenly distributed across areas. The inflow rate to area \( j \), relative to the local stock of workers, is then:

\[
\rho^\text{In}_{Cj}(\varepsilon) = \tilde{\pi} \lambda_j \int_{\varepsilon}^{\infty} \left[ F^\mu \left( \frac{V_j(\varepsilon') - V(\varepsilon)}{\sigma^\mu} \right) \right] dF^e(\varepsilon')
\]

(12)

and the outflow rate from area \( j \) is:

\[
\rho^\text{Out}_{Cj}(\varepsilon) = \tilde{\pi} \lambda_j \int_{\varepsilon}^{\infty} \left[ F^\mu \left( \frac{V(\varepsilon') - V_j(\varepsilon)}{\sigma^\mu} \right) \right] dF^e(\varepsilon')
\]

(13)

A labor demand shock may manifest in local returns \( \gamma_j \) to human capital\(^{13} \) in (1), or in local offers rates \( \lambda_j \). Though I have not offered an equilibrium model, one might expect the offer

\(^{13}\)Or even in area-specific dispersion in these returns, e.g. after introducing a \( j \) subscript in \( \sigma_j^e \).
rate $\lambda_j$ to respond more heavily if the “wage curve” (or labor supply relationship) is flatter.

What is the migratory response? Looking at (12) and (13), what matters is how the shock affects (1) local worker values $V_j(\varepsilon)$ over the support of $\varepsilon$ and (2) the offer rate $\lambda_j$. The migratory response to a given change in $V_j(\varepsilon)$ will depend purely on the size of migration costs $\sigma^\mu$: this determines how many workers are marginal residents. However, the response to $\lambda_j$ will depend additionally on the level of offer dispersion $\sigma^\varepsilon$. In particular I make the following claim:

**Proposition 3.** For sufficiently large moving costs (i.e. sufficiently large $\sigma^\mu$), the migratory response to the local offer rate $\lambda_j$ is increasing in offer dispersion $\sigma^\varepsilon$.

The response of inflows (12) to $\lambda_j$ can be expressed as:

$$\frac{d\rho_{Cj}^{\text{In}}(\varepsilon)}{d\lambda_j} = \frac{\bar{\pi} \lambda}{\sigma^\mu} \int_\varepsilon^\infty \left[ \frac{dV_j(\varepsilon')}{d\lambda_j} f^\mu \left( \frac{\sigma^\varepsilon}{\sigma^\mu} \Omega (\varepsilon' - \varepsilon | \varepsilon) \right) \right] dF^\varepsilon (\varepsilon') + \frac{\rho c(\varepsilon)}{\lambda} \quad (14)$$

Holding local values $V_j(\varepsilon)$ fixed, the response is equal to the final term $\frac{\rho c(\varepsilon)}{\lambda}$. Given Proposition 1, this must be increasing in $\sigma^\varepsilon$: there are more workers (outside area $j$) on the margin of moving, so offers from $j$ will elicit larger migratory inflows.

I also show in Appendix I.2 that, for sufficiently large moving costs $\sigma^\mu$, the response of local values $\frac{dV_j(\varepsilon)}{d\lambda_j}$ is positive and increasing (proportionally) with $\sigma^\varepsilon$. Intuitively, workers who expect larger match surplus will benefit more from a larger offer rate - which should elicit a larger migratory response. Similarly to Proposition 2, the vagaries of the moving cost distribution $F^\mu$ may in principle upset this result. But for the same reasons as before, the proposition will hold for sufficiently large migration costs $\sigma^\mu$.

Turning to outflows, the response to the offer rate $\lambda_j$ is:

$$\frac{d\rho_{Cj}^{\text{Out}}(\varepsilon)}{d\lambda_j} = -\frac{\bar{\pi} \lambda}{\sigma^\mu} \frac{dV_j(\varepsilon)}{d\lambda_j} \int_\varepsilon^\infty \left[ f^\mu \left( \frac{\sigma^\varepsilon}{\sigma^\mu} \Omega (\varepsilon' - \varepsilon | \varepsilon) \right) \right] dF^\varepsilon (\varepsilon') \quad (15)$$

Based on the same reasoning as above, $\frac{dV_j(\varepsilon)}{d\lambda_j}$ and therefore the outflow response will be increasing in offer dispersion $\sigma^\varepsilon$ (for sufficiently large $\sigma^\mu$). But holding $V_j(\varepsilon)$ fixed, there is no longer a direct effect of $\lambda_j$. This may help explain why, in practice, local population adjustment materializes largely through changes in inflows rather than outflows: see e.g. Monras (2015a), Dustmann, Schoenberg and Stuhler (2017), Amior (2018) and Amior and Manning (2018).

In this model, changes in the offer rate $\lambda_j$ play a crucial role in driving the large migratory response from better educated workers - though it is difficult to assess this empirically. As outlined above, the evidence has consistently shown that local employment shocks elicit larger migratory responses from better educated workers (Bound and Holzer, 2000; Wozniak, 2010; Notowidigdo, 2011; Amior and Manning, 2018). These studies have relied on initial local industrial composition to identify these shocks (as in Bartik, 1991): some industries have
consistently shed employment, while others have expanded rapidly. An important role for $\lambda_j$ (reflecting job creation or destruction) is at least consistent with this approach.

The purpose of this analysis is to offer a new theoretical rationale for existing evidence on the migratory response to local shocks. Certainly, it would be interesting to reassess this evidence in the light of my hypothesis, but I leave this to future work. In my empirical analysis, I focus instead on steady-state migratory flows - building on Propositions 1 and 2.

### 3.6 Other considerations

Before moving to the evidence, I briefly consider a number of other pertinent issues: (i) the distribution of match quality $\varepsilon$ among workers, (ii) the determinants of the offer distribution, (iii) the determinants of long-distance search intensity $\pi$, (iv) speculative migration, (v) migration distance, (vi) heterogeneous local preferences, and (vii) the implications of differential local returns to human capital.

**Distribution of match quality.** In the analysis above, I have focused on the determination of matching rates $\rho_L(\varepsilon)$ and $\rho_C(\varepsilon)$ for given $\varepsilon$. But the equilibrium distribution of $\varepsilon$ across workers may itself vary with education, and this will have implications for mean matching rates. In Appendix H, I show that mean matching rates are increasing in the separation rate $\delta$ and decreasing in the reservation match quality $\varepsilon_R$. Intuitively, larger $\delta$ and smaller $\varepsilon_R$ push workers down the jobs ladder, so a larger fraction of job offers become viable. Note that both $\delta$ and $\varepsilon_R$ may be sensitive to offer dispersion $\sigma^\varepsilon$: meager surpluses may make job matches more vulnerable to idiosyncratic shocks (thus raising $\delta$); and at least for given non-employment income $b$, the reservation offer $\varepsilon_R$ will depend on the offer spread (as in (3)).

**Determinants of offer distribution.** I have chosen to take the offer distribution as given. But in a more complete model, offer dispersion may originate from dispersion in idiosyncratic match productivity, driven in turn by complementarities between human capital and job attributes. This would generate larger total surpluses in job matches, which are shared by both firms and workers - according to the wages posted by firms or negotiated by the parties.\footnote{The particular multi-location environment of this model offers interesting implications for wage setting. To the extent that match complementarities encourage geographical mobility (by expanding job surpluses), outside options in distant locations become more viable. Firms would then be compelled to offer higher wages to compete for workers over longer distances (as in Caldwell and Danieli, 2018), and this would amplify any initial effect of these complementarities on the job surplus and mobility.} To the extent that the surplus is not fully captured by firms, dispersion in match productivity will pass through to wage offers. In an earlier version of this paper (Amior, 2015), I offer such a model with optimizing firms; and I also bring evidence that firms recruiting better educated workers invest more resources in the recruitment process (in terms of applications received, applicants interviewed and human resource hours). As I argue there, this is consistent with these firms securing larger match surplus themselves.

**Determinants of search intensity.** I have taken $\pi$, the intensity of long-distance job search,
as given. But there is good reason to believe that $\pi$ is increasing in education. Suppose workers (and firms in a more complete model) choose how much effort to invest in long-distance job search (and recruitment, e.g. by funding fly-outs of potential hires). In principle, the amount of effort should be increasing in the expected surplus net of moving costs in cross-area matches (i.e. the long-distance analogue of Lentz, 2010). Therefore, $\pi$ should be increasing in offer dispersion $\sigma^e$ (or in match productivity dispersion, in a model with firms) and decreasing in moving costs $\sigma^\mu$. In this way, endogenous search effort would serve to amplify any effects of $\sigma^e$ and $\sigma^\mu$ - which would reinforce my hypothesis. Still, I show below that observable differentials in job surplus by education may be sufficient to account for the mobility gap without resorting to endogenous search costs - at least under a particular parameterization of the costs distribution.

**Speculative migration.** Until now, I have ruled out the possibility of individuals moving speculatively to look for work: i.e. from unemployment in one area to unemployment in another. Of course, this is only relevant in an environment where local utilities differ; and few workers actually make speculative moves in practice (Figure 2). Suppose that, at rate $\xi$, unemployed workers are given the option of making a speculative move, conditional on a $\mu$ draw from the moving cost distribution $F^\mu$. Notice there is no incentive to make a costly speculative move if $\pi = 1$, i.e. if workers already have full access to job offers from all locations. As $\pi$ decreases though, workers are more likely to make speculative moves and less likely to engage in cross-area matching (i.e. $\rho_C(\epsilon)$ falls). In this way, speculative migration and cross-area matching are substitutes. And if $\pi$ is indeed larger for better educated workers (for the reasons discussed above), this may help explain why they make fewer speculative moves (Figure 2). See Amior (2015), an earlier version of this paper, for further discussion on this point.

**Migration distance.** I have not explicitly modeled distance above, but its effect can be interpreted through the lens of moving cost draws. Conditional on a higher cost draw, the cross-area matching rate will be more elastic to offer dispersion: a consequence of the monotone hazard rate of the offer distribution $F^\epsilon$. So, job surplus should matter more for longer-distance migration decisions - to the extent that these are more costly. This is consistent with evidence in Appendix A.3 that the (positive) education slope is proportionally steeper for cross-state than cross-county moves. See also Schwartz (1973) and Davis and Dingel (2019) for similar evidence on the effect of distance.

**Heterogeneous local preferences.** In the main exposition, I have modeled the migration friction as a fixed one-off cost $m$. But one can generate equivalent results by replacing this with heterogeneous preferences over locations (as in e.g. Moretti, 2011, Gyourko, Mayer and

---

15Consider a worker who draws a moving cost with annuitized value $\hat{m}$. I define the “annuitized cost” more formally in equation (21) below: it summarizes the minimum wage improvement required to justify a given long-distance move. The probability of accepting such a cross-area match (with ex ante unspecified wage offer) is $\left[1 - F^\epsilon(\epsilon + \frac{\hat{m}}{\sigma^e})\right]$, and the elasticity of this probability to offer dispersion $\sigma^e$ is $f^e(\epsilon +\frac{\hat{m}}{\sigma^e}) \frac{1 - F^\epsilon(\epsilon + \frac{\hat{m}}{\sigma^e})}{1 - F^\epsilon(\epsilon + \frac{\hat{m}}{\sigma^e})}$. Given the monotone hazard assumption, the probability is more elastic for larger cost draws $\hat{m}$.

---

18
Sinai, 2013, and Hilber and Robert-Nicoud, 2013). Suppose a worker $i$ residing in area $j$ receives a flow utility equal to $w(ε, X) + σ^α(X) α_{ij}$, where the $α_{ij}$ matches are i.i.d. random draws from a distribution $F^α$ with a maximum of 0, and where $σ^α$ determines the strength of preferences. Given human capital $X$, the value of a match can be summarized by the pair $(ε, α)$. A worker with current match quality $(ε, α)$ will accept a job offering $(ε', α')$ if $σ^ε(ε' − ε) + σ^α(α' − α) ≥ 0$. So, the cross-area matching rate is:

$$ρ_C(ε, α) = \tilde{π} λ \int_{−∞}^{∞} \left[ 1 − F^α(\alpha + \frac{σ^ε}{σ^α} (ε − ε')) \right] dF^ε(ε')$$

(16)

As $σ^α$ becomes large, the equilibrium distribution of workers’ amenity matches $α$ will converge to a mass point at zero (its maximum value). And so, the expected cross-area matching rate $E[ρ_C(ε, α) | ε]$ for given $ε$ will converge to $\tilde{π} λ \int_{−∞}^{ε} \left[ 1 − F^α(\frac{σ^ε}{σ^α} (ε − ε')) \right] dF^ε(ε')$. This is analogous to (10), and Propositions 1 and 2 follow immediately, though with $σ^µ$ (the size of moving costs) replaced by $σ^α$ (the strength of local preferences). One advantage of this specification is that it can motivate return migration: workers who leave home for a productive match (i.e. a high $ε$) may later return to restore their $α$. Kennan and Walker (2011) emphasize that return migration accounts for many long-distance moves, though I show in Appendix B and D that it does not drive education differentials in mobility.

Also, this extension can help explain why low-educated movers disproportionately report “non-job” (primarily family and housing) reasons: see Figure 2. Given a meager job surplus, low-educated workers are more likely to break their existing job match and move elsewhere for the sake of a large amenity draw $α$.

**Differential local returns.** The specification for wages in (1) permits local variation in the returns (encapsulated by $γ_j$) to a multi-dimensional index of human capital $X$. For example, software engineers may expect larger wage offers in California than New York, and the reverse for bankers. I have abstracted from this variation when deriving the main results, but I now briefly consider its implications. Interestingly, local dispersion in $γ_j$ will actually reduce the rate of cross-area matching in steady-state. This effect is analogous to the effect of larger $σ^α$ in the previous paragraph. Intuitively, individuals will be reluctant to leave locations which offer larger returns to their particular human capital. However, suppose workers retire and are replaced by new entrants with different $X$s. Differential local returns may then boost geographical mobility, as new entrants sort to the areas which offer them the largest returns.

Still, to the extent that differential returns account for the mobility gap between education groups, we should expect to see large net flows of high educated individuals - especially within detailed occupation groups. But this is not what the evidence shows (Section 2.3), even when I restrict the sample to new labor market entrants (Appendix D). In line with (1), we should also expect that the wage return to long-distance matching is largely driven by area-specific returns to human capital. But in Section 6 below, I find no evidence that the return is driven by differential local human capital premia - even for detailed occupation-defined tasks.
4 Evidence on job matching rates

In this section, I show that empirical patterns in within and cross-state matching rates are consistent with better educated workers facing larger offer dispersion $\sigma^\epsilon$ (and enjoying larger job surpluses). However, using matching rates alone, it is not possible to identify the effect of $\sigma^\epsilon$ separately from differences in moving costs $\sigma^{\mu}$: see Propositions 1 and 2 above. I attempt to disentangle these effects in Sections 5 and 6, using evidence on the wage returns to matching.

4.1 Data

I base my analysis on the Survey of Income and Program Participation (SIPP). The SIPP offers substantial samples and high-frequency waves, just four months apart. I study transitions between outcomes recorded at the end of each wave\(^{16}\) for individuals with 1 to 30 years of labor market experience and no business income, in SIPP panels beginning 1996, 2001, 2004 and 2008 (which cover the period 1996-2013).

A “job match” occurs when an individual works for an employer at the end of wave $t$ for whom they did not work in $t - 1$. For individuals with multiple jobs at the end of $t$, I restrict attention to cases where the new job is the “primary” job: that is, the job which occupies the most weekly hours.\(^{17}\) A cross-state job match is one which is accompanied with a change in state of residence. In principle, it would be preferable to use the state of the individual’s workplace (to account for workers who choose to commute), but this information is not available.

4.2 Mean job matching rates

Table 2 summarizes mean within-state and cross-state matching rates by education. Identifying the “areas” in Section 3 with states, these represent the mean $\rho_L(\epsilon)$ and $\rho_C(\epsilon)$ across the distribution of workers’ $\epsilon$, which I denote $\bar{\rho}_L$ and $\bar{\rho}_C$. Column 1 shows the within-state rate is decreasing in education: about 10 percent of high school dropouts begin a new local job in each wave, compared to 6 percent of postgraduate degree holders. The patterns are similar for the initially employed (column 2), though with smaller numbers. This is likely to reflect substantial churn in low educated markets, driven by large separation rates to non-employment (column 9) and subsequent progression up the jobs ladder (from low initial match quality $\epsilon$).\(^{18}\) See the discussion in Section 3.6 and Appendix H.

\(^{16}\)SIPP respondents report their earnings at the end of each month, but I do not exploit these monthly frequencies. The SIPP is known to suffer from severe seam bias (see e.g. Marquis and Moore, 2010), presumably due to poor recall: monthly changes in individuals’ outcomes tend to be larger between months at the seam of two waves than within the same wave.

\(^{17}\)Individuals report up to two jobs in each wave. In cases where there are two jobs with equal hours, I define the primary job as the first one reported by the individual.

\(^{18}\)Among the non-employed, better educated workers do find work somewhat more quickly (columns 3-4). In terms of the model above, this may reflect a lower reservation match quality $\epsilon_R$. See also Mincer (1991) on education differences in employment transitions.
Table 2: Job matching and separation rates: 4-month intervals

<table>
<thead>
<tr>
<th></th>
<th>Within-state matching rate (%)</th>
<th>Cross-state matching rate (%)</th>
<th>Separation rate to non-emp (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All indiv</td>
<td>By emp status in t−1</td>
<td>All indiv</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>HS dropout</td>
<td>10.42</td>
<td>8.38</td>
<td>26.03</td>
</tr>
<tr>
<td>HS graduate</td>
<td>10.57</td>
<td>7.15</td>
<td>30.48</td>
</tr>
<tr>
<td>Some college</td>
<td>10.40</td>
<td>6.97</td>
<td>32.83</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>8.25</td>
<td>5.51</td>
<td>36.77</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>6.18</td>
<td>4.03</td>
<td>36.33</td>
</tr>
</tbody>
</table>

This table reports within-state and cross-state matching rates across four-month waves, together with separation rates to non-employment, for individuals with 1-30 years of potential experience and with no business income, based on the SIPP panels of 1996, 2001, 2004 and 2008. The full sample consists of 1.2m individual wave transitions. A "job find" occurs when an individual works for an employer at the end of wave $t$ for whom they did not work in $t−1$. For individuals with multiple jobs at the end of $t$, I restrict attention to cases where the new job is the "primary" job: that is, the job which occupies the most weekly hours. A cross-state job find is one which is accompanied with a change in state of residence.

The mean cross-state matching rates $\bar{\rho}_C$ in columns 5-8 are of course much smaller. But the education slopes are consistently positive and proportionally much steeper. In light of Proposition 1, this is consistent with positive education differentials in offer dispersion $\sigma^\epsilon$ or negative differentials in moving costs $\sigma^\mu$.

Are these differentials larger at the bottom of the jobs ladder, as Proposition 2 predicts? A natural proxy for a worker’s rung is initial employment status. As one would expect, the unemployed form more cross-state matches: 0.39 percent over four-month waves on average, compared to 0.18 percent for the employed. But crucially, the education gradient is also much steeper for the unemployed. Again, this is consistent with better educated workers facing larger $\sigma^\epsilon$ or smaller $\sigma^\mu$.

4.3 Job matching rates within employment cycles

Table 2 sheds some light on Proposition 2 by focusing on one dimension of the jobs ladder: initial employment status. I now dig further into Proposition 2 by studying heterogeneity in match quality $\epsilon$ among the initially employed. I proxy initial $\epsilon$ with initial (i.e. $t−1$) job tenure. Intuitively, conditional on the offer rate $\lambda$, a longer tenure reflects a higher valuation of a particular match, relative to the outside options which may arise: see e.g. Burdett (1978).19

In Figure 5, I study how within-state and cross-state matching rates vary with initial match quality, within “employment cycles”20 (i.e. excluding transitions with intervening unemployment or layoff spells). Panel A plots the within-state rate on the log of initial job tenure (mea-

---

19This is a natural consequence of a frictional environment where jobs are modeled as “pure search goods” (as in Section 3 above) or “experience goods” (as in e.g. Jovanovic, 1979), or if tenure drives the accumulation of firm-specific human capital (e.g. Topel, 1991).

20In the language of Wolpin (1992) or Barlevy (2008).
Figure 5: Job matching rates within employment cycles, by tenure and experience

This figure plots the rate of within-state and cross-state job matching on both the log of initial (i.e. $t - 1$) job tenure (measured in months) and experience, separately by education. I restrict attention to job finds within employment cycles, i.e. without intervening unemployment or layoff spells. I have divided the support of both experience and tenure into decile bins, within each education group. Each data point identifies the mean log tenure (or experience) and the matching rate in a given decile bin.

Trivially, within-state matching rates are decreasing in initial tenure. Tenure is typically larger for better educated workers: see the decile markets along the x-axis. This is consistent with lower churn (see Table 2) and can partially account for the negative education differentials in mean matching rates: at least at high initial tenure, education has little effect in Panel A. The gaps are larger however at lower tenure: intuitively, tenure provides little information on match quality for new matches (see Appendix E).

The cross-state matching rates in Panel B look very different. As before, these are decreasing in initial tenure. But the education differentials are now clearly positive and strongly decreasing in tenure - reaching zero at the top of the support. Based on Propositions 1 and 2, these patterns are again consistent with better educated workers facing larger $\sigma^\varepsilon$ or smaller $\sigma^\mu$.

These effects can help account for patterns in similar in matching rates by labor market experience (Panels C and D): older workers have spent more years searching, so they should typically enjoy larger match quality $\varepsilon$ (e.g. Topel and Ward, 1992; Manning, 2003a; Gottfries and Teulings, 2016). Of course, experience (and by association, tenure) is likely to affect moving costs independently\(^{21}\), so these patterns are unlikely to be purely driven by match

---

\(^{21}\) In Figure 3 above, older workers report less willingness to move for work. A human capital explanation is
quality. Still, Appendix F does show that initial tenure can (statistically) account for a third of the experience effect on cross-state matching\(^{22}\) - which suggests match quality does play a significant role. Though the effect of experience on mobility is an interesting question, I do not pursue it further: my focus here is the effect of education.

To summarize, the evidence in this section offers empirical support for a jobs ladder model of migration. Based on Propositions 1 and 2, the patterns of within-state and cross-state job matching are consistent with education differentials in both offer dispersion \(\sigma^e\) and moving costs \(\sigma^u\). In what follows, I attempt to discriminate between these hypotheses.

5 Wage returns to within-state job matching

I now attempt to identify variation in expected job surplus (and offer dispersion \(\sigma^e\)) across education groups, by estimating the mean wage return to within-state job matching. I focus on within-state returns to exclude a contribution from moving costs. And I then consider whether this variation is sufficiently large to account for differentials in cross-state matching. One might alternatively identify \(\sigma^e\) by studying cross-sectional wage variances (or posted wages), but my approach offers the advantage of eliminating worker fixed effects.

5.1 Empirical specification

Using equation (1), the expected wage change between \(t - 1\) and \(t\) for individual \(i\) can be written as:

\[
\mathbb{E}(w_{it} - w_{it-1}) = \sigma^e \mathbb{E}(\epsilon_{it} - \epsilon_{it-1}) \cdot I[\text{NewJob}_{it} = 1] + \gamma' (X_{it} - X_{it-1})
\]

where \(\mathbb{E}(\epsilon_{it} - \epsilon_{it-1})\) is the expected change in match quality for job changers (for whom the indicator function \(I\) takes 1), and \(\gamma' (X_{it} - X_{it-1})\) is the contribution from changes in human capital. This yields the following empirical specification:

\[
\Delta w_{it} = \beta_0 + \beta_1 \text{NewJob}_{it} + \beta'_2 X_{it} + \beta_t + u_{it}
\]

where \(\Delta w_{it}\) is the change in worker \(i\)’s log wage between \(t - 1\) and \(t\), controlling for a vector \(X_{it}\) of demographic characteristics (see the notes under Table 3) and time effects \(\beta_t\). A log specification for wages may be justified by writing the model above in log utility.\(^{23}\) This may not be theoretically innocuous since workers cannot borrow or save in the model. However, that older workers have fewer years to benefit from the sunk cost of moving (see e.g. Kennan and Walker, 2011).

\(^{22}\)To the extent that tenure is an imprecise proxy for match quality, this presumably understates the true contribution of match quality \(\epsilon\) to the experience effect.

\(^{23}\)Grogger and Hanson (2011) show that a Roy model with linear utility and skill-invariant moving costs can better explain the observed selection of high and low-educated migrants across countries than an alternative specification with log utility and moving costs which are proportional to income. However, it is not clear whether this result is generalizable to internal migration in the US, where wage gains are much smaller.
the log specification will yield more conservative estimates of the education gradient in wage returns (since better educated workers earn more).

I restrict the sample to transitions which lie within employment cycles: that is, I exclude transitions which include unemployment or layoff spells. Without this restriction, the initial wage cannot be interpreted as a reservation - in which case I cannot identify the job surplus. I also exclude transitions which involve cross-state residential moves. So, $\beta_1$ identifies the expected return to a within-state match against the counterfactual of remaining in the same job. This analysis necessarily excludes individuals out of employment in $t-1$. But, it is worth emphasizing that education differentials in cross-state matching among the initially employed closely approximate those of the full population: compare columns 5 and 6 of Table 2.

I have no interest here in identifying a causal effect of an “exogenous” job change. Rather, the model makes predictions on the conditional mean wage change - and this is the moment that equation (18) identifies. Of course, this conditional mean is driven by selection on job offers, but it is precisely this selection which interests me. In particular, better educated workers should expect larger job surpluses if they face larger offer dispersion $\sigma^\epsilon$. To test this, I interact the $NewJob_{it}$ dummy with a set of education effects.

There is already a literature which estimates wage returns to job mobility, using similar specifications to (18). These wage returns are known to be increasing in education and decreasing in age: see Bartel and Borjas (1981), Mincer (1986), Topel and Ward (1992), Manning (2003a) and Gottfries and Teulings (2016). To my knowledge, my paper is the first to link these differential returns to geographical mobility.

5.2 Empirical estimates

I present estimates of (18) in Table 3, based on the SIPP sample described above. I identify $w_{it}$ with log hourly wages at the end of each four-month wave $t$, and I restrict the sample to jobs commanding at least 15 hours per week. Column 1 shows that the mean wage return to a new local job is 0.04. In the next column, I interact the $NewJob_{it}$ dummy with a set of education effects (all of which are included in the demographic controls). There is a steep education gradient, ranging from 0.01 for high school dropouts to 0.07 for postgraduate degree-holders. The interaction effects are precisely estimated, with standard errors of around 0.01.

As the model predicts, these wage returns appear to be larger for workers with lower initial match quality - which I proxy with experience and initial job tenure. Columns 3 to 5 show the education differentials are largest for the young: among those with under 10 years of experience, $\beta_1$ reaches 0.09 for postgraduates. Similarly, the differentials are largest for those with lower initial tenure.

In the model, I have assumed that match quality $\epsilon$ is fixed over the duration of a job. But to the extent that the wage return for better educated workers is delayed (due to seniority returns or the accumulation of firm-specific human capital), the $\beta_1$ coefficient will understate their true
### Table 3: Wage returns to within-state job matching

<table>
<thead>
<tr>
<th></th>
<th>All individuals</th>
<th>Experience groups</th>
<th>Initial tenure (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3) (4) (5)</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
</tr>
<tr>
<td>New job (NJ)</td>
<td>0.036***</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.012*</td>
<td>(0.006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.026***</td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.028***</td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.068***</td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.070***</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>Demog controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sample</td>
<td>657,610</td>
<td>190,965</td>
<td>220,762</td>
</tr>
<tr>
<td>% New job</td>
<td>5.785</td>
<td>8.444</td>
<td>8.838</td>
</tr>
</tbody>
</table>

This table reports estimates of (18), based on four-month transitions in SIPP panels beginning 1996, 2001, 2004 and 2008, for observations with no cross-state migration. Throughout, I control for a full set of wave effects and a detailed set of demographic characteristics, specifically: experience and experience squared; four education indicators (high school graduate, some college, undergraduate and postgraduate), each interacted with a quadratic in experience; black and Hispanic race dummies; foreign-born and native-born dummies (the omitted category contains the 7 percent of respondents who do not answer the relevant survey module); and a gender indicator which is also interacted with all previously mentioned variables. I use hourly wage data for workers paid by the hour, and I impute hourly wages for salaried workers using monthly earnings and hours. I exclude workers with multiple jobs or business income at the end of a wave, and I exclude wage observations below the 1st or above the 99th percentiles (within SIPP panels and education groups). The sample is restricted to individuals with 1-30 years of experience and working at least 15 hours per week, and it excludes wage transitions with intervening unemployment or layoff spells. Errors are clustered by individual, and robust SEs are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

5.3 Quantifying the effect of wage returns on geographical mobility

I now consider whether these estimates of high-educated wage returns (in local matches) are sufficiently large to explain the education differentials in cross-state matching rates. Such an exercise necessitates assumptions on the distribution of moving costs. But it can nevertheless offer some useful guidance.

Suppose the draws of $\mu \sim F^\mu$ are distributed uniformly between 0 and a maximum normalized to 1; so moving costs $m$ range from 0 to $\sigma^\mu$. And suppose also there are no wage offers which can justify moving at the maximum cost draw: that is, for every initial match quality $\epsilon$ and for every offer $\epsilon' \sim F^\epsilon$, $V(\epsilon') - V(\epsilon) < \sigma^\mu$. Using (6), (9) and (10), I show in Appendix...
Table 4: Quantifying the effect of wage returns

<table>
<thead>
<tr>
<th>Differences with high school</th>
<th>Coll grad v non-grad</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Some coll (1) Undergrad (2) Postgrad (3) difference (4)</td>
</tr>
<tr>
<td>A. Cross-state matching: log ( \bar{\rho}_C )</td>
<td>0.281 0.759 1.032 0.709</td>
</tr>
<tr>
<td>B. Within-state matching: log ( \bar{\rho}_L )</td>
<td>-0.052 -0.290 -0.582 -0.352</td>
</tr>
<tr>
<td>C. Mean returns to local search: log ( \bar{\rho}_L \beta_1 )</td>
<td>0.293 0.893 0.659 0.670</td>
</tr>
<tr>
<td>D. Discounted returns: log ( \bar{\rho}_L \beta_1 + \delta + \bar{\rho}_L + \bar{\rho}_C )</td>
<td>0.426 1.341 1.403 1.142</td>
</tr>
<tr>
<td>E. Standard errors on log ( \beta_1 )</td>
<td>(0.250) (0.229) (0.294) (0.164)</td>
</tr>
</tbody>
</table>

Each row of this table reports education differentials in the variable recorded on the left. Row A reports differentials in the mean cross-state matching rate, log \( \bar{\rho}_C \), among those employed in \( t-1 \). In columns 1-3, these are differences in log \( \bar{\rho}_C \) between the recorded education level and the omitted category, high school. The final column reports the difference in log \( \bar{\rho}_C \) between college graduates (i.e. undergraduate and postgraduate degree-holders) and non-graduates. Row B does the same for the mean within-state matching rate, log \( \bar{\rho}_L \). Row C reports education differentials in log \( \bar{\rho}_L \beta_1 \). I estimate the mean local wage returns \( \beta_1 \) using equivalent regressions to those of column 2 of Table 3, but with high school education (for columns 1-3) and non-graduate (for column 4) respectively as omitted categories. For row D, I calibrate \( \delta \) using the separation rate to non-employment, as in column 9 of Table 2; and I set the four-monthly interest rate \( r \) at 0.015. The final row offers approximate standard errors on the education differentials in the log of \( \beta_1 \), based on the delta method.

I now study whether differences in job surplus can account for the disparities in \( \bar{\rho}_C \) across education groups in (19), without resorting to \( \sigma^H \). The first row of Table 4 reports education differentials in log \( \bar{\rho}_C \) for the initially employed. In the first three columns, these are differences in log \( \bar{\rho}_C \) between the recorded education level and the omitted category, high school (I have combined high school dropouts and graduates, to reduce standard errors). The final column reports the difference in log \( \bar{\rho}_C \) between college graduates (i.e. undergraduate and postgraduate degree-holders) and non-graduates.

Row B reports the same data for the within-state matching rate, log \( \bar{\rho}_L \). Just as in Table 2, this is decreasing in education - unlike log \( \bar{\rho}_C \). I now successively account for the remaining components of (19), to check whether they can account for the log \( \bar{\rho}_C \) differentials. In row C, I add the log of \( \beta_1 \), the mean local wage returns (see table notes for computation details). The education differentials are now clearly positive and are close to the log \( \bar{\rho}_C \) in magnitude. The postgraduate differential is somewhat smaller than the undergraduate, though the standard
errors (in the bottom row) are too large to make a definitive statement on this point. When I collapse the number of groups to two (in the final column), the differential in log \((\hat{\rho}_L \beta_1)\) closely matches that of log \(\hat{\rho}_C\).

In row D, I account for education differentials in the discount rates, i.e. \(r + \delta + \hat{\rho}_L + \hat{\rho}_C\). I calibrate \(\delta\) using the separation rate to non-employment (based on data from Table 2, column 9), and I set the four-monthly interest rate \(r\) to 0.015. Better educated workers face lower job-to-job finding rates and separation rates to non-employment, so they discount their surplus less heavily. As a result, the education differentials are now even larger. Looking at column 4, I now over-explain the \(\hat{\rho}_C\) differential. Note however that the discount rate is only relevant for geographical mobility to the extent that moving costs are fixed, as in the baseline model above. If mobility is instead constrained by heterogeneous local preferences (as in Section 3.6), workers will not be deterred by a high separation rate: moving entails no sunk cost.

To summarize, the education differences in job surplus can indeed account for the gap in geographical mobility - without resorting to costs. Of course, these results are entirely contingent on my assumptions on moving costs\(^24\) - though at least they can offer a useful guide. Also, it is worth stressing that the \(\frac{\pi}{\sigma^\varepsilon}\) term in (19) accounts not only for moving costs, but also for long-distance search (or recruitment) intensity \(\pi\). As I have argued in Section 3.6, education differentials in \(\pi\) are themselves a plausible (endogenous) outcome of differences in offer dispersion \(\sigma^\varepsilon\). So even if there were systematic education differences in this imputed measure, one need not attribute this to moving costs \(\sigma^\mu\).

6 Wage returns to cross-state job matching

To test my hypothesis more directly, I now exploit information on the wage returns to cross-state matching. This analysis requires much weaker distributional assumptions. As I show below, the cross-state returns depend on the relative importance of selection on wage and moving cost draws. To the extent that workers select into migration because of large job surplus (and despite large costs), this will be manifested in larger returns. This insight allows me to identify bounds on the expected realized costs of moving. My analysis here builds on earlier studies which estimate the wage returns to geographical mobility (such as Lkhagvasuren, 2014, or Huttunen, Moen and Salvanes, 2018), but I implement an empirical specification and offer an interpretation based on my particular theoretical model.

\(^{24}\)For example, suppose the uniform distribution understates the true density of high cost draws. Then, I would be understating the impact of the \(\beta_1\) differentials on mobility: as I note in the discussion on migration distance in Section 3.6, mobility is more elastic to job surplus if costs are higher.
6.1 Annuitized moving costs

A useful concept in this exercise is the “annuitized cost” of a long-distance move: this can aid the interpretation of the empirical estimates which follow. Consider a worker with initial match quality \( \epsilon \) who receives a long-distance offer with a moving cost draw equal to \( m = \sigma \mu \). The (equivalent variation) annuitized cost, which I denote \( \tilde{m} \), is the wage increase (over the worker’s current wage) which would make him indifferent to accepting the offer. This is the \( \tilde{m} \) which satisfies:

\[
V \left( \epsilon + \frac{\tilde{m}}{\sigma \epsilon} \right) - V (\epsilon) = \sigma^\epsilon \Omega \left( \frac{\tilde{m}}{\sigma^\epsilon \mu} \right) = \sigma^\mu \mu
\]

where the first equality follows from (8), and the second describes the indifference relationship. \( \Omega \) is the job surplus in \( \epsilon \) units (associated with an increase of match quality of \( \frac{\tilde{m}}{\sigma \epsilon} \)), as defined by (9). Rearranging this expression for \( \tilde{m} \) yields:

\[
\tilde{m} (\mu | \epsilon) = \sigma^\epsilon \Omega^{-1} \left( \frac{\sigma^\mu \mu}{\sigma^\epsilon \mu | \epsilon} \right)
\]

This is equivalent to the concept of “mobility-compatible indifference wages” in Schmutz and Sidibé (forthcoming), and it offers a theoretical basis for the subjective annuitized costs of Section 2.2 above. Note that \( \Omega \) can be inverted because it is monotonically increasing: see equation (9). \( \Omega^{-1} \) is an implicit function; but to aid intuition, consider a first order approximation of \( \Omega \) in (9) around the initial match quality \( \epsilon \). Substituting this approximation into (21) gives:

\[
\tilde{m} (\mu | \epsilon) \approx \sigma^\mu \mu [ r + \delta + \rho_L (\epsilon) + \rho_C (\epsilon) ]
\]

where the fixed moving cost \( \sigma^\mu \mu \) is discounted by the interest rate \( r \) and the total separation rate. The annuitized cost \( \tilde{m} \) is increasing in the cost draw \( \mu \) and decreasing in match quality \( \epsilon \). Intuitively, workers with lower initial \( \epsilon \) will expect many suitable offers, so they are more likely to reject matches with high cost draws.

6.2 Selection on wage and cost draws

I now turn to the selection of long-distance movers. I begin by deriving the theoretical distribution of realized moving costs (and specifically of annuitized costs). Conditional on accepting a cross-area match, the probability of having drawn an annuitized moving cost exceeding \( \tilde{m} \) is:

\[
1 - Z (\tilde{m} | \epsilon) = \frac{\int_{\tilde{m}}^{\infty} \left[ 1 - F^\epsilon (\epsilon + \frac{x}{\sigma^\epsilon}) \right] f^\mu \left( \frac{\sigma^\epsilon}{\sigma^\mu \mu} \Omega \left( \frac{x}{\sigma^\mu | \epsilon} \right) \right] dx}{\int_{0}^{\infty} \left[ 1 - F^\epsilon (\epsilon + \frac{x}{\sigma^\epsilon}) \right] f^\mu \left( \frac{\sigma^\epsilon}{\sigma^\mu \mu} \Omega \left( \frac{x}{\sigma^\mu | \epsilon} \right) \right] dx}
\]

The numerator describes the probability of drawing and accepting a cross-area offer with annuitized cost exceeding \( \tilde{m} \), and the denominator describes the probability of accepting any cross-area offer. I make the following claim:
Proposition 4. For sufficiently large moving costs (i.e. sufficiently large $\sigma^\mu$), and given a worker’s initial match quality $\epsilon$, the expected realized annuitized cost is increasing in both $\sigma^\mu$ and $\sigma^\epsilon$.

This follows from my assumptions on the offer distribution (monotone hazard) and cost distribution (decreasing elasticity of density), and I leave the proof to Appendix I.4. Intuitively, a larger $\sigma^\mu$ implies larger unconditional cost draws, so realized costs will also be larger. And a larger $\sigma^\epsilon$ expands job surpluses, so workers are more likely to accept cross-area offers with high cost draws. The condition of a sufficiently large $\sigma^\mu$ ensures stability in the relationship between moving costs and their annuitized values in (21): again, see Appendix I.4.

If these realized costs could be observed, this would offer a useful test on the origins of the mobility gap. If high-educated mobility is driven by low costs (i.e. low $\sigma^\mu$), realized costs should be decreasing in education. But if it is driven by large offer dispersion (large $\sigma^\epsilon$) and job surplus, realized costs should be increasing. Intuitively, in the latter case, better educated workers would be moving because of large surplus - and despite the associated costs.

Clearly, these costs are unobserved. But the model does offer a way to set identify the expectation of realized annuitized costs:

Proposition 5. (i) The expected wage return to cross-area job matching identifies an upper bound on the expectation of realized annuitized costs. (ii) The differential between the expected returns to cross-area and local matches identifies a lower bound on the expected realized costs.

The intuition for the upper bound is simple. Let $E_C[\epsilon' - \epsilon|\epsilon]$ denote the expected gain in match quality on accepting a cross-area (subscript C) offer, for given initial match quality $\epsilon$. The associated wage return can then be expressed as:

$$\sigma^\epsilon E_C[\epsilon' - \epsilon|\epsilon] = \int_0^\infty E_L[\sigma^\epsilon (\epsilon' - \epsilon)|\sigma^\epsilon (\epsilon' - \epsilon) \geq \tilde{m}] dZ(\tilde{m}|\epsilon)$$  \hspace{1cm} (24)

where the $E_L$ term is the expected return to local matches, conditional on the return exceeding the annuitized moving cost $\tilde{m}$. To derive the expected return to cross-area matches, I integrate this expression over the distribution of realized annuitized costs $Z$. Since the $E_L$ term must exceed $\tilde{m}$, the expected return must exceed the expected annuitized costs $\int_0^\infty \tilde{m}dZ(\tilde{m}|\epsilon)$ associated with those matches. Intuitively, workers will only accept cross-area offers if the associated job surplus exceeds the cost of moving.

The lower bound is identified by the differential between the expected wage return to cross-area and local job matching:

$$\sigma^\epsilon E_C[\epsilon' - \epsilon|\epsilon] - \sigma^\epsilon E_L[\epsilon' - \epsilon|\epsilon]$$  \hspace{1cm} (25)

$$= \int_0^\infty \{E_L[\sigma^\epsilon (\epsilon' - \epsilon)|\sigma^\epsilon (\epsilon' - \epsilon) \geq \tilde{m}] - E_L[\sigma^\epsilon (\epsilon' - \epsilon) |\epsilon' - \epsilon \geq 0]\} dZ(\tilde{m}|\epsilon)$$

Given my assumption that the offer distribution $F^\epsilon$ has a monotonically increasing hazard rate,
the term in curly brackets must be less or equal to $\tilde{m}$ for all annuitized cost draws $\tilde{m} \geq 0$. See Appendix I.5 for a proof. And consequently, the cross-area/local differential in expected wage returns will identify a lower bound on the expected realized annuitized costs, $\int_{0}^{\infty} \tilde{m} dZ(\tilde{m}|\epsilon)$. Intuitively, the moving cost is not always binding: there are some local offers which would be sufficient to justify a cross-area match. And so, the differential in job surplus should understate the magnitude of realized costs. This is effectively a compensating differentials argument, but accounting for search frictions: workers cannot choose from the universe of jobs.

### 6.3 Empirical estimates

Motivated by Propositions 4 and 5, I now offer estimates of the wage returns to cross-state matching. Based on the sample of wage changes within employment cycles which involve job transitions (i.e. conditional on $NewJob_{it} = 1$), I estimate the following empirical specification:

$$\Delta \log w_{it} = \theta_0 + \theta_1 Move_{it} + \theta'_X X_{it} + \theta_t + v_{it} (26)$$

where $Move_{it}$ is a dummy variable taking 1 if the individual moved state between $t-1$ and $t$. Based on Proposition 5, a lower bound on the expected annuitized cost of movers can be identified by the coefficient $\theta_1$ (the premium to cross-state relative to local matching). The upper bound is $\theta_1 + \beta_1$, where $\beta_1$ is the return to local matching in (18): this sum describes the total wage return to a long-distance match. Given the specification is in log wages, these bounds will approximate the expected costs as a fraction of initial wages; and they will also be comparable to the subjective cost estimates in Figure 4. To study how the coefficients vary with education, I interact $Move_{it}$ with education effects. Given the small sample of high school dropout movers, I aggregate high school graduates and dropouts into a single category.

I present estimates of (26) in Table 5. My basic estimate of $\theta_1$ in column 1 is 0.057, which identifies the lower bound on expected annuitized costs (as a fraction of a worker’s initial wage). The upper bound is $\theta_1 + \beta_1 = 0.093$, where $\beta_1$ is taken from column 1 of Table 3.

Column 2 shows the cross-state return is almost entirely driven by college graduates. For high school workers, both $\theta_1$ (in Table 5) and $\beta_1$ (Table 3) are close to zero and statistically insignificant - which suggests they typically move with negligible costs. In contrast, the costs are bound by 0.09 and 0.16 for undegraduate degree-holders, and by 0.16 and 0.23 for post-graduates.25

Of course, these returns to cross-state matching may simply reflect local differentials in human capital premia (i.e. $\gamma_j$ in equation (1); see Lkhagvasuren, 2014), rather than the surplus accruing to progression up a jobs ladder. To study this further, I purge log wages of state-specific human capital premia. Specifically, I regress log wage levels on a full set of state effects (using the full wage sample), each interacted with a full set of education indicators (five

---

25 These numbers are based on column 2 of Table 3 and column 2 of Table 5.
Table 5: Wage returns to cross-state job matching

<table>
<thead>
<tr>
<th>Move</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move * High school</td>
<td>-0.050</td>
<td>-0.047</td>
<td>-0.045</td>
<td>-0.025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.031)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Move * Some coll</td>
<td>0.038</td>
<td>0.038</td>
<td>0.038</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(0.038)</td>
<td>(0.038)</td>
<td></td>
</tr>
<tr>
<td>Move * Undergrad</td>
<td>0.087**</td>
<td>0.093**</td>
<td>0.112**</td>
<td>0.074*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.043)</td>
<td>(0.046)</td>
<td>(0.041)</td>
<td></td>
</tr>
<tr>
<td>Move * Postgrad</td>
<td>0.164***</td>
<td>0.180***</td>
<td>0.176***</td>
<td>0.175***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.059)</td>
<td>(0.055)</td>
<td>(0.055)</td>
<td></td>
</tr>
</tbody>
</table>

Demog controls: Yes, Yes, Yes, Yes, Yes
Purged of state*HC: No, No, Yes, Yes, Yes
Purged of state*occ: No, No, No, 2-digit, 3-digit
Sample: 34,580, 34,580, 34,580, 34,580, 34,580
% Move: 2.397, 2.397, 2.397, 2.397, 2.397

This table reports estimates of (26). The sample is identical to Table 3, except I now restrict it to individuals who change job (NewJob = 1). See notes under Table 3 for details on the demographic controls. In column 3, I use individual wages which are purged of triple interactions between state effects, education effects and a quadratic in experience. For column 4, I purge individual wages additionally of triple interactions between state effects and 2-digit occupation effects; column 5 repeats the same exercise for 3-digit occupations. Errors are clustered by individual, and robust SEs are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

categories), and all preceding variables interacted with a quadratic in experience. I then use individual-level changes in the regression residuals as the dependent variable for (26). As it happens though, this actually slightly raises the $\theta_1$ effects for college graduates: see column 3.

Now, there may also be selection on local returns to particular abilities within education-experience groups: for example, software engineers may be better rewarded in California, and bankers in New York. To address this, I purge individual wages (again, using the full sample) additionally of triple interactions between state fixed effects and 2-digit occupation fixed effects (between 87 and 99 categories, depending on SIPP panel). Looking at column 4, this makes little difference to the estimates. And even when I use a very demanding specification with 3-digit occupations (481-496 categories) in column 5, the results are similar. Certainly, it is not possible to rule out selection on returns to unobserved components of human capital. But the evidence on observables casts doubt on their importance.

To summarize, the estimates here point to much larger realized costs among better educated movers. According to Proposition 4, these larger costs must be driven by larger offer dispersion $\sigma^e$ or larger moving costs $\sigma^\mu$. In other words, better educated workers typically move because of large job surplus and despite large costs; whereas low educated workers move because of low cost draws and despite meager surplus. This is consistent with the subjective reasons for moving reported in Figure 2. The results here also suggest that the mobility differentials
6.4 Comparison with existing estimates of moving costs

Are the moving costs implied by this exercise quantitatively plausible? Reassuringly, they are similar in magnitude to the subjective costs from Section 2. Conditional on being “willing to move” for work, the mean annuitized cost in my PSID sample is 0.37 - relative to workers’ wages. This is somewhat larger than the 0.23 upper bound implied by the SIPP wage returns for postgraduate degree-holders (who face the largest realized costs), but this should be expected: the PSID estimates condition on “willingness” to move, whereas the SIPP estimates condition on actually moving. So the latter should be selected from lower down the costs distribution.

I also show in Appendix G how my estimates of realized annuitized costs, i.e. \( \tilde{m}(\mu|\epsilon) \) in (22), can be converted to fixed one-off equivalents, i.e. \( m = \sigma^\mu \mu \). Based on the numbers in Table 5 and calibrated values of \( r, \delta, \rho_L \) and \( \rho_C \), I compute mean one-off costs of about $50,000 for postgraduate movers and $20,000 for undergraduate movers. The realized costs are negligible for lower educated workers.

How do these compare with the literature? Existing estimates vary substantially, partly because they identify different objects. Most studies do not allow for individual heterogeneity in costs, which rules out the selection effects described above: Davies, Greenwood and Li (2001) estimate a cross-state moving cost of $200,000, Bayer and Juessen (2012) estimate $34,000, Lkhagvasuren (2014) finds a cost of $28,000 to $54,000 between census divisions, and Schmutz and Sidibé (forthcoming) €15,000 between French metro areas. Allowing for individual heterogeneity, Kennan and Walker (2011) estimate a large (unconditional) average cross-state cost of $312,000; but in their high school graduate sample, the cost for actual movers is typically negative. This is consistent with my claim that low educated movers are typically selected on low costs, rather than large wage gains.

7 Conclusion

In this paper, I have argued that better educated workers are more mobile because they face larger dispersion in wage offers. In a frictional labor market, this generates larger match surplus as workers climb the jobs ladder, irrespective of geography. This makes the labor market better integrated spatially, particularly for younger workers who are just beginning their careers. While surplus is unimportant for local matching, it plays a crucial role in driving long-distance matching - given the associated moving costs. In a more complete model, this effect is amplified by both workers and firms investing in long-distance search and recruitment networks - which raise the long-distance contact rate.

Larger surplus not only drives more long-distance mobility in the steady-state, but it also makes migratory flows more responsive to local employment shocks. This can help explain why
the low educated face much more persistence in local jobless rates. Though I focus on education differentials, this paper offers new insights for understanding geographical immobility more generally. And though I apply these insights to internal mobility, it also offers a rationale for the surprising degree of positive selection among international migrants.

My hypothesis is attractive firstly because it is theoretically intuitive. Larger offer dispersion may be motivated by a notion of specialized skills or supermodularity between workers’ abilities and job attributes (such as task complexity or firm quality). And second, it has strong empirical foundations. Patterns in within and cross-state matching are consistent with better educated workers facing larger offer dispersion. And crucially, I estimate large education differentials in the wage returns to within-state matching. Under certain assumptions on the distribution of moving costs, these differentials can quantitatively account for the bulk of the mobility gap - without resorting to competing explanations.

Nevertheless, I do consider the competing explanations more directly. The first is that better educated workers face lower moving costs. This view has endured in the literature because it can account for differential migratory responses to local shocks. However, moving costs imputed from subjective willingness to move (which predict future mobility) are remarkably similar across education groups. I also show that wage returns are much larger for better educated workers in cross-state job matches. This suggests they typically select into migration because of large job surplus and despite steep moving costs. In contrast, among the low-educated, returns are similarly small for both local and cross-state matches.

The second possibility is that better educated workers face larger local differentials in expected utility, driven perhaps by differential local returns to human capital. But this view is difficult to reconcile with evidence on the direction of migratory flows: high educated mobility is not driven by large net flows to particular states, even within detailed occupation-defined labor markets. Moreover, I show that the large wage returns to cross-state matches (among college graduates) cannot be explained by differential local human capital premia - even for detailed occupation-defined tasks.

References


Contents

A CPS sample description and supplementary estimates 39
  A.1 Sample description 39
  A.2 Historical changes in mobility differentials 40
  A.3 Breakdown of migration by reported reasons for moving 40
  A.4 Robustness to top earner restriction 44
B Contribution of returning students 45
C Predictive power of imputed costs 46
D Robustness of net migration patterns 48
E Tenure as a proxy for job match quality 48
F Effect of tenure and experience on cross-state matching 50
G Imputing one-off migration costs from wage returns 52
H Distribution of match quality and mean matching rates 52
I Theoretical proofs and derivations 54
  I.1 Proof of Proposition 2 in Section 3.4 54
  I.2 Proof of Proposition 3 in Section 3.4 55
  I.3 Derivation of equation (19) in Section 5.3 56
  I.4 Proof of Proposition 4 in Section 6.2 57
  I.5 Proof of Proposition 5 in Section 6.2 59
J Appendix references 59

A CPS sample description and supplementary estimates

A.1 Sample description

My Current Population Survey (CPS) estimates in the main text are based on March waves between 1999 and 2018, taken from the IPUMS database (Flood et al., 2018). The March waves
include the Annual Social and Economic Supplement, which reports whether respondents lived in a different state 12 months previously. Since 1999, individuals have also given their primary reason for moving.

I consider five education groups: high school dropouts (less than 12 years of schooling), high school graduates (12 years), some college (less than an undergraduate degree), and undergraduate and postgraduate degree holders. Potential labor market experience is defined as age minus years of education minus 6 (or age minus 16, whichever is smaller). I set years of schooling to 13 for individuals with some college but no degree, 14 for associate degrees, 16 for undergraduate, 18 for Master’s, 19 for professional and 21 for doctorate degrees. I restrict the sample to individuals with 2-30 years of experience at the survey date: this excludes people with less than one year of experience at the time of moving. I also restrict attention to individuals living in the US one year previously.

Kaplan and Schulhofer-Wohl (2012) show there are inconsistencies in the CPS’s procedure for imputing migration status in non-response cases: the imputed data artificially inflate the cross-state migration rate between 1999 and 2005. As it happens, the non-response rate for migration status varies little with education: 13 percent for college graduates and 14 percent for non-graduates. I choose to drop all these observations.

### A.2 Historical changes in mobility differentials

The CPS analysis in the main text is restricted to the period 1999-2018, for which I have information on reasons for moving. But the mobility gap between education groups goes back many decades. In Figure A1, I plot annual cross-state migration rates using CPS March waves from 1964 to 2018.\(^{26}\)

As is well known, migration rates have declined over this period: see e.g. Molloy, Smith and Wozniak (2011). Kaplan and Schulhofer-Wohl (2017) argue this was driven by the declining geographical specificity of occupational returns, coupled with improvements in communications technology. Molloy, Smith and Wozniak (2017) attribute it to a declining rate of labor market transitions. Either way, the decline was fairly uniform across education groups. The ratio of graduate to non-graduate mobility has mostly hovered around 1.4, and there is no clear upward or downward trend over the period as a whole.

### A.3 Breakdown of migration by reported reasons for moving

In Table A1, I present detailed disaggregations of cross-state and cross-county migration in the CPS by reported reason for moving. The first column gives the percentage of the full sample who changed state (in the previous 12 months) for each recorded reason, and the second column expresses these numbers as a percentage of cross-state migrants. The final two columns repeat

---

\(^{26}\)I omit 1995 because the relevant migration question was not asked that year.
this exercise for cross-county moves: these consist of both moves across states and across counties within states.

The bottom row shows that, each year, 2.4 percent of the sample move across states and 5.4 percent across counties. About half of cross-state moves are motivated by a specific job, compared with a third of cross-county moves. These are mostly due to a job change or transfer, but some workers also report commuting reasons. The commuting motivation can be interpreted in the context of a long-distance match: after accepting a distant job offer (with a long associated commute), the worker eventually changes residence. In contrast, it is rare to move to look for work without a job lined up. This sort of speculative job search accounts for just 5 percent of cross-state and 4 percent of within-state moves. This is unsurprising: moving without a job in hand is a costly and risky strategy. In terms of non-job migration, family and housing motivations account for most moves.

In Table A2, I report the cross-state and cross-county migration rates (in columns 1 and 3 of Table A1) separately by education group: high school dropouts (HSD), high school graduates (HSG), some college (SC), undergraduate degree (UG) and postgraduate (PG). As before, the first row reports the rate of job-motivated migration. Notice the (positive) education slope is steeper in proportional terms for cross-state than cross-county moves. I offer a rationale for this result in Section 3.6: to the extent that cross-state migration is more costly, education differences in offer dispersion and job surplus should matter more. Also, consistent with Figure 2 in the main text, better educated individuals make fewer speculative moves to “look for work”.

On aggregate, there is also a mild negative education gradient in non-job migration, which is stronger for cross-county moves. This effect is driven by a broad range of motivations: mostly to “establish own household”, “other family reasons”, “cheaper housing”, “other hous-
Table A1: Breakdown of primary reasons for moving

<table>
<thead>
<tr>
<th>Primary reason</th>
<th>State moves</th>
<th>Country moves</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% full sample</td>
<td>% state migrants</td>
</tr>
<tr>
<td>DUE TO SPECIFIC JOB</td>
<td>1.11</td>
<td>47.02</td>
</tr>
<tr>
<td>New job or job transfer</td>
<td>0.94</td>
<td>39.82</td>
</tr>
<tr>
<td>Easier commute</td>
<td>0.05</td>
<td>2.30</td>
</tr>
<tr>
<td>Other job reasons</td>
<td>0.12</td>
<td>4.91</td>
</tr>
<tr>
<td>LOOK FOR WORK</td>
<td>0.13</td>
<td>5.48</td>
</tr>
<tr>
<td>NON-JOB REASONS</td>
<td>1.12</td>
<td>47.49</td>
</tr>
<tr>
<td>Family</td>
<td>0.55</td>
<td>23.48</td>
</tr>
<tr>
<td>Change in marital status</td>
<td>0.10</td>
<td>4.42</td>
</tr>
<tr>
<td>Establish own household</td>
<td>0.09</td>
<td>3.79</td>
</tr>
<tr>
<td>Other family reasons</td>
<td>0.36</td>
<td>15.28</td>
</tr>
<tr>
<td>Housing</td>
<td>0.26</td>
<td>10.86</td>
</tr>
<tr>
<td>Want to own home</td>
<td>0.04</td>
<td>1.69</td>
</tr>
<tr>
<td>New or better housing</td>
<td>0.06</td>
<td>2.56</td>
</tr>
<tr>
<td>Cheaper housing</td>
<td>0.06</td>
<td>2.56</td>
</tr>
<tr>
<td>Other housing reasons</td>
<td>0.10</td>
<td>4.05</td>
</tr>
<tr>
<td>Environment</td>
<td>0.13</td>
<td>5.01</td>
</tr>
<tr>
<td>Better neighborhood</td>
<td>0.04</td>
<td>1.50</td>
</tr>
<tr>
<td>Climate, health, retirement</td>
<td>0.08</td>
<td>3.51</td>
</tr>
<tr>
<td>Attend/leave college</td>
<td>0.10</td>
<td>4.39</td>
</tr>
<tr>
<td>Other reasons</td>
<td>0.09</td>
<td>3.75</td>
</tr>
<tr>
<td>ALL REASONS</td>
<td>2.35</td>
<td>100</td>
</tr>
</tbody>
</table>

This table presents migration rates by primary reason in CPS March waves between 1999 and 2018. The first column reports the percentage of the full sample who changed state, for each given reason, over the previous twelve months. The second column expresses these numbers as a percentage of state-movers. The final two columns repeat the exercise for cross-county moves. I include individuals moving because of foreclosure or eviction in the CPS’s “other housing reasons” category; and I include individuals moving because of natural disasters in the “other reasons” category. See Appendix A.1 for sample details.
Table A2: Education gradients by primary reasons for moving

<table>
<thead>
<tr>
<th>Primary reason</th>
<th>State moves</th>
<th></th>
<th>County moves</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HSD</td>
<td>HSG</td>
<td>SC</td>
<td>UG</td>
</tr>
<tr>
<td>DUE TO SPECIFIC JOB</td>
<td>0.43</td>
<td>0.74</td>
<td>0.92</td>
<td>1.63</td>
</tr>
<tr>
<td>New job or job transfer</td>
<td>0.33</td>
<td>0.59</td>
<td>0.77</td>
<td>1.42</td>
</tr>
<tr>
<td>Easier commute</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Other job reasons</td>
<td>0.06</td>
<td>0.10</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>LOOK FOR WORK</td>
<td>0.21</td>
<td>0.15</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>NON-JOB REASONS</td>
<td>1.15</td>
<td>1.18</td>
<td>1.16</td>
<td>1.11</td>
</tr>
<tr>
<td>Family</td>
<td>0.63</td>
<td>0.64</td>
<td>0.58</td>
<td>0.46</td>
</tr>
<tr>
<td>Change in marital status</td>
<td>0.08</td>
<td>0.12</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>Establish own household</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>Other family reasons</td>
<td>0.44</td>
<td>0.42</td>
<td>0.37</td>
<td>0.29</td>
</tr>
<tr>
<td>Housing</td>
<td>0.29</td>
<td>0.27</td>
<td>0.23</td>
<td>0.27</td>
</tr>
<tr>
<td>Want to own home</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>New or better housing</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Cheaper housing</td>
<td>0.08</td>
<td>0.08</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Other housing reasons</td>
<td>0.10</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>Environment</td>
<td>0.16</td>
<td>0.15</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>Better neighborhood</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Climate, health, retirement</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Attend/leave college</td>
<td>0.00</td>
<td>0.05</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>Other reasons</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>ALL REASONS</td>
<td>1.79</td>
<td>2.07</td>
<td>2.22</td>
<td>2.86</td>
</tr>
</tbody>
</table>

This table reports cross-state and cross-county migration rates by primary reason for moving (as in Table A1), but now disaggregated by education. I consider five education groups: high school dropouts (HSD), high school graduates (HSG), some college (SC), undergraduate degree (UG) and postgraduate (PG).
Table A3: Migration rates (%) for all individuals and household top earners

<table>
<thead>
<tr>
<th></th>
<th>Specific job</th>
<th>Look for work</th>
<th>Non-job</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All indiv</td>
<td>Top earners</td>
<td>All indiv</td>
</tr>
<tr>
<td>HS dropout</td>
<td>0.43</td>
<td>0.46</td>
<td>0.21</td>
</tr>
<tr>
<td>HS graduate</td>
<td>0.74</td>
<td>0.76</td>
<td>0.15</td>
</tr>
<tr>
<td>Some college</td>
<td>0.92</td>
<td>0.95</td>
<td>0.11</td>
</tr>
<tr>
<td>Undergrad</td>
<td>1.63</td>
<td>1.67</td>
<td>0.12</td>
</tr>
<tr>
<td>Postgrad</td>
<td>2.20</td>
<td>2.36</td>
<td>0.07</td>
</tr>
</tbody>
</table>

This table reports annual cross-state job migration rates by reported reason for moving, separately for all individuals (identical to Figure 1) and for household top earners, and based on CPS March waves between 1999 and 2018. See Appendix A.1 for further sample details.

ing reasons” and “better neighborhood”. There are just two non-job motivations with (largely) positive education slopes: the desire to purchase a home and attending or leaving college.

The final row reports total migration rates - for all motivations combined. Notice this is much flatter for cross-county migration. Mechanically, this reflects the flatter (positive) gradient of job-motivated migration, the steeper (negative) gradient of non-job migration, and the greater dominance of non-job motivations for cross-county movers.

A.4 Robustness to top earner restriction

Importantly, the CPS question on reasons for moving is addressed to individuals within households. But of course, migration decisions are made in the context of the household. This ambiguity may yield some problems for interpretation: for example, household dependents may choose to simply report the reasons of the breadwinners. This is most clearly illustrated for children (though they are excluded from my sample): in households with at least one adult moving for a specific job, 80 percent of under-16s also report moving for the same reason.

To address this concern, I recompute migration rates by reason for moving (and by education), but this time restricting the sample to those individuals with the greatest annual earnings in each household. In households with joint top-earners, I divide the person weights by the number of top-earners. This restriction excludes 44 percent of the original sample. But as Table A3 shows, it makes little difference to the education slopes of job-specific, speculative or non-job migration. The first, third and fifth columns replicate the cross-state migration rates from Table A2, and these look very similar to the remaining columns which impose the top-earner restriction.
This table reports annual cross-state migration rates by education group, based on all (annual) PSID waves between 1990 and 1997. Migration rates are constructed using reported state of residence 12 months previously. The first row gives the fraction of the sample who were recently students (in the current or previous wave). The second row reports cross-state migration rates for the full sample, and the third row reports these rates excluding recent students. The fourth and fifth rows disaggregate the migration rate (for the full sample) into return and non-return moves. Return moves include all moves to (i) states where the individual has resided previously in the panel or (ii) the state where the individual reports having grown up. The sample consists of all individuals with 2-10 years of potential labor market experience at the end of each 12-month interval.

### B Contribution of returning students

In this section, I check whether returning students may be contributing to education differentials in mobility. Table A2 shows that workers who report moving primarily to leave or attend college account for a negligible part of these differentials. But even if this is not the primary stated motivation, it may be an underlying factor for those who report job-related reasons - at least for the young. Indeed, Kennan and Walker (2011) emphasize that a large fraction of long-distance movers in the US are returning to former places of residence; and Kennan (2015) studies the tendency of individuals to return home after studying in another state.

The contribution of this return migration can be assessed in the Panel Study of Income Dynamics (PSID). Similarly to the CPS analysis, I define a migrant as somebody living in a different state 12 months previously. I restrict attention to individuals with 2-10 years of potential labor market experience at the end of each 12-month interval, in annual PSID waves between 1990 and 1997. I exclude waves after 1997 because these are biennial: it is not possible to track migration at annual frequencies. The first row of Table A4 reports the fraction of individuals in each education group who were recently students (either in the current or previous annual wave). This is largest for high school dropouts (18 percent) and smallest for individuals with undergraduate and postgraduate degrees (5 and 2 percent respectively).

The remaining rows report annual cross-state migration rates by education. The second
row computes these for the full sample, illustrating the familiar positive education gradient. I exclude recent students in the third row; and unsurprisingly, this makes little difference - given they comprise such a small fraction of the college graduate sample. Interestingly, the graduate migration rates are actually a little larger once recent students are excluded.

However, excluding recent students does not address the concerns entirely, because ex-students may yet return to their home state several years after completing their education. In the final two rows, I disaggregate the cross-state migration rate into “return” and “non-return” moves. Return moves consist of moves to any state where the individual has previously resided since the panel began (i.e. since 1968). The education gradient is clearly positive for both return and non-return rates, and the gradient is about twice as steep for the latter. This can be rationalized by the model: if non-return moves are more costly, the rate of non-return migration should be more elastic to job surplus (and hence to education). This is an equivalent idea to that of migration distance in Section 3.6. To summarize then, the evidence shows that returning students (and return migration in general) cannot account for the mobility gap.

C Predictive power of imputed costs

Given that the imputed costs in Section 2.2 are based on the subjective judgments of respondents, there may be doubts over their accuracy. But reassuringly, the cost measures do have significant predictive power for future migration decisions. Suppose the instantaneous cross-area matching rate for some individual \(i\) is constant within the time interval \(t-1\) to \(t\), and denote this matching rate as \(\rho_{C_{it}}\). The probability of moving within this interval is then:

\[
\Pr(Move_{it} = 1) = 1 - \exp(-\rho_{C_{it}}) \tag{A1}
\]

This motivates a complementary log-log model:

\[
\Pr(Move_{it} = 1) = 1 - \exp(-\exp(\beta_m \tilde{m}_{it-1} + \beta_X X_{it} + \beta_t)) \tag{A2}
\]

where I express \(\rho_{C_{it}}\) as a function of the initial annuitized moving cost \(\tilde{m}_{it-1}\) (as defined in (21)), human capital indicators\(^{27} X_{it}\), and a full set of year effects \(\beta_t\). The advantage of this specification is that \(\beta_m\) can intuitively be interpreted as the elasticity of the instantaneous migration rate with respect to \(\tilde{m}_{it-1}\). And assuming a constant hazard, this interpretation is independent of the time horizon associated with the migration variable.

I report my estimates in Table A5. I only have imputed costs for employed household heads between 1969 and 1973, so I restrict attention to the mobility decisions of these individuals in the annual intervals between 1969 and 1974. The first two columns report the elasticity of

\(^{27}\)Specifically: experience and experience squared; four education indicators (high school graduate, some college, undergraduate and postgraduate), each interacted with a quadratic in experience; black and Hispanic dummies; and a gender dummy interacted with all previously mentioned variables.
Table A5: Log point responses of cross-state migration to imputed costs

<table>
<thead>
<tr>
<th></th>
<th>Unconditional sample (1)</th>
<th>Conditional sample (willing to move) (2)</th>
<th>Conditional sample (willing to move) (3)</th>
<th>Conditional sample (willing to move) (4)</th>
<th>Conditional sample (willing to move) (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial willingness to move</td>
<td>0.987***</td>
<td>0.868***</td>
<td>-1.346***</td>
<td>-1.345***</td>
<td>-1.345***</td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td>(0.191)</td>
<td>(0.423)</td>
<td>(0.462)</td>
<td>(0.462)</td>
</tr>
<tr>
<td>Initial willingness to move * Grad</td>
<td>0.404</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.406)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial imputed cost</td>
<td>-0.206</td>
<td>-1.346***</td>
<td>-1.345***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.349)</td>
<td>(0.423)</td>
<td>(0.462)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial imputed cost * Grad</td>
<td></td>
<td></td>
<td></td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.992)</td>
<td></td>
</tr>
<tr>
<td>Initial log wage</td>
<td>-1.321***</td>
<td>-1.631***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.373)</td>
<td>(0.419)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial log wage * Grad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.353*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.796)</td>
</tr>
<tr>
<td>Demographic controls, year effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>8,836</td>
<td>8,836</td>
<td>2,904</td>
<td>2,904</td>
<td>2,904</td>
</tr>
<tr>
<td>Cross-state mig rate</td>
<td>0.035</td>
<td>0.035</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
</tr>
</tbody>
</table>

This table reports responses of cross-state migration (over 12-month intervals) to subjective costs and wages (at the beginning of each interval), based on complementary log-log regressions. I study the response to both a binary indicator of “willingness” to move for work; and conditional on being willing to move, the response to the imputed moving cost. These measures are described in greater detail in the notes under Figure 4. In columns 2 and 5, I also allow for interactions between the cost measures and a college graduate dummy. Coefficients should be interpreted as the log point effect of each measure on the instantaneous cross-state migration rate, conditional on the empirical model described by equation (A2). I only have imputed costs for employed household heads between 1969 and 1973, so I restrict attention to the mobility decisions of these individuals in the annual intervals between 1969 and 1974. I exclude individuals with less than 2 or more than 30 years of potential experience at the end of each interval. Household heads in the PSID are always male, unless there is no husband (or cohabiting partner) present or the husband is too ill to respond to the survey. All specifications control for a full set of year effects and demographic controls, specifically experience and experience squared, four education indicators (high school graduate, some college, undergraduate and postgraduate), each interacted with a quadratic in experience, black and Hispanic dummies, and a gender dummy interacted with all previously mentioned controls. Errors are clustered by individual, and robust SEs are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

cross-state migration (within each 12-month interval) to a binary indicator for “willingness to move” (at the beginning of each interval). Willingness to move adds about 100 log points to the cross-state migration rate, and an interaction with a college graduate dummy reveals no significant difference in the response by education.

In the final three columns, I restrict the sample to those who are “willing to move” and estimate elasticities to the imputed costs (again, at the beginning of each interval). The imputed cost is the difference between (i) the log of a worker’s stated reservation for accepting a long-distance offer and (ii) the log of the worker’s actual wage: see Section 2.2 in the main text for further details. Column 3 reports an elasticity of -0.2, but it is statistically insignificant. This estimate will presumably be attenuated by classical measurement error, but there is also
a more systematic problem. If the long-distance reservation is noisier than a worker’s wage, the imputed cost (the difference between the two) will be artificially negatively correlated with wages. But as the model shows, to the extent that wages reflect match quality, workers with higher wages will be less likely to move. This should bias the estimated effect of imputed costs towards zero. To address this problem, in column 4, I control additionally for the initial log wage. Both the imputed cost and the wage now take strong negative effects, with elasticities of -1.3 and standard errors of 0.4 in each case. In column 5, I allow for education heterogeneity in these effects, but the interactions are not statistically significant.

The key message here is that the subjective costs do have predictive power for future mobility - which suggests they are informative about the true costs of moving. This reinforces the validity of the claim in Section 2.2 that moving costs vary little with education.

D Robustness of net migration patterns

In Table A6, I reproduce the results in Table 1 in the main text - but this time separately for individuals with 2-10 and 11-30 years of potential labor market experience. At least for young college-educated individuals, there is a discernible positive effect of education on net migration rates. But this effect is relatively small: for all experience groups and occupation schemes, the ratio of net to gross migration is still decreasing in education. I conclude from this that the mobility differentials between education groups are not driven by large net flows to particular states, even within detailed occupation-defined markets and within distinct experience categories. This reinforces the general message in Section 2.3 in the main text.

One might also be concerned that gross mobility differentials are merely driven by churn: software engineers moving to California, and then returning home. To assess this possibility, I next study stocks of migrants, which are not conflated by such churn. In Table A7, I reproduce the analysis above, but now defining “migrants” as people living outside their birth state (rather than recent movers): this excludes return migrants. Looking at column 1, more than half of postgraduate degree-holders live outside their birth state, compared to a quarter of high school dropouts. Though local imbalances between worker “imports” and “exports” are increasing in education (column 2), there is little change relative to the gross stock of migrants (column 3). The same is true within occupation-defined markets. This suggests that churn and return migration cannot account for the gross mobility differentials: see also the analysis of returning students in Appendix B.

E Tenure as a proxy for job match quality

In this paper, I have used initial job tenure as a proxy for job match quality $\varepsilon$. The quality of this proxy is increasing in tenure: intuitively, for workers who have just begun their job, tenure
Table A6: Net cross-state migration rates by education and experience

<table>
<thead>
<tr>
<th></th>
<th>Gross mig rate (%)</th>
<th>Net mig rate (%)</th>
<th>Net-gross ratio</th>
<th>Within 2-digit occs</th>
<th>Gross mig rate (%)</th>
<th>Net mig rate (%)</th>
<th>Net-gross ratio</th>
<th>Within 3-digit occs</th>
<th>Gross mig rate (%)</th>
<th>Net mig rate (%)</th>
<th>Net-gross ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td></td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
</tr>
<tr>
<td>Individuals with 2-10 years of experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS dropout</td>
<td>3.37</td>
<td>0.45</td>
<td>0.13</td>
<td>3.24</td>
<td>1.55</td>
<td>0.48</td>
<td>3.24</td>
<td>1.96</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS graduate</td>
<td>4.29</td>
<td>0.37</td>
<td>0.09</td>
<td>3.78</td>
<td>1.13</td>
<td>0.30</td>
<td>3.78</td>
<td>1.64</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>4.42</td>
<td>0.36</td>
<td>0.08</td>
<td>3.91</td>
<td>1.22</td>
<td>0.31</td>
<td>3.91</td>
<td>1.83</td>
<td>0.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>6.52</td>
<td>0.51</td>
<td>0.08</td>
<td>5.82</td>
<td>1.51</td>
<td>0.26</td>
<td>5.82</td>
<td>2.27</td>
<td>0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postgraduate</td>
<td>7.50</td>
<td>0.66</td>
<td>0.09</td>
<td>7.04</td>
<td>1.84</td>
<td>0.26</td>
<td>7.04</td>
<td>2.65</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individuals with 11-30 years of experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS dropout</td>
<td>2.11</td>
<td>0.32</td>
<td>0.15</td>
<td>1.83</td>
<td>0.79</td>
<td>0.43</td>
<td>1.83</td>
<td>1.06</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS graduate</td>
<td>2.01</td>
<td>0.27</td>
<td>0.14</td>
<td>1.65</td>
<td>0.51</td>
<td>0.31</td>
<td>1.65</td>
<td>0.73</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>2.32</td>
<td>0.28</td>
<td>0.12</td>
<td>1.92</td>
<td>0.60</td>
<td>0.31</td>
<td>1.92</td>
<td>0.90</td>
<td>0.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>2.38</td>
<td>0.24</td>
<td>0.10</td>
<td>2.03</td>
<td>0.58</td>
<td>0.28</td>
<td>2.03</td>
<td>0.87</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postgraduate</td>
<td>2.64</td>
<td>0.24</td>
<td>0.09</td>
<td>2.34</td>
<td>0.66</td>
<td>0.28</td>
<td>2.34</td>
<td>0.94</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reports annual gross and net cross-state migration rates within education groups, separately for individuals with 2-10 and 11-30 years of potential experience. See notes under Table 1 in main text for sample details and construction of variables.

Table A7: Net migrant stocks by education

<table>
<thead>
<tr>
<th></th>
<th>Gross mig share (%)</th>
<th>Net mig share (%)</th>
<th>Net-gross ratio</th>
<th>Within 2-digit occs</th>
<th>Gross mig share (%)</th>
<th>Net mig share (%)</th>
<th>Net-gross ratio</th>
<th>Within 3-digit occs</th>
<th>Gross mig share (%)</th>
<th>Net mig share (%)</th>
<th>Net-gross ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td></td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
</tr>
<tr>
<td>HS dropout</td>
<td>26.19</td>
<td>6.41</td>
<td>0.24</td>
<td>28.03</td>
<td>9.30</td>
<td>0.33</td>
<td>28.03</td>
<td>11.27</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS graduate</td>
<td>29.54</td>
<td>7.85</td>
<td>0.27</td>
<td>29.34</td>
<td>8.60</td>
<td>0.29</td>
<td>29.34</td>
<td>9.36</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>35.24</td>
<td>9.32</td>
<td>0.26</td>
<td>34.62</td>
<td>10.06</td>
<td>0.29</td>
<td>34.62</td>
<td>10.99</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>44.92</td>
<td>13.18</td>
<td>0.29</td>
<td>44.15</td>
<td>13.69</td>
<td>0.31</td>
<td>44.15</td>
<td>14.58</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postgraduate</td>
<td>53.06</td>
<td>15.34</td>
<td>0.29</td>
<td>52.67</td>
<td>16.12</td>
<td>0.31</td>
<td>52.67</td>
<td>17.15</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reports gross stocks of migrants and net imbalances by education. In this exercise, I define "migrants" as individuals who live outside their state of birth. Column 1 reports the gross share of individuals living outside their birth state. Column 2 reports net imbalances of migrant stocks across states, which I compute in the same way as for net flows in Table 1 in the main text. And column 3 reports the ratio of net imbalances to gross stocks. The remaining columns repeat this exercise within occupation-defined labor markets, just as I do in Table 1. The sample excludes foreign-born individuals, but is otherwise identical to that of Table 1.
offers little information about match quality. In this appendix, I show this more formally.

For large moving costs, the overall matching rate $\rho (\varepsilon)$ converges to the local matching rate $\rho_L (\varepsilon) = \lambda [1 - F^\varepsilon (\varepsilon)]$, as defined in (6). Conditional on the offer arrival rate $\lambda$ (which I take as given throughout), match quality $\varepsilon$ is fully identified by the corresponding job matching rate $\rho$. So in this analysis, it is sufficient to study the informativeness of tenure as a proxy for the $\varepsilon$-specific matching rate $\rho$.

The overall separation rate from a job is equal to $\delta + \rho$, where $\delta$ is the exogenous transition rate to unemployment, and $\rho$ is the quit rate to better jobs. Given this is a constant hazard, the probability that a worker is still in a given job match at tenure $\tau$ (i.e. the survivor function) is equal to $\exp (-\tau (\delta + \rho))$. Now, let $G^\rho (\rho)$ be the unconditional distribution of matching rates $\rho$ across workers (which I take as given), where $\rho$ is bounded below by zero. And let $G^\rho (\rho | \tau)$ be the distribution of matching rates $\rho$ for workers with tenure $\tau$. By Bayes’ theorem, the density of the latter is:

$$g^\rho (\rho | \tau) = \frac{\exp (-\tau (\delta + \rho)) g^\rho (\rho)}{\int \exp (-\tau (\delta + x)) g^\rho (x) dx} = \frac{\exp (-\tau \rho) g^\rho (\rho)}{\int \exp (-\tau x) g^\rho (x) dx} \quad (A3)$$

As $\tau \to 0$, the conditional density $g^\rho (\rho | \tau)$ collapses to the unconditional density $g^\rho (\rho)$; so tenure $\tau$ offers no information on $\rho$ (or equivalently, on match quality). But as $\tau$ increases, the conditional distribution of $\rho$ becomes less dispersed. And in the limit, as $\tau \to \infty$, the conditional distribution collapses to a unit mass at $\rho = 0$, which corresponds to the maximum match quality.

To summarize, tenure can indeed serve as a proxy for match quality (conditional on the offer rate $\lambda$), and the precision of the proxy is increasing in the level of tenure.

### F Effect of tenure and experience on cross-state matching

Geographical mobility is strongly decreasing in experience. One intuition for this effect, arising from my model, is that older workers enjoy larger match quality - and hence are tempted by fewer long-distance job offers. In Section 4.3, I propose using job tenure as a proxy for match quality, $\varepsilon$. And indeed, I show there that the rate of cross-state matching is decreasing in initial job tenure.

In this appendix, I ask the following question: statistically, to what extent can initial tenure (as an imperfect proxy for match quality) account for the negative effect of experience on cross-state matching? Applying the complementary log-log model in (A2), the probability of forming a cross-state match within the time interval $t - 1$ to $t$ can be written as:

$$\Pr (Move_{it} = 1) = 1 - \exp (-\exp (\beta'X_{it})) \quad (A4)$$

where $X_{it}$ contain a set of education effects, as well as experience or initial log tenure. The $\beta$
Table A8: Elasticity of cross-state job matching to experience and initial job tenure

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS grad</td>
<td>0.416**</td>
<td>0.425**</td>
<td>0.542***</td>
<td>0.502***</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(0.191)</td>
<td>(0.191)</td>
<td>(0.191)</td>
</tr>
<tr>
<td>Some coll</td>
<td>0.601***</td>
<td>0.557***</td>
<td>0.737***</td>
<td>0.656***</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.182)</td>
<td>(0.182)</td>
<td>(0.183)</td>
</tr>
<tr>
<td>Undergrad</td>
<td>1.192***</td>
<td>1.151***</td>
<td>1.416***</td>
<td>1.305***</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.180)</td>
<td>(0.180)</td>
<td>(0.180)</td>
</tr>
<tr>
<td>Postgrad</td>
<td>1.656***</td>
<td>1.766***</td>
<td>1.989***</td>
<td>1.963***</td>
</tr>
<tr>
<td></td>
<td>(0.181)</td>
<td>(0.183)</td>
<td>(0.182)</td>
<td>(0.183)</td>
</tr>
<tr>
<td>Experience</td>
<td>-0.106***</td>
<td>-0.071***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience ^ 2</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial log tenure</td>
<td>-0.398***</td>
<td>-0.266***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>874,582</td>
<td>874,582</td>
<td>874,582</td>
<td>874,582</td>
</tr>
<tr>
<td>Matching rate (%)</td>
<td>0.129</td>
<td>0.129</td>
<td>0.129</td>
<td>0.129</td>
</tr>
</tbody>
</table>

This table reports complementary log-log regressions, of the form of (A4), for the incidence of cross-state job matching, based on four-month transitions in SIPP panels beginning 1996, 2001, 2004 and 2008. I restrict attention to job finds within employment cycles, i.e. without intervening unemployment or layoff spells. Errors are clustered by individual, and robust SEs are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Coefficients can be interpreted as elasticities of the cross-state matching rate.

I present the complementary log-log estimates in Table A8. Column 1 reports the basic education effects, which are increasing monotonically. A postgraduate degree adds 170 log points to the cross-state matching rate, relative to high school dropouts (the omitted category). In column 2, I control additionally for a quadratic in labor market experience. One extra year of experience reduces the cross-state matching rate by 11 log points; the squared term is statistically insignificant. In column 3, I replace the experience controls with initial log tenure: the elasticity of the matching rate is -0.4. Notice also that the education effects become somewhat larger in column 3, which is consistent with better educated workers enjoying higher match quality $\varepsilon$: see the discussion in Section 4.3.

In column 4, I control for experience and initial log tenure simultaneously. Comparing this with column 2, initial tenure accounts statistically for about one third of the effect of experience on cross-state matching. To the extent that initial tenure is an imprecise proxy for match quality (see Appendix E), this presumably understates the true contribution of match quality $\varepsilon$ to the experience effect. But of course, experience itself may contain independent information on moving costs. For example, in Figure 3 above, older workers report less willingness to move for work. A human capital explanation is that older workers have fewer years to benefit from
the sunk cost of moving (see e.g. Kennan and Walker, 2011).

G  Imputing one-off migration costs from wage returns

In this section, I show how my estimates of realized annuitized costs $\tilde{m}$ (based on the wage returns in Table 5) can be converted to fixed one-off equivalents $m = \sigma^\mu \mu$. Taking expectations of (22) over the distribution of match quality $\varepsilon$, and again abusing Jensen’s inequality:

$$
E[m] \approx \frac{E[\tilde{m}(\mu|\varepsilon)]}{r + \delta + \bar{\rho}_L + \bar{\rho}_C}
$$

(A5)

Based on my estimates above, the low-educated typically move with negligible realized costs. Consider instead the case of postgraduate degree-holders, who face the highest realized costs. Relative to wages, the lower and upper bounds on the annuitized expected costs are 0.16 and 0.23 respectively (see above), so take a mid-point of 0.195. Average monthly earnings for postgraduates in my SIPP sample are $5,783 (2015 prices). Taking 19.5 percent yields a monthly annuitized cost $E[\tilde{m}(\mu|\varepsilon)]$ of $1,127. Dividing $1,127 by a discount rate of 0.022 (based on job transition rates from Table 2) gives an expected one-off moving cost $E[m]$ (conditional on moving) of about $51,000. An equivalent exercise yields an expected cost of $19,000 for undergraduate degree-holders.29

H  Distribution of match quality and mean matching rates

In this section, following Burdett and Mortensen (1998), I derive the equilibrium distribution of match quality $\varepsilon$. And I then consider the implications for the mean local and cross-area matching rates. For any match quality $\varepsilon$, let:

$$
\rho(\varepsilon) = \rho_L(\varepsilon) + \rho_C(\varepsilon)
$$

(A6)

be the total job matching rate, i.e. the sum of the local and cross-area rates. This is equal to $\rho(\varepsilon_R)$ for the unemployed, so the steady-state unemployment rate is:

$$
u = \frac{\delta}{\delta + \rho(\varepsilon_R)}
$$

(A7)

28 I take a value of 0.029 for the separation rate $\delta$ (column 9 of Table 2) over four-month waves, 0.040 for the mean local matching rate $\bar{\rho}_L$ (among the initially employed: column 2), and 0.003 for the cross-state rate $\bar{\rho}_C$ (column 6). Summing these together with a four-month interest rate $r$ of 0.015 gives 0.087, or 0.022 in monthly terms.

29 Based on the analysis above, the lower and upper bounds on relative annuitized costs are 0.09 and 0.18 respectively, with a mid-point at 0.135. Average monthly earnings are $4,080, and taking 13.5 percent yields a monthly annuitized cost of $551. Using the job transition rates in Table 2 and a four-month interest rate of 0.015, I calibrate a total (monthly) discount rate of 0.029. And dividing the monthly annuitized cost by this discount rate gives a fixed cost of $19,000.
Now, consider the set of employed workers with match quality below $\varepsilon$. The inflow of workers to this set must equal the outflow in equilibrium:

$$u [\rho (\varepsilon_R) - \rho (\varepsilon)] = (1 - u) G (\varepsilon) [\delta + \rho (\varepsilon)]$$  \hfill (A8)

where $G (\varepsilon)$ is the distribution of $\varepsilon$ among employed workers. The inflow is composed entirely of the unemployed, who enter jobs with match quality below $\varepsilon$ at rate $[\rho (\varepsilon_R) - \rho (\varepsilon)]$. The outflow is composed of employed workers with match quality below $\varepsilon$ who (i) are separated to unemployment (at rate $\delta$) or (ii) find jobs yielding utility exceeding $\varepsilon$. Substituting (A7) for $u$ gives:

$$G (\varepsilon) = \frac{\delta}{\delta + \rho (\varepsilon)} \cdot \frac{\rho (\varepsilon_R) - \rho (\varepsilon)}{\rho (\varepsilon_R)}$$  \hfill (A9)

This equation demonstrates the importance of market frictions for my hypothesis. For all $\varepsilon$, $G (\varepsilon)$ converges to zero as the offer rate $\lambda$ (and therefore the matching rate $\rho (\varepsilon)$) becomes large relative to the separation rate $\delta$. So, in a frictionless world, all workers will benefit from the maximum match quality - and there will be no job surplus to justify geographical mobility.

The distribution $G (\varepsilon)$ accounts for employed workers only. To extend this to the unemployed, notice that they behave identically to workers with match quality $\varepsilon_R$. In this vein, I can define a distribution function $\hat{G} (\varepsilon)$: the fraction of all workers (irrespective of employment status) who receive effective match quality below $\varepsilon$. Specifically:

$$\hat{G} (\varepsilon) = \begin{cases} 0 & \text{if } \varepsilon < \varepsilon_R \\ u + (1 - u) G (\varepsilon) = \frac{\delta}{\delta + \rho (\varepsilon)} & \text{if } \varepsilon \geq \varepsilon_R \end{cases}$$  \hfill (A10)

with probability density:

$$\hat{g} (\varepsilon) = \begin{cases} 0 & \text{if } \varepsilon < \varepsilon_R \\ \frac{\delta}{\delta + \rho (\varepsilon_R)} & \text{if } \varepsilon = \varepsilon_R \\ \frac{\delta \rho' (\varepsilon)}{[\delta + \rho (\varepsilon)]^2} & \text{if } \varepsilon > \varepsilon_R \end{cases}$$  \hfill (A11)

where the unemployed are treated as receiving $\varepsilon_R$. This is effectively a left-censored distribution, with a discrete probability mass (i.e. the unemployed) at the censored value of $\varepsilon_R$.

In an elaboration of the discussion in Section 3.6, I now consider the determinants of the mean job matching rate across all workers, both employed and unemployed. The mean local and cross-area matching rates can be expressed as:

$$\bar{\rho}_X = \int_{\varepsilon_R}^{\infty} \rho_X (\varepsilon) d\hat{G} (\varepsilon)$$  \hfill (A12)

for $X = \{L, C\}$. Consider first the mean local rate $\bar{\rho}_L$. Based on (6), (A10) and (A12), it can be fully summarized by the offer rate $\lambda$, the separation rate $\delta$, and the the reservation match quality $\varepsilon_R$. Evidently, $\bar{\rho}_L$ is increasing in $\lambda$. It is also increasing in $\delta$: this is because a larger $\delta$ raises
\( \hat{G}(\epsilon) \) for all \( \epsilon \), and \( \rho_L(\epsilon) \) is decreasing in \( \epsilon \). Intuitively, workers have less time to rise up the ladder before they fall to the bottom (through a separation), so more offers will be acceptable to them on average. Finally, \( \bar{\rho}_L \) is decreasing in \( \epsilon_R \). As I note above, \( \epsilon_R \) can be interpreted as the censoring value of a left-censored distribution. Therefore, a larger reservation \( \epsilon_R \) causes \( \hat{G}(\epsilon) \) to decline for given \( \epsilon \) in the neighborhood of \( \epsilon_R \). Intuitively, if workers are more demanding, they will be located at higher \( \epsilon \) in equilibrium; and given \( \rho_L(\epsilon) \) is decreasing in \( \epsilon \), fewer offers will be acceptable to them on average. Based on Proposition 1, the mean cross-area matching rate \( \bar{\rho}_C \) additionally depends (positively) on the offer dispersion \( \sigma^e \) and (negatively) on the size of moving costs \( \sigma^\mu \).

I. Theoretical proofs and derivations

I.1 Proof of Proposition 2 in Section 3.4

Proposition 2 states that, for sufficiently large \( \sigma^\mu \), the (positive) response of cross-area job matching to \( \sigma^e \) is decreasing in initial match quality \( \epsilon \). In Section 3.4, I offer the following expression for \( \frac{d\rho_C(\epsilon)}{d\log \frac{\sigma^e}{\sigma^\mu}} \):

\[
\frac{d\rho_C(\epsilon)}{d\log \frac{\sigma^e}{\sigma^\mu}} = \pi \lambda \int_\epsilon^\infty \left\{ \left[ \frac{d\log \Omega(\epsilon' - \epsilon|\epsilon)}{d\log \sigma^e} + 1 \right] \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) f^\mu \left( \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) \right) \right\} dF^e(\epsilon')
\]

(A13)

I begin by considering the \( \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) f^\mu \left( \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) \right) \) term, and I return to the term in square brackets later. \( \Omega(\epsilon' - \epsilon|\epsilon) \) is unambiguously decreasing in \( \epsilon \). So, to ensure that \( \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) f^\mu \left( \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) \right) \) is decreasing in \( \epsilon \), a sufficient condition is that \( \mu f^\mu(\mu) \) is increasing in \( \mu \); or equivalently, that the elasticity of the density \( \frac{f(\mu)}{f^\mu(\mu)} > -1 \). In Proposition 2 though, I rely on an alternative sufficient condition: that \( \sigma^\mu \) is sufficiently large. Note I have assumed throughout that the elasticity of the density \( f^\mu \) is monotonically decreasing. Therefore, as \( \mu \to 0 \) from above, the elasticity of \( f^\mu \) grows and will eventually exceed -1. It follows that, for sufficiently large \( \sigma^\mu \), \( \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) f^\mu \left( \frac{\sigma^e}{\sigma^\mu} \Omega(\epsilon' - \epsilon|\epsilon) \right) \) must be decreasing in \( \epsilon \).

It remains to consider the implications of the \( \frac{d\log \Omega(\epsilon' - \epsilon|\epsilon)}{d\log \frac{\sigma^e}{\sigma^\mu}} \) term. Notice that:

\[
\frac{d\log \Omega(\epsilon' - \epsilon|\epsilon)}{d\log \frac{\sigma^e}{\sigma^\mu}} = - \int_\epsilon^{\epsilon'} \frac{d\rho_C(x)}{d\log \frac{\sigma^e}{\sigma^\mu}} \frac{1}{[r + \delta + \rho_L(x) + \rho_C(x)]^2} dx \left[ \int_\epsilon^{\epsilon'} \frac{1}{r + \delta + \rho_L(x) + \rho_C(x)} dx \right]^{-1}
\]

(A14)

This is a weighted average of \( \frac{d\rho_C(x)}{d\log \frac{\sigma^e}{\sigma^\mu}} \) across different values of initial match quality \( x \). With this in mind, the proposition can be demonstrated by contradiction. I begin by considering the response at the top of the support of match quality \( \epsilon \), and I then move down the distribution. Sup-
pose the response $\frac{d\rho_c(\epsilon)}{d\log \frac{\sigma_\mu}{\sigma^e}}$ is non-decreasing in $\epsilon$ at the top of the support of $\epsilon$. Based on (A13) and assuming that $\sigma^\mu$ is “sufficiently large”, such that $\frac{\sigma^e}{\sigma^\mu} \Omega(e' - \epsilon|\epsilon) f^{\mu} \left( \frac{\sigma^e}{\sigma^\mu} \Omega(e' - \epsilon|\epsilon) \right)$ is decreasing in $\epsilon$, it then follows that $\frac{d\log \Omega(e' - \epsilon|\epsilon)}{d\log \frac{\sigma_\mu}{\sigma^e}}$ must be increasing in $\epsilon$ at the top of the support. And using (A14), this in turn implies that $\frac{d\rho_c(\epsilon)}{d\log \frac{\sigma_\mu}{\sigma^e}}$ must be decreasing in $\epsilon$ (notice the negative sign in (A14)) at the top of the support of $\epsilon$. But this is a contradiction. One can then make a similar argument for every other $\epsilon$, moving sequentially down the distribution of $F^\epsilon$ - which implies that $\frac{d\rho_c(\epsilon)}{d\log \frac{\sigma_\mu}{\sigma^e}}$ is decreasing in $\epsilon$ over the full support of $\epsilon$.

I.2 Proof of Proposition 3 in Section 3.4

Proposition 3 states that, for sufficiently large $\sigma^\mu$, the responses of migratory outflows and inflows to changes in the local offer rate $\lambda_j$ are increasing (in magnitude) in offer dispersion $\sigma^e$ - around a steady-state where local areas are identical, i.e. $\gamma_j = \gamma$, $\lambda_j = \lambda$ and $V_j = V$ for all $j$.

Consider first the response of outflows. Using (15), this can be expressed as:

$$\frac{d\rho^{\text{Outflow}}_{C_j} (\epsilon)}{d\lambda_j} = -\pi \lambda \left( \frac{1}{\sigma^e} \frac{dV_j(\epsilon)}{d\lambda_j} \right) \int_\epsilon^\infty \left[ \sigma^e \frac{\sigma^e}{\sigma^\mu} f^\mu \left( \frac{\sigma^e}{\sigma^\mu} \Omega(e' - \epsilon|\epsilon) \right) \right] dF^\epsilon(e') \quad (A15)$$

Based on the same argument from the previous section (and due to the decreasing elasticity of the density $f^\mu$), the impact of $\sigma^e$ on $\frac{\sigma^e}{\sigma^\mu} f^\mu \left( \frac{\sigma^e}{\sigma^\mu} \Omega(e' - \epsilon|\epsilon) \right)$ must be positive for sufficiently large $\sigma^\mu$. To demonstrate that $\frac{d\rho^{\text{Outflow}}_{C_j} (\epsilon)}{d\lambda_j}$ is increasing (in magnitude) in $\sigma^e$, it is therefore sufficient to show that $\frac{d\rho^{\text{Outflow}}_{C_j} (\epsilon)}{d\lambda_j}$ is increasing (at least) proportionally with $\sigma^e$.

To see this, notice that local worker value (4) can be written as:

$$rV_j(\epsilon) = \gamma X + \sigma^e \epsilon + \delta \left[ V_j(\epsilon_R) - V_j(\epsilon) \right] + \lambda_j \int_\epsilon^\infty \left[ V_j(e') - V_j(\epsilon) \right] dF^\epsilon(e')$$

$$+ \pi \lambda \int_0^\infty \left[ \int_{-\infty}^\infty \max \left\{ V(e') - V_j(\epsilon) - \sigma^\mu \mu, 0 \right\} dF^\epsilon(e') \right] dF^\mu(\mu) \quad (A16)$$

This expression equates values and offer rates in all areas $k \neq j$ in equation (4). Differentiating with respect to the local offer rate $\lambda_j$, around a steady-state with $\lambda_j = \lambda$:

$$\frac{rV_j(\epsilon)}{d\lambda_j} |_{\lambda_j = \lambda} = \delta \left[ \frac{dV_j(\epsilon_R)}{d\lambda_j} |_{\lambda_j = \lambda} - \frac{dV_j(\epsilon)}{d\lambda_j} |_{\lambda_j = \lambda} \right] + \sigma^e \int_\epsilon^\infty \Omega(e' - \epsilon|\epsilon) dF^\epsilon(e')$$

$$+ \rho_L(\epsilon) \int_\epsilon^\infty \frac{dV_j(e')}{d\lambda_j} |_{\lambda_j = \lambda} dF^\epsilon(e') - \rho_L(\epsilon) \frac{dV_j(\epsilon)}{d\lambda_j} |_{\lambda_j = \lambda} - \rho_C(\epsilon) \frac{dV_j(\epsilon)}{d\lambda_j} |_{\lambda_j = \lambda} \quad (A17)$$

55
where the function $\Omega$ is defined by (9). Rearranging this expression:

$$
\frac{dV_j(\varepsilon)}{d\lambda_j}|_{\lambda_j=\lambda} = \frac{\delta}{r+\delta+\rho(\varepsilon)} \frac{dV_j(\varepsilon)}{d\lambda_j}|_{\lambda_j=\lambda} + \frac{\rho_L(\varepsilon)}{r+\delta+\rho(\varepsilon)} \int_{\varepsilon}^{\infty} \frac{dV_j(\varepsilon')}{d\lambda_j}|_{\lambda_j=\lambda} dF^\varepsilon(\varepsilon')
$$

(A18)

where $\rho(\varepsilon) \equiv \rho_L(\varepsilon) + \rho_C(\varepsilon)$. Both $\Omega(\varepsilon'-\varepsilon|\varepsilon)$ and $\rho(\varepsilon)$ are functions of the cross-state matching rate $\rho_C(\varepsilon)$, so they are both in principle sensitive to offer dispersion $\sigma^\varepsilon$. But as $\sigma^\mu$ becomes large, this sensitivity goes to zero. By inspection of (A18), $\frac{dV_j(\varepsilon)}{d\lambda_j}|_{\lambda_j=\lambda}$ must then be increasing proportionally with $\sigma^\varepsilon$.

Given this result for $\frac{dV_j(\varepsilon)}{d\lambda_j}$, a parallel argument can be made for the response of migratory inflows in equation (14). And as a result, the inflow response must also be increasing in $\sigma^\varepsilon$: see the discussion in Section 3.5.

### I.3 Derivation of equation (19) in Section 5.3

Equation (19) approximates the mean cross-area matching rate, $\bar{\rho}_C$, in terms of the returns to local job search. This approximation relies on two assumptions on the moving cost distribution. First, I assume the draws of $\mu \sim F^\mu$ are distributed uniformly between 0 and a maximum normalized to 1; so moving costs $m$ range from 0 to $\sigma^\mu$. Second, I assume there are no wage offers which can justify moving at the maximum cost draw: that is, for every initial match quality $\varepsilon$ and for every offer $\varepsilon' \sim F^\varepsilon, V(\varepsilon') - V(\varepsilon) < \sigma^\mu$.

Using (10), the cross-area matching rate then collapses to:

$$
\rho_C(\varepsilon) = \tilde{\pi} \lambda \frac{\sigma^\varepsilon}{\sigma^\mu} \int_{\varepsilon}^{\infty} \Omega(\varepsilon'-\varepsilon|\varepsilon) dF^\varepsilon(\varepsilon')
$$

(A19)

And using (9), the $\varepsilon$-unit surplus $\Omega(\varepsilon'-\varepsilon|\varepsilon)$ can be linearly approximated as:

$$
\Omega(\varepsilon'-\varepsilon|\varepsilon) \approx \frac{\varepsilon'-\varepsilon}{r+\delta+\rho_L(\varepsilon)+\rho_C(\varepsilon)}
$$

(A20)

Now, applying the approximation in (A20) to (A19), and combining with (6):

$$
\rho_C(\varepsilon) \approx \tilde{\pi} \lambda \frac{\sigma^\varepsilon}{\sigma^\mu} \int_{\varepsilon}^{\infty} (\varepsilon'-\varepsilon) dF^\varepsilon(\varepsilon') = \frac{\tilde{\pi}}{\sigma^\mu} \sigma^\varepsilon E_L[\varepsilon'-\varepsilon|\varepsilon' \geq \varepsilon] \rho_L(\varepsilon)
$$

(A21)

where $E_L[\varepsilon'-\varepsilon|\varepsilon' \geq \varepsilon]$ is the expected improvement in match quality $\varepsilon$ arising from a local

---

30 Looking at (9), the job surplus $\Omega(\varepsilon'-\varepsilon|\varepsilon)$ is discounted by the overall rate of job matching $\rho(\varepsilon)$. But intuitively, as $\sigma^\mu$ becomes large, the contribution of cross-area matching $\rho_C(\varepsilon)$ to the overall matching rate $\rho(\varepsilon)$ becomes small. More formally, this effect can be appreciated from the argument in the previous section. According to Proposition 2, $\frac{d\rho_C(\varepsilon)}{d \log \sigma^\mu}$ must be decreasing in $\varepsilon$ for sufficiently large $\sigma^\mu$. Looking at (A13), it must therefore also be true that $\frac{d\rho_C(\varepsilon)}{d \log \sigma^\mu}$ decreases to zero as $\sigma^\mu$ becomes large.

56
(subscript $L$) job match; so $\sigma^e \mathbb{E}_L [e' - e | e' \geq e]$ is the expected wage return. Now, taking the expectation of (A21) over the distribution of match quality $\varepsilon$ for initially employed workers, and abusing Jensen’s inequality:

$$\bar{\rho}_C \approx \frac{\bar{\pi}}{\sigma^\mu} \cdot \frac{\bar{\rho}_L \beta_1}{r + \delta + \bar{\rho}_L + \bar{\rho}_C} \quad \text{(A22)}$$

which is equation (19) in the main text. The $\bar{\rho}_X$ are the mean matching rates of the initially employed; and $\beta_1 = \sigma^e \mathbb{E}_L [e' - e | e' \geq e]$ are their expected local wage returns, as identified by the estimating equation (18).

### I.4 Proof of Proposition 4 in Section 6.2

Proposition 4 states that the expected annuitized cost is increasing in both $\sigma^e$ and $\sigma^\mu$, for given initial match quality $\varepsilon$ and for sufficiently large $\sigma^\mu$. It is sufficient to show that increases in $\sigma^e$ and $\sigma^\mu$ cause hazard rate dominating transformations of the realized annuitized cost distribution, $Z(\tilde{m}|\varepsilon)$. In this section, the distribution function $Z(\cdot|\varepsilon)$ and job surplus $\Omega(\cdot|\varepsilon)$ are both defined for given match quality $\varepsilon$, but I omit the “$|\varepsilon$” from here on to ease notation.

Importantly, the $\Omega$ relationship depends on the level of $\sigma^e$ and $\sigma^\mu$. This is because $\sigma^e$ and $\sigma^\mu$ affect the cross-area matching rate $\rho_C(\varepsilon)$, and this rate matters for the discounting of job surplus in (9). However, for a sufficiently large $\sigma^\mu$, cross-area matching contributes relatively little to job matching overall, so changes in $\sigma^\mu$ and $\sigma^e$ have little effect on the $\Omega$ relationship. See footnote 30 in Appendix I.2. From here on, I therefore neglect the impact of $\sigma^e$ and $\sigma^\mu$ on the $\Omega$ relationship.

Using (23), the hazard rate of $Z$ can be written as:

$$\frac{z(m)}{1 - Z(m)} = \left[ \int_m^\infty \frac{1 - F^e(\varepsilon + \frac{x}{\sigma^e})}{1 - F^e(\varepsilon + \frac{\tilde{m}}{\sigma^e})} \cdot \frac{f^\mu(\frac{\sigma^e}{\sigma^\mu} \Omega(\frac{\varepsilon}{\sigma^\mu}))}{f^\mu(\frac{\sigma^e}{\sigma^\mu} \Omega(\frac{\tilde{m}}{\sigma^\mu}))} dx \right]^{-1} \quad \text{(A23)}$$

The $\frac{1 - F^e(\varepsilon + \frac{x}{\sigma^e})}{1 - F^e(\varepsilon + \frac{\tilde{m}}{\sigma^e})}$ term is increasing in $\sigma^e$, conditional on $\varepsilon$, for all $x > \tilde{m}$. To see this, consider the derivative of its log with respect to $\sigma^e$

$$\frac{d}{d\sigma^e} \left\{ \log \frac{1 - F^e(\varepsilon + \frac{x}{\sigma^e})}{1 - F^e(\varepsilon + \frac{\tilde{m}}{\sigma^e})} \right\} = \frac{1}{(\sigma^e)^2} \left( x \cdot \frac{f^e(\varepsilon + \frac{x}{\sigma^e})}{1 - F^e(\varepsilon + \frac{x}{\sigma^e})} - \tilde{m} \cdot \frac{f^e(\varepsilon + \frac{\tilde{m}}{\sigma^e})}{1 - F^e(\varepsilon + \frac{\tilde{m}}{\sigma^e})} \right) > 0 \quad \text{(A24)}$$

which exceeds zero due to my assumption that $F^e$ has a monotonically increasing hazard rate.

I now show the second term in (A23), $\frac{f^\mu(\frac{\sigma^e}{\sigma^\mu} \Omega(\frac{\varepsilon}{\sigma^\mu}))}{f^\mu(\frac{\sigma^e}{\sigma^\mu} \Omega(\frac{\tilde{m}}{\sigma^\mu}))}$, is increasing in both $\sigma^\mu$ and $\sigma^e$. 57
Consider first the impact of \( \sigma^\mu \):

\[
\frac{d}{d \sigma^\mu} \left\{ \log \left( \frac{f^\mu \left( \frac{\sigma^\mu}{\sigma^\mu} \Omega \left( \frac{m}{\sigma^\mu} \right) \right)}{f^\mu \left( \frac{\sigma^\mu}{\sigma^\mu} \Omega \left( \frac{m}{\sigma^\mu} \right) \right)} \right) \right\} = \frac{1}{\sigma^\mu} \left[ \sigma^\epsilon \left( \frac{\sigma^\mu}{\sigma^\epsilon} \right) \frac{f^\mu \left( \frac{\sigma^\mu}{\sigma^\epsilon} \Omega \left( \frac{m}{\sigma^\epsilon} \right) \right)}{f^\mu \left( \frac{\sigma^\mu}{\sigma^\epsilon} \Omega \left( \frac{m}{\sigma^\epsilon} \right) \right)} - \frac{\sigma^\epsilon}{\sigma^\mu} \left( \frac{x}{\sigma^\epsilon} \right) \frac{f^\mu \left( \frac{\sigma^\epsilon}{\sigma^\mu} \Omega \left( \frac{m}{\sigma^\epsilon} \right) \right)}{f^\mu \left( \frac{\sigma^\epsilon}{\sigma^\mu} \Omega \left( \frac{m}{\sigma^\epsilon} \right) \right)} \right] > 0
\]

(A25)

which exceeds zero for \( x > \bar{m} \), due to my assumption that the density’s elasticity, i.e. \( \mu \frac{f^\mu (\mu)}{f^\mu (\mu)} \), is monotonically decreasing in \( \mu \). Now, consider the impact of \( \sigma^\epsilon \):

\[
\frac{d}{d \sigma^\epsilon} \left\{ \log \left( \frac{f^\mu \left( \frac{\sigma^\epsilon}{\sigma^\epsilon} \Omega \left( \frac{m}{\sigma^\epsilon} \right) \right)}{f^\mu \left( \frac{\sigma^\epsilon}{\sigma^\epsilon} \Omega \left( \frac{m}{\sigma^\epsilon} \right) \right)} \right) \right\} = \frac{d}{d \sigma^\epsilon} \left\{ \left[ \sigma^\mu \left( \frac{x}{\sigma^\epsilon} \right) \right] \right\} - \frac{d}{d \sigma^\epsilon} \left\{ \left[ \sigma^\epsilon \left( \frac{x}{\sigma^\epsilon} \right) \right] \right\} > 0
\]

(A26)

Given my assumption on the elasticity’s density, a sufficient condition for this expression to exceed zero is:

\[
\frac{d}{d \sigma^\epsilon} \left\{ \left[ \sigma^\mu \left( \frac{x}{\sigma^\epsilon} \right) \right] \right\} > \frac{d}{d \sigma^\epsilon} \left\{ \left[ \sigma^\epsilon \left( \frac{x}{\sigma^\epsilon} \right) \right] \right\}
\]

(A27)

which, using the definition of \( \Omega \) in (9), implies:

\[
\frac{\bar{m}}{(\sigma^\epsilon)^2} \frac{\sigma^\epsilon \Omega' \left( \frac{\bar{m}}{\sigma^\epsilon} \right)}{\sigma^\mu \Omega \left( \frac{\bar{m}}{\sigma^\epsilon} \right)} > \frac{x}{(\sigma^\epsilon)^2} \frac{\sigma^\epsilon \Omega' \left( \frac{x}{\sigma^\epsilon} \right)}{\sigma^\mu \Omega \left( \frac{x}{\sigma^\epsilon} \right)}
\]

(A28)

\[
\bar{m} \frac{\left[ r + \delta + \rho \left( \varepsilon + \frac{\bar{m}}{\sigma^\epsilon} \right) \right]^{-1}}{\Omega \left( \frac{\bar{m}}{\sigma^\epsilon} \right)} > x \frac{\left[ r + \delta + \rho \left( \varepsilon + \frac{x}{\sigma^\epsilon} \right) \right]^{-1}}{\Omega \left( \frac{x}{\sigma^\epsilon} \right)}
\]

\[
\frac{\sigma^\epsilon}{\bar{m}} \int_0^{\bar{m}} \frac{r + \delta + \rho \left( \varepsilon + \frac{\bar{m}}{\sigma^\epsilon} \right)}{r + \delta + \rho \left( \varepsilon + \frac{x}{\sigma^\epsilon} \right)} ds > \frac{\sigma^\epsilon}{x} \int_0^{\bar{m}} \frac{r + \delta + \rho \left( \varepsilon + \frac{\bar{m}}{\sigma^\epsilon} \right)}{r + \delta + \rho \left( \varepsilon + \frac{x}{\sigma^\epsilon} \right)} ds
\]

where \( \rho \left( \varepsilon \right) \equiv \rho_L \left( \varepsilon \right) + \rho_C \left( \varepsilon \right) \) is the total job matching rate. The final line must be true because of the monotonicity of \( \rho \left( \varepsilon \right) \). This confirms that \( \frac{f^\mu \left( \frac{\sigma^\epsilon}{\sigma^\epsilon} \Omega \left( \frac{x}{\sigma^\epsilon} \right) \right)}{f^\mu \left( \frac{\sigma^\mu}{\sigma^\epsilon} \Omega \left( \frac{m}{\sigma^\epsilon} \right) \right)} \) in (A23) is indeed increasing in both \( \sigma^\mu \) and \( \sigma^\epsilon \).

As a result, the hazard rate in (A23) must be decreasing in both \( \sigma^\mu \) and \( \sigma^\epsilon \) for given \( \bar{m} \) and \( \varepsilon \). This means that increases in \( \sigma^\mu \) and \( \sigma^\epsilon \) cause hazard rate dominating transformations of the realized annuitized cost distribution, \( Z(\bar{m}) \). And so, the expected annuitized costs must be increasing in both \( \sigma^\mu \) and \( \sigma^\epsilon \).
I.5 Proof of Proposition 5 in Section 6.2

Finally, I prove that (25) can serve as a lower bound on the expected realized annuitized costs. Following the argument given in Section 6.2, it suffices to show that the expression in the curly brackets in (25) is less or equal to $\tilde{m}$, i.e.:

$$\mathbb{E}_L \left[ \sigma^e (\varepsilon' - \varepsilon) \mid \sigma^e (\varepsilon' - \varepsilon) \geq \tilde{m} \right] - \mathbb{E}_L \left[ \sigma^e (\varepsilon' - \varepsilon) \mid \varepsilon' - \varepsilon \geq 0 \right] \leq \tilde{m} \quad \text{(A29)}$$

conditional on the initial match quality $\varepsilon$, where the operator $\mathbb{E}_L$ denotes the expected improvement in match quality $\varepsilon$ arising from a local (subscript $L$) match. This can be rewritten as:

$$\mathbb{E}_L \left[ \sigma^e (\varepsilon' - \varepsilon) - \tilde{m} \mid \sigma^e (\varepsilon' - \varepsilon) - \tilde{m} \geq 0 \right] \leq \mathbb{E}_L \left[ \sigma^e (\varepsilon' - \varepsilon) \mid \varepsilon' - \varepsilon \geq 0 \right] \quad \text{(A30)}$$

Since I have assumed the annuitized cost $\tilde{m}$ always exceeds zero, it is sufficient to show that:

$$\frac{d}{dx} \log \mathbb{E}_L [\varepsilon' - x | \varepsilon' - x \geq 0] \leq 0 \quad \text{(A31)}$$

for all $x \equiv \varepsilon + \frac{\tilde{m}}{\sigma^e}$. Writing this in terms of the offer distribution $F^e$:

$$\frac{d}{dx} \log \mathbb{E}_L [\varepsilon' - x | \varepsilon' - x \geq 0] = \frac{d}{dx} \log \frac{\int_x^{\infty} \varepsilon f^e (\varepsilon) d\varepsilon}{1 - F^e (x)} = \frac{d}{dx} \log \frac{1 - F^e (x)}{1 - F^e (x)} + \frac{f^e (x)}{1 - F^e (x)} \quad \text{(A32)}$$

where the second line follows from integration by parts. Now, I have assumed that $F^e$ has a monotonically increasing hazard rate; that is, $\frac{f^e (\varepsilon)}{1 - F^e (\varepsilon)} \geq \frac{f^e (x)}{1 - F^e (x)}$ for all $\varepsilon \geq x$. Therefore:

$$\frac{d}{dx} \log \mathbb{E}_L [\varepsilon' - x | \varepsilon' - x \geq 0] \leq - \frac{1 - F^e (x)}{\int_x^{\infty} \frac{1 - F^e (\varepsilon)}{f^e (\varepsilon)} f^e (\varepsilon) d\varepsilon} + \frac{f^e (x)}{1 - F^e (x)} = 0 \quad \text{(A33)}$$

so equation (A31) is satisfied.

J Appendix references


