Selection of Job-to-job Migrants on Match Quality*

Juho Alasalmi†
University of Konstanz / Pellervo Economic Research

September 2023
Latest version here

Abstract

Many migrants search on the job and move only after having found a job in their destination. The migration choices of such job-to-job migrants are not based on source and destination location wage distributions but on realizations from these distributions. With wage dispersion, migration choices then also depend on job match qualities. A Roy-Borjas model extended to capture the selection of job-to-job migrants predicts negative selection on source and positive selection on destination job match quality. For empirical support, I compare selection on residual wages between job-to-job migrants and workers who similarly contract a job outside their location of residence but choose to commute. Mobility costs amplify selection, and comparing job-to-job migrants and commuters, two groups similar in their unobservable skills incurring different mobility costs, identifies migrants' wage residual selection consistent with the predicted selection on job match quality.

JEL classification: J61; R23; D83
Keywords: Labor mobility, internal migration, migrant selection, job match quality

1 Introduction

Migration is often modelled as a risky investment: workers compare their labor market prospects in the source and destination locations and potentially relocate hoping to find employment in their destination. Migration choice is then based on the source and destination location wage distributions. However, migrants often relocate only after successful on-the-job search, after a job in the destination has been found. Migration choice is then based on realizations from source and destination location wage distributions. These two migration strategies may differ in the migrant selection they generate. While the earlier literature on migrant selection has focused on the former, this paper studies the selection of *job-to-job* migrants: migrants who migrate after having accepted a job offer in the destination.

*I thank Luna Bellani, Wolfgang Dauth, Sebastian Findeisen, Susanne Goldlücke, Janne Huovari, Giuseppe Ippedico, Elina Jussila, Stephan Maurer, Panu Poutvaara, and many others at PTT, Helsinki 12/2020; European Public Choice Society Meeting, Lille 4/2021; International Institute of Public Finance Annual Congress, Reykjavík 8/2021; CEMIR Junior Economist Workshop on Migration Research, Munich 7/2022; VATT, Helsinki 2/2023; Annual Meeting of the Finnish Economic Association, Espoo 2/2023; 38th Meeting of the European Economic Association, Barcelona 8/2023 for valuable comments and discussions.

†juho.alasalmi@uni-konstanz.de

Selection in cases where opportunities are searched and secured in other markets before entering them is complicated. The first selective behavior is to search: only those who expect search to increase their utility net search and mobility costs search for opportunities in other markets. The second hurdle is not up to the searcher: the market provides opportunities only to those it judges to fit its goals best. For instance, in the labor market, employers select their employees and the hiring of profit-maximizing employers is not random. Finally, if a searcher receives an offer, she judges whether, net relocation costs, the offer is worth accepting and chooses whether to relocate her labor supply or not.

I reduce this complexity by studying selection among individuals that have passed the first two hurdles. Taking the selection generated by the first two hurdles as exogenous allows for a tractable model while taking account the nonrandom selection that the choice to search and employers' hiring choices create. This modelling approach is accompanied by an empirical setting that similarly abstracts from the first two hurdles by studying selection among workers who have all chosen to search for jobs interregionally and have received a job offer.

I capture the essential aspects of the selection of job-to-job migrants by extending the Roy-Borjas (Roy, 1951; Borjas, 1987) model of migrant selection in three ways. First, I allow those who are in a position to choose whether to migrate or not to be a nonrandomly selected subset of source location workers. Second, I allow within-skill wage dispersion: given skill, the wage is not deterministic, but each worker faces a distribution of potential wages. Deviations from the expected compensation for skills are interpreted as job match quality. Third, I allow observability of both the current wage in the source and the potential wage in the destination prior to the migration choice. The choice of job-to-job migration is then not based on the mean and variance of source and destination location wage distributions but on specific realizations from these distributions.

The resulting model of selection of job-to-job migrants, while nests the familiar results on selection on skills, also implies selection on job match quality. Given current wage, low wage offers may not be enough to compensate for relocation costs while high offers may be. Thus, the deviation of the offered wage from the expected wage, job match quality in the destination, becomes a factor of relocation choice and there is positive selection on destination location job match quality. On the other hand, given offered wage, the lower the current wage is, the larger are the gains from relocation. Thus, job match quality in the source becomes a factor of relocation choice and there is negative selection on source location job match quality. Negative selection on job match quality in the source and positive selection on job match quality in the destination can be seen as a reinterpretation of the Borjas' (1987) "refugee sorting": here workers are fleeing bad job matches for good job matches.

To identify the predicted selection on job match quality, I study selection on wage residuals using high-quality administrative data on internal migration and job mobility in Finland. Identifying selection on job match quality is not straightforward since selection on wage residuals may also reflect selection on unobservable skills: First, in the source, job-to-job migrants may not be negatively selected relative to stayers on pre-migration residuals if they are positively selected relative to stayers on their unobservable skills. Potential positive selection of those who receive job offers may mask negative selection job match quality when this group is compared to stayers. Second, in the destination, job-to-job migrants may be positively selected relative to destination region workers on post-migration residuals not only due to their higher job match quality but also due to their more valuable unobservable skills. Third, migrants may increase their residuals when relocating their labor supply not only due to a good new job match in comparison to the current job match but due to higher compensation for unobservable skills in their destination than in their source.

Identifying selection on job match quality by comparing migrants to stayers is thus difficult. An alternative comparison group for the job-to-job migrants are those who also contract a job outside their location of residence but who choose to commute. These two groups of mobile workers make a relevant comparison for three reasons: First, the commuters, like job-to-job migrants, have received a job offer, and, thus have both self-selected to search interregionally and have been selected by employers. Second, when comparing migrants and commuters within source-destination pairs, migrants and commuters experience the same source and destination rate of return for unobservable skills. Third, those who choose to commute and those who choose to migrate incur different relocation costs. With a model of mobility mode choice, I show how heterogeneity in costs of employing different mobility modes with optimal mobility mode choices makes the employed mobility mode informative of incurred costs and, thus, a proxy for mobility costs. Relocation costs, on the other hand, magnify selection. Comparison of migrants and commuters can thus be interpreted as studying variation in relocation costs and, thus, helps us discern whether selection on job match quality or selection on unobservable skills is magnified. I find that migrants have 1-2 percent lower pre-mobility residuals and 2-3 percent higher post-mobility residuals than commuters. These findings are consistent with the model's predicted selection of job-to-job migrants on job match quality.

These results have implications for interpreting migrant selection on wage residuals. As a large fraction of variation in wages cannot be explained by observable determinants of productivity (e.g. Mortensen (2003)), it can be expected that a large fraction of selection occurs on unobservable determinants of productivity. Indeed, Borjas et al. (2019) assess that 70 (50) percent of positive selection on source location earnings is due to unobservable determinants of productivity among (fe)male emigrants. Selection on unobservable skills, as measured by wage residuals, is thus an important part of migrant selection. However, observed selection on residuals does not always align with the predicted selection on unobservable skills (Chiquiar and Hanson, 2005; Fernandez-Huertas Moraga, 2011; Kaestner and Malamud, 2014; Birgier et al., 2022). For job-to-job migrants, wage residuals may not only reflect selection on unobservable skills but also contain systematic variation due to selection on job match quality. While the role of job match quality in explaining selection on wage residuals has been discussed before (Nakosteen et al., 2008; Borjas et al., 2019; Birgier et al., 2022), selection on job match quality has so far not been modelled nor empirically identified. The findings here suggest that interpreting results on selection on residuals without taking job match quality into account underestimates positive selection on unobservable skills in the source and overestimates positive selection on unobservable skills in the destination. In other words, this may lead to underestimation of brain drain in the source, but also to overestimation of incoming migrants' skills.

The next section positions the work into the related literature. Section 3 presents the model of selection of job-to-job migrants. Section 4 outlines the empirical approach. Section 5 introduces the data and the empirical definitions. Section 6 provides evidence on selection on job match quality. Section 7 concludes. All proofs are in the appendix.

2 Related Literature

Since Hicks (1932), Schultz (1961) and Sjaastad (1962), migration has been modelled as determined by economic incentives. Borjas (1987) noted the heterogeneity in incentives arising from differences in individual productivities and modelled the consequences for migrant selection. Since then, a large empirical literature has studied the role of incentives not only in inducing migration but also in selecting migrants.

Selection on unobservable skills, as measured by wage residuals, is an important part of migrant selection (Borjas et al., 2019). Selection on unobservable determinants of productivity has been studied by studying selection on wage residuals (Borjas et al., 1992; Abramitzky, 2009; Fernandez-Huertas Moraga, 2011; Kaestner and Malamud, 2014; Borjas et al., 2019; Birgier et al., 2022). These studies have interpreted residuals to reflect location-variant compensation for time-invariant unobservable skills. However, time-invariant unobservable skills may not be the only component of residuals. Gould and Moav (2016) decompose unobservable skills into location-invariant and location-specific components and show how selection on these two types of unobservable skills is qualitatively different. But with wage dispersion given skills (Mortensen, 2003), not all variation in residuals is due to unobservable skills. I decompose the variation in wages not explainable by observable determinants of productivity into a location and time-invariant skill component and a job (location) and time-variant job match quality component.

Migrant selection is often studied by comparing out-migrants (Chiquiar and Hanson, 2005; Fernandez-Huertas Moraga, 2011; Kaestner and Malamud, 2014; Borjas et al., 2019; Rosso, 2019; Birgier et al., 2022) or in-migrants (Chiswick, 1978; Carliner, 1980; Abramitzky, 2009) to stayers. Also, incoming migrants from different locations (Borjas, 1987; Abramitzky, 2009), outgoing migrants to different locations (Hunt and Mueller, 2004; Dostie and Léger, 2009; Parey et al., 2017), and also migrants moving between the same locations but working in different industries and occupations (Gould and Moav, 2016) have been compared. On the other hand, in estimating the labor market returns to migration, wage changes due to job changes and wage changes due to location changes have been separated by comparing migrants to job movers (Bartel, 1979; Yankow, 2003; Ham et al., 2011; Emmler and Fitzenberger, 2020). The comparison of migrants to commuters adds to this set of settings by comparing two groups of workers who relocate their labor supply, thus controlling for job changes and changes in location-specific compensation for skills.

The distinction between *contracted*, that is, job-to-job migration, and *speculative* migration, that is, migration to search for work in the destination, was introduced by Silvers (1977).[1] The evidence on the respective roles of these two forms of migration is scarce. Saben (1964) reports that 62 percent of high-skilled intercounty migrants moved after having accepted a job in the destination whereas 38 percent of other migrants had a job at hand when migrating within the US in 1962. Detang-Dessendre and Molho (1999), using a small survey of young first-time migrants from rural regions in France in 1993, report job-to-job migration to be more common than migration without a contracted job in the destination. Since the collection of these data, the share of job-to-job migration has likely increased and will likely be increasing in the future. As job search more and more often occurs online, job opportunities can more easily be searched for and secured in distant labor markets. Moreover, policies regarding international immigration have been gearing toward favoring high-skilled migration, and such policies often contain requirements of a job contract at arrival (Kerr et al., 2017). It is thus likely that job-to-job migration is the dominant form of labor-related migration, at least in developed countries.

While not made explicit, earlier research has often likely studied job-to-job migration, for instance, by only allowing short gaps between source and destination job spells (Ham et al., 2011) or by defining migration as a change in job location (Emmler and Fitzenberger, 2020). Moreover, even without restrictions that increase the prevalence of job-to-job migration in the data, job-to-job migration has likely been common in the data used by many studies given the likely dominant role of job-to-job migration in labor-related migration. Thus, to assist the interpretations of empirical results, models of labor-related migration would benefit in making the possibility of job-to-job migration explicit.

3 Selection of Job-to-job Migrants

Consider two locations or labor markets indexed by $h = j, k$. Worker i works in location j and potentially searches for a job in location k. Let φ_{ik} be the probability that i receives a job offer from location k. If i searches for a job in location k, $\varphi_{ik} > 0$ and otherwise $\varphi_{ik} = 0$. I call the subset I of workers working in location j that search for location k jobs and receive a job offer the *population at risk of job-to-job migration*. Thus, the probability that i enters I is φ_{ik}. Only the workers in the population at risk of job-to-job migration can relocate their

[1] The terms *contracted* and *speculative* migration coined by Silvers (1977) have been later used at least by Rogerson (1982), Pickles and Rogerson (1984), Molho (1986) and Detang-Dessendre and Molho (1999).

labor supply.[2] Each worker has a skill ν_i and the mean skill in location $h = j, k$ is denoted μ_h^ν. Skills are time-invariant and perfectly transferable across locations.[3] Let F_{ih} denote the distribution of wages i can potentially earn in location h.

Let each worker have a location l with respect to which labor supply costs are determined such that the per-period cost for a worker supplying labor in location h is π_{lh}. If we focus on workers who reside in location j, to whom supplying labor there has no cost, and to whom relocation of labor supply to location k is accompanied with migration to location k, cost π_{jk} can be interpreted as periodized migration cost. More generally, $l \neq j$ allows location j workers to already incur mobility costs, say in the form of commuting from l to j or having previously migrated from l to j.

While defining labor supply costs as periodized and with respect to third location l allows for general pre-migration labor supply costs, it also simplifies the model by making the job search problem stationary. The worker i maximizes a discounted stream of per period net income $u_{ilh} = u(w_{ih}, \pi_{lh})$, where w_{ij} is the current wage and job offers w_{ik} from location k are sampled from F_{ik}. The asset value of search is

$$rV_i(u_{ilj}) = u_{ilj} + \varphi_{ik} \int_{-\infty}^{\infty} \max\{0, V_i(u_{ilk}) - V_i(u_{ilj})\} dF_{ik}(w_{ik}), \tag{1}$$

where r is the discount rate and the second term on the right-hand side is the value of optimal job acceptance and migration behavior. A job offer is accepted if and only if $V_i(u_{ilk}) > V_i(u_{ilj})$, which, since V_i is strictly increasing, is equivalent to $u_{ilk} > u_{ilj}$. Letting $u(w_{ih}, \pi_{lh}) = w_{ih} - \pi_{lh}$, we have the job acceptance rule

$$w_{ik} > w_{ij} + \pi_{ljk}, \tag{2}$$

where $\pi_{ljk} = \pi_{lk} - \pi_{lj}$ is the cost of relocating labor supply from j to k.[4] For now, for simplicity, suppose l is the same for all location j workers and denote $\pi_{ljk} = \pi_{jk}$.

Decompose worker i's wage in location h as $w_{ih} = \bar{\mu}_h + \rho_h(\nu_i - \mu_h^\nu) + q_{ih}$ where $\bar{\mu}_h$ is the expected compensation for location h mean skill, $\int w dF_{ih}(w) = \bar{\mu}_h + \rho_h(\nu_i - \mu_h^\nu)$ is the expected compensation for skill ν_i, and the deviation q_{ih} from i's expected wage allows within skill wage dispersion. Plugging this wage decomposition into (2), we have the migration

[2] Allowing migration without an accepted job in the destination, a parameter restriction ensuring all migration is job-to-job requires modelling job search in the destination. For instance, if they search in continuous time with unemployment income b and discount rate r and accept the first job offer that arrives at rate φ, the condition is $\frac{\int w dF_{ik}(w) - w_{ij}}{r} < \frac{w_{ij} - b + \pi_{ilk} - \pi_{ilj}}{\varphi}$ (See Lemma 1 in Appendix A). With risk aversion, migration without a job in the destination is less attractive.

[3] This region-invariance of unobservable skills is plausible especially when studying internal migration and generalization to imperfect transferability of skills is straightforward. Note however, that while I assume that the absolute level of skill is location-invariant, I do not assume that the ranking of a worker in skill distribution is location-invariant. A migrant with above-mean skill in source location may have below-mean skill in the destination if the destination location mean skill μ_k^ν is higher than the source location mean skill μ_j^ν.

[4] Interpreting w as the logarithm of wage, a formally equivalent model follows from a time-equivalent labor supply cost which specifies $u(w, \pi) = e^w(1 - \pi)$ as $ln[e^w(1 - \pi)] \approx w - \pi$.

condition

$$\mu_j - \mu_k + \pi_{jk} < (\rho_k - \rho_j)\nu_i + q_{ik} - q_{ij}, \tag{MC}$$

where $\mu_h := \bar{\mu}_h - \rho_h \mu_h^\nu$ is the compensation paid for zero skill level in location h.

The wage dispersion may have many sources, such as firm heterogeneity in productivity or wage-setting power due to search frictions or worker-firm match effects (see e.g., Mortensen (2003)), but here it is taken as exogenous. I call q_{ih} job match quality. Job match quality is thus interpreted broadly as capturing all variation in wages that, within location, is not due to variation in workers' skills. Thus, also firm effects (as in Abowd et al. (1999)) are part of job match quality. Also, while each job offer may come with an initially unobservable match specific productivity as in Jovanovic (1979), here job match quality is defined at the level of wage and is thus observable to the worker.

To study the selection that the migration condition (MC) generates, specify heterogeneity in I on the different components of wages and see what types of workers satisfy the migration condition. Let

$$\nu_i | i \in I \sim \mathcal{N}(\mu_\nu, \sigma_\nu^2). \tag{3}$$

As those who are in a position to choose whether to relocate or not may be nonrandomly selected, allow $\mu_\nu \neq \mu_j^\nu$. The values of skills in the source and destination regions are then distributed in I as[5]

$$\begin{bmatrix} \rho_j \nu_i \\ \rho_k \nu_i \end{bmatrix} | i \in I \sim \mathcal{N}\left(\begin{bmatrix} \rho_j \\ \rho_k \end{bmatrix} \mu_\nu, \begin{bmatrix} \rho_j^2 & \rho_j \rho_k \\ \rho_j \rho_k & \rho_k^2 \end{bmatrix} \sigma_\nu^2 \right). \tag{4}$$

Specify within skill wage dispersions as,

$$\begin{bmatrix} q_{ij} \\ q_{ik} \end{bmatrix} | i \in I \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_j^2 & 0 \\ 0 & \sigma_k^2 \end{bmatrix} \right). \tag{5}$$

The choices of interregional search and employers' hiring choices may also select on job match quality. Low job match quality increases the relative payoff from job search. Also, if only job-seeker-employer meetings with high job match quality lead to job offers, then the population at risk of job-to-job migration is negatively selected on source location job match quality and positively selected on destination region job match quality. As we will see, such selection is qualitatively similar to the selection that (MC) generates and, thus, abstracting from the selection on job match quality generated by job search and hiring choices simplifies without affecting the qualitative results. The selection into the population at risk of job-to-job migration may also generate correlation between job match quality and skills. For tractability,

[5] $Var[\rho_h \nu_i] = \rho_h^2 Var[\nu_i] = (\rho_h \sigma_\nu)^2$, $Cov(\rho_k \nu_i, \rho_j \nu_i) = E[\rho_k \nu_i \rho_j \nu_i] - E[\rho_k \nu_i] E[\rho_j \nu_i] = \rho_k \rho_j (E[\nu_i^2] - E[\nu_i]^2) = \rho_k \rho_j Var[\nu_i] = \rho_k \rho_j \sigma_\nu^2$.

these potential correlations are ignored here.

Proposition 1. *Given (3), (4), and (5), the expected source location wages of migrants are*

$$
\begin{aligned}
E[w_{ij}|(\mathrm{MC})] &= \mu_j + E[\rho_j \nu_i|(\mathrm{MC})] + E[q_{ij}|(\mathrm{MC})] \\
&= \mu_j + \rho_j \mu_\nu + \frac{\sigma_\nu^2}{\sigma_\Delta}(\rho_k - \rho_j)\rho_j \lambda(z_{jk}) - \frac{\sigma_j^2}{\sigma_\Delta}\lambda(z_{jk}),
\end{aligned} \tag{6}
$$

and the expected destination location wages of migrants are

$$
\begin{aligned}
E[w_{ik}|(\mathrm{MC})] &= \mu_k + E[\rho_k \nu_i|(\mathrm{MC})] + E[q_{ik}|(\mathrm{MC})] \\
&= \mu_k + \rho_k \mu_\nu + \frac{\sigma_\nu^2}{\sigma_\Delta}(\rho_k - \rho_j)\rho_k \lambda(z_{jk}) + \frac{\sigma_k^2}{\sigma_\Delta}\lambda(z_{jk}),
\end{aligned} \tag{7}
$$

where $\sigma_\Delta^2 := \sigma_k^2 + \sigma_j^2 + ((\rho_k - \rho_j)\sigma_\nu)^2$, $\lambda(\cdot) := \phi(\cdot)/(1 - \Phi(\cdot))$, *and where* ϕ *and* Φ *denote the density and distribution functions of the standard normal, respectively, and*

$$
z_{jk} := \frac{1}{\sigma_\Delta}\left(\mu_j - \mu_k + \pi_{jk} - (\rho_k - \rho_j)\mu_\nu\right). \tag{8}
$$

The second terms in the expressions (6) and (7) capture the potential exogeneously taken selection on skills due the choices of interregional search and employers' hiring choices. The third terms capture the already familiar selection on skills that the migration choice generates. The fourth terms capture selection on job match quality.

The model predicts negative selection on job match quality in the source. Given a wage offer, the lower the current job match quality is, the larger is the gain of acceptance and relocation and, thus, the more likely relocation costs are covered. Hence, those with low current job match quality are more likely to relocate than those with high current job match quality. The model predicts positive selection on job match quality in the destination. Low wage offers may not be enough to compensate for the costs of relocation, whereas high job offers may be. Hence, those realizing job offers with high job match quality are more likely to migrate. Figure 1 illustrates. Since the difference between destination and source wages have to compensate for the migration cost, a typical job-to-job migrant is negatively selected in her skill-specific wage distribution in the source and positively selected in her skill-specific wage distribution in the destination.

The model nests the selection on skills as in Borjas (1987).[6] The selection of job-to-job migrants on skills is formally equivalent to selection of migrants in the Roy-Borjas model. However, the interpretation of the mechanism of selection is slightly different. High-skilled job-to-job migrants are more likely to receive acceptable job offers from locations with wide wage distributions, while low-skilled job-to-job migrants are more likely to receive acceptable job offers from locations with narrow wage distributions.

[6] For $\mu_\nu = \mu_j^\nu = \mu_k^\nu = q_{ik} = q_{ij} = 0$. See Corollary 2 in the Appendix.

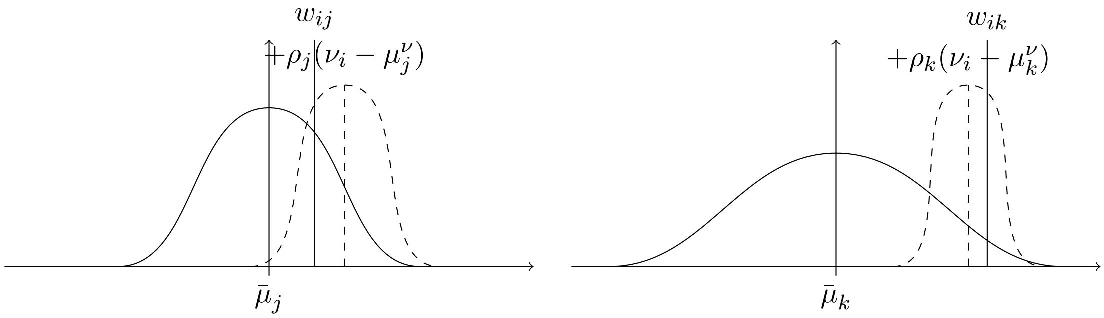

Figure 1: **Source and destination wages of job-to-job migrants.** *Notes:* Depiction of source and destination location distributions of compensation for skills and wage dispersions for a worker with skill ν_i, current wage w_{ij} and an offered wage w_{ik}.

Whether selection on skills or on job match quality dominates selection depends on the relative prices of skills ρ_h and the spreads of within-skill wage dispersions σ_h^2: If the variation in skill prices across locations is low, locational variation in ρ_h may not be enough to induce relocation. With enough within skill wage dispersion in both locations, however, a change in job match quality may be enough to incentivize relocation and selection on job match quality dominates. On the other hand, with large locational differences in returns to skills relative to within-skill wage dispersion, selection on skills dominates.

Borjas (1987) categorizes the possible selection patterns into positive selection, negative selection, and refugee sorting (inverse sorting (Borjas, 2014)). Positive and negative selection have been used in further theoretical and empirical work. The pattern of refugee sorting, however, in its somewhat narrow interpretation of high-skilled but low-wage emigrants suppressed in communist countries immigrating to market economies and earning above-average wages there, has remained a mere mathematical possibility. Selection on job match quality with zero correlation between source and destination location match qualities produces a pattern that combines negative selection in the source and positive selection in the destination. A new interpretation of this pattern arises as we view workers fleeing their bad job matches for better job matches.

We can interpret the model's relation to the Roy-Borjas model in two ways. First, the extension can be seen as allowing within-skill wage dispersion. Without within-skill wage dispersion, each worker is always compensated exactly the value of their skills, and there is no selection on job match quality. Second, if we interpret the wages in Roy-Borjas model as means of the within skill wage dispersions, then the extension is the observability of wage realizations in the source and destination before relocation choice. If the wages are not observed, then the choices are made on expected wages, which with risk neutrality is equivalent to the model without within-skill dispersion.

The expected wages of the Roy-Borjas model can also be interpreted more in line of the human capital approach to migration of Sjaastad (1962) as the expectations of discounted income streams. Such an interpretation allows the wage expectation to contain the option value of further job mobility in the destination. In contrast, interpreting the offered destination wage w_{ih} as an intertemporal utility stream may seem to restrict the model of job-to-job migration to no further job mobility in the destination. However, the mobility condition (MC) is derived from on-the-job search model (1) that clearly allows job mobility in the destination. While Sjaastad's human capital approach and the Roy-Borjas model typically assume irreversibility of migration choices, containing the choice of staying, here staying does not preclude migration later. Later job-to-job migration to destination remains possible even after declining a job offer and, thus, in comparing the choices of staying and migration, the on-the-job mobility prospects in the destination cancel out.

4 Empirical Strategy

I now proceed to attempt to provide evidence for the predicted selection on job match quality. I start by reformulating the model of selection of job-to-job migrants as a model of selection on wage disturbances where wage disturbances are defined as the the sum of the values of unobservable skills and job match quality. This model is used to interpret selection on wage residuals and is the empirically relevant model as the value of unobservable skills and job match quality are not separately observed. I then discuss how due to this confounding of unobservable skills and job match quality, the comparison of job-to-job migrants to stayers may not identify selection on job match quality and describe an alternative strategy comparing job-to-job migrants to commuters.

4.1 Selection on Disturbances

To model selection on disturbances, partial out the observable factors of productivity such that $w_{ih} = \bar{\mu}_{ih} + u_{ih}$ where $\bar{\mu}_{ih}$ is the component of wage that can be predicted by i's observable characteristics including location fixed effect. Reinterpret ν_i as the unobservable skills of i, ρ_h as the price of unobservable skills in location h, μ_h^ν as the mean unobservable skill in location h, and $\mathcal{N}(\mu_\nu, \sigma_\nu^2)$ as the distribution of unobservable skills in the population at risk of job-to-job migration. Disturbances are assumed to be the sum of the value of (demeaned) unobservable skills and job match quality.[7]

Assumption 1. $u_{ih} = \rho_h(\nu_i - \mu_h^\nu) + q_{ih}$.

Corollary 1. *The expected source location disturbances of migrants are*

$$E[u_{ij}|(\mathrm{MC})] = \rho_j(\mu_\nu - \mu_j^\nu) + \frac{1}{\sigma_\Delta}\big(\sigma_\nu^2(\rho_k - \rho_j)\rho_j - \sigma_j^2\big)\lambda(z_{ijk}), \tag{9}$$

and the expected destination location disturbances of migrants are

$$E[u_{ik}|(\mathrm{MC})] = \rho_k(\mu_\nu - \mu_k^\nu) + \frac{1}{\sigma_\Delta}\big(\sigma_\nu^2(\rho_k - \rho_j)\rho_k + \sigma_k^2\big)\lambda(z_{ijk}), \tag{10}$$

[7] This residual wage decomposition was introduced by Flinn (1986) and Garen (1989). Similar decomposition has been used in migration literature (Borjas et al., 1992; Gould and Moav, 2016; Bartolucci et al., 2018), where the component corresponding to q_{ih} has been interpreted as a location-specific effect.

where $\sigma_\Delta^2 := \sigma_k^2 + \sigma_j^2 + ((\rho_k - \rho_j)\sigma_\nu)^2$, $\lambda(\cdot) := \phi(\cdot)/(1 - \Phi(\cdot))$, *and where* ϕ *and* Φ *denote the density and distribution functions of the standard normal, respectively, and*

$$z_{ijk} := \frac{1}{\sigma_\Delta}\left(\mu_{ij} - \mu_{ik} + \pi_{jk} - (\rho_k - \rho_j)\mu_\nu\right). \tag{11}$$

Selection on disturbances is thus due to both selection on unobservable skills and selection on job match quality. Unobservable skills confound the relationship between job match quality and disturbances in two ways: First, as disturbance is the sum of job match quality and unobservable skills, those with higher unobservable skills have higher disturbances. Second, if the destination region compensates for the unobservable skills well (little) relative to the source region, it generates acceptable job offers disproportionally for those with high (low) unobservable skills selecting them into mobility toward this region. Then there is positive selection on disturbances due to positive selection on unobservable skills.

The population at risk of job-to-job migration is likely not a random sample of location j workers. Prior to being in a position to choose between job-to-job mobility and staying, these workers have chosen to search for jobs interregionally and have received a job offer. From (9), we can see how even if there is negative selection on disturbances in the source at the job acceptance and migration choice, i.e. the second term is negative, the potential positive selection of the population at risk of mobility, $\mu_\nu > \mu_j^\nu$, may mask this. In this case, the positive selection on unobservable skills of those who are in a position to choose whether to migrate or not confounds the effect of job match quality when comparing the mobile to the stayers. Hence, positive selection of migrants relative to stayers on residuals is not evidence against negative selection on job match quality. Looking at (10), on the other hand, positive selection on disturbances in the destination may be due to both selection on job match quality and selection on unobservable skills. Thus, observing positive selection on residuals in the destination is not evidence for positive selection on job match quality. If migrants and stayers are very different in their unobservable skills, then comparing migrants' and stayers' disturbances does not identify selection on job match quality.

4.2 Migrants and Commuters

To control for the nonrandom selection into the possibility of job-to-job migration, I compare the job-to-job migrants to those that also change the location of their workplace, but instead of migrating, start commuting to the new location. Commuting may contain telecommuting. I call these two groups of mobile workers *(job-to-job) migrants* and *commuters*. To the union of these two groups, I refer as the *(job-to-job) mobile*.

Commuters make a relevant group of comparison for three reasons: First, as migrants, commuters are at risk of job-to-job migration: they have both chosen to search for jobs outside their region of residence and have both received a job offer. They may thus be similarly selected

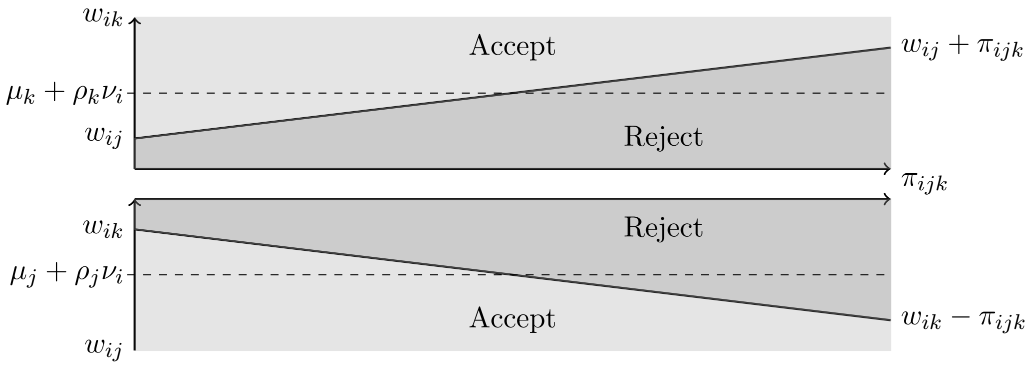

Figure 2: **Reservation and inverse reservation wages.** *Notes:* Reservation wage $w_{ij} + \pi_{ijk}$ and inverse reservation wage $w_{ik} - \pi_{ijk}$ as a function of relocation cost π_{ijk} for a worker with current wage w_{ij} and offered wage w_{ik} and skill ν_i.

on their unobservable skills. Second, when restricting comparisons to source-destination pairs, the migrants and commuters experience the same change in the compensation for unobservable skills ρ_h and the same change in the unobservable location mean skill level effects $\rho_h \mu_h^\nu$. Thus, in the within source-destination location comparison of migrants and commuters, these unobservable location mean skill level effects and their changes cancel out. Third, commuters incur different costs in relocating their labor supplies than migrants.

Relocation costs create what Borjas (1987) calls the scale effect: costs magnify selection effects. Here, higher relocation costs can be covered in two ways: by larger changes in the compensation for unobservable skills or by larger changes in job match quality. Relocation costs enter the degrees of selection (9) and (10) via the inverse Mill's ratio λ and the comparison of migrants and commuters is interpreted as a source of variation in λ to reveal whether variation in mobility costs magnifies selection on job match quality or on unobservable skills.

Figure 2 illustrates how mobility costs magnify selection on job match quality. Looking at the upper half of the figure, the vertical axis tracks the support of i's within skill wage dispersion in location k. The worker i's reservation wage $w_{ij} + \pi_{ijk}$ is an element on this support and increasing in mobility cost. Given current wage, w_{ij} the larger is the mobility cost, the smaller subset of possible location k wages are acceptable, and the larger is job match quality required for i to accept a job offer and migrate. Looking at the lower half of the figure, the vertical axis tracks the support of i's within skill wage dispersion in location j. Worker i's inverse reservation wage $w_{ik} - \pi_{ijk}$ is an element on this support and decreasing in mobility cost. Given a job offer, the higher is the mobility cost, the lower is current job match quality required for this offer to be acceptable. Summing up, the gap in source and destination location job match qualities required for relocation increases in relocation costs.

4.2.1 Mobility Mode Choice

To understand how migrants and commuters may systematically differ in their incurred mobility costs and, thus, how the chosen mobility mode may work as a proxy for mobility costs, I extend the model with mobility mode choice. Suppose there are available two mobility technologies, migration (m) and commuting (c), for relocation of labor supply. Allow heterogeneity in relocation costs such that each worker in the population at risk of job-to-job migration has a cost type $(\pi^c_{ijk}, \pi^m_{ijk})$, a cost of employing c and the cost of employing m, respectively. These costs depend on a variety of factors such as access to public transport or a car, housing, family and social ties, and while known to the workers, may be unobservable to the researcher.

Each worker prefers the least costly technology. These preferences divide the population at risk of job-to-job migration into *potential migrants*, those who would migrate if they accepted their job offer, and *potential commuters*, those who would commute if they accepted their job offer. We can now see how potentially mobile workers' optimal mobility technology choices generate systematic variation in the mobility costs across those preferring different technologies:

Proposition 2. *Let π^c_{ijk} and π^m_{ijk} be independently and normally distributed with means $\bar{\pi}^m_{jk}$ and $\bar{\pi}^c_{jk}$ and equal standard deviations. If and only if $\bar{\pi}^m_{jk} > (<)\ \bar{\pi}^c_{jk}$, we have $\lambda^m > (<)\ \lambda^c$, where $\lambda^m := E[\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}]$ and $\lambda^c := E[\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}]$.*

When one mobility technology tends to be more costly than the other, the preference over mobility technologies is informative of potential mobility costs. If, say, migration tends to be more costly than commuting, then the distribution of mobility costs of potential migrants dominates the distribution of mobility costs of potential commuters. Importantly, this holds despite everyone choosing the least costly technology. Potential migrants have lower costs of migrating than commuting. Nevertheless, if migration costs tend to dominate commuting costs in the population at risk of job-to-job mobility, those that prefer migration tend to have high commuting costs rather than small migration costs. The optimal mobility mode choice preserves the dominance of migration costs over commuting costs in the sense that the potential migrants' relocation costs dominate the potential commuters' relocation costs. With the distributional assumptions of Proposition 2, this dominance is in the sense of monotone likelihood ratio.[8] Thus, an observation that a mobile worker prefers to migrate is informative of high mobility costs in the sense of *more favorable than* -relation of Milgrom (1981). Monotone likelihood ratio dominance then also implies that the mobility costs of potential migrants first order stochastically dominate the mobility costs of potential commuters. This first order stochastic dominance then implies the variation in the expected nonlinear but increasing scale term λ across potential migrants and commuters.

Proposition 2 thus establishes how the preferred mobility technology is a relevant proxy for

[8]Noe (2020) studies the preservation of distributional rankings in *competitive selection*, that is, in optimal choice among realizations of two random variables. He shows how monotone likelihood ratio dominance of random variables is preserved when conditioning on the optimal choice over the realizations of random variables for a large class of distributions, such as normal, lognormal, gamma and Pareto distributions. The assumption of normal distribution in Proposition 2 is thus not crucial. Also, for normal distribution, monotone likelihood ratio dominance survives competitive selection for dependent random variables. Independence of mobility costs in Proposition 2 could thus be relaxed.

mobility costs. As workers with different costs take different actions, these actions, as revealed preferences, convey information about mobility costs and so become proxies of mobility costs. As we will see, only by assuming that migration costs tend to be larger than commuting costs, will the model produce predictions that are consistent with the data. Presumably, migrants, who relocate both their residence and work locations, incur on average higher costs than commuters who relocate only their work locations. To fix ideas, assume migration costs tend to be larger than commuting costs:

Assumption 2. $\bar{\pi}^m_{jk} > \bar{\pi}^c_{jk}$.

The preferred mobility technology is observable only for those who accept a job offer and reveal their preference. Define an indicator function $D : \{i \in \mathcal{I} : (\text{MC})\} \mapsto \{0, 1\}$ among the mobile as

$$D_i = \mathbb{1}(\pi^c_{ijk} - \pi^m_{ijk} + \omega_i > 0), \tag{12}$$

where ω_i is a mean zero error. This error may be an error in worker's optimization or a measurement error. For migrants, $D_i = 1$; for commuters, $D_i = 0$. In an equation relating residuals and employed mobility technology, the employed mobility technology D_i is an exogenous proxy for mobility costs if the error ω_i is independent of the components of wage disturbances.

Assumption 3. $(\nu_i, q_{ij}, q_{ik})|i \in \{i \in \mathcal{I} : (\text{MC})\} \perp\!\!\!\perp \omega_i$.

Assumption 3 clearly rules out the dependence of ω_i on the offered wage. Hence, the preferred mobility technology is assumed to be independent of both source and destination location wages.[9] We thus think of workers, given a potential destination location, having committed to a mobility mode such that the offered wage does not determine the mobility mode.

Finally, to formalize the core idea of the empirical strategy that potential migrants and potential commuters, having all passed the first two selective hurdles to become potential job-to-job migrants, are homogeneous in their unobservables:

Assumption 4. $(\nu_i, q_{ij}, q_{ik})|i \in \mathcal{I} \perp\!\!\!\perp (\pi^c_{ijk}, \pi^m_{ijk})$.

Assumption 4 implies that potential migrants and potential commuters are similar in their disturbances. Assumption 4 is not necessary in identifying selection patterns on residuals that

[9] This follows from both wage-independent and time-equivalent mobility costs. For additive wage-independent mobility costs, those with $\pi^m_{ijk} > \pi^c_{ijk}$ prefer commuting while those with $\pi^m_{ijk} < \pi^c_{ijk}$ prefer migration. Time-equivalent mobility costs are $w_{ih}\pi^c_{ijk}$ and $w_{ih}\pi^m_{ijk}$. Again, those with $\pi^m_{ijk} > \pi^c_{ijk}$ prefer commuting while those with $\pi^m_{ijk} < \pi^c_{ijk}$ prefer migration. It also does not matter whether mobility cost is specified time-equivalent with respect to source or destination location wages. This applies to the interpretation of w as a logarithm of wage as well. Specifying the relocation costs as time-equivalent with respect to the source location wage, the condition to migrate $e^{w_{ik}} - \pi_{ijk}e^{w_{ij}} > e^{w_{ij}}$ is, using the approximation $\ln(1 + \pi) \approx \pi$, equivalent to $w_{ik} - w_{ij} > \pi_{ijk}$. Specifying the mobility cost as time equivalent with respect to the destination location wage, the condition $e^{w_{ik}} - \pi_{ijk}e^{w_{ik}} > e^{w_{ij}}$ is, using the approximation $\ln(1-\pi) \approx -\pi$, again equivalent to $w_{ik} - w_{ij} > \pi_{ijk}$.

are likely due to selection on job match quality: for that, we only need migrants and commuters to be similar enough. However, it allows a structural interpretation of the different degrees of selection across migrants and commuters in terms of the model.

5 Data and Empirical Definitions

This section describes the data, gives an empirical definition of job-to-job migrants and commuters, and describes the computation of wage residuals.

5.1 Data

I use total population annual individual level data compiled from various administrative registers provided by Statistics Finland.[10] For each observation, I assemble data from three periods: the year before the potential mobility event $t-1$, the year of the potential mobility event, t, and one year after the potential mobility event $t+1$.[11] The sample for a year t consists of those aged weakly between 30 and 60 in year t and who were salaried employees, alive and in Finland in the last week of years $t-1$, t and $t+1$, who had positive earnings and zero registered unemployment days in these years and to whom all variables used in the analysis are observed. Students and retirees defined by the longest principal activity during any of the years $t-1$, t, or $t+1$ are excluded. The age restriction aims to remove mobility of students, first-time movers and mobility that may occur with retirement from the sample. I also, for reasons explained in Section 5.3.1, require that employer and establishment do not change between years t and $t+1$. I pool data such that $t \in \{t \in \mathbb{N} : 2006 \leq t \leq 2014\}$.[12]

5.2 Empirical Mobility

The *mobile* are defined as those who change the location of their employment. This captures both labor related migration and changes in commuting destinations. The *stayers* are those who are not mobile and do not change the location of their residence. The mobile are further categorized as migrants and commuters according to their post-mobility residential locations. I do not aim to separate job-to-job migration from other types of migration, but to restrict the sample such that the observed mobility is likely job-to-job.

First, I define *job movers* as those whose postal code area or municipality of workplace changes between years $t-1$ and t and, to capture job-to-job transitions, who have zero days in registered unemployment in year t.

Next, for all job-movers, I compute the distance from the location of their residence in year $t-1$ to the location of their workplace in year t. The distance to the location of the

[10]The data are available in data sets called FOLK modules: FOLK Basic, FOLK Employment and FOLK Cohabitation modules are used (Statistics Finland, 2023d).

[11]Figure 4 uses data on periods $t-5, t-4, ..., t+1$.

[12]The empirical analysis uses R Statistical Software (R Core Team, 2022) and R-packages tidyverse (Wickham et al., 2019) for data wrangling, plm (Croissant and Millo, 2008) for fixed-effects regression, and lmtest (Zeileis and Hothorn, 2002) and sandwich (Zeileis, 2004, 2006) for uncertainty estimation.

new job is defined differently for those who eventually commute and those who eventually migrate: For the commuters, the data contains information on the commuting distance in year t as a Euclidean distance between job and residence locations with an accuracy of 250m by 250m squares computed in Statistic Finland. As commuters, by definition, do not change their residence, the distance between their residence and their new job equals their commuting distance in year t.

For the migrants, the distance to new job location does not equal the commuting distance in year t. Rather, the distances to new job locations are computed as the Euclidean distances between the centroids of the postal code area they resided in year $t-1$ and the postal code area of new job in year t.[13] The different accuracies used in measuring the distance to new job location are unlikely to be an issue, as the distances to new job location are typically longer for migrants than for commuters. Hence, relatively, distances measured using postal code areas are probably not subject to larger measurement errors than distances computed using the 250m by 250m squares. On the other hand, Euclidean distances between postal code area centroids are subject to the most severe measurement errors for small distances. Using information on commuting distances for these distances avoids this problem. All location information is from the last week of the year, similarly to employment information.

Next, the job-movers are partitioned into mobile and nonmobile. The extent of dislocation that qualifies as mobility is defined by a distance threshold of 50 kilometers. This number is somewhat arbitrary, chosen to ensure that there are both migrants and commuters in the sample. However, the results are robust to variation in this threshold. *Stayers* are now defined as those who are not job-movers nor mobile. To remove potential residential migration from the sample, all observations that are not classified as stayers or mobile are discarded.

The mobile are then categorized into migrants and commuters based on their post-mobility residential locations relative to pre-mobility residential locations. If the residential location does not change, a mobile individual is defined as commuter. Otherwise, she is a migrant. Those mobile who migrate such that the distance between the location of their residence and their job increases are discarded. Such moves are likely motivated by factors unrelated to labor market.

A concern with this categorization is that commuting is often temporary. Rigidities in housing markets may force an individual to commute for a while even if she prefers migration. Thus, defining migration and commuting by the residential locations at the end of the year of mobility may classify willing migrants as commuters due to lags in residential adjustment. Migrants could be distinguished from commuters by studying their residential locations in all years after the mobility event. However, requiring commuting, say, until after n years after the mobility event, would drop mobile observations whose employment lasted fewer than n

[13] I use the information on the coordinates of the centroids of postal code areas of job and residence in ETRS-35TMFIN format (Universal Transverse Mercation) for which a reasonable approximation of the distance between a pair of coordinates is the Euclidean distance. The data on centroids of postal code areas are from Statistic Finland's Paavo postal code area statistics database (Statistics Finland, 2020).

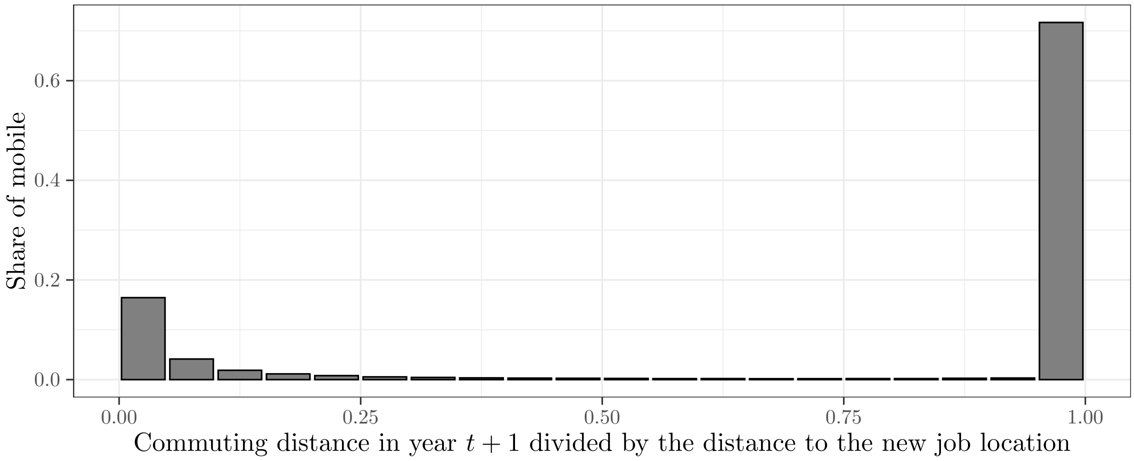

Figure 3: **New commuting distance as a fraction of distance to new job location.** *Notes:* Distribution of commuting distances in year $t+1$ relative to the distance between the location of residence in year $t-1$ and location of work in year t among the mobile. See definitions of the mobile and the distances in Section 5.2.

years from the analysis. The loss of observations would be nonrandom. As a compromise, commuters are defined to be those that have still not changed their original residence location in year $t+1$.[14]

The choice between migration and commuting is not truly binary: migrants may have a strictly positive post-mobility commuting distance as well. Figure 3[15] displays a non-binary classification of migrants and commuters by the ratio of their post-mobility commuting distance to the distance between their pre-mobility residential location and the new job location. This measure describes the fraction of commuting distance left after potential migration. For commuters, this ratio is one as their new commuting distance is the distance to the new job. Moving closer to the new job makes this ratio smaller. We see that the fraction of approximate corner solutions is large enough for the binary classification of migrants and commuters to be a reasonable approximation.

Table 1 presents descriptive statistics. The commuters are more often males and older than the migrants. The migrants are clearly less constrained by family: they less often have a working spouse and children, and they more often live alone. The migrants are also less constrained by housing: they more often rent and less often own their dwelling. Job-to-job migrants tend to have higher education than commuters. Choices between migration and commuting seem to autocorrelate as well: previous migration experience is more common among the migrants, whereas previous commuting experience is more prevalent among the commuters. We also observe large migration costs for those who chose to commute.

[14] Post-mobility earnings are computed using the earnings information of year $t+1$ and by requiring that the employer does not change between years t and $t+1$. See Section 5.3. Given this constraint, the definition of mobility mode does not drop observations from the analysis.

[15] Figures depicting empirical results are created with R-packages ggplot2 (Wickham, 2016) and tikzDevice (Sharpsteen and Bracken, 2023).

	Mobile		Stayers
	Commuters	**Migrants**	
Female	0.32	0.45	0.52
Age, year t	44.6	39.6	45.6
Education, year t			
Basic education	0.10	0.06	0.12
Secondary education	0.34	0.28	0.40
Tertiary education	0.53	0.63	0.46
Doctoral or equivalent	0.03	0.03	0.02
Work, year $t-1$			
Tenure in current job, days	2,203.7	1,579.0	2,869.6
Employment days	359.91	357.98	362.36
Unemployment days	0	0	0
Log wage	4.67	4.55	4.52
Mobility experience, year $t-1$			
Migration experience	0.18	0.39	0.10
Commuting experience	0.78	0.58	0.43
Family, year $t-1$			
Spouse working	0.67	0.45	0.65
Living alone	0.05	0.08	0.06
Living with spouse	0.80	0.58	0.77
Children	0.60	0.43	0.58
Housing, year $t-1$			
Right of occupancy dwelling	0.01	0.02	0.02
Rents the dwelling	0.11	0.37	0.15
Owns the dwelling	0.86	0.59	0.82
Distance to (new) job, km			
Mean	148.34	216.04	14.57
Median	104.52	159.93	6.50
Observations	72,073	29,181	7,287,307

Table 1: **Descriptive statistics by mobility group.** *Notes:* Means for continuous variables unless otherwise mentioned and shares for categorical variables. Commuting experience is defined as having had residence and workplace in different municipalities at least once during the last 5 years. Migration experience is defined having changed the municipality of residence at least once during the last 5 years. A spouse is a married different sex person with cohabitation at least 90 days. A spouse is working if (s)he is employed in the end of the year. For the data, see Section 5.1. See Section 5.2 for the definitions of migrants, commuters and stayers.

5.3 Residuals

Disturbances are estimated as residuals from wage regressions. Residuals are computed from the following wage models estimated in the full sample for each year $t-1$, t and $t+1$:

$$\hat{w}_{ih} = E_{ih}[w] + u_{ih} = x_i'\beta_h + \bar{\mu}_h + u_{ih}, \tag{13}$$

where $\hat{w}_{ih}$ is the estimated log wage (see Section 5.3.1), x_i' contains observable determinants of wages (see Section 5.3.2), β_h is a vector of returns to these observables, and $\bar{\mu}_h$ is the location effect. Let $\hat{u}_{ih}$ be the computed residual of i in location h and $\hat{E}_{ih}[w]$ the predicted wage of i in location h. Wage regressions are estimated within the subsamples of location h workers where here locations are municipalities (LAU 2 regions).

5.3.1 Wages

The data have information on annual earnings: wages are computed as the ratio of annual earnings to annual working days.[16] As earnings and working days are observed only annually, earnings cannot be allocated to jobs in source and destination locations that the mobile hold in the year of mobility t. Thus, year t information cannot be used to measure the pre- or post-mobility wages. Pre-mobility wages are measured using year $t-1$ information. To measure the post-mobility wage and accepted wage offer, I use year $t+1$ information. The latter does not come without potential problems. While the wage of job offer is determined as mobility occurs, the wages in year $t+1$ may partly be a consequence of a certain mobility mode choice. This may happen if source and destination location labor markets differ in their on-the-job search possibilities putting migrants and commuters in different positions with respect to their potential on-the-job search outcomes. I remove this problem by restricting the analysis to those mobile workers who do not change their employer or establishment before the end of year $t+1$. Within firm career advancement and wage growth likely do not depend on the residence location and are, thus, not suspect to this concern. Also, the whole sample, including the stayers, is restricted to those who do not change their employer between years t and $t+1$ to avoid any conditioning of mobility classification on specific employment paths.

The interpretation of year $t+1$ wage as a determinant of mobility follows from the underlying assumption that all studied migration is job-to-job such that the wage in the destination is observed prior to migration choice. As described in Section 5.2, the migrants are defined to highly likely be job-to-job migrants. The requirement of zero days in registered unemployment in the year of mobility is likely to exclude other than job-to-job migrants in the sample of workers with solid labor market histories to whom claiming unemployment benefits in case of unemployment is well incentivized. Nevertheless, it is possible that the sample contains migrants who quickly gained employment only once in destination and without drawing unemployment benefits. For these workers, the destination wage is determined after migration. However, these workers transition from unemployment to employment and are, thus, in a weaker position to realize high job match qualities than the commuters who transition from employment to employment. Hence, these workers would likely bias the estimated positive selection on residuals in the destination toward zero.

5.3.2 Wage Predictors

In computing the residuals, x_i contains gender, whether born in Finland, age, age squared, indicators for the level of education, the field of education, occupation and industry.[17]

[16] Working days are computed from the information of start and end dates of employment spells. Unfortunately, the data do not have information on whether these working days are part-time or full-time employment.

[17] The field of education is classified by the uppermost level of International Standard Classification of Education 2016 (10 categories + 1 for unknowns) (Statistics Finland, 2023b). Occupations are classified by the uppermost level of Classification for Occupation AML 2010 (10 categories +1 for unknowns) (Statistics Finland, 2023a). Industries are classified by the uppermost level of Standard Industrial Classification TOL 2008 (21 categories +1 for unknowns) (Statistics Finland, 2023c).

If commuting time and labor supply competed from the same finite endowment of time, then keeping the wage fixed, the daily income should decrease in commuting time. As the empirical measure of a wage defined above is daily income, wages may be underestimated for those who commute. If commuting costs were then left in the residuals, the commuters would have lower post-mobility residuals than the migrants because they commute. However, if wages compensate for commuting costs, then removing commuting costs from the residuals would rather remove variation in job match quality that we aim to explain. That is, if the measured wages associate negatively with commuting costs, we have the first case, and we should include commuting costs in the wage regression. However, if the association is positive, we have the second case, and we should not include commuting costs in the wage regression. A measure of commuting costs in the data is commuting distance. The estimated coefficient for the commuting distance in the wage regression in the full sample is significantly positive (not reported). Thus, in computing the residuals, no measures of commuting costs are included in the wage regressions.

6 Empirical Results

6.1 Selection on Residuals

Figure 4 Panel A presents the wage paths of migrants, commuters, and stayers. While the wages of stayers are growing along a stable path, the wage growth of the mobile exceed this stable growth path during the period of potential mobility. Moreover, the migrants increase their wages more than the commuters. The decomposition of wages into their predicted and residual components, depicted in Panels B and C, respectively, shows that the residuals drive the pattern observed in Panel A. The migrants have lower residuals than the commuters prior to mobility in the source and higher residuals than the commuters after mobility in the destination. This is consistent with negative selection on job match quality in the source and positive selection on job match quality in the destination. The first column in Table 2 shows that the unconditional differences in source and destination location residuals, that is, residuals at times $t-1$ and $t+1$, respectively, are statistically different from zero with signs as in Figure 4.

Recall indicator D_i among the mobile with $D_i = 1$ if i is a migrant and $D_i = 0$ if i is a commuter. The conditional differences in residuals across the migrants and commuters, for $h = j, k$ are modelled as

$$\hat{u}_{ih} = \kappa_t + \mu_{ljk} + \alpha_1 \Pi_i + \alpha_2 \hat{E}_{ij}[w] + \alpha_3 \hat{E}_{ik}[w] + \tau D_i + \gamma v_i + \varepsilon_i, \tag{14}$$

where $\tau := E[\hat{u}_{ih}|D_i = 1] - E[\hat{u}_{ih}|D_i = 0]$. The triadic residence-source-destination municipality (LAU 2 region) fixed effect μ_{ljk} restricts the comparison to migrants and commuters within residence-source-destination location triplets, where residence location refers to year

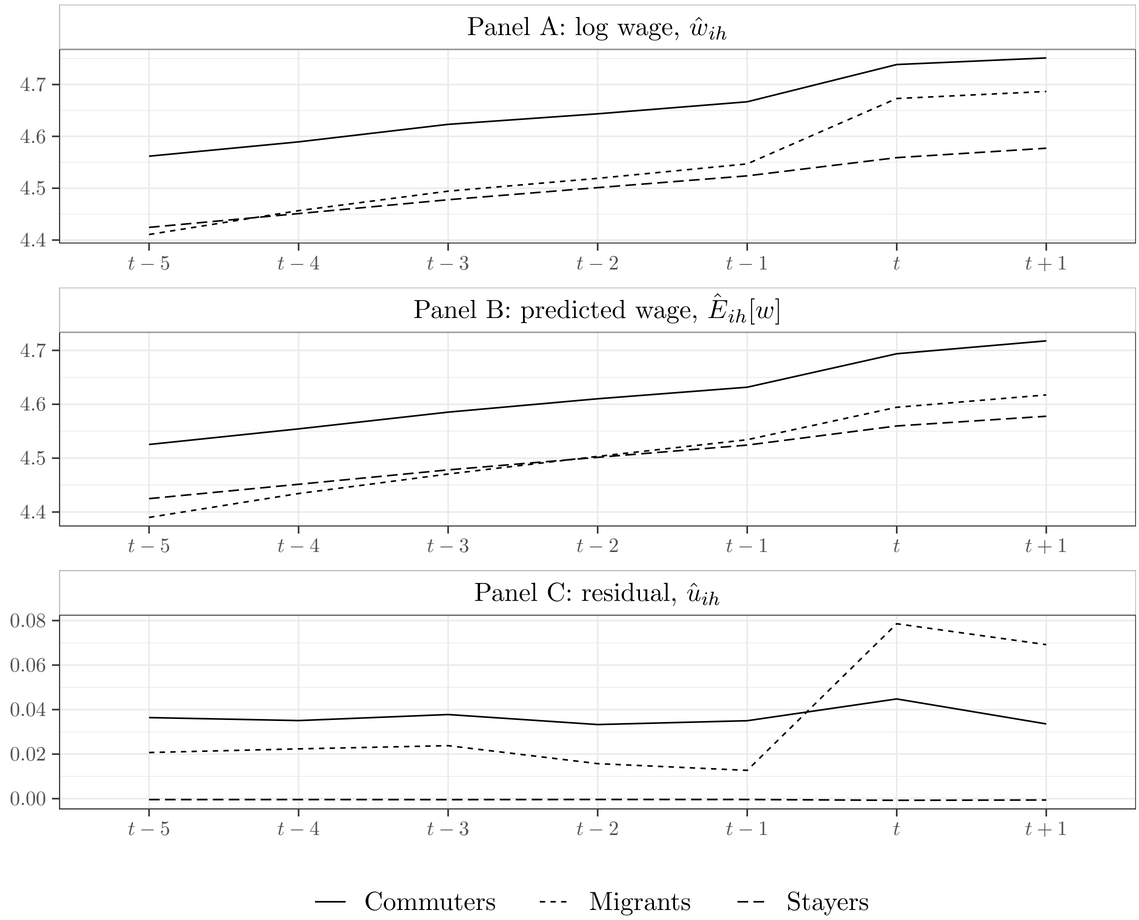

Figure 4: **Wage decompositions of commuters, migrants, and stayers.** *Notes:* Decomposition of the wages (Panel A) of commuters, migrants and stayers into predicted wages (Panel B) and residuals (Panel C). Mobility potentially occurs between time points $t-1$ and t. For the data, see Section 5.1. For the definitions of migrants, commuters and stayers, see Section 5.2. For the wage models, see Section 5.3.

$t-1$ location of residence. Selection on unobservable skills may create dependence of the pre-mobility residuals of the outgoing workers on the destination location and dependence of the post-mobility residuals of the incoming workers on the source location. Selection on unobservable skills also depends on the relative returns to skills between two locations. Thus, to control for the effect of destination region in the model of pre-mobility residuals and the effect of source location in the model of post-mobility residuals, the triadic fixed effect that interacts the source and destination locations is more appropriate than monadic source and destination location fixed effects.

The triadic fixed effect also controls for the differences in residence, source, and destination region characteristics, e.g., prices and availability of housing, amenities, and transport infrastructures connecting the residence and destination locations. The triadic fixed effect μ_{ljk} also controls for potential commuting across municipal borders in year $t-1$. To further control for potential pre-mobility mobility costs, Π_i contains the difference between the distance to new job and the distance to old job and a dummy for recent migration experience.

The time fixed effect κ_t controls for common year of mobility effects. To control for po-

Dependent variable:	$\hat{u}_{ij}$			
Migrant	-0.0223^{***}	-0.0217^{***}	-0.0209^{***}	-0.0134^{**}
(ref: Commuter)	(0.0035)	(0.0052)	(0.0053)	(0.0051)
Dependent variable:	$\hat{u}_{ik}$			
Migrant	0.0356^{***}	0.0310^{***}	0.0311^{***}	0.0276^{***}
(ref: Commuter)	(0.0031)	(0.0047)	(0.0047)	(0.0047)
Dependent variable:	$\hat{u}_{ik} - \hat{u}_{ij}$			
Migrant	0.0578^{***}	0.0525^{***}	0.0518^{***}	0.0410^{***}
(ref: Commuter)	(0.0036)	(0.0056)	(0.0057)	(0.0053)
Spatial controls				
Π_i	No	No	Yes	Yes
ljk-triad FE μ_{ljk}	No	Yes	Yes	Yes
$\hat{E}_{ik}[w]$, $\hat{E}_{ij}[w]$	No	No	No	Yes
Year FE , κ_t	No	Yes	Yes	Yes
v_i	No	No	Yes	Yes
Constant term	Yes	No	No	No
Observations	101,254	101,254	101,254	101,254
Migrants	29,181	29,181	29,181	29,181
Commuters	72,073	72,073	72,073	72,073
R^2, $\hat{u}_{ij}$	0.0005	0.0009	0.0011	0.0317
R^2, $\hat{u}_{ik}$	0.0012	0.0016	0.0019	0.0249
R^2, $\hat{u}_{ik} - \hat{u}_{ij}$	0.0027	0.0020	0.0020	0.0828

Table 2: **Selection on residuals.** *Notes:* *p<0.05; **p<0.01; ***p<0.001. White heteroskedasticity robust standard errors clustered at residence-source-destination municipality level in parenthesis. Spatial controls: d_i is the distance to the new job substracted from the commuting distance in year $t-1$; migration experience is defined as having changed the municipality of residence at least once during years from $t-5$ to $t-1$; μ_{ljk} is a dummy interacting year $t-1$ municipality of residence, year $t-1$ municipality of workplace and year $t+1$ municipality of workplace. $\hat{E}_{ik}[w]$, $\hat{E}_{ij}[w]$ are the predicted wages in destination and source locations, respectively. κ_t is FE for year t. v_i contains all the main and interaction effects of indicators of employer, industry and occupation changes between year $t-1$ and t.

tential losses in firm, industry or occupation specific human capital, the vector v_i contains all the main and interaction effects of indicators of employer, industry or occupation changes between years $t-1$ and t and γ is a conformable vector of coefficients. The predicted wages $\hat{E}_{ij}[w]$ and $\hat{E}_{ik}[w]$ control for the potential differences in migrants' and commuters' productivity that can be explained by observable characteristics and restrict comparison to migrants and commuters with similar values of observable skills.

Table 2 presents the results of estimation of (14) for source and destination location residuals. Migrants are negatively selected relative to the commuters on source location residuals and positively selected relative to the commuters on destination location residuals. The pattern of residuals in Panel C of Figure 4 holds within residence-source-destination-location triplets comparing similar workers with respect to their observable determinants of wages and previously incurred mobility costs and controlling for potential industry, employer, and occupation changes.

6.2 Selection on Job Match Quality

Given Assumption 1, selection on residuals may reflect either selection on job match quality or selection on unobservable skills. This section studies the identification of selection on job match quality. After discussing identification and the assumptions required, I pit selection on unobservable skills and on job match quality against each other by deriving predicted coefficient sign restrictions from the theory. I then discuss the conditions under which difference-in-difference estimation identifies the effect of job match quality.

While the statistical relationship studied is between mobility mode and residuals, the structural relationship of interest is from mobility costs to selection on residuals. Hence, first, neither the antecedent nor the consequent of the structural relationship of interest is observed. Second, the dependent variable is the sum of the variable of interest, job match quality, and another variable. Third, mobility mode is a proxy for the independent variable of interest, mobility costs. Fourth, the consequent of the structural relationship of interest is not the dependent variable but selection on dependent variable that the mean of the dependent variable measures. The interpretation of estimated difference in residuals across migrants and commuters depends on the assumptions we are willing to assume.

Without assumptions, we can interpret the estimated differences in residuals across migrants and commuters only as reflecting differences in residuals across migrants and commuters. With Assumption 1, these estimates may be interpreted as differences in the sum of job match quality and value of unobservable skills. Since time-invariant unobservable skills cannot produce the observed pattern of migrants having lower residuals in the source location and higher residuals in the destination location, the results point to selection on job match quality.

The comparison of source and destination location residuals of migrants and commuters reflects the effect of job match quality alone only if migrants and commuters are equal in their mean unobservable skills. To see this, consider model (14):

$$\begin{aligned}\hat{u}_{ih} &= controls + \tau D_i + \varepsilon_i \\ &= controls + (E[q_{ih}|D_i = 1] - E[q_{ih}|D_i = 0])D_i + \epsilon_i \qquad (15)\end{aligned}$$

with $\epsilon_i := \rho_h(E[\nu_i|D_i = 1] - E[\nu_i|D_i = 0])D_i + \varepsilon_i$. If $E[D_i\varepsilon_i] = 0$, that is if *controls* contain all factors with a simultaneous effect on $\hat{u}_{ih}$ and D_i excluding unobservable skills, unbiased estimation of $E[q_{ih}|D_i = 1] - E[q_{ih}|D_i = 0]$ requires $E[D_i\epsilon_i] = 0$ which requires

$$\rho_h \left(E[\nu_i|D_i = 1] - E[\nu_i|D_i = 0]\right) = 0 \iff E[\nu_i|D_i = 1] = E[\nu_i|D_i = 0]. \qquad (16)$$

There is no direct way of knowing whether (16) holds in the estimation of (14). The theory suggests that if migrants and commuters face different mobility costs, they are selected on their unobservable skills to different degrees speaking against (16). Thus, (16) is not implied by Assumption 4, where there is no conditioning on (MC). However, if the selection on unobservable skills played a relatively minor role in selection, (16) might be plausible owing to the sample choice ensuring high homogeneity with respect to labor market outcomes and the similar hurdles of search and job finding that the mobile pass.

Proposition 3 shows how Assumptions 4 and 3 allow a more structural interpretation of the estimated difference in the selection of migrants and commuters in terms of the model of selection of job-to-job migrants extended with the model of mobility mode choice. These assumptions ensure that the employed mobility mode is an exogeneous and relevant proxy for incurred mobility costs and, thus, that the estimated differences in residuals are due to variation in the scale term λ. This allows us to interpret the differences in residuals across migrants and commuters as different degrees of selection due to variation in mobility costs. The independence of the full distribution is important as the potential degrees of selection are truncated expectations with treatment as the truncation point. To identify the effect of a change in the truncation point on the degree of selection, we need the underlying distribution that is truncated to stay constant as the truncation point varies.

We also need that the total relative selection of migrants and commuters is due to the last hurdle of choosing whether to accept a job offer or not. Interpreting the differences in the selection of migrants and commuters as due to the last hurdle of job acceptance and mobility decision follows from the modelling approach of abstracting from the first two hurdles of job-to-job migration and focusing on the last one. Clearly, if migrants and commuters differ in their mobility costs, they may also face different incentives for interregional job search. On the other hand, the selection that the first two hurdles of becoming a job-to-job mobile worker create is likely qualitatively similar to the selection that the last hurdle creates. Those with low job match qualities are more likely to search for jobs and, if worker-firm match quality is observed during the hiring process, then worker-firm meetings with high job match quality are more likely to lead to job offers.

Proposition 3. *Suppose Assumptions 1, 4, and 3. Then*

(i) the expected difference in the source location disturbances of migrants and commuters is

$$\begin{aligned}\tau_j =&E[u_{ij}|D_i = 1] - E[u_{ij}|D_i = 0]\\ =&\rho_j(E[\nu_i|D_i = 1] - E[\nu_i|D_i = 0]) + E[q_{ij}|D_i = 1] - E[q_{ij}|D_i = 0]\\ =&\frac{1}{\sigma_\Delta}\big(\sigma_\nu^2(\rho_k - \rho_j)\rho_j - \sigma_j^2\big)(\lambda^m - \lambda^c),\end{aligned}$$

(ii) the expected difference in the destination location disturbances of migrants and commuters is

$$\begin{aligned}\tau_k =&E[u_{ik}|D_i = 1] - E[u_{ik}|D_i = 0]\\ =&\rho_k(E[\nu_i|D_i = 1] - E[\nu_i|D_i = 0]) + E[q_{ik}|D_i = 1] - E[q_{ik}|D_i = 0]\\ =&\frac{1}{\sigma_\Delta}\big(\sigma_\nu^2(\rho_k - \rho_j)\rho_k + \sigma_k^2\big)(\lambda^m - \lambda^c),\end{aligned}$$

(iii) the expected difference in difference in residuals of migrants and commuters is

$$\begin{aligned}\tau_k - \tau_j =& E[u_{ik} - u_{ij}|D_i = 1] - E[u_{ik} - u_{ij}|D_i = 0] \\ =& (\rho_k - \rho_j)(E[\nu_i|D_i = 1] - E[\nu_i|D_i = 0]) \\ &\qquad + E[q_{ik} - q_{ij}|D_i = 1] - E[q_{ik} - q_{ij}|D_i = 0] \\ =& \left(\sigma_\nu^2(\rho_k - \rho_j)^2 + \sigma_k^2 + \sigma_j^2\right)^{\frac{1}{2}} (\lambda^m - \lambda^c),\end{aligned}$$

where $\lambda^m := E[\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}]$ and $\lambda^c := E[\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}]$. In each part (i), (ii), and (iii), the second equality follows from Assumption 1 and the third equality from Assumptions 4 and 3.

6.2.1 Model Restrictions

Selection on job match quality and selection on unobservable skills differ in predictions for the differences in wageg residuals across migrants and commuters. We can thus pit the models against each other and see whether selection on job match quality or selection on unobservable skills is more consistent with the data. Intuitively, this theoretical identification is based on the observation that if the selection is driven by unobservable skills, then whichever group has more valuable unobservable skills in source location has to have more valuable unobservable skills in destination location as well. Selection on job match quality, on the other hand, predicts the group with higher mobility costs to have lower residuals in source location and higher residuals in destination location.

There is no direct measure of the location's compensation for unobservable skills. I proxy ρ_h with two measures of wage variation: the standard deviation of wages, denoted $sd(\hat{w}_{ih})$, and standard deviation of residuals, denoted $sd(\hat{u}_{ih})$, among location h workers.[18] These measures are computed at municipal level and in the analysis sample to capture wage variation in the labor markets that the workers in the sample face. Since there is substantial year-to-year variation, for each t, for each municipality, the average over the years $t-1$, t, and $t+1$ is computed. Figure 5 depicts the variation in estimated returns to unobservable skills $\hat{\rho}_h$ across municipalities and years of potential mobility t. The standard deviation of residuals (standard deviation of wages) has a mean 0.389 (0.491) and standard deviation of 0.0602 (0.0535) across municipalities for the (randomly chosen) subset $t = 2010$. These numbers are similar for $t \neq 2010$.

Consider the expected differences in residuals across migrants and commuters given in Proposition 3 and note that by Assumption 2, $\lambda^m > \lambda^c$. Removing within skill wage dispersion from the model, $\sigma_j^2 = \sigma_k^2 = 0$, so that there is no selection on job match quality yields, for

[18] Previous research has proxied ρ_h with various inequality measures of location h wages such as 90-20 income share ratio (Borjas, 1987), ratio of 75th to 25 the percentile of the earnings distribution (Parey et al., 2017), Gini coefficient (Liebig and Sousa-Poza, 2004) standard deviation of log wage (Borjas et al., 1992), standard deviation of residuals from wage regressions (Borjas et al., 1992; Gould and Moav, 2016) and with various measures of returns to education (Gould and Moav, 2016; Fernández-Huertas Moraga, 2013).

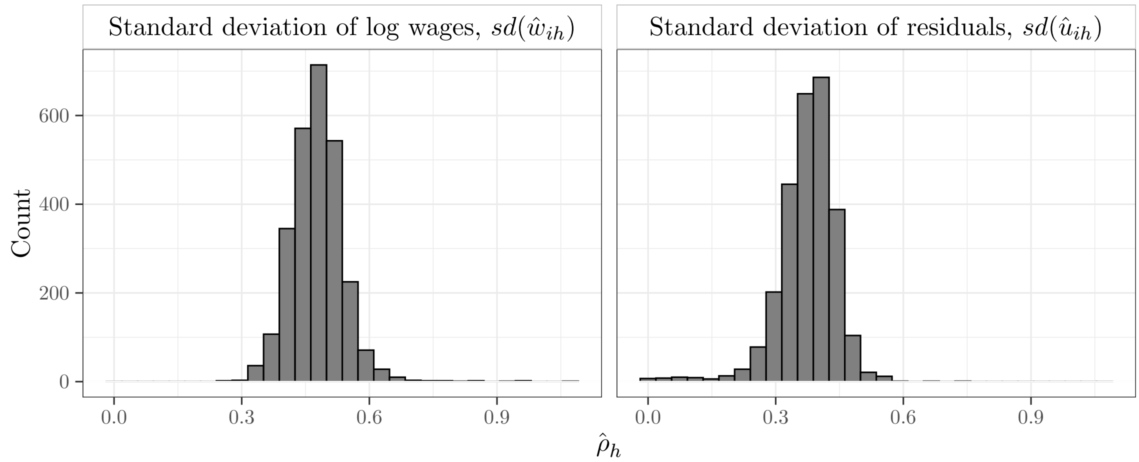

Figure 5: **Distributions of $\hat{\rho}_h$.** *Notes:* Distributions of estimated municipal-year t level standard deviations of log wages and standard deviations of residuals. Standard deviations of residuals and log wages are computed for each municipality in a set of workers working in the municipality for each year $t-1$, t, and $t+1$ and then averaged over years $t-1$, t, and $t+1$. For the computation of wages, see Section 5.3.1. For the computation of residuals, see Section 5.3.

$\rho_k - \rho_j > (<)\ 0$

$$\tau_j = \frac{\sigma_\nu^2}{\sigma_\Delta}(\rho_k - \rho_j)\rho_j(\lambda^m - \lambda^c) > (<)\ 0,$$
$$\tau_k = \frac{\sigma_\nu^2}{\sigma_\Delta}(\rho_k - \rho_j)\rho_k(\lambda^m - \lambda^c) > (<)\ 0.$$

Restricting $\sigma_\varepsilon^2 = 0$ so that there is no heterogeneity in unobservable skills and thus no selection on unobservable skills yields

$$\tau_j = -\frac{\sigma_j^2}{\sigma_\Delta}(\lambda^m - \lambda^c) < 0, \qquad \tau_k = \frac{\sigma_k^2}{\sigma_\Delta}(\lambda^m - \lambda^c) > 0.$$

Table 3 presents estimated differences in the degree of selection on residuals across migrants and commuters by subsamples where the samples are restricted by the sign of $\hat{\rho}_k - \hat{\rho}_j$. As seen in Table 3, we estimate $\hat{\tau}_j < 0$ and $\hat{\tau}_k > 0$ for both $\hat{\rho}_k - \hat{\rho}_j > 0$ and $\hat{\rho}_k - \hat{\rho}_j < 0$. For source location residuals, the estimates are imprecise, but the coefficients are similar to the coefficients estimated in the full sample of mobile workers.

Consider the differences in changes in residuals across migrants and commuters. Removing within skill wage dispersion from the model, $\sigma_j^2 = \sigma_k^2 = 0$, so that there is no selection on job match quality yields for $\rho_k - \rho_j > (<)\ 0$

$$\tau_k - \tau_j = \sigma_\nu(\rho_k - \rho_j)(\lambda^m - \lambda^c) > (<)\ 0.$$

Restricting $\sigma_\nu^2 = 0$ so that there is no heterogeneity in unobservable skills and thus no selection on unobservable skills yields

$$\tau_k - \tau_j = \left(\sigma_k^2 + \sigma_j^2\right)^{\frac{1}{2}}(\lambda^m - \lambda^c) > 0.$$

ρ measure:	$sd(\hat{u}_{ih})$	$sd(\hat{w}_{ih})$	$sd(\hat{u}_{ih})$	$sd(\hat{w}_{ih})$
Sample:	$\hat{\rho}_k - \hat{\rho}_j > 0$		$\hat{\rho}_k - \hat{\rho}_j < 0$	
Dependent variable:	$\hat{u}_{ij}$			
Migrant	-0.0123	-0.0134	-0.0107	-0.0087
(ref: Commuter)	(0.0066)	(0.0069)	(0.0076)	(0.0078)
Dependent variable:	$\hat{u}_{ik}$			
Migrant	0.0391^{***}	0.0463^{***}	0.0161^{**}	0.0134
(ref: Commuter)	(0.0063)	(0.0065)	(0.0074)	(0.0074)
Dependent variable:	$\hat{u}_{ik} - \hat{u}_{ij}$			
Migrant	0.0513^{***}	0.0597^{***}	0.0268^{***}	0.0221^{**}
(ref: Commuter)	(0.0071)	(0.0079)	(0.0079)	(0.0084)
Spatial controls				
Π_i	Yes	Yes	Yes	Yes
ljk-triad FE μ_{ljk}	Yes	Yes	Yes	Yes
$\hat{E}_{ik}[w]$, $\hat{E}_{ij}[w]$	Yes	Yes	Yes	Yes
Year FE , κ_t	Yes	Yes	Yes	Yes
v_i	Yes	Yes	Yes	Yes
Observations	44,487	44,523	37,144	37,108
Migrants	13,953	13,945	13,370	13,378
Commuters	30,534	30,578	23,774	23,730
R^2, $\hat{u}_{ij}$	0.0403	0.0444	0.0475	0.0386
R^2, $\hat{u}_{ik}$	0.0411	0.0360	0.0253	0.0280
R^2, $\hat{u}_{ik} - \hat{u}_{ij}$	0.1070	0.1050	0.0816	0.0778

Table 3: **Selection on residuals in subsamples by sign of $\rho_k - \rho_j$.** *Notes:* *p<0.05; **p<0.01; ***p<0.001. White heteroskedasticity robust standard errors clustered at residence-source-destination municipality level in parenthesis. Spatial controls: d_i is the distance to the new job substracted from the commuting distance in year $t-1$; migration experience is defined as having changed the municipality of residence at least once during years from $t-5$ to $t-1$; μ_{ljk} is a dummy interacting year $t-1$ municipality of residence, year $t-1$ municipality of workplace and year $t+1$ municipality of workplace. $\hat{E}_{ik}[w]$, $\hat{E}_{ij}[w]$ are the predicted wages in destination and source locations, respectively. κ_t is FE for year t. v_i contains all the main and interaction effects of indicators of employer, industry and occupation changes between year $t-1$ and t.

As seen in Table 3, we estimate $\hat{\tau}_k - \hat{\tau}_j > 0$ regardless of the sign of $\hat{\rho}_k - \hat{\rho}_j$. Hence, the overall pattern of signs of the residual differences between the migrants and commuters is consistent with selection on job match quality and inconsistent with selection on unobservable skills.

6.2.2 Difference-in-difference

Job-to-job migrants have lower source location residuals and higher destination location residuals than commuters. Clearly then the change in residuals among the migrants should be greater than among the commuters. Modelling the change in residuals is, however, interesting in its own right since if the compensations for unobservable skills in the source and destination region equal, a difference-in-difference estimator identifies the change in job match quality.

To see this, consider model (14) but with the change in residuals $\hat{u}_{ik} - \hat{u}_{ij}$ as the dependent variable. The coefficient τ then captures the difference in the residual differences as given in part (iii) of Proposition 3, where the first term is the difference in trends due to changes in the compensation for unobservable skills. Note that even if, as visible in Figure 4, the residuals of the migrants and commuters evolve similarly prior to mobility, this is not good

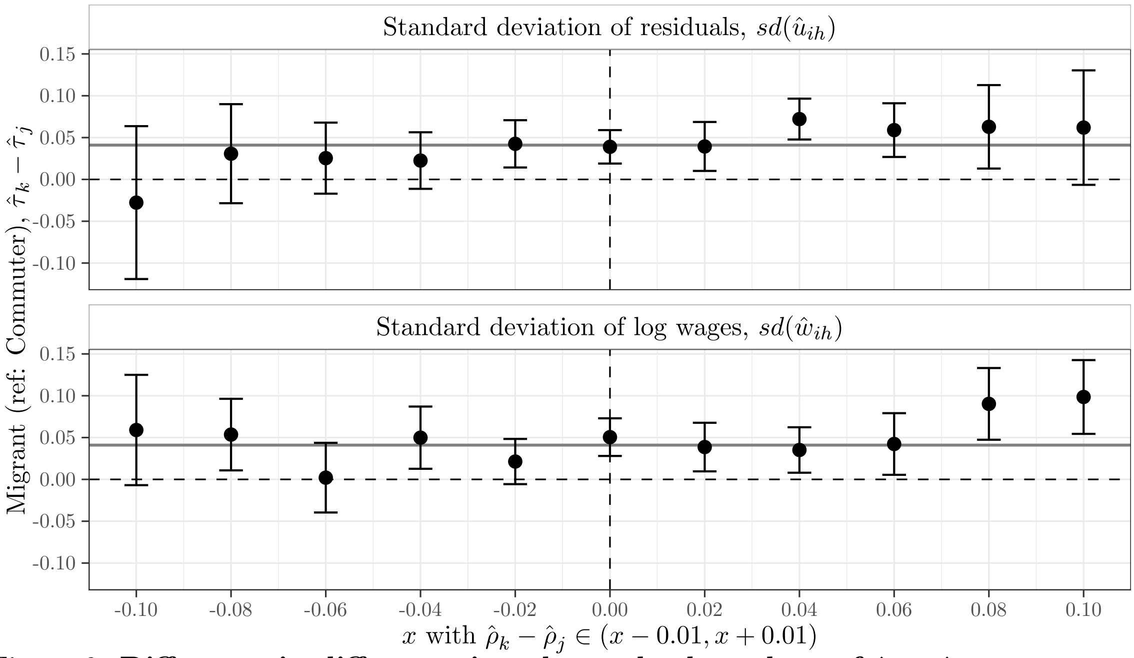

Figure 6: **Difference-in-difference in subsamples by values of $\hat{\rho}_k - \hat{\rho}_j$.** *Notes:* Estimate of $\tau_k - \tau_j$ from model of column 4 in Table 2 in each subsample defined by $\hat{\rho}_k - \hat{\rho}_j \in (x - 0.01, x + 0.01)$ with $x \in \{-0.1, -0.08, ..., 0.08, 0.1\}$. Horizontal solid line depicts the estimate in full sample as in Table 2 row 3 column 4. 95-percent confidence intervals are based on white heteroskedasticity robust standard errors clustered at residence-source-destination municipality level.

evidence for common trends in values of unobservable skills in the year of mobility. In the year of mobility, the compensation for the unobservable skills changes so that the group-specific unobservable effects do not cancel out in the before-after comparison.[19] Thus, if the migrants and commuters are on average different in their unobservable skills, then the compensations for their unobservable skills evolve differently in the year of mobility violating common trend.[20]

However, as can be seen from part (iii) of Proposition 3, there is a common trend with respect to unobservable skills for $\rho_k = \rho_j$. I thus study subsamples restricted by the values the difference in compensation for unobservable skills $\hat{\rho}_k - \hat{\rho}_j$ takes. Figure 6 presents the estimates in these samples. The upper (lower) panel uses as a measure of ρ_h the standard deviation of wages (standard deviation of residuals) of workers in location h. The interesting sample is the one where $\hat{\rho}_k - \hat{\rho}_j$ is restricted to be close to zero. In this sample, the trends due to values of unobservable skills among the migrants and commuters are close to parallel and the group specific time-invariant effects are expected to cancel out. Nevertheless, also in this sample, migrants' changes in residuals seem larger than commuters' changes in residuals.

[19] Since the group unobservable effect $E[\nu_i|D_i]$ and the effect of changing location $\rho_k - \rho_j$ are multiplicative, the common trend assumption is violated for both levels and logarithmic transformation of wages. See e.g. (Lechner, 2011, page 186).

[20] Many have compared migrants' and stayers' wage growth in order to cancel out the individual (or group) fixed effects (Bartel, 1979; Yankow, 2003; Ham et al., 2011). However, the individual fixed effect in a wage regression is the compensation paid for the underlying individual fixed skills. If the price of these skills changes, as might be if the location of labor supply changes, then the individual fixed effect is not invariant in time nor location and does not cancel out in a before-after comparison.

Dependent variable:	$\hat{u}_{ik} - \hat{u}_{ij}$					
Cost proxy, $t-1$:	Lives alone	Has children	Spouse working	Lives rental	Owns house	Owns a car
Migrant * Cost proxy (ref: Commuter * Cost proxy)	-0.0284^{**} (0.0094)	0.0288^{***} (0.0084)	0.0275^{**} (0.0091)	-0.0084 (0.0106)	0.0066 (0.0098)	0.0258^{**} (0.0096)
Spatial controls						
Π_i	Yes	Yes	Yes	Yes	Yes	Yes
ljk-triad FE μ_{ljk}	Yes	Yes	Yes	Yes	Yes	Yes
$\hat{E}_{ik}[w]$, $\hat{E}_{ij}[w]$	Yes	Yes	Yes	Yes	Yes	Yes
Year FE , κ_t	Yes	Yes	Yes	Yes	Yes	Yes
v_i	Yes	Yes	Yes	Yes	Yes	Yes
Observations	101,254	101,254	101,254	101,254	101,254	101,254
Migrants	29,181	29,181	29,181	29,181	29,181	29,181
Commuters	72,073	72,073	72,073	72,073	72,073	72,073
R^2	0.0831	0.0856	0.0831	0.0829	0.0829	0.0833

Table 4: **Difference-in-difference interacted with cost proxies.** *Notes:* *p<0.05; **p<0.01; ***p<0.001. White heteroskedasticity robust standard errors clustered at residence-source-destination municipality level in parenthesis. Spatial controls: d_i is the distance to the new job substracted from the commuting distance in year $t-1$; migration experience is defined as having changed the municipality of residence at least once during years from $t-5$ to $t-1$; μ_{ljk} is a dummy interacting year $t-1$ municipality of residence, year $t-1$ municipality of workplace and year $t+1$ municipality of workplace. $\hat{E}_{ik}[w]$, $\hat{E}_{ij}[w]$ are the predicted wages in destination and source locations, respectively. κ_t is FE for year t. v_i contains all the main and interaction effects of indicators of employer, industry and occupation changes between year $t-1$ and t. A spouse is a married different sex person with cohabitation at least 90 days. A spouse is working if (s)he is employed in the end of the year. Model of column 4 in Table 2 with D_i interacted with a cost proxy.

6.3 Mobility Costs

The differences between the selection on residuals across migrants and commuters are interpreted as reflecting differences in the mobility costs across these two groups. To support the argument, I now study the potential mediation of migrants' and commuters' different degrees of selection by observable proxies of mobility costs.

If mobility costs drive the results, then when migration costs are smaller relative to commuting costs, the differences in residuals between migrants and commuters should be smaller. Hence, I study the interaction effects of mobility mode and observable proxies of mobility cost. Table 4 presents the estimated coefficients for the interaction terms. In the first column, the mobility cost proxy is whether the worker lives alone prior to mobility. Living alone reduces the relative migration costs. Thus, the difference in changes in residuals across migrants and commuters should be smaller. This seems to be the case. Children or a working spouse, on the other hand, increase the relative migration cost. Correspondingly, we observe positive coefficients in the second and third columns. Living rental reduces the relative cost of migration, while owning a house increases the relative cost of migration. The estimated interaction effects, however, do not statistically differ from zero. Owning a car reduces the cost of commuting relative to cost of migration, and as expected, we observe a positive interaction effect.

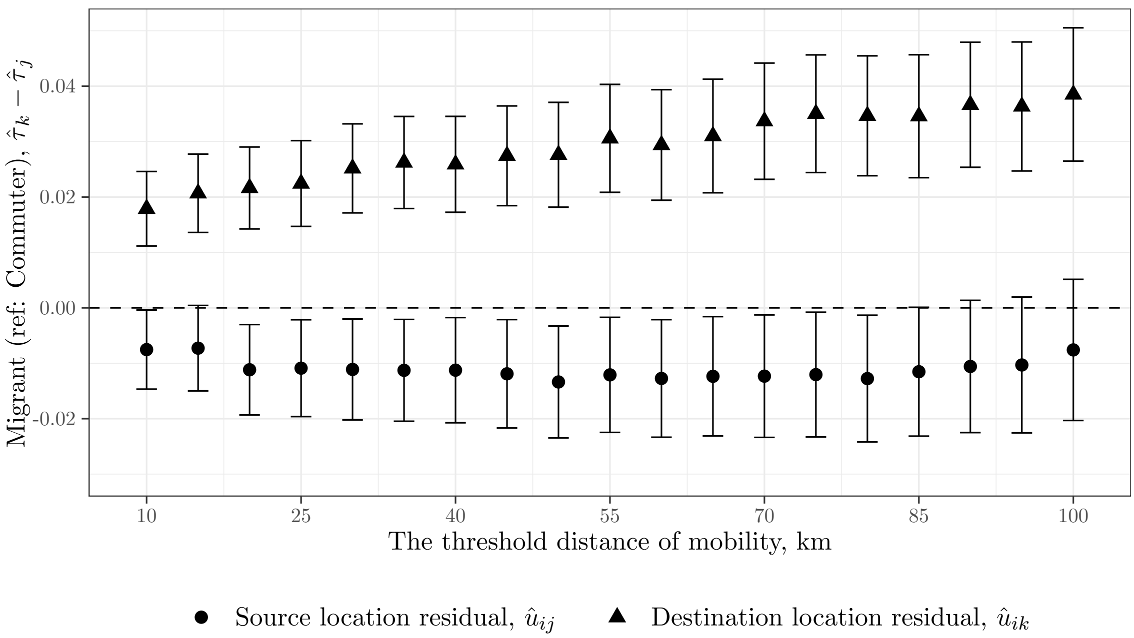

Figure 7: **Robustness with respect to the threshold distance of mobility.** *Notes:* Estimate of $\tau_k - \tau_j$ from model of column 4 in Table 2 with distance thresholds 10, 15, 20, ... 95, 100. 95-percent confidence intervals are based on white heteroskedasticity robust standard errors clustered at residence-source-destination municipality level.

6.4 Robustness

The threshold distance of 50km in defining mobility is somewhat arbitrary, motivated to ensure that the choice between commuting and migration is not trivial and that there are both migrants and commuters in the sample. Figure 7 shows that the estimated difference in job match qualities among the migrants and commuters is robust to variation in this threshold distance.

It could be argued that those with larger wage gains are more likely to afford housing in destination and thus more likely to be able to migrate and this is why migrants are observed to have larger wage gains. However, as seen in Figure 4, commuters on average have higher wages than migrants, and they should, thus, be better able to afford housing in the destination.

Larger wage gains among the migrants could be explained by an autoregressive wage process where the migrants' wages suffer a larger Ashenfelter's dip in their wages prior to mobility than commuters and then regress toward the mean during the period of mobility event. It is, however, clear from Figure 4 that the wage paths of migrants and commuters are similar prior to the mobility event. While Ashenfelter's dip's logic is similar to the negative selection on job match quality in the source here, it does not explain the positive selection in the destination.

7 Conclusion

Migrants often move after having accepted a job in their destinations. This has consequences for migrant selection. First, no self-selection is required for job-to-job migrants to be positively selected when there is competition for jobs. Job-to-job migrants are selected by employers and profit-maximizing employers for each wage select the most productive applicants. Hence, even if a random sample of workers searches for jobs interregionally, those who receive a job offer are not randomly selected. Second, as job-to-job migrants base their migration choice on a realized wage offer, their observed wage distribution in the destination is not their wage offer distribution in the destination but their wage offer distribution truncated below at some threshold wage. The job-to-job immigrants may thus be drawing wage offers from the very same wage distribution as the natives and so be equally able. However, as the job-to-job migrants only accept wages that compensate for their migration costs, they realize on average higher wages than the natives. There is positive selection on destination job match quality. Third, since at the time of migration choice, the destination wage of a job-to-job migrant is fixed, it is independent of the current wage. Thus, given the destination wage, lower current wages more likely induce migration, and the observed wage distribution of job-to-job migrants in the source is not their expected wage distribution but their expected wage distribution truncated above. There is negative selection on source job match quality.

Selection on job match quality can be expected to dominate selection on unobservable skills when within skill wage dispersion is wide relative to differences in returns to skills across locations. This may especially be the case for internal migration. However, while the empirical findings of this paper concern internal migration, job-to-job migration occurs across countries as well. The theoretical findings of this paper are thus relevant also for international migration.

The observed selection on wage residuals among migrants and commuters is consistent with the selection on job match quality predicted by the model. This supports the relevance of the model. Nevertheless, the structural identification of the selection mechanisms here is based on strong ignorability assumptions. The research on migrant selection still lacks attempts to create causal evidence for the determinants of selection. Mobility costs is one place where exogenous variation may plausibly be found. The formalization of the causal effect of relocation costs on selection in the framework of potential outcomes in this paper may support such attempts.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Abramitzky, R. (2009). The effect of redistribution on migration: Evidence from the Israeli kibbutz. *Journal of Public Economics*, 93(3-4):498–511.

Bartel, A. P. (1979). The migration decision: What role does job mobility play? *American Economic Review*, 69(5):775–786.

Bartolucci, C., Villosio, C., and Wagner, M. (2018). Who migrates and why? Evidence from Italian administrative data. *Journal of Labor Economics*, 36(2):551–588.

Birgier, D. P., Lundh, C., Haberfeld, Y., and Elldér, E. (2022). Movers and stayers: A study of emigration from Sweden 1993–2014. *European Journal of Population*, 38:1033–1064.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 77(4):531–553.

Borjas, G. J. (2014). *Immigration Economics*. Harvard University Press, Cambridge.

Borjas, G. J., Bronars, S. G., and Trejo, S. J. (1992). Self-selection and internal migration in the United States. *Journal of Urban Economics*, 32(2):159–185.

Borjas, G. J., Kauppinen, I., and Poutvaara, P. (2019). Self-selection of emigrants: Theory and evidence on stochastic dominance in observable and unobservable characteristics. *The Economic Journal*, 129(617):143–171.

Carliner, G. (1980). Wages, earnings and hours of first, second, and third generation American males. *Economic Inquiry*, 18(1):87–102.

Chiquiar, D. and Hanson, G. H. (2005). International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113(2):239–281.

Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy*, 86(5):897–921.

Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2):1–43.

Detang-Dessendre, C. and Molho, I. (1999). Migration and changing employment status: a hazard function analysis. *Journal of Regional Science*, 39(1):103–123.

Dostie, B. and Léger, P. T. (2009). Self-selection in migration and returns to unobservables. *Journal of Population Economics*, 22(4):1005–1024.

Emmler, J. and Fitzenberger, B. (2020). The role of unemployment and job change when estimating the returns to migration. *IZA Discussion Paper No. 13740*.

Fernandez-Huertas Moraga, J. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics*, 93(1):72–96.

Fernández-Huertas Moraga, J. (2013). Understanding different migrant selection patterns in rural and urban mexico. *Journal of Development Economics*, 103:182–201.

Flinn, C. J. (1986). Wages and job mobility of young workers. *Journal of Political Economy*, 94(3, Part 2):S88–S110.

Garen, J. E. (1989). Job-match quality as an error component and the wage-tenure profiler a comparison and test of alternative estimators. *Journal of Business & Economic Statistics*, 7(2):245–252.

Gould, E. D. and Moav, O. (2016). Does high inequality attract high skilled immigrants? *The Economic Journal*, 126(593):1055–1091.

Ham, J. C., Li, X., and Reagan, P. B. (2011). Matching and semi-parametric IV estimation, a distance-based measure of migration, and the wages of young men. *Journal of Econometrics*, 161(2):208–227.

Heckman, J. J. and Honoré, B. E. (1990). The empirical content of the Roy model. *Econometrica*, 58(5):1121–1149.

Hicks, J. R. (1932). *The Theory of Wages*. Macmillan, New York.

Hunt, G. L. and Mueller, R. E. (2004). North American migration: returns to skill, border effects, and mobility costs. *Review of Economics and Statistics*, 86(4):988–1007.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of political economy*, 87(5, Part 1):972–990.

Kaestner, R. and Malamud, O. (2014). Self-selection and international migration: New evidence from Mexico. *Review of Economics and Statistics*, 96(1):78–91.

Kerr, S. P., Kerr, W., Özden, Ç., and Parsons, C. (2017). High-skilled migration and agglomeration. *Annual Review of Economics*, 9(1):201–234.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, 4(3):165–224.

Liebig, T. and Sousa-Poza, A. (2004). Migration, self-selection and income inequality: An international analysis. *Kyklos*, 57(1):125–146.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391.

Molho, I. (1986). Theories of migration: a review. *Scottish Journal of Political Economy*, 33(4):396–419.

Mortensen, D. (2003). *Wage dispersion: why are similar workers paid differently?* MIT press, Cambridge.

Nakosteen, R. A., Westerlund, O., and Zimmer, M. (2008). Migration and self-selection: Measured earnings and latent characteristics. *Journal of Regional Science*, 48(4):769–788.

Noe, T. (2020). Comparing the chosen: Selection bias when selection is competitive. *Journal of Political Economy*, 128(1):342–390.

Parey, M., Ruhose, J., Waldinger, F., and Netz, N. (2017). The selection of high-skilled emigrants. *Review of Economics and Statistics*, 99(5):776–792.

Pickles, A. and Rogerson, P. (1984). Wage distributions and spatial preferences in competitive job search and migration. *Regional Studies*, 18(2):131–142.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rogerson, P. (1982). Spatial models of search. *Geographical Analysis*, 14(3):217–228.

Rosso, A. (2019). Emigrant selection and wages: The case of poland. *Labour Economics*, 60:148–175.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Saben, S. (1964). Geographic mobility and employment status, March 1962-March 1963. *Monthly Lab. Rev.*, 87:873.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1):1–17.

Sharpsteen, C. and Bracken, C. (2023). *tikzDevice: R Graphics Output in LaTeX Format*. R package version 0.12.4.

Silvers, A. L. (1977). Probabilistic income-maximizing behavior in regional migration. *International Regional Science Review*, 2(1):29–40.

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, 70(5, Part 2):80–93.

Statistics Finland (2020). Paavo (open data by postal code area) 12f7 –9. all data groups. https://www.stat.fi/tup/paavo/index_en.html. Referred on 27th March 2020.

Statistics Finland (2023a). Classification of occupations 2010. https://www.stat.fi/en/luokitukset/am matti/. Referred on 15th April 2023.

Statistics Finland (2023b). National classification of education 2016. https://www.stat.fi/en/luokituk set/koulutus/. Referred on 15th April 2023.

Statistics Finland (2023c). Standard industrial classification tol 2008. https://www.stat.fi/en/luok itukset/toimiala/. Referred on 15th April 2023.

Statistics Finland (2023d). Taika - research data catalogue. https://taika.stat.fi/en/. Referred on 15th April 2023.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Use R! Springer International Publishing, Cham, Switzerland, 2 edition.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Yankow, J. J. (2003). Migration, job change, and wage growth: a new perspective on the pecuniary return to geographic mobility. *Journal of Regional Science*, 43(3):483–516.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, 16(9):1–16.

Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3):7–10.

A Proofs

Lemma 1 (Only job-to-job migration)**.** *Let a migrant from j to k search for a job in k in continuous time with unemployment income b and accept the first offered job arriving at rate φ. Let she discount future with discount rate r. Then worker i does not migrate without having accepted a job in the destination iff* $\frac{\mu_k+\rho_k\nu_i-w_{ij}}{r} < \frac{w_{ij}-b+\pi_{ijk}}{\varphi}$.

Proof of Lemma 1. Let $rV_k = b - \pi_{ilk} + \varphi(\frac{\mu_k+\rho_k\nu_i}{r} - V_k)$ be the asset value of migrating to k and $rV_j = w_{ij} = w_{ij} - \pi_{ilj} + \varphi(\frac{w_{ij}}{r} - V_j)$ the asset value of staying the in the current job in j. The values can then be written as

$$(r+\varphi)V_k = b - \pi_{ilk} + \frac{\varphi}{r}(\mu_k + \rho_k\nu_i) \tag{17}$$

$$(r+\varphi)V_j = w_{ij} - \pi_{ilj} + \frac{\varphi}{r}w_{ij} \tag{18}$$

and thus $V_k < V_j \iff \frac{\mu_k+\rho_k\nu_i-w_{ij}}{r} < \frac{w_{ij}-b+\pi_{ik}-\pi_{ij}}{\varphi}$. □

Lemma 2. *Let $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $\tilde{\Delta}_i \sim \mathcal{N}(E[\tilde{\Delta}_i], \sigma_\Delta^2)$ and $\Delta_i = \tilde{\Delta}_i - E[\tilde{\Delta}_i]$. Then*

$$E[x|\frac{\Delta_i}{\sigma_\Delta} > z] = \mu_x + \frac{E[x\Delta_i]}{\sigma_\Delta}\lambda(z), \tag{19}$$

where $\lambda(\cdot) := \phi(\cdot)/(1-\Phi(\cdot))$, where ϕ and Φ denote the density and distribution functions of the standard normal, respectively.

Proof of Lemma 2. First note that

$$E\left[\frac{x}{\sigma_x}\Big|\frac{\Delta_i}{\sigma_\Delta}\right] = E\left[\frac{x}{\sigma_x}\right] + \frac{Cov[\frac{x}{\sigma_x}, \frac{\Delta_i}{\sigma_\Delta}]}{Var[\frac{\Delta_i}{\sigma_\Delta}]}\frac{\Delta}{\sigma_\Delta} = \frac{1}{\sigma_x}\mu_x + \frac{E[x\Delta_i]}{\sigma_x\sigma_\Delta^2}\Delta_i \tag{20}$$

Thus,

$$E[x|\frac{\Delta_i}{\sigma_\Delta} > z] = \sigma_x E[\frac{x}{\sigma_x}|\frac{\Delta_i}{\sigma_\Delta} > z] = \mu_x + \frac{E[x\Delta_i]}{\sigma_\Delta}E[\frac{\Delta_i}{\sigma_\Delta}|\frac{\Delta_i}{\sigma_\Delta} > z] \tag{21}$$

$$= \mu_x + \frac{E[x\Delta_i]}{\sigma_\Delta}\frac{\phi(z)}{1-\Phi(z)}, \tag{22}$$

where the second equality follows from Lemma 3 and (20). Defining $\lambda(\cdot) := \phi(\cdot)/(1-\Phi(\cdot))$ gives the result. □

Lemma 3. *Let $r, a \in \mathbb{R}$ and $z \sim f_z$, $x \sim f_x$. Then $E[z|x] = c + rx \implies E[z|x > a] = c + rE[x|x > a]$.*

Proof of Lemma 3. Suppose $E[z|x] = c + rx$. Then

$$
\begin{aligned}
E[z|x > a] &= \int_z z f_z(z|x > a)dz = \int_z z \int_{x>a} f_z(z|X = x) f_x(x|x > a)dxdz \\
&= \int_{x>a} \int_z z f_z(z|X = x)dz f_x(x|x > a)dx = \int_{x>a} E[z|x] f_x(x|x > a)dx \\
&= \int_{x>a} (c + rx) f_x(x|x > a)dx = c + r \int_{x>a} x f_x(x|x > a)dx = c + rE[x|x > a].
\end{aligned}
$$

$\square$

Lemma 4. *Let π^m_{ijk} and π^c_{ijk} be independently and normally distributed with means $\bar{\pi}^m_{jk}$, $\bar{\pi}^c_{jk}$, respectively, and identical standard deviations σ_π. If and only if $\bar{\pi}^m_{jk} > (<) \bar{\pi}^c_{jk}$, then $\pi^m_{ijk}|\pi^m_{ijk} < \pi^c_{ijk}$ (is) strictly first order stochastically dominates (dominated by) $\pi^c_{ijk}|\pi^c_{ijk} < \pi^m_{ijk}$.*

Proof of Lemma 4. Note first that for $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ with $X \perp\!\!\!\perp Y$ we have

$$
Pr(X|X < Y) = \frac{Pr(X < Y|X)Pr(X)}{Pr(X < Y)} = \frac{\left[1 - \Phi\left(\frac{x-\mu_y}{\sigma_y}\right)\right] \phi\left(\frac{x-\mu_x}{\sigma_x}\right)}{Pr(X < Y)},
$$

where ϕ and Φ denote the density and distribution functions of the standard normal, respectively. Thus, letting, without loss of generality, $\bar{\pi}^c_{jk} = 0$ and $\sigma_\pi = 1$, we have

$$
f^m(\pi) = \frac{[1 - \Phi(\pi)]\phi(\pi - \bar{\pi}^m_{jk})}{Pr(\pi^m_{ijk} < \pi^c_{ijk})}, \quad f^c(\pi) = \frac{[1 - \Phi(\pi - \bar{\pi}^m_{jk})]\phi(\pi)}{Pr(\pi^m_{ijk} > \pi^c_{ijk})}.
$$

where f^m denotes the density of $\pi^m_{ijk}|\pi^m_{ijk} < \pi^c_{ijk}$ and f^c denotes the density of $\pi^c_{ijk}|\pi^c_{ijk} < \pi^m_{ijk}$. We thus have a likelihood ratio

$$
\frac{f^m(\pi)}{f^c(\pi)} = \frac{Pr(\pi^m_{ijk} > \pi^c_{ijk})}{Pr(\pi^m_{ijk} < \pi^c_{ijk})} \frac{\phi(\pi - \bar{\pi}^m_{jk})}{[1 - \Phi(\pi - \bar{\pi}^m_{jk})]} \frac{[1 - \Phi(\pi)]}{\phi(\pi)}
$$

and, thus, a monotone likelihood ratio if

$$
\frac{d}{d\pi} \frac{f^m(\pi)}{f^c(\pi)} \gtreqless 0 \iff \frac{d}{d\pi} \left(\frac{\phi(\pi - \bar{\pi}^m_{jk})}{[1 - \Phi(\pi - \bar{\pi}^m_{jk})]} \frac{[1 - \Phi(\pi)]}{\phi(\pi)} \right) = \frac{d}{d\pi} \frac{\lambda(\pi - \bar{\pi}^m_{jk})}{\lambda(\pi)} \gtreqless 0, \tag{23}
$$

where $\lambda(\cdot) = \phi(\cdot)/[1 - \Phi(\cdot)]$ is the inverse Mill's ratio.

Suppose $\bar{\pi}^m_{jk} > 0$ ($\bar{\pi}^m_{jk} < 0$). Then, the likelihood ratio (23) is strictly increasing (decreasing), if and only if,

$$
\lambda'(\pi - \bar{\pi}^m_{jk})\lambda(\pi) - \lambda'(\pi)\lambda(\pi - \bar{\pi}^m_{jk}) > (<)\, 0 \iff \frac{\lambda'(\pi - \bar{\pi}^m_{jk})}{\lambda(\pi - \bar{\pi}^m_{jk})} > (<)\, \frac{\lambda'(\pi)}{\lambda(\pi)},
$$

that is, if and only if λ'/λ is strictly decreasing. Since $\phi'(\pi) = -\pi\phi(\pi)$, we have $\lambda'(\pi)/\lambda(\pi) = \lambda(\pi) - \pi$, which is strictly decreasing in π since $\lambda'(\pi) < 1$ (e.g. Heckman and Honoré (1990) (R-2)). Strictly increasing (decreasing) likelihood ratio (23) then implies that f^m (is) strictly first order stochastically dominates (dominated by) f^c. The converse is now clear from contraposition as $\bar{\pi}^m_{jk} < 0$ ($\bar{\pi}^m_{jk} > 0$) implies that f^m is strictly first order stochastically dominated by (dominates) f^c. $\square$

Proof of Proposition 1. Let $\tilde{\Delta}_i = (\rho_k - \rho_j)\nu_i + q_{ik} - q_{ij}$ be the right-hand side of (MC), $\Delta_i = \tilde{\Delta}_i - (\rho_k - \rho_j)\mu_\nu$ and

$$z_{jk} \coloneqq \frac{1}{\sigma_\Delta}\left(\mu_j - \mu_k + \pi - (\rho_k - \rho_j)\mu_\nu\right) \tag{24}$$

such that (MC) can be written as $\frac{\Delta_i}{\sigma_\Delta} > z_{jk}$. Then by Lemma 2,

$$E[\rho_h\nu_i|(\text{MC})] = E[\rho_h\nu_i|\frac{\Delta_i}{\sigma_\Delta} > z_{jk}] = \rho_h\mu_\nu + \frac{\sigma_\nu^2}{\sigma_\Delta}(\rho_k - \rho_j)\rho_h\lambda(z_{jk}) \tag{25}$$

and

$$E[q_{ij}|(\text{MC})] = -\frac{\sigma_j^2}{\sigma_\Delta}\lambda(z_{jk}), \quad E[q_{ik}|(\text{MC})] = \frac{\sigma_k^2}{\sigma_\Delta}\lambda(z_{jk}). \tag{26}$$

The result follows from the decomposition $w_{ih} = \mu_h + \rho_h\nu_i + q_{ih}$. $\square$

Proof of Corollary 1. As Proposition 1 but the condition for relocation is now

$$w_{ij} < w_{ik} - \pi \iff \mu_{ij} - \mu_{ik} + \pi_{jk} < (\rho_k - \rho_j)\nu_i + q_{ik} - q_{ij} \tag{27}$$

and thus let

$$z_{ijk} \coloneqq \frac{1}{\sigma_\Delta}\left(\mu_{ij} - \mu_{ik} + \pi_{jk} - (\rho_k - \rho_j)\mu_\nu\right). \tag{28}$$

The result follows from the decomposition $u_{ih} = \rho_h(\nu_i - \mu_h^\nu) + q_{ih}$. $\square$

Corollary 2. *Proposition 1 nests the Roy-Borjas (Borjas, 1987) model of migrant selection.*

Proof of Corollary 2. Setting $\mu_\nu = \mu_j^\nu = \mu_k^\nu = q_{ik} = q_{ij} = 0$, (6) can be written as

$$E[w_{ij}|(\text{MC})] = \mu_j + \frac{sd(\rho_k\nu_i)sd(\rho_j\nu_i)}{sd(\rho_k\nu_i - \rho_j\nu_i)}\left(\frac{sd(\rho_k\nu_i)}{sd(\rho_j\nu_i)} - \frac{Cov[\rho_k\nu_i, \rho_j\nu_i]}{sd(\rho_k\nu_i)sd(\rho_j\nu_i)}\right)\lambda(z), \tag{29}$$

and (7) can be written as

$$E[w_{ik}|(\text{MC})] = \mu_k + \frac{sd(\rho_k\nu_i)sd(\rho_j\nu_i)}{sd(\rho_k\nu_i - \rho_j\nu_i)}\left(\frac{Cov[\rho_k\nu_i, \rho_j\nu_i]}{sd(\rho_k\nu_i)sd(\rho_j\nu_i)} - \frac{sd(\rho_k\nu_i)}{sd(\rho_j\nu_i)}\right)\lambda(z), \tag{30}$$

where $sd(\cdot) \coloneqq \sqrt{Var[\cdot]}$ and

$$z = \frac{1}{sd(\rho_k\nu_i - \rho_j\nu_i)}\left(\mu_j - \mu_k + \pi_{jk}\right). \tag{31}$$

giving us the conditional expectations as formulated by Borjas (1987). $\square$

Proof of Proposition 2. Suppose π^m_{ijk} and π^c_{ijk} are independently and normally distributed with means $\bar{\pi}^m_{jk}$, $\bar{\pi}^c_{jk}$ and equal standard deviations. Since λ is strictly increasing (e.g. Heckman and Honoré (1990) (R-2)) and z_{ijk} is strictly increasing in π, $\lambda \circ z_{ijk}$ is strictly increasing in π, the result follows from Lemma 4. □

Proof of Proposition 3. Assume Assumptions 1, 4, and 3. Then for parts (i) and (ii) we have

$$
\begin{aligned}
&E[u_{ih}|D_i = 1] - E[u_{ih}|D_i = 0] \\
&= E[\rho_h \nu_i + q_{ih}|D_i = 1] - E[\rho_h \nu_i + q_{ih}|D_i = 0] \\
&= E[E[\rho_h \nu_i + q_{ih}|(\mathrm{MC}), \pi^m_{ijk} < \pi^c_{ijk}]] - E[E[\rho_h \nu_i + q_{ih}|(\mathrm{MC}), \pi^m_{ijk} > \pi^c_{ijk}]] \\
&= \begin{cases} E\left[\rho_j \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_j - \sigma^2_j\right)\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}\right] \\ E\left[\rho_k \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_k + \sigma^2_k\right)\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}\right] \end{cases} \\
&\qquad - \begin{cases} E\left[\rho_j \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_j - \sigma^2_j\right)\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}\right] \\ E\left[\rho_k \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_k + \sigma^2_k\right)\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}\right] \end{cases} \\
&= \begin{cases} \left(\rho_j \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_j - \sigma^2_j\right)\right) E\left[\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}\right] \\ \left(\rho_k \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_k + \sigma^2_k\right)\right) E\left[\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}\right] \end{cases} \\
&\qquad - \begin{cases} \left(\rho_j \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_j - \sigma^2_j\right)\right) E\left[\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}\right] \\ \left(\rho_k \mu_\nu + \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_k + \sigma^2_k\right)\right) E\left[\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}\right] \end{cases} \\
&= \begin{cases} \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_j - \sigma^2_j\right)(\lambda^m - \lambda^c) \\ \frac{1}{\sigma_\Delta}\left(\sigma^2_\nu(\rho_k - \rho_j)\rho_k + \sigma^2_k\right)(\lambda^m - \lambda^c) \end{cases},
\end{aligned}
$$

where the first equality follows from Assumption 1, the second from Assumption 3 and the third and fourth from Assumption 4 and where $\lambda^m := E[\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}]$ and $\lambda^c := E[\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}]$. Part (iii) follows from parts (i) and (ii). □