

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Alasalmi, Juho

# Conference Paper Self-selection of Job-to-job Migrants on Match Quality

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage"

**Provided in Cooperation with:** Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Alasalmi, Juho (2023) : Self-selection of Job-to-job Migrants on Match Quality, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage", ZBW -Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at: https://hdl.handle.net/10419/277633

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



# WWW.ECONSTOR.EU

# Self-selection of Job-to-job Migrants on Match Quality<sup>\*</sup>

Juho Alasalmi<sup>†</sup> University of Konstanz / Pellervo Economic Research

> February 2023 Latest version here

#### Abstract

Many migrants migrate after having found a job in the destination. Relocation choices are then not based on source and destination location wage distributions but on specific realizations from these distributions. I extend the Roy-Borjas migrant selection model with job search, wage dispersion and observability of source and destination wage realizations prior to relocation choice. This model of selection of job-to-job migrants, while nesting the benchmark results on selection on skills, predicts negative selection on source and positive selection on destination job match quality. Using high quality administrative data, 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 facing different costs of mobility, identifies migrants' residual selection (in)consistent with selection on job match quality (unobservable skills) predicted by the theory.

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

# 1 Introduction

Labor related migration is often modelled as a risky investment: migrants compare their labour market prospects in the source and destination locations and potentially relocate in the hopes of finding employment in their destination. Thus, migration choice

<sup>\*</sup>I thank Susanne Goldlücke, Janne Huovari, Giuseppe Ippedico, Elina Jussila, Panu Poutvaara, and many other participants at workshops and conferences at European Public Choice Society Meeting, Lille, France 4/2021; International Institute of Public Finance Annual Congress, Reykjavik, Iceland 8/2021; CEMIR Junior Economist Workshop on Migration Research, Münich 7/2022; VATT, Helsinki 2/2023; Annual Meeting of the Finnish Economic Association, Espoo 2/2023 for valuable comments and discussions.

<sup>&</sup>lt;sup>†</sup>juho.alasalmi@uni-konstanz.de

is based on the source and destination location wage distributions. However, migrants often relocate only after successful on-the-job search, after a job in destination has been found. Thus, migration choice is based on realizations from source and destination location wage distributions. The processes that select migrants with these two different migration strategies are different. While the earlier literature 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.

The selection mechanisms in cases where opportunities are searched and secured in other markets before entering them are complicated, and have received little attention in the literature of market self-selection following Roy (1951). 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 the use of tractable techniques in studying selection while respecting the nonrandom selection that the choice to search and employers' hiring choices create. This modeling 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 selection of job-to-job migrants by extending the Roy-Borjas (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 nonrandomly selected subset of source location workers. Second, I allow within skill wage dispersion: given skill, wage is not deterministic but each worker faces a distribution of potential wages. Third, I allow observability of both the current wage in the source and the potential wage in the destination prior to the migration choice. Choice of job-to-job migration is thus not only based on the mean and variance of source and destination region wage distributions but on specific realizations from these distributions.

The resulting model of selection of job-to-job migrants, while nests the benchmark results on selection on skills, generates novel selection patterns on job match quality. Given current wage, low wage offers may not be enough to compensate for relocation costs whereas 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.

I then provide support for the relevance of the selection model of job-to-job migrants by studying selection on wage residuals to identify the selection patterns on job match quality that the model predicts. Clearly, selection on residuals may be due to selection on unobservable skills. This complicates the identification of job match quality in at least three ways. 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. The positive selection of the subpopulation that receives job offers may mask low 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 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 good new job match in comparison to current job match but due to higher compensation for unobservable skills in their destination than in their source.

It is thus difficult to identify effects of job match quality by comparing migrants to stayers. An alternative comparison group for the job-to-job migrants are those who contract a job outside their location of residence, but who choose to (tele)commute.<sup>1</sup> These two groups of mobile workers defined by their different technology to supply labor outside their current regions of residence make a relevant comparison for three reasons: First, the commuters, like job-to-job migrants, have received and accepted a job offer, and, thus have both self-selected to search interregionally and have been selected by employers. Second, when comparing commuters and migrants within source-destination pairs, commuters and migrants experience the same change in the compensation paid for unobservable skills due to the change of labor market. Third, those who choose to commute and those who choose to migrate incur different relocation costs. Relocation costs, on the other hand, magnify selection effects. Comparison of commuters and migrants thus allows us study how the selection effects change in costs and helps us discern whether, selection on job match quality or selection on unobservable skills is magnified.

This paper contributes to the migrant selection literature in two ways. First, I

 $<sup>^1\</sup>mathrm{In}$  the following, I use the term commuting but commuting throughout the paper may contain telecommuting as well.

connect the Roy-Borjas migrant selection model to job search models and capture jobto-job migration by within skill wage dispersions and the observability of source and destination wage realizations prior to migration. The ensuing patterns of 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" pattern. This interpretation is relevant for market economies: refugees are workers fleeing bad job matches. Second, I establish the relevance of the theoretical extension by providing the first explicit evidence on selection on job match quality both in the source and destination locations. Migrants have 2-3 percent lower pre-mobility residuals and 2-3 percent higher post-mobility residuals than commuters. Selection on job match quality explains these results while selection on unobservable skills does not. The findings imply that selection on job match quality plays a role in selection of job-tojob migrants and 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.

My exploration of these ideas proceeds as follows. The next section positions the work into the existing 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 chosen empirical counterparts of theoretical concepts. Section 6 provides evidence of the selection of migrants on job match quality. Section 7 concludes. All proofs of propositions are in the Appendix.

# 2 Related Literature

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

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 as well. Indeed, Borjas, Kauppinen, and Poutvaara (2019) assess that 70 (50) percent of positive selection on source location earnings is due to unobservable determinants of productivity among (fe)male emigrants. The selection on unobservable determinants of productivity has been studied by studying selection on wage residuals (Borjas, 1991; Abramitzky, 2009; Moraga, 2011; Kaestner & Malamud, 2014; Bartolucci, Villosio, & Wagner, 2018; Borjas et al., 2019). Usually, residuals are interpreted to reflect location-variant compensation for time-invariant unobservable skills. The observed selection on residuals, however, does not always align with the predicted selection on unobservable skills (Chiquiar & Hanson, 2005; Moraga, 2011; Kaestner & Malamud, 2014; Birgier, Lundh, Haberfeld, & Elldér, 2022).

The difficulties of modeling selection on residuals with a model of selection on timeinvariant skills may be due to value of time-invariant unobservable skills not being the only component of residuals. Gould and Moav (2016) decompose unobsevable skills into a location-invariant and location specific components and show how these two types of unobservable skills affect selection differently. However, if there is wage dispersion given skills (Mortensen, 2003) then not all variation in residuals is due to unobservable skills. I decompose the variation in wages not explainable by observables into a locationinvariant skill component and a job match quality component. While the role of job match quality in explaining selection on residuals has been discussed before (Nakosteen, Westerlund, & Zimmer, 2008; Borjas et al., 2019; Birgier et al., 2022), selection on job match quality has so far not been modelled nor empirically identified.

Migrant selection is often studied by comparing out-migrants (Moraga, 2011; Kaestner & Malamud, 2014; Borjas et al., 2019; Rosso, 2019; Birgier et al., 2022) or inmigrants (Chiswick, 1978; Carliner, 1980; Abramitzky, 2009) to stayers. In controlling for the heterogeneity of migrants and stayers, the literature studying migrant selection has also compared incoming migrants from different locations (Borjas, 1987; Abramitzky, 2009), outgoing migrants to different locations (Hunt & Mueller, 2004; Dostie & Léger, 2009; Parey, Ruhose, Waldinger, & Netz, 2017) and also migrants moving between the same locations but working in different industries (Gould & Moav, 2016). On the other hand, the literature aiming to estimate the labor market returns to migration has distinguished wage changes due to job changes and wage changes due to location changes by comparing migrants to job movers (Bartel, 1979; Yankow, 2003; Ham, Li, & Reagan, 2011; Emmler & Fitzenberger, 2020). The comparison of migrants and commuters in this paper adds to this selection of settings by controlling for the effect of job change but still comparing two groups that are both interregionally mobile.

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).<sup>2</sup> The evidence on the respective roles of these two forms of labor related migration is scarce. Saben (1964) reports that 62 percent of high-skilled intercounty migrants moved having accepted a job in the destination whereas 38 percent of other

<sup>&</sup>lt;sup>2</sup>The terms *contracted* and *speculative* migration were coined by Silvers (1977) and later used, for instance, by Pickles and Rogerson (1984); Molho (1986); Detang-Dessendre and Molho (1999).

migrants had a job at hand when migrating in the US in 1962. Detang-Dessendre and Molho (1999), using a small survey of young first-time migrants from rural regions, report job-to-job migration to be more common than migration without a contracted job in the destination in France in 1993. Since the collection of the data used in these studies 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 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, Kerr, Özden, & Parsons, 2017). It is thus likely that job-to-job migration is the dominant form of labor related migration, at least in the developed countries.

While not made explicit, earlier research has often likely studied job-to-job migration for instance, by only allowing short gaps between job spells (Ham et al., 2011) or by defining migration as a change in job location (Emmler & 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, migration literature would benefit in making the possibility of job-to-job migration in the data explicit.

# 3 Selection of Contracted Migrants

Consider three locations or labor markets indexed by h = j, k, l. Let worker *i* reside in location *l*, work in location *j*, and potentially search for a job in location *k*. If worker *i* searches for a job in *k*, she may receive a job offer for a job in *k*. 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*. Only the workers in the population at risk of job-to-job migration can relocate.<sup>3</sup> The set of workers working in location h = j, kis defined by a skill distribution  $\nu_i \sim \mathcal{N}(\mu_h^{\nu}, 1)$ . Skills are time-invariant and perfectly transferable across locations.<sup>4</sup> Let  $F_{ih}$  denote the distribution of wages *i* can potentially

<sup>&</sup>lt;sup>3</sup>Allowing migration without accepted job in the destination, a parameter restriction ensuring all migration is job-to-job requires specifying 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  $\varphi$ , the condition is  $\frac{\int wdF_{ik}(w) - w_{ij}}{r} < \frac{w_{ij} - b + \pi_{ilk} - \pi_{ilj}}{\varphi}$  (See Lemma 1 in Section A.1). Adding risk aversion makes migration without a job in the destination less attractive.

<sup>&</sup>lt;sup>4</sup>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  $\mu_k^{\nu}$  is higher than

earn in location h.

The per-period cost for a worker residing in l of supplying labor in location h is  $\pi_{lh}$ . The worker i maximizes a discounted stream of per period net income  $e_{ilh} = e(w_{ih}, \pi_{lh})$ , h = j, k, 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 for  $i \in I$  is

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

where the second term on the right-hand side is the option value of optimal job acceptance and mobility behavior. Job offers are accepted if and only if  $V_i(e_{ilk}) > V_i(e_{ilj})$ which, since  $V_i$  is strictly increasing, is equivalent to  $e_{ilk} > e_{ilj}$ . Letting  $e(w_{ih}, \pi_{lh}) = w_{ih} - \pi_{lh}$ , we have the job acceptance and mobility condition

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

where  $\pi_{ljk} = \pi_{lk} - \pi_{lj}$  is the cost of relocating labor supply from j to k.<sup>5</sup> For now, for simplicity, suppose all location j workers are residing in the same location l 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  $\nu_i$  and the deviation  $q_{ih}$  from *i*'s expected wage allows within skill wage dispersion. This wage dispersion may have many sources (see e.g. Mortensen (2003)) but here it is taken as exogenous. From the perspective of a worker, the source of such variation is likely of little importance. I interpret  $q_{ih}$  as job match quality. Plugging this wage decomposition into (2), we have the mobility condition

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

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

Even if worker i when evaluating a job offer does not care how her current and offered wage can be decomposed, the decompositions determine the probability of worker ireceiving an acceptable job offer. Thus, to study the selection that the migration condition (MC) generates, specify heterogeneity in I on the different components of the

the source location mean skill  $\mu_j^{\nu}$ .

<sup>&</sup>lt;sup>5</sup>Interpreting w as the logarithm of wage, a formally equivalent model follows from a time-equivalent labor supply cost  $e^w \pi_{lh}$  as  $ln[e^w(1-\pi_{lh})] \approx w - \pi_{lh}$ .

wage and see what types of agents 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  $\mu_{\nu} \neq \mu_{j}^{\nu}, \sigma_{\nu}^{2} \neq 1$ . The values of skills in the source and destination regions in I are then distributed as<sup>6</sup>

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

Specify within skill wage dispersions as,

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

The population at risk of job-to-job migration I may be selected 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 equivalent 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, I ignore these potential correlations here.

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

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

and the expected destination location wages of migrants are

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

 $<sup>\</sup>overline{{}^{6}Var[\rho_{h}\nu_{i}] = \rho_{h}^{2}Var[\nu_{i}] = (\rho_{h}\sigma_{\nu})^{2}, Cov(\rho_{k}\nu_{i},\rho_{j}\nu_{i}) = E[\rho_{k}\nu_{i}\rho_{j}\nu_{i}] - E[\rho_{k}\nu_{i}]E[\rho_{j}\nu_{i}] = \rho_{k}\rho_{j}(E[\nu_{i}^{2}] - E[\nu_{i}]^{2}) = \rho_{k}\rho_{j}Var[\nu_{i}] = \rho_{k}\rho_{j}\sigma_{\nu}^{2}.$ 



Figure 1: Source and destination wages of job-to-job migrants.

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

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

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

Job match quality enters the expected pre-mobility wage (6) with a negative sign indicating 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.

Job match quality enters the expected post-mobility wage (7) with a positive sign indicating 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 the effect of wage dispersion. Since the difference between destination and source wages have to compensate for the migration cost, a typical jobto-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).<sup>7</sup> While the selection of job-to-job migrants on skills is formally equivalent to selection of migrants in the

<sup>7</sup>Setting  $\mu_{\nu} = \mu_{j}^{\nu} = \mu_{k}^{\nu} = q_{ik} = q_{ij} = 0$ , (6) can be written as

$$E[w_{ij}|(\mathrm{MC})] = \mu_j + \frac{sd(\rho_k\nu_i)sd(\rho_j\nu_i)}{sd(\rho_k\nu_i - \rho_j\nu_i)} \left(\frac{sd(\rho_k\nu_i)}{sd(\rho_j\nu_i)} - \frac{Cov[\rho_k\nu_i, \rho_j\nu_i]}{sd(\rho_k\nu_i)sd(\rho_j\nu_i)}\right)\lambda(z),\tag{9}$$

Roy-Borjas model, the interpretation of the mechanism of selection on skills 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 location with narrow wage distributions.

Whether selection on job match quality or on skills dominate selection, depends on the relative magnitudes of the spreads of within skill wage dispersions  $\sigma_h^2$  and distributions of expected wages  $\rho_h \sigma_{\nu}^2$ : If the variation in the distributions of expected wages across locations is low, the locational variation in  $\mu_h$  and  $\rho_h$  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. This is especially likely to be the case in internal migration. On the other hand, if there are large locational differences in wage distributions relative to within skill wage dispersion, then 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 inspired further theoretical and empirical work. However, the pattern of refugee sorting, in its somewhat narrow interpretation of high-skilled but low wage emigrants suppressed in communist countries immigrating to non-communist countries and earning above average wages in market economies has remained a mere mathematical possibility. Selection on job match quality here with the zero correlation between source and destination location match qualities, corresponds to the case of Borjas' refugee sorting. Here, we reinterpret refugee sorting as workers leaving jobs that do not pay them what they would expect to earn in the markets into jobs that do, as if refugees fleeing low paying jobs to high paying jobs. This interpretation makes the refugee sorting pattern relevant also in studies focusing on market economies.

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

and (7) can be written as

$$E[w_{ik}|(\mathrm{MC})] = \mu_k + \frac{sd(\rho_k\nu_i)sd(\rho_j\nu_i)}{sd(\rho_k\nu_i - \rho_j\nu_i)} \left(\frac{Cov[\rho_k\nu_i, \rho_j\nu_i]}{sd(\rho_k\nu_i)sd(\rho_j\nu_i)} - \frac{sd(\rho_k\nu_i)}{sd(\rho_j\nu_i)}\right)\lambda(z),\tag{10}$$

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

$$z = \frac{1}{sd(\rho_k \nu_i - \rho_j \nu_i)} \left( \mu_j - \mu_k + \pi_{jk} \right).$$
(11)

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

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 again equivalent to the model without within skill dispersion.

The expected wages of Roy-Borjas model can, however, 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 mgiration 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 the 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 in later time. Later job opportunities in the potential destination location remain available even after declining a job offer and, thus, in comparing staying and migration, the on-the-job mobility prospects in the destination cancel out.

# 4 Empirical Strategy

I now reformulate the model of selection of job-to-job migrants as a model for wage disturbances. This is the empirically relevant model as the value of unobservable skills and job match quality are not separately observed. I then discuss how this confounding of unobservable skills and job match quality makes the comparison of migrants to stayers obsolete in identifying the selection on job match quality and describe an alternative strategy based on variation in mobility costs of otherwise arguably similar groups of mobile workers, commuters and migrants.

### 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  $\nu_i$  as the unobservable skills of *i*,  $\rho_h$  as the price of unobservable skills 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. Disturbance is assumed to be sum of the value of the (demeaned) unobservable skills and job match quality.<sup>8</sup>

<sup>&</sup>lt;sup>8</sup>This residual wage decomposition was introduced by Flinn (1986) and Garen (1989). Similar decomposition has been used in migration literature (Borjas, Bronars, & Trejo, 1992; Gould & Moav,

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

**Proposition 2.** The theoretical conditional expectation of migrants' source location disturbances is

$$E[u_{ij}|(\mathrm{MC})] = \rho_j(\mu_\nu - \mu_j^\nu) + \frac{1}{\sigma_\Delta} \big(\sigma_\nu^2(\rho_k - \rho_j)\rho_j - \sigma_j^2\big)\lambda(z_{ijk}),$$
(12)

and destination location disturbances is

$$E[u_{ik}|(\text{MC})] = \rho_k(\mu_\nu - \mu_k^\nu) + \frac{1}{\sigma_\Delta} \big(\sigma_\nu^2(\rho_k - \rho_j)\rho_k + \sigma_k^2\big)\lambda(z_{ijk}),$$
(13)

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

$$z_{ijk} \coloneqq \frac{1}{\sigma_{\Delta}} \left( \mu_{ij} - \mu_{ik} + \pi_{jk} - (\rho_k - \rho_j) \mu_{\nu} \right).$$
 (14)

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 relative to the source region, it generates acceptable job offers disproportionally for those whose skill composition has a high weight on 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. Both of these hurdles are likely to select positively on unobservable skills. Looking at (12) we can see how even if there is negative selection on disturbances in the source at the job acceptance and migration choice, the positive selection of the population at risk of mobility,  $\mu_{\nu} > \mu_{j}^{\nu}$ , may mask this. Thus, the positive selection on disturbances 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

<sup>2016;</sup> Bartolucci et al., 2018), where the component corresponding to  $q_{ih}$  has been interpreted as a location specific effect.

against negative selection on job match quality. Looking at (13), on the other hand, positive selection on disturbances in the destination may be due to both job match quality and unobservable skills and, thus, observing positive selection on residuals in the destination is not evidence for positive selection on job match quality. Thus, if migrants and stayers are very different in their unobservable skills, then comparison of migrants and stayers 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 study two groups of workers who both relocate their labor supply but employ different mobility technologies. I compare those that migrate to those that also change the location of their workplace, but instead of migrating, start commuting to the new location. 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 since migrants and commuters are similar in their unobservable skills and in the compensation for their unobservables skills but different in incurred mobility costs. First, as migrants, commuters have passed the first to hurdles and 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. Second, when restricting comparisons to source-destination pairs, the migrants and commuters experience the same change in the compensation for unobservable skills  $\rho_h$  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, the unobservable location mean skill level effects  $\rho_h \mu_h^{\nu}$  and their changes cancel out. Third, commuters incur different costs in relocating their labor supplies than migrants.

#### 4.2.1 Mobility Mode Choice

To understand how migrants and commuters differ in their incurred mobility costs, suppose now there are available two technologies for relocation of labour supply denoted m and c. We now allow heterogeneity in migration costs such that each worker in the population at risk of job-to-job migration has a cost type  $(\pi_{ijk}^c, \pi_{ijk}^m)$ . These costs depend on a variety of factors such as access to public transport or a car, housing status and family. The choice between migration and commuting is based on the relative costs of these two mobility modes with everyone choosing the least costly technology to relocate labor supply. These choices 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. Consider workers who are identical in their observable skills.

The preferred mobility mode is assumed to be independent of source and destination location wages. This follows from both wage-independent and time-equivalent mobility costs. For additive wage-independent mobility costs, those with  $\pi_{ijk}^m > \pi_{ijk}^c$  prefer commuting whereas those with  $\pi_{ijk}^m < \pi_{ijk}^c$  prefer migration migration. Time-equivalent mobility costs are  $w_{ih}\pi_{ijk}^c$  and  $w_{ih}\pi_{ijk}^m$ . Again, those with  $\pi_{ijk}^m > \pi_{ijk}^c$  prefer commuting whereas those with  $\pi_{ijk}^m < \pi_{ijk}^c$  prefer migration.<sup>9</sup> We thus imagine workers, given a potential destination location, having committed to a mobility mode such that the offered wage does not determine the mobility mode.

# Assumption 2. $(\nu_i, q_{ij}, q_{ik})|i \in \mathcal{I} \perp (\pi_{ijk}^c, \pi_{ijk}^m).$

Assumption 2 implies that potential migrants and potential commuters are similar in their disturbances. It formalizes the core idea of the empirical strategy that the commuters and migrants, having all passed the first two selective hurdles to become a potential job-to-job migrant, are homogeneous in their unobservables. Assumption 2 is not necessary in identifying selection patterns on residuals that can only be explained in terms of selection on job match quality: we only need migrants and commuters to be similar enough. However, it allows a structural interpretation of the different selection of commuters and migrants in terms of the model.

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

$$D_{i} = \mathbb{1}(\pi_{ijk}^{c} - \pi_{ijk}^{m} + \omega_{i} > 0), \tag{15}$$

where  $\omega_i$  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,  $D_i = 0$ . Mobility mode  $D_i$  is an exogeneous proxy for mobility costs if the error  $\omega_i$  is independent of the components of wage disturbances.

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

<sup>&</sup>lt;sup>9</sup>It also does not matter whether mobility cost are 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  $w_{ik} - w_{ij} > \pi_{ijk}$ .

Assumption 3 makes mobility mode  $D_i$  an exogenous proxy for the preference to migration relative to commuting in an equation that relates disturbances to the mobility mode.

**Proposition 3.** Suppose Assumption 2. Let  $\pi_{ijk}^c$  and  $\pi_{ijk}^m$  be independently and normally distributed with means  $\bar{\pi}_{jk}^m$  and  $\bar{\pi}_{jk}^c$  and equal standard deviations. If and only if  $\bar{\pi}_{jk}^m > \bar{\pi}_{jk}^c$ ,

- (i) there are more commuters than migrants,
- (*ii*)  $E[\pi_{ijk}^c | D_i = 0] < E[\pi_{ijk}^m | D_i = 1] < E[\pi_{ijk}^c | D_i = 1] < E[\pi_{ijk}^m | D_i = 0],$
- (iii)  $\lambda^m > \lambda^c$ ,

where  $\lambda^c \coloneqq E[\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}]$  and  $\lambda^m \coloneqq E[\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}].$ 

It also establishes how mobility mode  $D_i$  is a relevant proxy for mobility costs. Proposition 3 is used to calibrate the relative costs of potential migrants and commuters in the model. The preference reveal over mobility mode choice is informative of potential mobility costs. Given the cost distributions among the population at risk of contracted migration as in Proposition 3, if  $\bar{\pi}_{jk}^m > \bar{\pi}_{jk}^c$ , the posterior distributions of mobility costs of potential migrants and commuters satisfy the monotone likelihood ratio property. This informativeness is thus in the sense of more favorable than -relation of Milgrom (1981). Monotone likelihood ratio property 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 scale terms  $\lambda$  across potential migrants and commuters.

The employed mobility mode can thus be seen as a statistic summarizing different cost structures. As workers with different costs take different actions, these actions, as revealed preferences convey information and 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 larger costs than the commuters who relocate only their work locations.

#### 4.2.2 Cost's Scaling Effect

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 changes in the compensation for unobservable skills or as changes in job match quality. The comparison of migrants and commuters thus reveals the sign of the multiplier of  $\lambda^m - \lambda^c$  and so



Figure 2: Reservation and inverse reservation wages as a function of relocation cost.

reveals whether variation in mobility costs magnify 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, 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, 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 larger 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.

# 5 Data and Empirical Definitions

This section describes the data I use and gives an empirical definition of the two mobility groups of comparison, migrants and commuters, and the outcome variables of interest, wages, predicted wages and residuals.

### 5.1 Data

I use total population annual individual level data compiled from various administrative registers provided by Statistics Finland.<sup>10</sup> For each observation, I assemble data from

<sup>&</sup>lt;sup>10</sup>The data are available in data sets called FOLK modules. To build the data used I combine information from FOLK Basic data, FOLK Employment and FOLK Cohabitation modules. See

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 (Figure 4 uses data all the way to the period t - 5). The theory in Section 3 models the behavior of salaried workers who are well attached to labor market. I thus use a sample of such workers. 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. All residents and workers in and all migrants and commuters to and from the Åland Islands and students and retirees defined by the longest principal activity during the year are excluded. The age restriction is to reduce mobility of students, the first-time movers and mobility that may occur with retirement in 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  $t \in \{2011, 2012, 2013, 2014\}$ .

### 5.2 Empirical Mobility

The mobile are defined as those who change the location of their employment. This definition of mobility 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. As I study job-to-job migration, I do not aim to separate job-to-job from other types of migration, but to restrict the sample such that the mobility events in the data are likely job-to-job.

First, I define *job movers* as those whose postal code area or municipality of work place changes between years t - 1 and t and, to capture employment-to-employment 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 work place in year t. The distance to the location of the new job is defined differently for those who eventually commute and those who eventually migrate: For the commuters, the data contains information of the commuting distance in year t as an Euclidean distance between job and residence locations with 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 <u>https://taika.stat.fi/en/.</u>

distance in year t. For them, 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.<sup>11</sup> 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 the migrants than to 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 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. The threshold distance for mobility is set to 50 kilometers in the main analysis. This number is somewhat arbitrary, chosen to ensure that there are both commuters and migrants in the sample. The main results are robust to variation in this threshold. *Stayers* are now defined as those who are not job-movers nor mobile. To remove some 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 postmobility 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. I leave out those mobile who migrate such that the distance between the location of their residence and their job increases as these moves are likely motivated by factors unrelated to labor market.

A concern with this categorization is that commuting often is 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 on the year of mobility may classify willing migrants as commuters due to the lags in residential adjustment. Ideally, migrants would 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 years from the analysis. The loss of observations would

<sup>&</sup>lt;sup>11</sup>I use 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 area are from Statistic Finland's Paavo postal code area statistics database, see https://www.stat.fi/tup/paavo/index\_en.html



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. For data, see Section 5.1

be nonrandom leading to sample selection problems. As a compromise, commuters are defined to be those that have still not changed their original residence location in year t + 1.<sup>12</sup>

The choice between migration and commuting is not truly binary: The migrants have a strictly positive post-mobility commuting distance as well. Figure 3 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. 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 to render the binary classification of migrants and commuters a reasonable approximation. We also, as predicted by Proposition 3, observe a relatively larger share of commuters than migrants. This is consistent with migrants tending to incur larger relocation costs than commuters.

Table 1 presents the descriptive statistics of the sample by subsample. 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

<sup>&</sup>lt;sup>12</sup>Note that the post-mobility earnings are computed using the earnings information of year t + 1and 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.

	Mob	Stayers	
	Commuters	Migrants	v
Female	0.32	0.47	0.52
Age, year $t$	44.7	39.7	45.5
Born abroad	0.03	0.06	0.04
Education, year $t$			
Basic education	0.08	0.06	0.10
Secondary education	0.34	0.28	0.41
Tertiary education	0.55	0.62	0.48
Doctoral or equivalent	0.03	0.03	0.02
Work, year $t-1$			
Tenure in current job, days	2,209.6	$1,\!601.6$	2,878.1
Employment days	360.83	359.05	363.22
Unemployment days	0	0	0
Log wage	4.75	4.60	4.59
Mobility experience, year $t-1$			
Migration experience	0.18	0.38	0.10
Commuting experience	0.79	0.59	0.45
Family, year $t-1$			
Spouse working	0.68	0.45	0.65
Living alone	0.05	0.08	0.06
Living with spouse	0.80	0.57	0.76
Children	0.60	0.42	0.58
Housing, year $t-1$			
Right of occupancy dwelling	0.01	0.02	0.02
Rents the dwelling	0.10	0.36	0.15
Owns the dwelling	0.88	0.60	0.82
Distance to (new) job, km			
Mean	147.77	216.68	14.86
Median	104.38	159.60	6.86
Observations	30,896	12,172	3,131,309

Table 1: Descriptive statistics by mobility group. Notes: Means for continuous variables unless otherwise mentioned and shares for categorical variables. For the data, see Section 5.1. See Section 5.2 for the definitions of migrants, commuters and stayers. See Section 5.1 for the sample restrictions.

more often rent and less often own their dwelling. Contracted migrants tend to have higher education than the commuters. Choices between migration and commuting seem to autocorrelate as well: previous migration experience is higher among the migrants whereas previous commuting experience is more prevalent among the commuters. We also, as predicted by Proposition 3, observe large migration costs for those who chose to commute. This is consistent with migrants tending to incur larger relocation costs than commuters.

## 5.3 Empirical Job Match Quality

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

$$\hat{w}_{ih} = E_{ih}[w] + u_{ih} = x'_i \gamma + \bar{\mu}_h + u_{ih}.$$
(16)

where  $\hat{w}_{ih}$  is the estimated wage (see Section 5.3.1),  $x'_i$  contains observable determinants of wages (see Section 5.3.2), 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*.

#### 5.3.1 Wages

The data have information on annual earnings: wages are computed as the ratio of annual earnings to annual employment days. As earnings and employment 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 construct 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. This does not come without potential problems. Whereas the wage in 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-thejob 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 is, 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 mobile are defined so as to highly likely be job-to-job migrants. Especially, 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 mobile 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 mobile workers would likely bias the estimated positive selection on job match quality 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 level of education, field of education, occupation and industry.<sup>13</sup>

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 strictly speaking daily income, the concern is that wages are underestimated for those who commute. If commuting costs were then left in the error term when computing the residuals, the commuters would have lower post-mobility residuals than the migrants simply because they take more time commuting. 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 empirical measure of wage has a negative association with commuting costs, we have the first case and we should include commuting costs into the wage regression but if the the association is positive, we have the second case and we should not include commuting costs into the wage regression. The most straightforward measure of commuting costs in the data is the commuting distance. The estimated coefficient for the commuting distance in the wage regression in the whole sample of the mobile and stayers is significantly positive (not reported). Thus, when 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 we observe the wages of stayers growing along a stable path, for the mobile the growth

 $<sup>^{13}</sup>$ Field education is classified by the uppermost level of of International Stanof Education 2016 (10 categories + 1 for dard Classification unknowns). See https://www.stat.fi/en/luokitukset/koulutusala/koulutusala\_3\_20180101/. Occupations are classified by the uppermost level of Classification for Occupation (AML 2010) (10 categories +1 for unknows). See https://www.stat.fi/en/luokitukset/ammatti/ammatti\_1\_20100101/. Industries are classified by the uppermost level of Standard Industrial Classification TOL 2008 (21 categories +1 for unknowns). See https://www.stat.fi/en/luokitukset/toimiala/toimiala\_1\_20080101/.



Figure 4: Wage decompositions of commuters, migrants and stayers. Notes: For the data, see Section 5.1. See definitions in Section 5.2. For the wage models, see Section 5.3. Mobility potentially occurs between time points t - 1 and t.

of the wages exceeds 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 the Panel A. While the migrants have lower residuals than the commuters prior to mobility in the source, they have 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 residuals pre- and post mobility are statistically different from zero with signs as in Figure 4.

Recall indicator  $D_i$  on the set of 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 modeled as

$$\hat{u}_{ih} = \mu_{jkl} + \alpha_3 d_i + \kappa_t + \alpha_1 \hat{E}_{ij}[w] + \alpha_2 \hat{E}_{ik}[w] + \tau D_i + \gamma v_i + \varepsilon_i, \tag{17}$$

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  $\mu_{jkl}$  restricts the comparison to commuters and migrants within residence-source-destination location triplets.<sup>14</sup> 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. 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, and for transport infrastructures connecting the residence and destination locations. The difference between the distance to new job and the distance to old job  $d_i$  controls for commuting costs prior to mobility. The time fixed effect  $\kappa_t$  controls for common year effects. To control for potential 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 of occupation changes between periods t-1 and t and  $\gamma$  is a conformable vector of coefficients.

Table 2 presents the results of estimation of (17) 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. Thus, the pattern of residuals in Panel C holds within residence-sourcedestination-location triplets comparing similar workers in terms of their observable determinants of wages and controlling for the net distance to the new job and potential industry, employer, and occupation changes.

### 6.2 Selection on Job Match Quality

Selection on residuals may reflect either selection on job match quality or selection on unobservable skills. I now argue that the results identify selection on job match quality. I first show how balance on unobservable skills among the observed commuters and migrants identifies the effect of job match quality and discuss its plausibility. I then discuss the conditions under which difference-in-difference estimation identifies the effect of job match quality and study the differences in residual changes. Lastly, I complement the empirical identification with theoretical identification. I pit selection on unobservables and job match quality against each other by deriving corresponding coefficient sign restrictions from the theory. I show that the data is consistent with selection on job match quality while the data is not consistent with selection on unobservable skills

<sup>&</sup>lt;sup>14</sup>Residence location refers to year t - 1 location of residence.

Dependent variable:		$\hat{u}$	ij	
Migrant (ref: Commuter)	-0.0239 $(0.0055)^{***}$	-0.0214 (0.0091)*	-0.0228 (0.0092)*	-0.0203 (0.0092)*
Dependent variable:		$\hat{u}$	ik	
Migrant (ref: Commuter)	$0.0316 \\ (0.0051)^{***}$	$0.0249 \\ (0.0082)^{**}$	0.0239 $(0.0083)^{**}$	0.0244 $(0.0083)^{**}$
Dependent variable:	$\hat{u}_{ik} - \hat{u}_{ij}$			
Migrant (ref: Commuter)	$0.0554 \\ (0.0030)^{***}$	$0.0462 \\ (0.0090)^{***}$	$0.0468 \\ (0.0091)^{***}$	0.0447 $(0.0088)^{***}$
Spatial controls				
$d_i$	No	No	Yes	Yes
$jkl$ -triad FE $\mu_{jkl}$	No	Yes	Yes	Yes
$\hat{E}_{ik}[w], \hat{E}_{ij}[w]$	No	No	No	Yes
Year FE , $\kappa_t$	No	Yes	Yes	Yes
$v_i$	No	No	Yes	Yes
Constant term	Yes	No	No	No
Observations	43,068	43,068	43,068	43,068
Migrants	12,172	12,172	12,172	12,172
Commuters	30,896	30,896	30,896	30,896
$\mathbf{R}^2,  \hat{u}_{ij}$	0.0005	0.0010	0.0013	0.0215
$\mathbf{R}^2, \hat{u}_{ik}$	0.0008	0.0027	0.0029	0.0267
$\mathbf{R}^2,  \hat{u}_{ik} - \hat{u}_{ij}$	0.0024	0.0020	0.0022	0.0695

Table 2: Selection on residuals. Notes: \*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.001. Data pooled from years 2011-2014. See Section 5.2 for the definitions of commuters, migrants and stayers. See Section 5.1 for sample construction. For the computation of residuals, see Section 5.3. White heteroskedasticity robust standard errors clustered at residence-source-destination municipality level in parenthesis.

alone.

#### 6.2.1 Inbalanced Unobservable Skills Among Mobile

The comparison of pre- and post-mobility residuals of the commuters and migrants in Table 2 identifies the effect of job match quality if the commuters and migrants are equal in their mean unobservable skills. To see this, consider the conditional difference in the residuals of migrants and commuters:

$$\hat{u}_{ih} = controls + \tau D_i + \varepsilon_i$$
  
= controls + (E[q\_{ih}|D\_i = 1] - E[q\_{ih}|D\_i = 0])D\_i + \epsilon\_i (18)

with  $\epsilon_i \coloneqq \rho_h(E[\nu_i|D_i=1] - E[\nu_i|D_i=0])D_i + \varepsilon_i$ . 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

$$\rho_h \left( E[\nu_i | D_i = 1] - E[\nu_i | D_i = 0] \right) = 0 \iff E[\nu_i | D_i = 1] = E[\nu_i | D_i = 0].$$
(19)

There is, of course, no direct way of knowing whether (19) holds in the estimation of (17). The theory suggests that everything else held constant, migrants are selected more strongly on unobservable skills if they incur larger relocation costs than commuters speaking against (19). Thus, (19) is not implied by Assumption 2, where there is no conditioning on (MC). However, if the selection on unobservable skills played a relatively minor role in selection, (19) 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. If (19) holds then the results in Table 2 identify selection on job match quality.

### 6.2.2 Identification

By imposing more structure, we can interpret the difference in selection on residuals between migrants and commuters in terms of the model.

**Proposition 4.** Suppose Assumptions 1, 2, and 3. Then

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

$$\tau_{j} = E[u_{ij}|D_{i} = 1] - E[u_{ij}|D_{i} = 0]$$
  
=  $\rho_{j}(E[\nu_{i}|D_{i} = 1] - E[\nu_{i}|D_{i} = 0]) + E[q_{ij}|D_{i} = 1] - E[q_{ij}|D_{i} = 0]$   
=  $\frac{1}{\sigma_{\Delta}} (\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{j} - \sigma_{j}^{2})(\lambda^{m} - \lambda^{c}),$ 

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

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

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

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

where  $\lambda^c := E[\lambda(z_{ijk}(\pi_{ijk}^c))|\pi_{ijk}^m > \pi_{ijk}^c]$  and  $\lambda^m := E[\lambda(z_{ijk}(\pi_{ijk}^m))|\pi_{ijk}^m < \pi_{ijk}^c]$ . In each part (i), (ii), and (iii), the first equality follows from Assumption 1 and the second equality from Assumptions 2 and 3.

While the statistical relationship of interest is between mobility mode and residuals, the causal relationship of interest is from mobility costs to selection on residuals. Hence, first, both the antecedent and the consequent of the causal relationship of interest are unobserved. 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 causal relationship of interest is not the dependent variable but selection on dependent variable that is the mean of the dependent variable measures.

Assumption 2 allows a more structural interpretation of the estimated difference in the relative selection of commuters and migrants. However, this structure comes with the cost of further assumptions. First, the independence of unobservables on costs among the potentially mobile allows to interpret the differences in the residuals across commuters and migrants as due to different degrees of selection. 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 to be truncated to stay constant as the truncation point varies.

Second, to use the model of selection of job-to-job migrants in Section 3, we have to assume that the total relative selection of commuters and migrants is due to the last hurdle of choosing whether 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 modeling approach of abstracting from the first two hurdles 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. However, 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.

#### 6.2.3 Model Restrictions

Selection on job match quality and selection on unobservable skills differ in predictions for the differences in 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 pre-mobility has more valuable unobservable skills post-mobility as well. Selection on job match quality on the other hand predicts the higher cost group to go having lower residuals pre-mobility to having higher residuals post-mobility.

There is no direct measure of the location's compensation for unobservable skills. I proxy  $\rho_h$  with two measures of wage variation: the standard deviation of wages and standard deviation of residuals among location h workers.<sup>15</sup> All these measures are computed in the analysis sample to capture wage variation in the labor markets that the workers in the sample face. As there is substantial year-to-year variation, for each municipality an average over the sample years is computed.

Consider the expected differences in residuals across migrants and commuters given in Proposition 4 and note that by Proposition 3,  $\lambda^m > \lambda^c$ . 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  $\rho_k - \rho_j > (<) 0$ 

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

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

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

As seen in Table 3, we estimate  $\hat{\tau}_j < 0$  and  $\hat{\tau}_k > 0$  for both  $\rho_k - \rho_j > 0$  and  $\rho_k - \rho_j < 0$ . For pre-mobility residuals the estimates are imprecise, but the coefficients correspond largely to the coefficients estimated in the whole sample. Thus, selection on job match

<sup>&</sup>lt;sup>15</sup>Previous research has proxied  $\rho_h$  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 & Sousa-Poza, 2004) standard deviation of log wage (Borjas et al., 1992), standard deviation of residuals from wage regressions (Borjas et al., 1992; Gould & Moav, 2016) and with various measures of returns to education (Gould & Moav, 2016; Moraga, 2013).

quality explains the pattern in observed residuals whereas selection on unobservable skills does not.

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

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

As seen in Table 3, we estimate  $\hat{\tau}_k - \hat{\tau}_j > 0$  regardless of the sign of  $\rho_k - \rho_j$ . Hence, again, 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.

#### 6.2.4 Difference-in-difference

Contracted migrants have lower pre-mobility residuals and higher post-mobility residuals than commuters. Clearly then the change in residuals among the migrants should be greater than among the commuters. Modeling the change in residuals is, however, interesting in its own right since if the unobservable skills balance in the two mobility groups or 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 (17) but let now the residual change  $u_{ik} - u_{ij}$  be the dependent variable. The coefficient  $\tau$  then identifies the difference in the residual differences as given in part (iii) of Proposition 4, 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 very similarly prior to mobility, these common trends cannot be extrapolated to the year of mobility as in that year the compensation for the unobservable skills changes so that the group specific unobservable effects do not cancel out in the before-after comparison.<sup>16</sup> This means that if the commuters and migrants are on average different in their unobservable skills, then the compensations for their unobservable skills evolve differently in the year

<sup>&</sup>lt;sup>16</sup>Since the group unobservable effect  $E[\nu_i|D_i = d]$  and the effect of changing location  $\rho_k - \rho_j$  are multiplicative, common trend is violated for both levels and logarithmic transformation of wages. See e.g. Lechner (2011) page 186.

$\rho$ measure:	$sd(\hat{u}_{ih})$	$sd(w_{ih})$	$sd(\hat{u}_{ih})$	$sd(w_{ih})$	
Sample:	$\rho_k - \rho_j > 0$		$ ho_k -  ho_k$	$\rho_k - \rho_j < 0$	
Dependent variable:		$\hat{u}_{i}$	ij		
Migrant	-0.0201	-0.0196	-0.0108	-0.0166	
(ref: Commuter)	(0.0107)	(0.0110)	(0.0129)	(0.0124)	
Dependent variable:	$\hat{u}_{ik}$				
Migrant	0.0329	0.0445	0.0385	0.0185	
(ref: Commuter)	$(0.0095)^{***}$	$(0.0105)^{***}$	$(0.0123)^{**}$	(0.0122)	
Dependent variable:	$\hat{u}_{ik} - \hat{u}_{ij}$				
Migrant	0.0531	0.0641	0.0494	0.0351	
(ref: Commuter)	$(0.0109)^{***}$	$(0.0110)^{***}$	$(0.0143)^{***}$	$(0.0130)^{**}$	
Spatial controls					
$d_i$	Yes	Yes	Yes	Yes	
$jkl$ -triad FE $\mu_{jkl}$	Yes	Yes	Yes	Yes	
$\hat{E}_{ik}[w], \hat{E}_{ij}[w]$	Yes	Yes	Yes	Yes	
Year FE, $\kappa_t$	Yes	Yes	Yes	Yes	
$v_i$	Yes	Yes	Yes	Yes	
Observations	22,797	23,069	$21,\!455$	19,999	
Migrants	6,345	6,314	5,827	5,858	
Commuters	$16,\!452$	16,755	$14,\!444$	$14,\!141$	
$\mathbf{R}^2,  \hat{u}_{ij}$	0.0261	0.0267	0.0236	0.0222	
$\mathbf{R}^2, \hat{u}_{ik}$	0.0263	0.0226	0.0248	0.0250	
$\mathbf{R}^2,  \hat{u}_{ik} - \hat{u}_{ij}$	0.0771	0.0689	0.0660	0.0652	

Table 3: Residual differences and difference-in-differences by sign of  $\rho_k - \rho_j$ . Notes: \*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.001. Data pooled from years 2011-2014. See Section 5.2 for the definitions of mobile, migrants and commuters. For the computation of residuals, see Section 5.3. For the computation of residuals, see Section 5.3. White heteroskedasticity robust standard errors clustered at residence-source-destination municipality level in parenthesis.

of mobility violating common trend.<sup>17</sup>

However, as can be seen from part (iii) of Proposition 4 common trend with respect to unobservable skills is satisfied when  $\rho_k = \rho_j$ . I thus study subsamples restricted by the values the difference  $\rho_k - \rho_j$  can take. Figure 5 presents the estimated coefficients in these samples. The upper panel uses as a measure of  $\rho_h$  the standard deviation of wages of location h workers whereas the lower panel uses the standard deviation of residuals of workers in location h. The interesting sample is the one where  $\rho_k = \rho_j$  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 timeinvariant effects are expected to cancel out. Migrants have larger changes in residuals

<sup>&</sup>lt;sup>17</sup>Many have compared migrants' wage growth to 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 should be interpreted as a the price paid for the underlying individual fixed characteristics. If this price 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.



Figure 5: Difference-in-difference with sample splits. Notes: For the data, see Section 5.1. See definitions in Section 5.2. Mobility potentially occurs between time points t - 1 and t. Models of column 4 in Table 2.

than commuters also in the sample where the common trend assumption is likely to be satisfied to a reasonable degree. The difference in the residual changes across migrants and commuters is thus likely due to selection on job match quality rather than selection on unobservable skills.

### 6.3 Mobility Costs

The differences between the residuals of migrants and commuters are interpreted as reflecting differences in the mobility costs across these two groups. To support this reasoning, I now study if the selection effects are mediated by more direct measures 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 different mobility cost proxies. Table 4 presents the results. 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. Working spouse, on the other hand, increases the relative migration cost. Correspondingly, we observe a positive coefficient. Living rental reduces the relative cost of migration and as expected, we observe a negative effect, albeit not significantly different from zero. Owning a car

Dependent variable:			$\hat{u}_{ik} - \hat{u}_{ij}$		
Cost proxy, $t - 1$ :	Lives alone	Spouse working	Lives rental	Owns a car	Distance to new job
Migrant * Cost proxy (ref: Commuter * Cost proxy)	-0.0482 $(0.0209)^*$	$0.0470 \\ (0.0195)^*$	-0.0148 (0.0224)	$0.0302 \\ (0.0203)$	-0.0009 (0.0077)
Spatial controls $d_i$ $jkl$ -triad FE $\mu_{jkl}$ $\hat{E}_{ik}[w], \hat{E}_{il}[w]$ Year FE, $\kappa_t$ $v_i$ Observations Migrants Commuters	Yes Yes Yes Yes 43,068 12,172 30,896	Yes Yes Yes Yes 43,068 12,172 30,896	Yes Yes Yes Yes 43,068 12,172 30,896	Yes Yes Yes Yes 43,068 12,172 30,896	Yes Yes Yes Yes 43,068 12,172 30,896

Table 4: Residual difference-in-differences interacted with cost proxies. Notes: \*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.001. Data pooled from years 2011-2014. See Section 5.2 for the definitions of mobile, migrants and commuters. For the computation of residuals, see Section 5.3. For the computation of residuals, see Section 5.3. White heteroskedasticity robust standard errors clustered at residence-source-destination municipality level in parenthesis.

reduces the cost of commuting and as expected, we observe a positive interaction effect, albeit not significantly different from zero. Longer distance to new job increases the cost of commuting relative to migration. As would be predicted, we observe a negative effect, albeit not significantly different from zero.

### 6.4 Robustness

### 6.4.1 The Threshold Distance of Mobility

The threshold distance of 50km in defining spatial 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 6 shows that the estimated difference in job match qualities among the migrants and commuters is robust to this threshold.

### 6.4.2 Ashenfelter's Dip

Larger wage gains among the migrants could be explained by an autoregressive wage process where the migrants 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



Figure 6: Robustness with respect to the threshold distance of mobility. Notes: For the data, see Section 5.1. See definitions in Section 5.2. For the computation of residuals, see Section 5.3. Models of column 4 in Table 2. 95-percent confidence intervals are based on white heteroskedasticity robust standard errors clustered at residence-source-destination municipality level.

event. It is, however, clear from Figure 4 that the wage paths of commuters and migrants are very 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.

#### 6.4.3 Housing Constraints

It could be argued that those with larger wage gains are more likely to be able to afford migration and thus 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.

# 7 Conclusion

Labor related migration is often job-to-job: migrants often migrate after having accepted a job in the destination. Ignoring the possibility of migrants searching and accepting jobs prior to migration and migrating with known destination location wage and interpreting migration as a risky human capital investment may bias interpretations of results about 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 choice of migration 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, but as the job-to-job migrants only accept wages that compensate for their migration costs, they realize 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 lead to 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.

This paper takes a step toward including the possibility of job-to-job migration into the analyses of migrant selection. The Roy-Borjas model extended with observability of source and destination location wages derived from a job search model that naturally highlights the wage dispersions in the source and destination location provides a framework to study job-to-job migration with many results and methods already familiar from the literature. While there is selection on unobservable skills depending on source and destination location difference in returns to skills, this selection is confounded by negative selection on job match quality in the source and positive selection on job match quality in the destination.

Selection on job match quality can be expected to dominate selection on unobservable skills when within skill wage dispersion is wide and location differences in returns to skills small. This is the case for internal migration. I study internal migration and compare migrants and commuters, two groups of workers that similarly have searched for jobs interregionally and have been selected to receive a job offer and thus are both in a position to choose between job-to-job migration or staying. This controls for the self-selection into interregional job search and employers' selection. These two groups of mobile workers, however, differ in their incurred mobility costs and the magnification of selection effects in costs identifies stronger selection on job match quality for migrants than for commuters.

A model outperforming the benchmark model in explaining data is support for the relevance of the model. Nevertheless, the more structural identification of selection mechanisms here is based on strong ignorability assumptions. The research on migrant selection still lacks attemps to create credible causal evidence for the selection mechanisms. Variation in mobility costs is one place where exogenous variation may plausibly be found.

Adding job match quality with zero autocorrelation to the Roy-Borjas model also provides a relevant interpretation for Borjas' (1987) refugee sorting selection pattern that combines negative selection in the source and positive selection in the destination. However, job-to-job migration together with job match quality arising from wage dispersion generates this selection pattern as well. Rather than the narrow interpretation of migration from communist to non-communist regimes, a new and potentially more useful interpretation arises as we view workers fleeing their bad job matches for better job matches.

# References

- 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., & 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., & Elldér, E. (2022). Movers and stayers: A study of emigration from sweden 1993–2014. European Journal of Population, 1–32.
- Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 531–553.
- Borjas, G. J. (1991). Immigration policy, national origin, and immigrant skills: A comparison of Canada and the United States. National Bureau of Economic Research Cambridge, Mass., USA.
- Borjas, G. J. (2014). *Immigration economics*. Harvard University Press.
- Borjas, G. J., Bronars, S. G., & 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., & 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., & 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 earnings of immigrants. Journal of Political Economy, 86, 897–921.
- Detang-Dessendre, C., & Molho, I. (1999). Migration and changing employment status: a hazard function analysis. *Journal of Regional Science*, 39(1), 103–123.
- Dostie, B., & Léger, P. T. (2009). Self-selection in migration and returns to unobservables. *Journal of Population Economics*, 22(4), 1005–1024.
- Emmler, J., & Fitzenberger, B. (2020). The role of unemployment and job change when estimating the returns to migration.
- 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., & Moav, O. (2016). Does high inequality attract high skilled immigrants? The Economic Journal, 126(593), 1055–1091.
- Ham, J. C., Li, X., & 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., & Honoré, B. E. (1990). The empirical content of the Roy model. *Econometrica*, 58(5), 1121–1149.
- Hicks, J. R. (1932). The theory of wages. New York: Macmillan.
- Hunt, G. L., & Mueller, R. E. (2004). North american migration: returns to skill, border effects, and mobility costs. *Review of Economics and Statistics*, 86(4), 988–1007.
- Kaestner, R., & 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., Ozden, Ç., & Parsons, C. (2017). High-skilled migration and agglomeration. Annual Review of Economics, 9, 201–234.
- Liebig, T., & 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*, 380–391.
- Molho, I. (1986). Theories of migration: a review. Scottish Journal of Political Economy, 33(4), 396–419.
- Moraga, J. F.-H. (2011). New evidence on emigrant selection. *The Review of Economics* and Statistics, 93(1), 72–96.
- Moraga, J. F.-H. (2013). Understanding different migrant selection patterns in rural and urban mexico. *Journal of Development Economics*, 103, 182–201.
- Mortensen, D. (2003). Wage dispersion: why are similar workers paid differently? MIT press.
- Nakosteen, R. A., Westerlund, O., & Zimmer, M. (2008). Migration and self-selection: Measured earnings and latent characteristics. *Journal of Regional Science*, 48(4), 769–788.
- Parey, M., Ruhose, J., Waldinger, F., & Netz, N. (2017). The selection of high-skilled emigrants. *Review of Economics and Statistics*, 99(5), 776–792.
- Pickles, A., & Rogerson, P. (1984). Wage distributions and spatial preferences in competitive job search and migration. *Regional Studies*, 18(2), 131–142.
- 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. American Economic Review, 51(1), 1–17.
- 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.
- 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.

# A Appendix

### A.1 Proofs of Propositions and Lemmas

**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  $\varphi$ . 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

$$(r+\varphi)V_k = b - \pi_{ilk} + \frac{\varphi}{r}(\mu_k + \rho_k\nu_i)$$
(20)

$$(r+\varphi)V_j = w_{ij} - \pi_{ilj} + \frac{\varphi}{r}w_{ij}$$
(21)

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

$$E[x|\frac{\Delta_i}{\sigma_{\Delta}} > z] = \mu_x + \frac{E[x\Delta_i]}{\sigma_{\Delta}}\lambda(z), \qquad (22)$$

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

Proof of Lemma 2. First note that

=

$$E\left[\frac{x}{\sigma_x}\Big|\frac{\Delta_i}{\sigma_\Delta}\right] = E\left[\frac{x}{\sigma_x}\right] + \frac{Cov[\frac{x}{\sigma_x},\frac{\Delta_i}{\sigma_\Delta}]}{Var[\frac{\Delta_i}{\sigma_\Delta}]}\frac{\Delta}{\sigma_\Delta} = \frac{1}{\sigma_x}\mu_x + \frac{E[x\Delta_i]}{\sigma_x\sigma_\Delta^2}\Delta_i$$
(23)

Thus,

$$E[x|\frac{\Delta_i}{\sigma_{\Delta}} > z] = \sigma_x E[\frac{x}{\sigma_x}|\frac{\Delta_i}{\sigma_{\Delta}} > z] = \mu_x + \frac{E[x\Delta_i]}{\sigma_{\Delta}}E[\frac{\Delta_i}{\sigma_{\Delta}}|\frac{\Delta_i}{\sigma_{\Delta}} > z]$$
(24)

$$= \mu_x + \frac{E[x\Delta_i]}{\sigma_\Delta} \frac{\phi(z)}{1 - \Phi(z)},\tag{25}$$

where the second equality follows from Lemma 3 and (23). Defining  $\lambda(\cdot) \coloneqq \phi(\cdot)/(1 - \Phi(\cdot))$  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

$$\begin{split} E[z|x > a] &= \int_{z} z f_{z}(z|x > a) dz = \int_{z} z \int_{x > a} f_{z}(z|X = x) f_{x}(x|x > a) dx dz \\ &= \int_{x > a} \int_{z} z f_{z}(z|X = x) dz f_{x}(x|x > a) dx = \int_{x > a} E[z|x] f_{x}(x|x > a) dx \\ &= \int_{x > a} (c + rx) f_{x}(x|x > a) dx = c + r \int_{x > a} x f_{x}(x|x > a) dx = c + r E[x|x > a]. \end{split}$$

**Lemma 4.** Let  $\pi_{ijk}^m$  and  $\pi_{ijk}^c$  be independently and normally distributed with means  $\bar{\pi}_{jk}^m$ ,  $\bar{\pi}_{jk}^c$ , respectively, and identical standard deviations  $\sigma_{\pi}$ . If and only if  $\bar{\pi}_{jk}^m > \bar{\pi}_{jk}^c$ , then  $\pi_{ijk}^m |\pi_{ijk}^m < \pi_{ijk}^c$  strictly first order stochastically dominates  $\pi_{ijk}^c |\pi_{ijk}^c < \pi_{ijk}^m$ .

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

$$P(X|X < Y) = \frac{P(X < Y|X)P(X)}{P(X < Y)} = \frac{\left[1 - \Phi\left(\frac{x - \mu_y}{\sigma_y}\right)\right]\phi\left(\frac{x - \mu_x}{\sigma_x}\right)}{\Phi\left(\frac{\mu_y - \mu_x}{\sqrt{\sigma_y^2 + \sigma_x^2}}\right)}$$

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

$$f^{m}(\pi) = \frac{[1 - \Phi(\pi)]\phi(\pi - \bar{\pi}_{jk}^{m})}{\Phi(-\bar{\pi}_{jk}^{m}(\sqrt{2})^{-1})}, \quad f^{c}(\pi) = \frac{[1 - \Phi(\pi - \bar{\pi}_{jk}^{m})]\phi(\pi)}{\Phi(\bar{\pi}_{jk}^{m}(\sqrt{2})^{-1})}.$$

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

$$\frac{f^m(\pi)}{f^c(\pi)} = \frac{\Phi(\bar{\pi}_{jk}^m(\sqrt{2})^{-1})}{\Phi(-\bar{\pi}_{jk}^m(\sqrt{2})^{-1})} \frac{\phi(\pi - \bar{\pi}_{jk}^m)}{[1 - \Phi(\pi - \bar{\pi}_{jk}^m)]} \frac{[1 - \Phi(\pi)]}{\phi(\pi)}$$

and, thus, a monotone likelihood ratio if

$$\frac{d}{d\pi} \frac{f^m(\pi)}{f^c(\pi)} > 0 \iff \frac{d}{d\pi} \left( \frac{\phi(\pi - \bar{\pi}_{jk}^m)}{[1 - \Phi(\pi - \bar{\pi}_{jk}^m)]} \frac{[1 - \Phi(\pi)]}{\phi(\pi)} \right) = \frac{d}{d\pi} \frac{\lambda(\pi - \bar{\pi}_{jk}^m)}{\lambda(\pi)} > 0, \quad (26)$$

where  $\lambda(\cdot) = \phi(\cdot)/[1 - \Phi(\cdot)]$  is the inverse Mill's ratio. Thus the likelihood ratio is strictly increasing, if and only if

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

that is, if and only if  $\lambda'/\lambda$  is strictly decreasing. Since  $\phi'(x) = -x\phi(x)$ , we have  $\lambda'(x)/\lambda(x) = \lambda(x) - x$ , which is strictly decreasing in x since  $\lambda'(x) < 1$  (e.g. Heckman and Honoré (1990) (R-2)). Strictly increasing likelihood ratio (26) then implies strict first order stochastic dominance of  $\pi_{ijk}^m | \pi_{ijk}^m < \pi_{ijk}^c$  over  $\pi_{ijk}^c | \pi_{ijk}^c < \pi_{ijk}^m$  if and only if  $\bar{\pi}_{jk}^m > 0$ .

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

$$z_{jk} \coloneqq \frac{1}{\sigma_{\Delta}} \left( \mu_j - \mu_k + \pi - (\rho_k - \rho_j) \mu_\nu \right)$$
(27)

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

$$E[\rho_h \nu_i | (\text{MC})] = E[\rho_h \nu_i | \frac{\Delta_i}{\sigma_\Delta} > z_{jk}] = \rho_h \mu_\nu + \frac{\sigma_\nu^2}{\sigma_\Delta} (\rho_k - \rho_j) \rho_h \lambda(z_{jk})$$
(28)

and

$$E[q_{ij}|(\mathrm{MC})] = -\frac{\sigma_j^2}{\sigma_\Delta}\lambda(z_{jk}), \quad E[q_{ik}|(\mathrm{MC})] = \frac{\sigma_k^2}{\sigma_\Delta}\lambda(z_{jk}).$$
(29)

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

Proof of Proposition 2. As Proposition 1 but the condition for relocation is now

$$w_{ij} < w_{ik} - \pi \iff \mu_{ij} - \mu_{ik} + \pi_{jk} < (\rho_k - \rho_j)\nu_i + q_{ik} - q_{ij}$$
 (30)

and thus let

$$z_{ijk} := \frac{1}{\sigma_{\Delta}} \left( \mu_{ij} - \mu_{ik} + \pi_{jk} - (\rho_k - \rho_j) \mu_{\nu} \right).$$
(31)

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

Proof of Proposition 3. Suppose  $\pi_{ijk}^m$  and  $\pi_{ijk}^c$  are normally distributed with means  $\bar{\pi}_{jk}^m$ ,  $\bar{\pi}_{jk}^c$  and equal standard deviations. Then by Lemma 4,  $\pi_{ijk}^m | \pi_{ijk}^m < \pi_{ijk}^c$  first oder stochastically dominates  $\pi_{ijk}^c | \pi_{ijk}^c < \pi_{ijk}^m$  if and only if  $\bar{\pi}_{jk}^m > \bar{\pi}_{jk}^c$ 

(i) We have

$$\bar{\pi}_{jk}^m > \bar{\pi}_{jk}^c \iff E[1 - \Phi(z_{ijk}(\pi_{ijk}^m)) | \pi_{ijk}^m < \pi_{ijk}^c] < E[1 - \Phi(z_{ijk}(\pi_{ijk}^c)) | \pi_{ijk}^m > \pi_{ijk}^c],$$

which, since  $1 - \Phi(z_{ijk}(\pi))$  is decreasing in  $\pi$ , follows from the first order stochastic dominance of  $\pi_{ijk}^m | \pi_{ijk}^m < \pi_{ijk}^c$  over  $\pi_{ijk}^c | \pi_{ijk}^c < \pi_{ijk}^m$ .

(ii) To be proven (with stochastic dominance)

(iii) Since  $\lambda \circ z$  is increasing in  $\pi$ , follows from the first order stochastic dominance of  $\pi_{ijk}^m | \pi_{ijk}^m < \pi_{ijk}^c$  over  $\pi_{ijk}^c | \pi_{ijk}^c < \pi_{ijk}^m$ .

*Proof of Proposition 4.* Assume Assumptions 1, 2, and 3. Then for parts (i) and (ii) we have

$$\begin{split} 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_{ijk}^{m} < \pi_{ijk}^{c}]] - E[E[\rho_{h}\nu_{i} + q_{ih}|(\mathrm{MC}), \pi_{ijk}^{m} > \pi_{ijk}^{c}]] \\ &= \begin{cases} E\left[\rho_{j}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{j} - \sigma_{j}^{2})\lambda(z_{ijk}(\pi_{ijk}^{m}))|\pi_{ijk}^{m} < \pi_{ijk}^{c}\right] \\ E\left[\rho_{k}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{k} + \sigma_{k}^{2})\lambda(z_{ijk}(\pi_{ijk}^{c}))|\pi_{ijk}^{m} > \pi_{ijk}^{c}\right] \\ - \begin{cases} E\left[\rho_{j}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{j} - \sigma_{j}^{2})\lambda(z_{ijk}(\pi_{ijk}^{c}))|\pi_{ijk}^{m} > \pi_{ijk}^{c}\right] \\ E\left[\rho_{k}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{k} + \sigma_{k}^{2})\lambda(z_{ijk}(\pi_{ijk}^{c}))|\pi_{ijk}^{m} > \pi_{ijk}^{c}\right] \\ \\ \left(\rho_{i}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{j} - \sigma_{j}^{2})\right)E\left[\lambda(z_{ijk}(\pi_{ijk}^{m}))|\pi_{ijk}^{m} < \pi_{ijk}^{c}\right] \\ - \begin{cases} \left(\rho_{j}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{j} - \sigma_{j}^{2})\right)E\left[\lambda(z_{ijk}(\pi_{ijk}^{m}))|\pi_{ijk}^{m} < \pi_{ijk}^{c}\right] \\ \\ - &\left\{\left(\rho_{j}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{j} - \sigma_{j}^{2})\right)E\left[\lambda(z_{ijk}(\pi_{ijk}^{c}))|\pi_{ijk}^{m} > \pi_{ijk}^{c}\right] \\ \\ - &\left\{\left(\rho_{\mu}\mu_{\nu} + \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{k} + \sigma_{k}^{2}\right)\right)E\left[\lambda(z_{ijk}(\pi_{ijk}^{c}))|\pi_{ijk}^{m} > \pi_{ijk}^{c}\right] \\ \\ - &\left\{\left(\frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{j} - \sigma_{j}^{2})(\lambda^{m} - \lambda^{c}) \\ \\ \frac{1}{\sigma_{\Delta}}(\sigma_{\nu}^{2}(\rho_{k} - \rho_{j})\rho_{k} + \sigma_{k}^{2})(\lambda^{m} - \lambda^{c}) \\ \end{array}\right.\right\}$$

where the first equality follows from Assumption 1, the second from Assumption 3 and the third and fourth from Assumption 2 and where we define  $\lambda^m := E[\lambda(z_{ijk}(\pi^m_{ijk}))|\pi^m_{ijk} < \pi^c_{ijk}]$  and  $\lambda^c := E[\lambda(z_{ijk}(\pi^c_{ijk}))|\pi^m_{ijk} > \pi^c_{ijk}]$ . Part (iii) follows from parts (i) and (ii).  $\Box$