WORKING PAPER 2025:24

Labor market size and occupational skill match

Erika Forsberg

IFAU INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Labor market size and occupational skill match [a]

Erika Forsberg[b]

November 27, 2025

Abstract. Individuals working in larger labor markets tend to earn more than those working in smaller labor markets, but the reason for this is still unclear. This paper studies whether larger cities provide better occupational skill matches by combining machine learning techniques with data on individuals' productive skills matched with employer data to construct a novel measure of match quality. I show that occupational skill-match quality is higher for individuals living in large local labor markets. Conditional on skills, differences in match quality explain around 30 percent of the city-size wage gap. The higher match quality in larger labor markets is related to a more diversified occupation structure and more learning possibilities in these markets.

Keywords: Matching, Agglomeration, Occupational choice
JEL Codes: R12, R23, J24, J31

[a] I am grateful to my previous supervisors Martin Nybom, Lena Hensvik, and Hans Grönqvist for their support. I also thank Petter Berg, Matz Dahlberg, Jorge De la Roca, Mitchell Downey, Anders Forslund, Simon Franklin, Peter Fredriksson, Adam Gill, Felipe Gonzalez, Georg Graetz, Rafael Lalive, Sang Yoon Tim Lee, Matthew Lindquist, Marco Manacorda, Raoul van Maarseveen, Federica Meluzzi, Enrico Moretti, Oskar Nordström Skans, Luca Repetto, Anna Raute, Olof Rosenqvist, Michael Simmons, Tom Zohar and participants at presentations at workshops/seminars at the poster session at HCEO-FAIR 2022 Summer School on Socioeconomic Inequality Bergen, Queen Mary University of London, Uppsala brown bag, Uppsala labor group, the Uppsala Urban lab workshop, the 5th Queen Mary University of London Economic and Finance workshop, Second Nordic meeting in urban economics, seminar at Linnaeus University, seminar at the Institute for Evaluation of Labour Market and Education Policy (IFAU), and seminar at Umeå University for valuable comments.

[b] IFAU, erika.forsberg@ifau.uu.se

1 Introduction

A large empirical literature has shown that individuals who work in larger labor markets earn more than individuals who work in smaller labor markets (see for example Rosenthal and Strange, 2008; Papageorgiou, 2022; Eliasson and Westerlund, 2022). This city-size wage premium should partly be explained by productivity differences (see for example Glaeser and Maré, 2001), otherwise firms in the tradable sector would increase profits by relocating to smaller cities where both wage and land costs are lower (Moretti, 2011). From the workers' perspective, productivity differences could remain due to mobility frictions, such as higher cost of living in larger cities or individuals having preferences to reside in certain areas.

To be able to address geographical inequalities, it is important to understand the mechanism behind the city-size wage differential. Theoretically, competing mechanisms such as increased opportunities for individuals to augment human capital and higher match quality in larger labor markets have been suggested as explanations for the wage difference (see for example Puga, 2010). While research has provided evidence on learning and city size (see for example De La Roca and Puga, 2016), the empirical evidence on how match quality differs between large and small labor markets is still limited. The main reason is the challenges involved in measuring match quality. Most of the existing literature instead relies on indirect evidence. For instance, Wheeler (2008), Bleakley and Lin (2012), and Korpi and Clark (2019), show job change patterns that are consistent with higher match quality in larger cities.

In this study, I use rich population-based data to construct a novel measure of occupational skill-match quality that allows me to provide direct evidence of the importance of this mechanism in explaining the city-size wage differential. The measure draws from advances in the literature on multidimensional mismatch (see Guvenen et al., 2020; Fredriksson et al., 2018), which highlights that different combinations of skills might be useful in different jobs. To construct the match quality measure, I use Swedish data on eight different types of skills from the military enlistment test. The tests are conducted at age 18 or 19 and include both cognitive (inductive, verbal, spatial, and technical ability) and non-cognitive skills (social maturity, intensity, psychological energy, and emotional stability). As a proxy for skill requirements in each occupation, I estimate the return to skills in different occupations. Because theory gives little guidance on how to model the relationship between skills and wages, I pursue a non-parametric approach to estimation. Specifically, I use recent advancements in machine learning to improve on the earlier match quality measure and estimate a random forest model trained on tenured workers with data on the eight skills.[1] Analogously, I use a random forest to estimate the return to skills on the whole market. The match quality measure is then constructed as the estimated return to skills in the occupation where the individual works, minus the market return to skills. Thus, individuals are well-matched if they work in occupations with high returns to their skills compared to what they could receive on average based on their skills.

[1]See recent work by Almgren et al. (2022) who also use a random forest to proxy for skill requirements in occupations.

My results show that match quality is higher in larger compared to smaller labor markets. This result holds conditional on skills and in a subset of workers who move across labor markets of different sizes. Conditional on skills, the difference in match quality explains around 30 percent of the city-size wage gap. The magnitude is in line with the result of Papageorgiou (2022), who uses a calibrated model and shows that occupational match quality explains around 35 percent of the city-size wage premium in the United States. The difference in match quality between individuals in larger and smaller labor markets is especially large for high-skilled workers, measured either by their cognitive skills, non-cognitive skills, or education. This is well in line with the fact that the city-size wage premium is higher for highly skilled and highly educated individuals (see Bacolod et al., 2009; Autor, 2019).

At the beginning of the career, match quality is similar in labor markets of different sizes, and occupational match quality then increases over the career. This result is consistent with initial uncertainty about the optimal match that decreases with experience (see Guvenen et al., 2020; Fredriksson et al., 2018). The increase in match quality over the life-cycle is also in line with job-ladders models, where search frictions can lead to variation in jobs and job-ladders (see Burdett and Mortensen, 1998; Christensen et al., 2005; Manning, 2003), and where individuals gradually are able to climb to better occupation matches. However, the increase in match quality over the life-cycle is substantially larger in large labor markets, leading to an increased city-size match quality gap for older individuals. The increase in the city-size match quality gap over the life-cycle seems to come from both more frequent, and better, occupation switches in larger labor markets. The finding of a steeper increase in match quality in larger cities over the life-cycle is also in line with the finding by Eckert et al. (2022). The authors find that refugees placed in a large labor market initially have similar wages as refugees placed in smaller labor markets and experience faster wage growth through sorting to better (urban) jobs.

I examine two mechanisms behind the higher match quality in larger cities: occupation diversity and learning possibilities. If a more diverse set of occupations exists in large labor markets, this could give individuals more occupations to choose from and, therefore, could increase the likelihood that individuals find a good match (see Papageorgiou, 2022). Theoretically, higher learning possibilities in larger cities could come from a lower cost of switching occupations in larger cities (Wheeler, 2008). This allows individuals to explore different occupations and learn about what their optimal match is. Moreover, working in larger cities gives workers more valuable knowledge (Glaeser and Maré, 2001; De La Roca and Puga, 2016), which in some cases might be needed to be able to enter occupations with high match quality.

The analysis gives suggestive evidence that both of these mechanisms are at work. By constructing a Herfindahl-Hirschman Index (HHI) of occupation concentration in the labor markets, I show that larger labor markets have a more diversified occupation structure. Moreover, the difference in match quality in large and small labor markets partly seems to be explained by the difference in occupation diversity. Following the method in De La Roca and Puga (2016), I also show that experience obtained in larger labor markets is more valuable for future match quality than experience obtained in smaller labor markets. In addition, prior experience from large labor markets is valuable for individuals who work in small labor markets, suggesting that learning might be an important mechanism for higher match quality in larger labor markets.

This paper is related to several strands of literature. First, this paper is related to the strand of research that employs AKM models (see Abowd et al., 1999) to investigate mismatch between individuals and firms in relation to city size. These studies estimate assortative matching based on one-dimensional skills and tend to find that larger cities have a higher degree of assortative matching (see for example Dauth et al., 2022; Andersson et al., 2007; Leknes et al., 2022; Card et al., 2025).[2] While these studies have focused on the mismatch between workers and firms, I shift the focus to occupational mismatch. Occupational match quality plays a vital role in explaining workers' earnings (Guvenen et al., 2020). Consequently, investigating the role of occupational match quality in explaining the city-size wage premium is crucial. Furthermore, while prior studies employing AKM models have examined matching in terms of one-dimensional skills, my research utilizes data on eight distinct types of skills and estimates a multidimensional match quality measure. The match quality measure allows for both horizontal dimensions of mismatch, in terms of workers being overqualified and underqualified for their jobs, and vertical dimensions of mismatch, in terms of workers having the wrong set of skills for their jobs.

Secondly, this paper complements the literature exploring the mismatch between individuals' education and the educational requirements of their jobs, which tends to find higher education mismatch in smaller cities (see Abel and Deitz, 2015; Berlingieri, 2018; Boualam, 2014; Koster and Ozgen, 2021). Moreover, smaller cities seem to have more mismatch in terms of earlier industry experience (Harmon, 2013). Using survey questions, Andini et al. (2013) find small effects of density on match quality and Corradini et al. (2025) find a positive effect from industrial density on match quality. In contemporaneous work, Moretti and Yi (2024) show that after displacement, workers in larger labor markets are more likely to find a job with higher match quality, as indicated by the fact that they are more likely to work in an industry relevant to their college major, be employed in the same industry as before, and that the new job is likely to last longer. Moreover, the paper complements studies that use calibarted models to study occupational matching depending on city size (Papageorgiou, 2022), by providing a direct measure of occupational match quality.

In contrast to earlier studies, this paper looks at mismatch in terms of skills that are measured at age 18 or 19, when the individuals are unlikely to have entered the labor market. Unlike education choices, which may be influenced by local labor market conditions, pre-labor market skills are unlikely to be shaped by local labor market conditions. Consequently, this study enriches the existing literature by shedding light on match quality of pre-determined skills, and suitable occupations in markets of different sizes. Moreover, I provide novel evidence on the life-cycle dynamics of match quality in relation to labor market size and direct evidence on how the city-size match quality gap is related to learning possibilities and occupation diversity.

[2] In contrast Mion and Naticchioni (2009) find a negative association between market size and assortative matching in Italy and Figueiredo et al. (2013) find limited evidence of more associative matching with more firm clustering within the same industry in Portugal.

The paper is structured as follows: Section 2 presents the conceptual framework and section 3 describes the data. In section 4, the empirical approach is explained, and section 5 presents background information about how wage differences and labor market size look in the Swedish setting. The main results are presented in section 6 and the mechanisms behind the higher match quality in larger labor markets are examined in section 7.Section 8 explores the robustness of the results and the conclusions are discussed in section 9.

2 Conceptual framework

This section presents a simple framework to illustrate the idea behind the difference in match quality between large and small labor markets. Before entering the labor market, the individual has a set of different skills that are useful in different occupations. The more similar the skills obtained by the individual are to the skills valued in the occupation, the higher the match quality. Higher match quality increases productivity, which is shared with the worker in terms of higher wages.

In a world without frictions, individuals would move to labor markets with higher match quality until the match quality between different labor markets equalizes. However, in the presence of mobility frictions, this might not happen. To get the idea behind the mobility friction, consider a utility function similar to Moretti (2011).

$$U_{iL} = W(MQ, X)_{iL} - r_L + A_L + e_{iL}$$

The utility for individual (i) living in the local labor market (L) depends on the wage (W), the cost of living (r), amenities (A) and idiosyncratic preferences e_{iL}. The idiosyncratic preferences e_{iL} show how much the individual values the labor market holding the wage and amenity constant. For example, Moretti (2011) claims that being born in a city or having relatives in the city might give a higher value of e_{iL}. If e_{iL} is high, regional mobility will be low. Yagan (2019) shows that individuals do not seem to migrate to new places to any great extent as a response to economic shocks, indicating that mobility frictions might be high in practice.

Match quality (MQ) is assumed to have an impact on utility by affecting the wage. Productive skills also depend on a vector of other worker characteristics X, which, for example, can be the age of the worker. If workers get direct utility from working in an occupation that matches their skill set, match quality can also enter the utility function through the amenity. If idiosyncratic preferences e_{iL} are high, individuals might not move even in the case of large differences in real wages and amenities. Thus, the implication from the equation above is that in the presence of mobility frictions, match quality might differ between different labor markets.

In the presence of mobility frictions, the possibility of forming good matches in the labor market has been theorized to be higher in large labor markets (Puga, 2010). There are at least three mechanisms for why match quality might be higher in larger labor markets: occupation diversity, frictions, and learning possibilities.

2.1 Occupation diversity

Larger labor markets have been shown to have a more diversified occupation structure in the sense that more different occupations exist in large labor markets (Papageorgiou, 2022; Korpi, 2007), where especially scarce specialist occupations are more common in larger cities (Duranton and Jayet, 2011). A more diversified occupation structure gives the individual more occupations to choose from which could increase the likelihood that individuals find the optimal match given their skills.

2.2 Search frictions

The cost of switching jobs and occupations has been theorized to be lower in larger cities since larger cities have more job openings (Puga, 2010; Bleakley and Lin, 2012). With more job openings, individuals could apply to many jobs at the same time and be able to change jobs when they like instead of waiting for a job opening. Thus, lower search frictions speed up the matching process.

2.3 Learning possibilities

Earlier research about match quality has shown that work experience and job switches are important for learning about the optimal match (Guvenen et al., 2020; Fredriksson et al., 2018). A lower cost of switching occupations in larger labor markets will allow the individual to change occupations more frequantly and thus explore more different occupations (Wheeler, 2008). Thus, less search frictions in larger labor markets might give individuals the opportunity to learn about what occupations are a good match for them, given their skills. Moreover, working in larger cities has been shown to provide larger human capital learning possibilities compared to working in a smaller city (Glaeser and Maré, 2001; De La Roca and Puga, 2016). Thus, experience from working in larger cities might give workers valuable knowledge, which might open up possibilities to enter some new occupations with higher match quality. Finally, larger labor markets have more education possibilities (Frenette, 2006). For some occupations, a university education might be necessary for entering the occupation. Thus, more learning through university education could open doors to be able to enter more occupations.

Taken together, from theory, we could expect higher match quality in larger labor markets, driven by more occupation diversity in larger cities, lower search frictions, and more learning possibilities. To implement policies to increase match quality in small labor markets, it is vital to know why match quality differs between labor markets of different sizes. Thus, in section 7 I try to separate between the mechanisms. If occupation diversity drives the result, controlling for a measure of occupation diversity is expected to reduce the estimated city-size match quality gap. If learning possibilities in larger labor markets are driving the result, we would expect that the gain in match quality from working in a large labor market should not be immediate; instead, match quality is expected to increase over time when working in a large compared to a small labor market.

3 Data

I use administrative wage data collected by Statistics Sweden. Occupation data are available for a large part of the Swedish population, covering almost 50 percent of private and all public sector workers.[3] Skills are measured with scores from tests during the military enlistment for those who enlisted between the years 1969 and 1994 when almost all males participated in the military drafting. The enlistment is done at age 18 or 19. Since military enlistment was only mandatory for males, the data is limited to only include males.

The cognitive measures include four cognitive skills: inductive, verbal, spatial, and technical ability. The non-cognitive measures also capture four skills: social maturity, intensity, psychological energy, and emotional stability. The evaluation of the non-cognitive score is done by a psychologist in a 20-minute interview (see Mood et al., 2012); before the interview, the psychologist has access to a form where the tested has answered questions about friends and family. In some years, the cognitive tests are graded on a scale of 0-25, and in other years, on a scale of 0-40. The non-cognitive scores are measured on a scale of 1-5. To make the test scores comparable, all test scores have been standardized within each year of enlistment to have a mean of 0 and a standard deviation of 1.

Occupation data is used on the three-digit level for the years 1996-2013, with the ssyk96 definition, in turn based on the international isco88. Some individuals have multiple occupation observations in a given year (around 4 percent of the sample). For these individuals, only the main occupation is used, where the main occupation is defined as the occupation with the highest wage expressed in full-time equivalent wages. In the sample, there exist 111 occupations.[4] Wage is measured as full-time equivalent wages. Furthermore, the data is linked to data on education, firms, and the municipality the individual lives in.

To define labor markets, I rely on Statistics Sweden's definition of local labor markets, defined from commuting patterns. The local labor market definition in 2013 is used, and thus, municipalities are assumed to belong to the same local labor market during the whole sample period. This definition gives 73 local labor markets. In the main specification, the log of the number of inhabitants in the local labor market is used to measure labor market size. In some cases, a categorical definition of the size of the local labor market is used. The labor market is then divided into three size categories. Large labor markets are defined as labor markets with more than 500,000 inhabitants, consisting of three labor markets: Stockholm-Solna, Malmö-Lund, and Göteborg. Medium-sized labor markets are defined as labor markets with more than 100,000 and less than 500,000 inhabitants, consisting of 20 local labor markets. The small labor markets are labor markets with less than 100,000 inhabitants, consisting of 50 local labor markets. The number of inhabitants used is the number of inhabitants in the local labor market in 2013, defined by SCB.

To be able to use the military and occupation data, the sample is restricted to men born between

[3] The wage and occupation information is collected during a measurement week (in September–November) each year. To be included in the sampling, the individual needs to be employed for at least one hour during the sampling week. Sampling is stratified by firm size and industry, and small firms in the private sector are underrepresented.

[4] In total, there exist 113 occupations at the three-digit level. However, two occupations are excluded because they have fewer than one tenured worker (the excluded occupations are photo models and street market salespersons).

1951 and 1976, with non-missing occupation data from 1996 to 2013. The sample is then restricted to those with data on the military test scores, which is available for 80 percent of the sample. Of this sample, 17 percent miss test scores data on at least one of the cognitive or non-cognitive sub-tests. The sample further excludes those individuals. To train the random forest, 20 percent of all tenured individuals are used, where tenured are defined as individuals who have worked in the occupation for at least three years. The test data consists of the remaining 80 percent of the tenured individuals and all individuals who never have tenure. The results in the paper use the individuals in the test data.

Table 1 summarizes the data by labor market size. As can be seen in the table, even if most labor markets are small, the number of observations is larger in large labor markets. Thus, a large part of the population lives in large labor markets. In panel B, population characteristics are shown divided by labor market size. The population in large labor markets is slightly younger. Cognitive skills, non-cognitive skills, and education levels are higher in larger labor markets, in line with working sorting being an important explanation for earnings differences between different labor markets (see Card et al., 2025). The observed skill differences are also consistent with sorting incentives, where it has been shown that individuals with higher skill levels receive a greater city-size premium (Bacolod et al., 2009; Andersson et al., 2014; Koster and Ozgen, 2021; Carlsen et al., 2016; Neves et al., 2017) and are more likely to migrate to large labor markets (see Bacolod et al., 2021). The summary statistics also show that descriptively, wages and match quality are highest in large labor markets and smallest in small labor markets.

Table 1: Summary statistics

	Large labor markets (1)	Medium sized labor markets (2)	Small labor markets (3)
Panel A. Labor markets			
Number of labor markets	3	20	50
Number of observations	2,660,719 (44%)	2,407,657 (40%)	937,253 (15%)
Panel B. Population characteristics			
Age	41.85	41.90	42.49
Cognitive skill	0.29	0.09	-0.036
Non-cognitive skills	0.14	0.07	-0.01
Share University education	49.40	37.70	28.86
LN Wage	10.35	10.22	10.18
Match quality	0.097	-0.06	-0.11

Notes: This table provides summary statistics separately by the size of the local labor market. I rely on Statistics Sweden's definition of local labor markets based on commuting patterns. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Wages are measured in 2013 year price levels. The skill and match quality measure is standardized to have mean 0 and standard deviation 1.

4 Empirical approach

This section describes how the match quality measure is constructed and presents the method for studying the difference in match quality in labor markets of different sizes.

4.1 Match quality measure

To construct the match quality measures, I use data on wages and eight different types of skills, both cognitive (inductive, verbal, spatial, and technical ability) as well as non-cognitive skills (social maturity, intensity, psychological energy, and emotional stability). The skills measures have been shown to be important predictors of wages (Lindqvist and Vestman, 2011). The idea behind the match quality measure is that different combinations of skills might be differently useful in different occupations, and individuals have high match quality if they work in an occupation where their skills are highly valued. I follow Fredriksson et al. (2018) and use tenured workers to proxy for skill requirements. The conclusions remain similar when using O*Net data to proxy for skill requirements in the occupation.[5]

The match quality measure is then constructed by estimating returns to skills in each occupation. The functional form of the match quality measure is theoretically unclear. I thus deviate from the linear match quality measure used in Fredriksson et al. (2018) and instead use recent advancements in the machine learning literature and estimate the match quality measure using a random forest (in similarity with Almgren et al., 2022).[6]

The random forest is trained on data on each of the eight skills and residualized wages, where log wages are residualized on age and year-fixed effects. The training sample is constructed from a random sample of 20 percent of all tenured individuals. Individuals from the training sample are then excluded from the test data. All results shown in the paper only include individuals in the test data. The random forest is trained with residualized wages as the outcome and the eight skills as the independent variables. Thus, the random forest predicts wages based on skills. Separate random forest algorithms are estimated for each occupation, allowing different interactions of skills to be differently important for returns in each occupation. Appendix A shows the relationship between predicted wages based on skills from the random forest for the test data and actual residualized wages. As can be seen in Figure A.1 there is a clear positive relationship between the predicted wages from the random forest and actual wages.

Highly skilled individuals will have higher returns in all occupations. To take away the direct

[5]Section 8.3 provides results where the match quality measure instead are constructed using O*NET data to proxy for skill requirement in each occupation, following Guvenen et al. (2020).

[6]Fredriksson et al. (2018) use two different match quality measures, one using the skills of tenured workers and one using the estimated return to skills. The match quality measure in this paper is more similar to the estimated return to skills match quality measure, where match quality is defined as $MQ_{ij} = \sum_{s=1}^{8}(\hat{\beta}_{js} - \hat{\beta}_s)X_{is}$, where $\hat{\beta}_{js}$ is the estimated return to skill s in job j, $\hat{\beta}_s$ is the estimated return to skill s on the market, and X_{is} is the skill level of skill s for individual i. In difference to Fredriksson et al. (2018), this study focuses on match quality on the occupation level instead of job level and uses a random forest instead of assuming a specific functional form of the match quality measure. Section 8.3 presents results when using the linear version of the match quality measure.

impact of skills, I construct the match quality measure as the return to skill for the occupation the individual works in minus the market return to skill. To estimate the market return, I train the random forest on all workers in the training sample. Thus, the market return measures the returns to skill on average on the whole labor market, and is not restricted to the labor market the individuals live in, and aims to capture the average return the individual could receive given their skills. Thus, the idea behind the match quality measure is that the match quality is higher if an individual works in an occupation with higher returns to skills than they could receive on average on the labor market. Match quality is then defined as predicted earnings based on skills in the occupation the individual works in minus market returns:

$$MQ_{io} = Return_{io} - Return_{iM}$$

where $Return_{io}$ is the predicted return estimated with the random forest based on the skills for individual i working in occupation o, and $Return_{iM}$ is the estimated market return for individual i. To ease interpretation, the match quality measure is then normalized to have a mean zero and standard deviation equal to one. Figure A.2 in appendix A shows match quality plotted against skills. The figure shows a negative relationship between match quality and skills, indicating that higher match quality in larger cities can not be driven by higher skill levels in larger cities. Instead, it seems like individuals with higher skill levels have a harder time finding a good match compared to individuals with lower skill levels.

Since match quality is predicted using only skills, this implies that two individuals with the same skills who work in the same occupation will have the same match quality, even if they live in different labor markets. Moreover, the only way for an individual to change match quality is by changing occupation.

4.2 Methodology

To estimate the difference in match quality depending on labor market size, I estimate equation (1) below, using OLS.

$$MQ_{ito} = \beta Log(Citysize)_{itL} + X_{it} + \varepsilon_{it} \ (1)$$

The dependent variable of interest is MQ_{it}, the match quality for individual i in time period t working in occupation o. The independent variable of interest is $Log(Citysize)_{itL}$, the log of the number of inhabitants in the local labor market.

In the main specification, X_{it} includes age-fixed effects, year-fixed effects, and second-order polynomials in each of the eight skills to compare individuals with similar skill levels. Furthermore, alternative specifications are presented where X_{it} includes controls for additional individual characteristics such as education, and, in some specifications also individual fixed effects. It should be noted that, while controlling for education allows comparing more similar individuals, more education possibilities in larger labor markets could also be a mechanism for why individuals in larger labor markets can obtain higher match quality. Individual fixed effects control for all factors that are constant within the individual over time. However, when using individual fixed effects, the effect of local labor market size on match quality is identified from individuals who move between local labor markets of different sizes. If individuals dislike moving, we can expect that individuals who still choose to move do this for a reason, for example, to improve their occupational match. Moreover, when including individuals fixed effect match quality will only change for individuals who change occupation. Since many individuals might have already invested in an occupation before they move, this indicates that the effect on match quality from moving might be smaller than for individuals who live in labor markets of different sizes. Thus, we can expect the estimate for the sub-population of movers to differ from the effect identified for all individuals.

5 Wage differences and labor market size

Research has shown that individuals living in larger labor markets earn more than those living in smaller labor markets. This pattern has been shown to hold in the United States (Papageorgiou, 2022), and in different European countries (see for example Rosenthal and Strange, 2008) including Sweden (see Eliasson and Westerlund, 2022). Table 2 provides evidence on wage premiums in larger labor markets, also in the setting studied here, by estimating regression 1 with log wage as the dependent variable. Conditional on only year and age fixed effects, the city-size wage elasticity is 0.053, indicating that a 1 percent larger city is associated with 0.05 percent higher wages. Adding controls for skills, the wage elasticity is still 0.039, which means that conditional on skills wages are 12 percent higher in large compared to small labor markets (see table B.2 in appendix). Adding controls for education in addition to skills only reduces the estimate slightly, and the wage elasticity of city size is still 0.036.

The magnitude of the effect changes somewhat when individual fixed effects are included. This may be because the individual fixed effects control for time-invariant characteristics of the individual that influence both wages and the choice of where to live. However, the change in the estimated relationship may also reflect that the coefficient is now identified from individuals who move between labor markets of different sizes. This group may differ from the average population, for example because they may choose to relocate in order to improve their career opportunities. However, even when including individual fixed effects, wages are still higher in larger compared to smaller labor markets.

Figure 1 below shows a map of the distribution of labor market size in Sweden. As can be seen in the map, smaller and larger labor markets exist in all parts of Sweden.

Table 2: Wage and labor market size

VARIABLES	(1) Ln Wage	(2) Ln Wage	(3) Ln Wage	(4) Ln Wage
Log population size	0.053***	0.039***	0.036***	0.015***
	(0.000)	(0.000)	(0.000)	(0.000)
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Skill controls	No	Yes	Yes	No
Education control	No	No	Yes	No
Individual FE	No	No	No	Yes
Observations	6,005,629	6,005,629	6,002,946	6,005,629

Notes: This table provides results estimating equation 1, with log wage as the dependent variable, estimating the difference in log wages for individuals who live in labor markets of different sizes. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. In column 4, where individual fixed effects are included, the age dummies are normalized to be constant between 45 and 54 to avoid multicollinearity between age and year-fixed effects. Standard errors are adjusted for 659,206 clusters at the individual level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

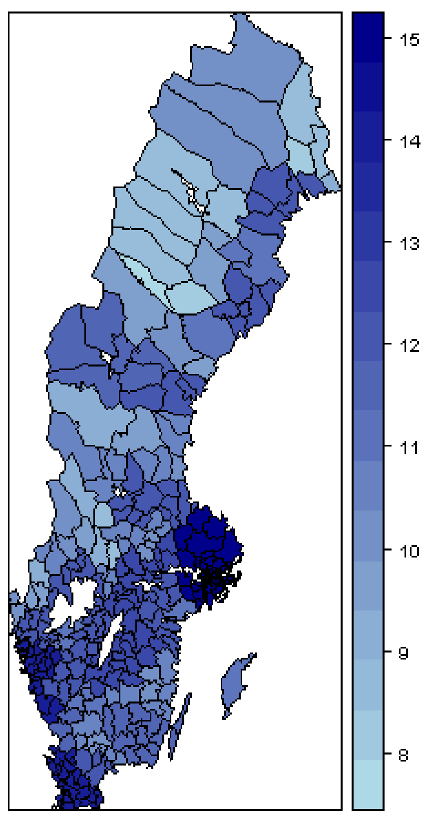


Figure 1: Size of local labor markets in Sweden

Notes: The figure plots the log size of local labor markets in Sweden. I rely on Statistics Sweden's definition of local labor markets based on commuting patterns. Darker colors represent larger labor markets.

6 Results

This section presents the key results on the city-size match quality gap. Section 6.1 presents the main results on how match quality differs depending on labor market size estimated with regression 1. It is possible that not all individuals gain equally in terms of match quality from living in a large labor market. Section 6.2, therefore presents how match quality differs with labor market size separately depending on an individual's skill level. Section 6.3 presents evidence on how match quality develops over the life-cycle for individuals who live in small, medium, and large labor markets.

6.1 Main results

Table 3 presents results for how match quality varies with local labor market size. In column 1, we see that individuals living in larger labor markets have higher match quality than individuals in smaller labor markets. Column 2 controls for second-order polynomials in each of the eight skills. Controlling for skills has a small and positive impact on the estimate, indicating that the difference in match quality depending on labor market size is not driven by different skill compositions of workers depending on local labor market size. When controlling for skills, the estimates suggest that an increase in city size by 10 percent is associated with 0.007 standard deviations higher match quality. Thus, the results in the table are in line with the theory of higher match quality in larger labor markets.

Table 3: Main result

VARIABLES	(1) Match quality	(2) Match quality	(3) Match quality	(4) Match quality
Log population size	0.064***	0.070***	0.059***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Skill controls	No	Yes	Yes	Yes
Control education	No	No	Yes	No
Individual FE	No	No	No	Yes
Observations	6,005,629	6,005,629	6,002,946	6,005,629

Notes: This table provides results estimating equation 1, estimating the difference in match quality for individuals who live in labor markets of different sizes. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. In column 4, where individual fixed effects are included, the age dummies are normalized to be constant between 45 and 54 to avoid multicollinearity between age and year-fixed effects. Standard errors are adjusted for 659,206 clusters at the individual level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Column 3 in addition controls for education. The estimate falls somewhat, indicating that one mechanism for higher match quality in larger labor markets might be more education possibilities, which might increase the potential to realize the optimal match. However, even conditional on education match quality is higher in larger local labor markets. Column 4 includes individual fixed effects, and thus the estimated effect of local labor market size on match quality comes

Table 4: Quantify the contribution of match quality to city-size wage premium

VARIABLES	(1) Ln Wage	(2) Ln Wage	(3) Ln Wage	(4) Ln Wage
Log population size	0.039***	0.028***	0.053***	0.043***
	(0.000)	(0.000)	(0.000)	(0.000)
Match quality		0.169***		0.163***
		(0.000)		(0.000)
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Skill controls	Yes	Yes	No	No
Observations	6,005,629	6,005,629	6,005,629	6,005,629

Notes: This table provides results estimating equation 1, with log wage as the dependent variable adding match quality as a control comparing the coefficient for local labor market size with and without controls for match quality. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Standard errors are adjusted for 659,206 clusters at the individual level. *** p<0.01, ** p<0.05, * p<0.1

from individuals who move between local labor markets of different sizes. When controlling for individual fixed effects the effect decreases substantially. As discussed before, this could be because the individual fixed effect captures individual characteristics that affect match quality. On the other hand, it could also be because the effect is now identified by movers. If individuals dislike to move, they might, for example, only choose to move if they receive higher match quality. Nevertheless, even when including individual fixed effects, match quality is higher in larger compared to smaller local labor markets.

Table B.1 in Appendix B provides results for alternative specifications of labor market size, providing results with labor market size as a categorical variable, where labor markets are divided into large, medium, and small. Moreover, Table B.1 provides results using labor market density instead of labor market size. Finally, Table B.1 also presents results using divisions of municipalities into rural areas, mid-sized areas, and large cities. The conclusion from the alternative definitions of labor market size is similar, with higher occupational match quality in larger labor markets. From Figure B.1 it is clear that the relationship between match quality and city size is relatively linear, where match quality gradually increases with log city size.

The result in Table 3 speaks to the potential role of match quality as one explanation for why larger labor markets have higher wages. However, the question still remains if match quality is an important contribution factor to the city-size wage premium or not. In Table 4 I quantify how much of the wage premium in large local labor markets that can be attributed to match quality. This is done by estimating the city-size wage premium regression and comparing the coefficient for local labor market size with and without controls for match quality. Thus, the coefficient in column 2 shows how much of the city-size wage premium remains when match quality is held constant, and the difference between the coefficients in columns 1 and 2 shows how much of the wage gap between large and small labor markets that can be attributed to match quality. Note that controls for skills are included in both columns 1 and 2. Thus, the difference between the coefficients in columns 1 and 2 shows how much match quality contributes to the conditional wage premium, when the fact that the population in small and large labor markets has different skill composition has already been taken into account [7].

Comparing the coefficients in columns 1 and 2 suggests that match quality explains around 30 percent of the city-size wage gap. Occupational match quality thus explains an important part of the wage difference between smaller and larger labor markets. The magnitude is in line with the result of Papageorgiou (2022), who uses a calibrated model and shows that occupational match quality explains around 35 percent of the city-size wage premium in the United States.

When using match quality as a control instead of as the dependent variable, it is important to get as accurate measure of match quality as possible. If the match quality measure is estimated with error, controlling for match quality might not capture the true match quality effect, and the contribution of match quality to the city-size wage gap might be underestimated. When using the linear version of match quality as a control instead of the non-parametric match quality measure estimated with the random forest, match quality is estimated to explain around 10 percent, instead of 30 percent of the city-size wage gap (see table A.1 in appendix). Thus, even if the conclusion of higher match quality in larger labor markets holds with the linear match quality measure (see table 10), using the recent advancements in machine learning to estimate a non-parametric version of the match quality measure seems to allow for a more accurate quantification of the importance of match quality for explaining the city-size wage gap.

6.2 Heterogeneity analysis

Table 5: Heterogeneity analysis

	Dependent variable: Match quality					
	(1) Cognitive ability under median	(2) Cognitive ability over median	(3) Non-cognitive ability under median	(4) Non-cognitive ability over median	(5) High school or less	(6) University
Log population size	0.041***	0.099***	0.046***	0.094***	0.042***	0.085***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Control skills	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,004,065	3,001,564	3,068,475	2,936,693	3,409,100	2,594,733

Notes: This table provides results estimating equation 1, estimating the difference in match quality for individuals who live in labor markets of different sizes, separately for different skill groups. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1

[7] Comparing column 3 and 4 gives how much match quality contributes to the city-size wage gap when skills have not been taken into account. When not conditioning on skills, occupational match quality explains 20 percent of the city-size wage gap.

Earlier research has shown that the city-size wage premium is higher among high-skilled individuals (see for example Bacolod et al., 2009). For policy makers who may want to decrease the city-size match quality gap, understanding if the city-size match quality gap exists for everybody or also is concentrated to highly skilled individuals might be important to implement policy effectively. Table 5 therefore explores whether the city-size match quality gap is also higher among high-skilled individuals. Heterogeneity analysis is presented by the level of cognitive skills, non-cognitive skills and education.

Table 6: Contribution to city-size wage premium by skill group

	Dependent variable: Ln Wage					
	(1) Cognitive ability under median	(2) Cognitive ability over median	(3) Non-cognitive ability under median	(4) Non-cognitive ability over median	(5) High school or less	(6) University
Panel A.						
Log population size	0.026***	0.053***	0.028***	0.050***	0.025***	0.053***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel B.						
Log population size	0.019***	0.036***	0.021***	0.035***	0.019***	0.038***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Match quality	0.168***	0.168***	0.169***	0.167***	0.148***	0.171***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year Fe	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Control skills	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,004,065	3,001,564	3,068,475	2,936,693	3,409,100	2,594,733
Panel C. Share of city size wage premium explained by match quality						
Share explained	27%	32%	25%	30%	24%	28%

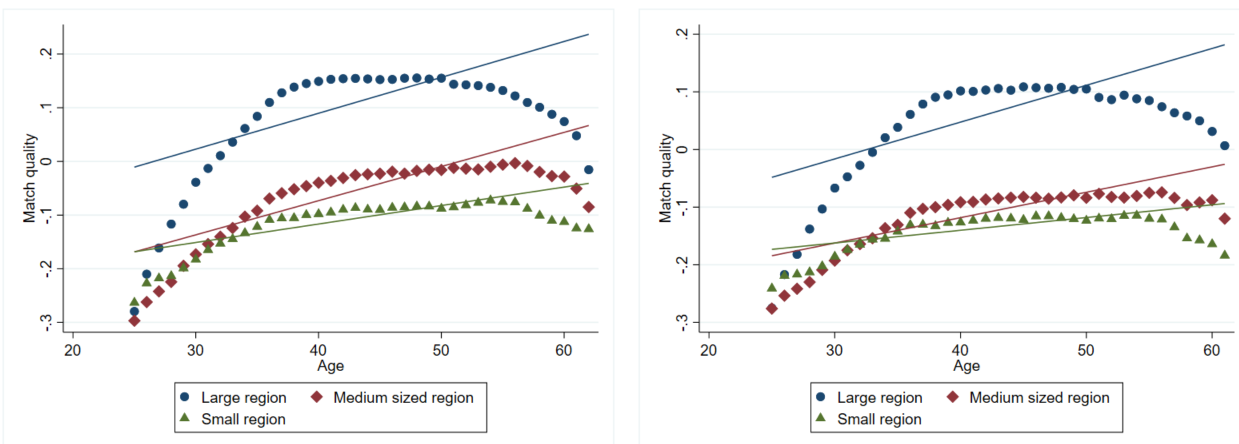
Notes: This table provides results estimating equation 1, separately by skill group, with log wage as the dependent variable adding match quality as a control comparing the coefficient for local labor market size with and without controls for match quality. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Panel A shows results for the city-size wage premium without controls for match quality, Panel B shows results for the city-size wage premium with controls for match quality and Panel C shows how much match quality contributes to the city-size wage premium. Standard errors are clustered at the individual level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

There is a positive association between match quality and living in a larger labor market across the skill distribution. However, the correlation between labor market size and match quality is especially large for highly skilled individuals, a pattern that holds when measuring skills in all dimensions: cognitive skills, non-cognitive skills, and education. Table 6 quantifies how much of the city-size wage premium that can be explained by match quality separately for all skill groups. Match quality is an important explanation behind the city-size wage premiums for all groups and explains between 24% to 32% of the city-size wage premium. Match quality explains a slightly higher part of the wage premium for high-skilled workers, indicating that one potential reason for higher wage premiums for high-skilled individuals might be that high-skilled individuals benefit more in terms of match quality from living in a large compared to a small local labor market. Nevertheless, it should be noted that both individuals with higher and lower skill levels seem to benefit in terms of higher match quality from living in larger labor markets.

6.3 Life-cycle pattern

The previous section showed that individuals living in large labor markets have higher match quality than individuals in smaller labor markets, and that occupational match quality explains a substantial part of the wage difference between labor markets of different sizes. This section explores the life-cycle dynamics of match quality depending on labor market size, showing new evidence on how occupational skill match quality evolves over the life cycle depending on labor market size.

Figure 2: Match quality over the life-cycle



(a) Match quality

(b) Match quality stayers

Notes: The figures plot match quality against age separately for individuals living in small, medium, and large labor markets. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Figure (a) includes all individuals and figure (b) only includes individuals who always stay in the local labor market they were born in.

Figure 2 plots match quality against age separately depending on labor market size, where a categorical definition of labor market size is used and labor markets are divided into small, medium and large.[8] Figure 2a plots this for all individuals, whereas figure 2b plots this for individuals who always stay in the local labor market in which they were born. At age 25, match quality is similar in local labor markets of different sizes. Match quality then increases over age in all labor markets. The increase in match quality is in line with the idea that individuals have imperfect

[8]Large labor markets are defined as labor markets with more than 500,000 inhabitants, medium-sized labor markets with between 100,000-500,000 inhabitants and small labor markets with less than 100,000 inhabitants.

information about their optimal match but learn about their match quality over time as they work (see Guvenen et al., 2020; Fredriksson et al., 2018). Moreover, it is in line with job-ladders models, where search frictions can lead to variation in jobs and job-ladders (see Burdett and Mortensen, 1998; Christensen et al., 2005; Manning, 2003), where individuals gradually are able to climb to better occupation matches.

Even if match quality increases with age across all labor markets, the magnitude of the increase is greater in large labor markets. Thus, the city-size match quality gap widens over the life cycle. The gradual increase in the city-size match quality gap is in line with the fact that workers in larger cities face a lower cost of switching jobs which might allow them to explore more different occupations and learn about what is a good match for them (see Wheeler, 2008).

However, the increase could also be in line with an occupation diversity explanation. If entry-level jobs are available in most cities but some high-skilled jobs are only available in larger cities, this could also explain the pattern observed in the data. At the beginning of the life-cycle match quality is similar between different labor markets, and when individuals in larger labor markets continue to better matches in occupations that only exist in the larger labor markets this would result in a larger increase in match quality in larger labor markets. Thus, individuals in large labor markets might be able to climb higher in the occupational job ladder.

Since Figure 2a shows match quality by the location of the individual at a specific age, the pattern could also be affected by a changed composition of individuals in large and small labor markets over the life cycle, where individuals move to local labor markets of different sizes over the life cycle. However, the pattern looks similar in Figure 2b, which only includes individuals who always remain in the local labor market they were born in, indicating that the life-cycle dynamics in match quality are not driven by changed composition of individuals in different labor markets.

Since the mean education level is higher in larger labor markets (see Table 1), individuals in larger labor markets might enter the labor market later, which could give different life-cycle patterns for individuals in large and small labor markets. To take this into account, Figure C.1 in Appendix C shows the results using potential experience instead of age.[9] The results look similar when using potential experience, indicating that the increase in match quality over the life cycle in large labor markets compared to small labor markets is not driven by the fact that individuals in large and small labor markets enter the labor market at different ages.

To examine the dynamics behind the life-cycle pattern, I provide evidence for how the likelihood of changing occupation as well as moving up in the occupation match-quality distribution is related to labor market size. This is done by running equation (1) with occupation change and higher occupation match quality as the dependent variable. Changing occupation is defined as working in another occupation than the individual was last observed in. Results are presented for all occupation changes, and separately for changes within and between firms. An occupation change within the firm is defined as a change of occupation where the firm remains the same and an occupation change

[9] Potential experience is defined as age minus (years of education+six), where six is the age the individual usually starts school.

between firms is defined as a change of both occupation and firm. Higher occupation match quality is constructed as a dummy variable equal to one if the individual changes to an occupation with higher match quality than before and zero if the individual changes to an occupation with lower match quality than before. When estimating the result for higher occupation match quality, the sample is limited to individuals who change occupation.

Table 7: Match quality dynamics

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Occupation switch			Switch to higher match quality		
	All	Within firm	Between firm	All	Within firm	Between firm
Large labor markets	0.018***	0.009***	0.009***	0.017***	0.002	0.014***
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Medium labor markets	0.002***	0.003***	-0.000	0.008***	0.009***	-0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Control skills	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,346,423	5,346,423	5,346,423	790,525	790,525	790,525

Notes: This table provides results estimating regressions with changing occupation and good switch as the dependent variables, and dummy variables for living in a large and medium-sized labor market compared to the omitted category living in a small labor market as the independent variable of interest. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Changing occupation is defined as working in another occupation than the individual did in the last observation. Switch to higher match quality is defined as a dummy variable equal to one for individuals who change occupation to an occupation with higher match quality and zero for individuals who change occupation to an occupation with worse match quality. When estimating the regression for the outcome switching to a higher match quality occupation, the sample is limited to individuals who switch occupations. Standard errors are clustered at the individual level.*** p<0.01, ** p<0.05, * p<0.1

Table 7 shows that individuals in larger labor markets are more likely to change occupations. The result of more occupation switches is in line with the theory of lower costs of changing occupation in larger labor markets (see Puga, 2010; Bleakley and Lin, 2012). The increased likelihood of changing occupation in large labor markets is equally driven by within and between-firm changes. Thus, the lower cost of changing occupations in larger labor markets could both come from more job openings, making it easier to change firms and more career opportunities within the firm and thus a lower cost of changing occupations without having to change employers. For individuals in medium-sized labor markets, the higher likelihood of changing occupation compared to small labor markets is entirely driven by within-firm changes.

Furthermore, Table 7 shows that individuals in larger labor markets are more likely to move up in the occupation match quality distribution when they change occupation. For individuals in large compared to small labor markets, the move up the occupation match-quality distribution is mainly driven by between-firm occupation changes. However, for individuals in medium-sized local labor markets, the pattern looks different. The increase in match quality for medium-sized compared to small local labor markets is entirely driven by within-firm changes.

Thus, the analysis shows that the improvement of match quality over the life cycle in larger compared to smaller labor markets comes from both more occupation switches and better occupation switches conditional on switching occupations.

7 Mechanisms

Section 6 showed that match quality is higher in larger labor markets. This section examines two different theoretical explanations for higher match quality in larger labor markets: more occupation diversity and more learning possibilities about match quality in larger labor markets.

7.1 Occupation diversity

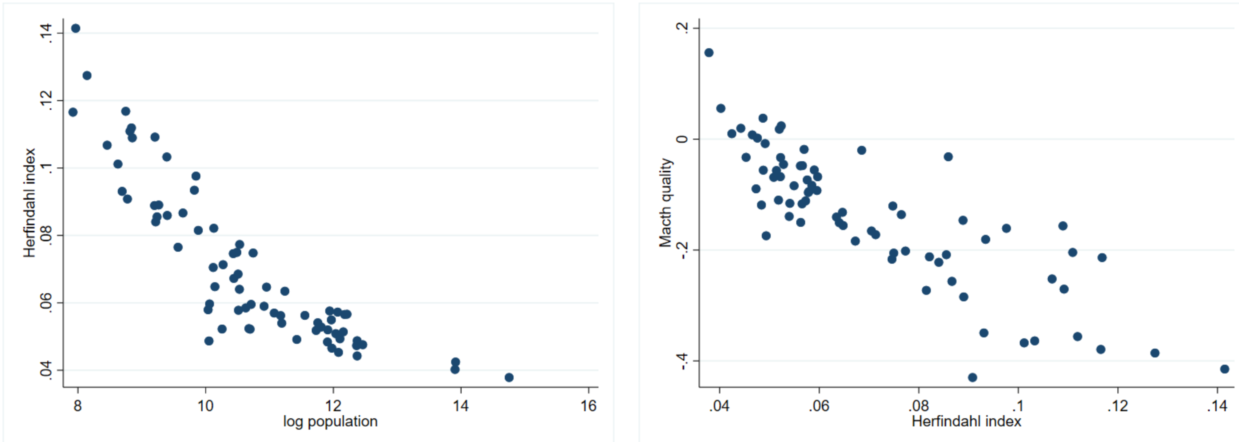
One theoretical mechanism behind the difference in match quality between large and small local labor markets is occupation diversity. If there are mobility costs, and the occupation where the individual would have the best match quality is not available in the labor market, the individual might not be able to reach the best occupation match. Thus, a more diversified occupation structure gives individuals more occupations to choose from and increases the likelihood that individuals find an occupation that matches their skill set.

To study if occupation diversity is a mechanism behind higher match quality in larger labor markets, I construct a Herfindahl-Hirschman Index (HHI). The HHI index is a measure widely used to measure employment concentration (see for example Thoresson, 2024; Benmelech et al., 2022), and have been used to show that larger labor markets have higher employment concentration (Halvarsson and Korpi, 2025). Here I construct a HHI measure of occupation concentration in the labor market, defined as: $HHI_L = \sum_{o=1}^{o}(s_o)^2$, where L denotes the labor market, s_o is the employment share in the labor market of occupation o. Thus, the HHI index is calculated by squaring the employment share of each occupation in the labor market and summing the numbers over all occupations. The HHI index can take values between 0 and 1, where 1 means there is only one occupation in the labor market and zero means less occupation concentration and, thus a more diversified occupation structure.

Furthermore, match quality might not only differ depending on labor market size due to larger labor markets having more different occupations. It might also be the case that larger labor markets have better occupations. To check this, I study whether the occupation where the individual would have the highest match quality exists in the labor market where the individual lives. The occupation is defined as existing if anybody in the occupation sample, in the relevant labor market, works in the occupation.

Figure 3a plots occupation diversity, defined with the HHI index, against labor market size. The figure shows a negative relationship between labor market size and the HHI index. Thus, larger labor markets seem to have less occupation concentration and more occupation diversity. The conclusions of higher occupation diversity in larger labor markets remain similar if occupation diversity is instead defined as the share of all occupations that exist in the local labor market fol-

Figure 3: Occupation diversity



(a) Labor market size

(b) Match quality

Notes: The HHI index is a measure of occupation concentration in the labor market, defined as: $HHI_L = \sum_{o=1}^{o}(s_o)^2$, where L denotes the labor market, s_o is the employment share in the labor market of occupation o. The HHI index can take values between 0 and 1, where 1 means it is only one occupation in the labor market and zero means less employment concentration and thus a more diversified occupation structure. Figure (a) plots the HHI index against labor market size and figure (b) plots match quality against the HHI index.

lowing Papageorgiou (2022) (see Table E.1 and E.2 in Appendix E). The finding of more occupation diversity in larger labor markets is in line with the result of more occupation diversity in larger labor markets in the United States (see Papageorgiou, 2022) and more industry diversity in larger labor markets in Sweden (see Korpi, 2007).

Figure 3b shows that there is a positive association between occupation diversity and match quality. Thus, the descriptive pattern supports the theory of occupation diversity as one mechanism for why match quality is higher in larger local labor markets.

To study how much occupation diversity contributes to the higher match quality in larger labor markets, I run equation (1) and include a control for the HHI index. Column 1 in Table 8 shows the main result for how match quality is related to labor market size, and column 2 shows the same result but controls for the HHI index. From the table, it is clear that the HHI index is negatively correlated to match quality (and thus, occupation diversity is positively correlated with match quality). The coefficient on labor market size becomes smaller when including controls for occupation diversity, indicating that occupation diversity might explain part of the city-size match quality gap. The coefficient on labor market size is reduced by approximately 23 percent when controls for occupation diversity are included. Column 3 shows that the occupation in which the individual would have the highest match quality is more likely to exist in larger compared to smaller labor markets. Thus, the results in this section suggest that one reason for higher match quality in larger labor markets is more occupation diversity in larger labor markets.

Table 8: Occupation diversity

VARIABLES	(1) Match quality	(2) Match quality	(3) Best match exist in labor market
log population size	0.070***	0.054***	0.003***
	(0.001)	(0.002)	(0.000)
Herfindahl-Hirschman Index		-2.370***	
		(0.209)	
Year FE	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Control skills	Yes	Yes	Yes
Observations	6,005,629	6,005,629	6,005,629

Notes: This table provides results estimating equation 1, estimating the difference in match quality for individuals who live in labor markets of different sizes. Column 2 includes controls for the HHI. The HHI is a measure of occupation concentration in the labor market, defined as: $HHI_L = \sum_{o=1}^{o}(s_o)^2$, where L denote the labor market, s_o is the employment share in the labor market of occupation o. The HHI can take values between 0 and 1, where 1 means it is only one occupation in the labor market and zero means less employment concentration and thus a more diversified occupation structure. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

7.2 Match quality and learning

Learning and match quality have been discussed as different mechanisms behind the city-size wage premium (Puga, 2010), and research has shown that more learning possibilities in larger cities are one important explanation for the city-size wage premium (De La Roca and Puga, 2016; D'Costa and Overman, 2014). However, learning and match quality are not necessarily disconnected explanations. Since individuals seem to be unsure about their match quality from the beginning, one possible explanation for higher match quality in larger labor markets is more learning opportunities about which occupations are a good match in larger labor markets. As noted by Wheeler (2008) a lower cost of switching jobs might allow the worker to explore more different jobs, making it easier to find the job where they are the most productive.

To estimate if learning is one channel for higher match quality in larger labor markets, I follow De La Roca and Puga (2016) and estimate the return to experience in labor markets of different sizes but with match quality instead of wages as the dependent variable:

$$MQ_{it} = \alpha_L + \lambda_i + \sum_{k=1}^{k=3} \mu_k \theta_{ikt} + \sum_{k=1}^{k=3} \rho_{kL} \theta_{ikt} * \alpha_L + \varepsilon_{it} \ (2)$$

In equation (2) α_L are separate dummy variables for living in a large or medium labor market, with the omitted category being living in a small labor market, λ_i is individual fixed effects, θ_{ikt} is years of experience working in a labor market of size k for worker i and time period t. Years of experience are calculated as years of working in a labor market of size L since age 20.[10] The

[10] To be able to calculate the experience of the worker, I use data on where the worker lives and if they have earnings from 1990. To be able to observe the worker from the age 20 the sample is limited to workers born 1970-1976. Since occupation data is available from 1996-2013, years from 1990 are used to calculate experience, but only years after 1996 are used to calculate the outcomes. One year of experience in labor market L is defined as a year of living in

value of experience is allowed to vary depending on both in which labor market the experience was obtained (k) and where the individual lives now (L). The specification also includes second-order polynomial terms of experience. Since experience in labor markets of different sizes now are included, α_L captures the static match quality premium of moving to labor market size L.

The idea behind the exercise is that if learning is one important mechanism behind higher match quality in larger labor markets, not only the static premium should matter. Instead, match quality should increase gradually and years of experience in large labor markets should be important. Moreover, if learning is the primary mechanism driving the difference in match quality, the knowledge of what is a good match should also be portable. Thus, if learning is important, we could expect years of experience in a large city to also be beneficial for individuals who now live in a small city.

The learning mechanism about match quality could either be driven by labor market explanations or more education possibilities in larger cities. Thus, it is also possible that the learning possibilities in larger cities are connected to learning through a more direct education channel. By moving to a large city, the worker can get access to a university education, which could increase the knowledge of the individual and give access to more occupations. To examine if the learning channel about match quality seems to be driven by education possibilities or labor market factors, I estimate regression (2) separately for individuals with and without a university education. The idea is that for individuals without a university education, the remaining mechanisms are related to labor market factors. As stated above, labor market factors driving the effect could be initial uncertainty about match quality and a lower cost of switching occupations in larger labor markets, allowing workers to explore more occupations and learn more about what occupation is a good match. However, another possible labor market driver of the effect could be more learning possibilities on the job in larger cities (see Glaeser and Maré, 2001; De La Roca and Puga, 2016). According to this channel, experience in large cities gives valuable knowledge that might open up possibilities to enter other occupations.

Table 9, column 1 shows the results for all workers, column 2 shows the results for workers with high school or less, and column 3 shows the results for workers with a university education. The regression includes experience in large and medium-sized labor markets as well as overall experience. Thus, the term experience now captures having experience in a small labor market and experience in large and medium labor markets captures the extra benefit of having experience in a large and medium relative to a small labor market. When studying the results for all workers, it is clear that the static premium becomes insignificant. Experience in large and medium relative to a small labor market is positive and significant, indicating that learning is one important mechanism behind the higher match quality in large labor markets. The interaction term of having experience in a large labor market and living in a large labor market is negative, indicating that the experience from a large labor market is actually more valuable for individuals now living in a small labor market.

labor market L and having positive earnings. Table D.1 in Appendix D presents results when experience is instead defined as years of living in a labor market since age 20. The results remain similar whether years of experience are defined conditional on working or not.

Thus, the knowledge the individual has obtained about what occupations are a good match seems to be portable.

Table 9: Learning

	Dependent variable: Match quality		
	(1)	(2)	(3)
VARIABLES	All	High school or less	University
Large labor market	-0.001	0.049	-0.074**
	(0.026)	(0.037)	(0.035)
Medium labor market	-0.032	-0.001	-0.108***
	(0.023)	(0.032)	(0.033)
Experience large	0.071***	0.038***	0.060***
	(0.006)	(0.010)	(0.008)
(Experience *large*)^2	-0.002***	-0.001*	-0.002***
	(0.000)	(0.001)	(0.001)
Experience medium	0.038***	0.020***	0.029***
	(0.005)	(0.007)	(0.008)
(Experience *medium*)^2	-0.002***	-0.001**	-0.001**
	(0.000)	(0.000)	(0.001)
Experience	0.022***	0.010***	0.049***
	(0.002)	(0.002)	(0.004)
Experience^2	-0.000**	0.000	-0.001***
	(0.000)	(0.000)	(0.000)
Large labor market#experience large	-0.041***	-0.014	-0.048***
	(0.006)	(0.009)	(0.008)
Large labor market#(experience *large*)^2	0.002***	0.001	0.002***
	(0.000)	(0.001)	(0.001)
Medium labor market#experience large	-0.000	0.013	-0.011
	(0.006)	(0.010)	(0.008)
Medium labor market#(experience *large*)^2	-0.001	-0.001	0.000
	(0.000)	(0.001)	(0.001)
Large labor market#experience medium	-0.000	-0.001	-0.004
	(0.006)	(0.008)	(0.008)
Large labor market#(experience *medium*)^2	-0.000	0.000	-0.000
	(0.000)	(0.001)	(0.001)
Medium labor market#experience medium	-0.029***	-0.017**	-0.038***
	(0.005)	(0.007)	(0.007)
Medium labor market#(experience *medium*)^2	0.001***	0.001*	0.002***
	(0.000)	(0.000)	(0.001)
Large labor market#experience	0.017***	0.004	0.033***
	(0.004)	(0.005)	(0.006)
Large labor market#*experience*^{2}	-0.001***	-0.000**	-0.001***
	(0.000)	(0.000)	(0.000)
Medium labor market#experience	0.006*	0.003	0.026***
	(0.003)	(0.005)	(0.005)
Medium labor market#*expeirence*^{2}	0.000	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)
Individual FE	Yes	Yes	Yes
Observations	1,351,753	679,803	671,950

Notes: This table provides the results of estimating equation 2. The sample includes males born between 1970 and 1976, with outcomes for the years 1996-2013. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1

When studying the result separately for individuals with high school education and individuals with university education, it is clear that for both groups experience in large and medium is more valuable for match quality than experience in small labor markets. The point estimates are larger for those with a university education, which could indicate that part of the learning effect found in column 1 can be learning through university education. However, experience in large labor markets is also more valuable than experience in small labor markets for individuals with a high school education, indicating that parts of the learning effect also come through the labor market. Thus, the result suggests that one explanation for higher match quality in larger labor markets is less frictions and more learning possibilities of what occupations to work in to have high match quality in larger labor markets.

8 Robustness

Earlier sections have shown that match quality is higher in larger labor markets. This section examines the robustness of the results. First, results are presented for individuals who move between labor markets of different sizes, making it possible to control for individual fixed effects. It should be noted that even if the movers' design allows for control for individual characteristics that are constant over time, the decision to move is a choice, and individuals might move because they have gotten a job with higher occupational match quality. Thus, the result from this group might differ from the general population. Nevertheless, studying movers is a good complement to the main specification. Finally, the section also presents alternative specifications of the match quality measure.

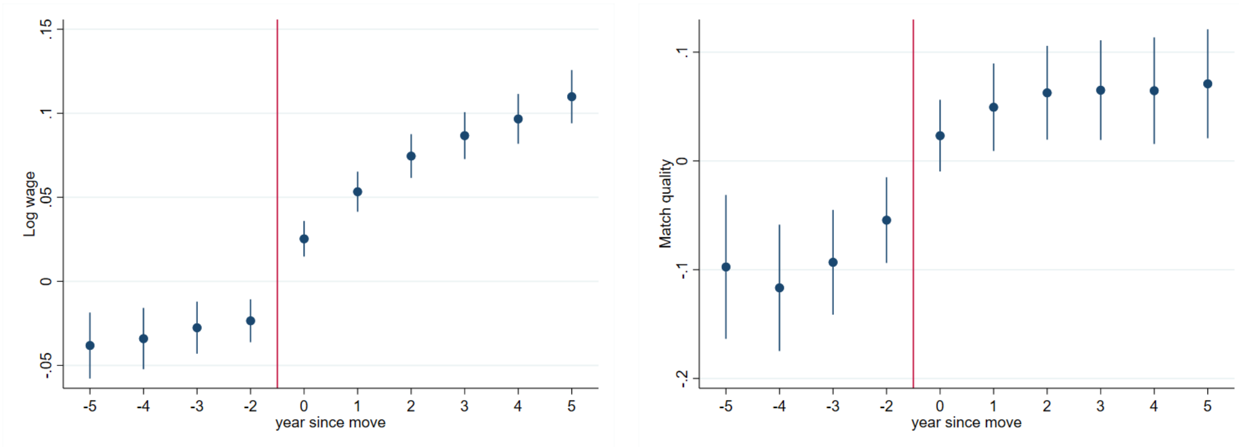
8.1 Comparing movers and stayers

This section presents results comparing individuals who move from the small local labor market they were born in to a large labor market with individuals who always stay in the small labor, by estimating:

$$y_{it} = \alpha_i + \lambda_t + \sum_l \mu_l D^l_{it} + X_{it} + \rho_{it} \quad (3)$$

where α_i is individual fixed effects, λ_t is year-fixed effects, and X_{it} is age dummies, and D^l_{it} is time since the move. Including both year and age-fixed effects together with the individual fixed effect would give perfect multicollinearity, so to overcome this problem the age dummies are normalized to be constant between 45 and 54. To avoid the problem with two-way fixed effects models and heterogeneous treatment effects, I use the procedure from Sun and Abraham (2021), and use never treated as the controls, and estimate a weighted average of the cohort-time-specific treatment effects, where weights are set to the estimated cohort shares. The result is estimated for the first move the individual makes. First moves are defined as the first move in the data, observed between the years 1996 and 2013 and that the move is from the local labor in which the individual was born. The control group is individuals who always stay in the small local labor market they were born in.

Figure 4: Move large from small



(a) Wage

(b) Match quality

Notes: The figures plot event graphs for individuals who move from a small labor market to a large labor market estimated with Sun and Abraham (2021) estimator, using never movers who always stay in the small local labor market they where born in as the control. The x-axis shows normalized time since the move, where period zero is the time of the move. Figure (a) plots the result for log wages and figure (b) plots the result for match quality. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013.

Figure 4 shows the results for moving to a large labor market where year zero is the year of the move. Figure 4a shows the results for wages, and Figure 4b shows the results for match quality. From Figure 4 it is clear that both wages and match quality on average increase after an individual moves to a large local labor market. However, the figure also shows a pre-trend before moving, indicating that individuals who are on a positive trend in match quality are more likely to move to a large local labor market. Thus, it seems like match quality increases for individuals who choose to move from a small to a large labor market. However, if individuals dislike moving, they may only choose to move if match quality increases. Thus, if individuals only move to increase match quality the higher match quality might come from moving itself, and not from the larger labor market. To take this into account, Section 8.2 instead compares movers to labor markets of different sizes.

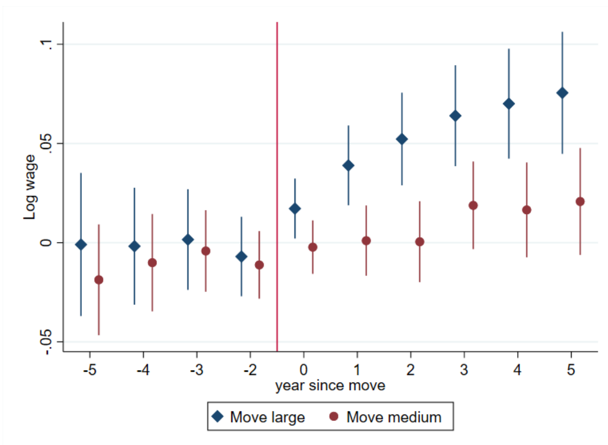
8.2 Comparing movers

Moving could in itself have a direct effect on match quality, and individuals may only choose to move if they currently have a bad match. Odio Zúñiga and Yuen (2020) have for example shown that individuals who have bad match quality are more likely to move. To take away the direct effect of the move, this section therefore estimates results using only movers and comparing individuals who move to local labor markets of different sizes. This is done by estimating the regression:

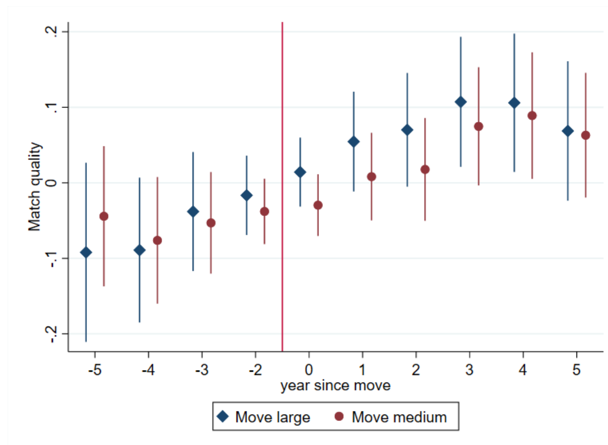
$$y_{it} = \sum_{l=j}^{l=k} \beta_l N_l + \beta_T treated_i * \sum_{l=j}^{l=k} N_l + treated_i + \lambda_t + X_{it} + e_{it} \quad (3)$$

where N is normalized time since the move (j years before the move to k years after the move) and treated are separate dummy variables for moving to a large or medium labor market with the omitted category being moving to a small local labor market. X_{it} includes age controls, and second-order polynomials in each of the eight skills, controls for the local labor market the individual is born in, and a fixed effect for the occupation the individual works in the year before the move. The parameter of interest β_T gives the difference between moving to a large and medium labor market compared to moving to a small local labor market.

Figure 5: Movers



(a) Wage



(b) Match quality

Notes: The figure plots the difference between moving to a large or medium-sized labor market compared to the omitted category, moving to a small labor market. The x-axis shows normalized time since the move, where period zero is the time of the move. Figure (a) plots the result for log wages and figure (b) plots the result for match quality. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013.

Figure 5 shows what happens to wages and match quality for individuals who move to a medium-sized and large labor market compared to the omitted category moving to a small labor market. Figure 5a shows that after the move, wages increase over time for individuals who move to a large compared to a small labor market. The estimates also suggest that over time, wages increase somewhat for individuals who move to a medium compared to a small labor market, even though the estimates are not statistically different from zero. Figure 5b shows a similar pattern for match quality, where match quality increases for individuals who move to large compared to small labor markets. The point estimates suggest that match quality increases somewhat in the first two years after the move and then is relatively constant over time. Overall, the results from the movers support the earlier result of higher match quality in large compared to small labor markets.

8.3 Robustness of the match quality measure

This section provides robustness tests of the match quality measure by presenting results with alternative specifications of the match quality measure.

One potential worry with the match quality measure is that the return in small occupations might be estimated with more uncertainty. Figure A.3 in Appendix A plots match quality against occupation size, and the figure indicates that there might be a relationship between match quality and occupation size. Thus, to study the robustness of the results Table 10 presents results following Guvenen et al. (2020), using the external O*Net data to proxy for skill requirements in each occupation.[11] Reassuringly, the conclusion of higher match quality in larger labor markets remains when using O*Net data to measure skill requirements in each occupation.

Column 2 estimates the return to skills linearly instead of using a random forest, again the conclusions remain with this specification.

One other potential worry with the match quality measure is that individuals might be well-matched even if they are not correctly priced for their skills. To take this into account column 3 uses the random forest to predict the probability that a worker works in an occupation instead of estimating return to skills; again, the conclusions remain similar.[12] Thus, from Table 10 it is clear that the conclusion of higher match quality in larger compared to smaller labor markets also holds with the alternative definitions of match quality.

One other potential concern with the construction of the match quality measure is that since more individuals live in large labor markets, the training data has more individuals from large labor markets. If returns to skills in specific occupations are different in large and small labor markets, the estimated returns might be more similar to the returns in large labor markets. To study if this drives the results, Column 4 in Table 10 presents results when the returns to skills in occupations are allowed to vary with labor market size by estimating separate random forest regressions for small, medium, and large labor markets. Also, with this specification, the conclusions of higher match quality in larger labor markets remain.

Another potential concern with the match quality measure is that some occupations might be over-represented in large or small labor markets and that the return might be higher in large cities for other urban wage reasons, such as a lower cost of living in smaller labor markets. To take this into account column 5 presents the result when only training the random forest in large labor markets. Reassuringly, the results remain very similar, indicating that the result of higher match quality in larger labor markets is not driven by different returns in labor markets of different sizes.

In the main specification, the match quality measure takes away the average return to the skills to take away the fact that some individuals will have higher returns everywhere. However, there exist other measures to compare with than the average return on the whole labor market. Column 6 instead presents an alternative that compares the returns to skills in the occupation the individuals work in, compared to the best occupational match for the individual.[13] The conclusion

[11] To construct the match quality measure, I use five different occupation skill requirements from O*Net chosen to match the skills for the individuals from the military enlistment test: inductive skills, verbal skills, spatial skills, technical skills, and social skills. The mismatch measure is constructed as the sum of the absolute difference between the rank of the standardized skill from the military enlistment test and the rank of the occupation requirement for that skill for each of the five skills. Match quality is then defined as the negative of the mismatch measure.

[12] The match quality measure is then constructed by estimating separate classification random forest for each occupation, allowing different interactions of skills to be differently important in different occupations, predicting the likelihood that an individual works in an occupation given their skill. Thus, with this match quality specification, an individual is well matched if they work in an occupation with a high probabilityty that they should work in given their skills.

[13] The match quality measure is then constructed as $Mismatch_{io} = |Return_{io} - Return_{i,best}|$, where $Return_{io}$ is

about higher occupational match quality in larger labor markets remains similar when comparing the returns to the optimal match, as compared to the average return on the market. Thus, from Table 10 it is clear that the conclusion of higher match quality in larger compared to smaller labor markets also holds with the alternative definitions of match quality.

Table 10: Alternative match quality measures

VARIABLES	(1) O*Net	(2) Linear return	(3) Random forest probability	(4) Return vary with labor market	(5) Random forest trained on large labor market	(6) Compared to optimal match
log population size	0.005***	0.015***	0.019***	0.168***	0.075***	0.057***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Control skills	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,820,642	6,005,629	6,005,629	6,005,629	6,005,629	6,005,629

Notes: This table provides results estimating equation 1, estimating the difference in match quality for individuals who live in labor markets of different sizes, using alternative measures of match quality. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1

9 Conclusion

While earlier research found that wages are higher in larger than smaller labor markets, the mechanism behind this has been less studied. One theoretical mechanism for the city-size wage premium is that workers in larger cities are better matched to their occupations. However, even though this mechanism has been discussed theoretically, the empirical evidence for how match quality differs in small and large labor markets has been limited. Building on the recent literature on multidimensional skill mismatch, this paper contributes to the literature by providing direct evidence on how occupational skill match quality differs depending on the labor market size in Sweden.

The key result from the paper is that match quality indeed is higher in larger compared to smaller local labor markets. This finding holds conditional on skills, when including individual fixed effects for individuals who move from a small to a large labor market, and when using movers comparing individuals who move to a large labor market compared to a small labor market. In the preferred specification, match quality explains around 30 percent of the city-size wage gap. The difference in match quality between individuals in large and small labor markets is especially large for individuals with high skills, measured as cognitive skills, non-cognitive ability, or education,

the return in the occupation the individuals works in and $Return_{i,best}$ is the occupation where the individual, given the skills would have the highest return. The match quality measure is then constructed as the negative of this mismatch measure and is similar to the main match quality measure standardized to have mean zero and standard deviation 1.

in line with the fact that the city-size wage premium is higher for highly skilled individuals (see Bacolod et al., 2009). The large differences in match quality are consistent with high mobility frictions, in line with research showing that people are unlikely to move in response to economic shocks (see Yagan, 2019).

I examine two different mechanisms behind the higher match quality in larger labor markets: a more diversified occupation structure in larger labor markets, increasing the likelihood that individuals find an occupation that matches their skills and more learning possibilities about match quality in larger cities. When exploring the empirical support for these mechanisms, suggestive evidence emerges that higher match quality in larger labor markets comes from both a more diversified occupation structure and higher learning possibilities.

It should be noted that the finding that match quality is higher in larger labor markets does not necessarily mean that individuals would be better off by moving to large labor markets since the utility of the individual is also affected by the cost of living and their preferences about where to live. Thus, while the result here highlights the limited possibility to find a good match in smaller labor markets, research is needed on how to best solve this problem. Possible solutions could be increasing mobility or increasing the size of small labor markets. Policies for increasing the size of the local labor market could be lowering the commuting cost by investing in infrastructure or increasing the possibility of working from home. Moreover, since this study has shown that match quality is an important factor behind the city-size wage premium, policies directly focused on increasing match quality in small labor markets could be useful in decreasing the city-size wage gap. Such policies could be increasing occupation diversity in small cities, for example by moving government agencies to smaller cities or increasing incentives for entrepreneurship in different occupations. Another way to increase match quality in smaller labor markets would be to decrease the cost of exploring different occupations in small labor markets by lowering search costs. All of these possible solutions probably have different pros and cons, and more research is needed to decide on the best solutions.

It should also be noted that the results only include males due to the fact that military enlistment was only mandatory for males, and thus, the skill data is only available for this sample. While the mechanism for the city-size match quality gap is likely to hold for broader populations, the magnitude of the effect might differ for other populations. Comparing the city-size wage gap for men in this paper with a similar sample with women for example shows a smaller city-size wage gap for women.[14] Research has shown that couples are more likely to move for men's career (see for example Jayachandran et al., 2024), where it has been argued that restrictions on women's geographical mobility will give smaller wage penalties in large labor markets (Nisic, 2017). In terms of match quality, this implies that if couples move to a small labor market because the man has a job there, the match quality penalty for females in small labor markets might be larger. Research has also shown that the gender segregation between different occupations differs depending on labor

[14] Results are available in table 2 for males and table F.1 for woman. Note that city-size wage differences conditional on skills is not available for woman, since skill data is not available for this population.

market size (Elass et al., 2024), which might affect the match quality gap. Thus, while the match quality mechanism studied are likely to be valid for a broader population, the magnitude of the effect might differ.

References

Abel, J. R. and Deitz, R. (2015). Agglomeration and job matching among college graduates. *Regional Science and Urban Economics*, 51:14–24.

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

Almgren, M., Kramer, J. V., and Sigurdsson, J. (2022). It runs in the family: Occupational choice and the allocation of talent. In *Essays on Macroeconomics, Monetary Policy and Mobility*, number 117 in Monograph series / Institute for International Economic Studies, University of Stockholm, page 276. Department of Economics, Stockholm University.

Andersson, F., Burgess, S., and Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61(1):112–128.

Andersson, M., Klaesson, J., and Larsson, J. P. (2014). The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? *Papers in Regional Science*, 93(4):727–747.

Andini, M., de Blasio, G., Duranton, G., and Strange, W. C. (2013). Marshallian labour market pooling: Evidence from italy. *Regional Science and Urban Economics*, 43(6):1008–1022.

Autor, D. H. (2019). Work of the past, work of the future. *AEA Papers and Proceedings*, 109:1–32.

Bacolod, M., Blum, B. S., and Strange, W. C. (2009). Urban interactions: soft skills versus specialization. *Journal of Economic Geography*, 9(2):227–262.

Bacolod, M., De la Roca, J., and Ferreyra, M. M. (2021). In search of better opportunities: Sorting and agglomeration effects among young college graduates in colombia. *Regional Science and Urban Economics*, 87:103656.

Benmelech, E., Bergman, N. K., and Kim, H. (2022). Strong employers and weak employees. *Journal of Human Resources*, 57(S):S200–S250.

Berlingieri, F. (2018). Local labor market size and qualification mismatch. *Journal of Economic Geography*, 19(6):1261–1286.

Bleakley, H. and Lin, J. (2012). Thick-market effects and churning in the labor market: Evidence from us cities. *Journal of Urban Economics*, 72(2):87–103.

Boualam, B. (2014). Getting a first job: Quality of the labor matching in French cities.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2):257–273.

Card, D., Rothstein, J., and Yi, M. (2025). Location, location, location. *American Economic Journal: Applied Economics*, 17(1):297–336.

Carlsen, F., Rattsø, J., and Stokke, H. E. (2016). Education, experience, and urban wage premium. *Regional Science and Urban Economics*, 60:39–49.

Christensen, B., Lentz, R., Mortensen, D., Neumann, G., and Werwatz, A. (2005). On-the-job search and the wage distribution. *Journal of Labor Economics*, 23(1):31–58.

Corradini, C., Morris, D., and Vanino, E. (2025). Marshallian agglomeration, labour pooling and skills matching. *Cambridge Journal of Economics*, page beaf010.

Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2022). Matching in Cities. *Journal of the European Economic Association*, 20(4):1478–1521.

D'Costa, S. and Overman, H. G. (2014). The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, 48:168–179.

De La Roca, J. and Puga, D. (2016). Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1):106–142.

Duranton, G. and Jayet, H. (2011). Is the division of labour limited by the extent of the market? evidence from french cities. *Journal of Urban Economics*, 69(1):56–71.

Eckert, F., Hejlesen, M., and Walsh, C. (2022). The return to big-city experience: Evidence from refugees in Denmark. *Journal of Urban Economics*, 130:103454.

Elass, K., García-Peñalosa, C., and Schluter, C. (2024). Gender Gaps in the Urban Wage Premium. Technical report.

Eliasson, K. and Westerlund, O. (2022). The urban wage premium and spatial sorting on observed and unobserved ability. *Journal of Economic Geography*, 23(3):601–627.

Figueiredo, O., Guimarães, P., and Woodward, D. (2013). Firm–worker matching in industrial clusters. *Journal of Economic Geography*, 14(1):1–19.

Fredriksson, P., Hensvik, L., and Skans, O. N. (2018). Mismatch of talent: Evidence on match quality, entry wages, and job mobility. *American Economic Review*, 108(11):3303–38.

Frenette, M. (2006). Too far to go on? distance to school and university participation. *Education Economics*, 14(1):31–58.

Glaeser, E. L. and Maré, D. C. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–342.

Guvenen, F., Kuruscu, B., Tanaka, S., and Wiczer, D. (2020). Multidimensional skill mismatch. *American Economic Journal: Macroeconomics*, 12(1):210–44.

Halvarsson, D. and Korpi, M. (2025). City size, employer concentration, and wage income inequality. Working Paper No. 2025:4, Institute for Evaluation of Labour Market and Education Policy (IFAU).

Harmon, N. (2013). Are workers better matched in large labor markets? Technical report.

Jayachandran, S., Sundberg, E., Nassal, L., Notowidigdo, M., Paul, M., and Sarsons, H. (2024). Moving to opportunity, together. NBER Working Paper 32970, National Bureau of Economic Research.

Korpi, M. (2007). Does size of local labour markets affect wage inequality? a rank-size rule of income distribution. *Journal of Economic Geography*, 8(2):211–237.

Korpi, M. and Clark, W. A. (2019). Migration and occupational careers: The static and dynamic urban wage premium by education and city size. *Papers in Regional Science*, 98(1):555–574.

Koster, H. R. and Ozgen, C. (2021). Cities and tasks. *Journal of Urban Economics*, 126:103386.

Leknes, S., Rattsø, J., and Stokke, H. E. (2022). Assortative labor matching, city size, and the education level of workers. *Regional Science and Urban Economics*, 96:103806.

Lindqvist, E. and Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, 3(1):101–28.

Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

Mion, G. and Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue canadienne d'économique*, 42(1):28–55.

Mood, C., Jonsson, J. O., and Bihagen., E. (2012). Socioeconomic gradients in children's outcomes. In John Ermisch, M. J. and Smeeding, T. M., editors, *From Parents to Children: The Intergenerational Transmission of Advantage*, pages 32–52. Russell Sage Foundation.

Moretti, E. (2011). Chapter 14 - local labor markets. volume 4 of *Handbook of Labor Economics*, pages 1237–1313. Elsevier.

Moretti, E. and Yi, M. (2024). Size matters: Matching externalities and the advantages of large labor markets. Working Paper 32250, National Bureau of Economic Research.

Neves, E., Azzoni, C., and Chagas, A. (2017). Skill wage premium and city size. Working papers, department of economics, University of São Paulo (FEA-USP).

Nisic, N. (2017). Smaller differences in bigger cities? assessing the regional dimension of the gender wage gap. *European Sociological Review*, 33(2):292 – 304. Cited by: 27.

Odio Zúñiga, M. and Yuen, C. Y. K. (2020). Moving for Better Skill Match. *SSRN*.

Papageorgiou, T. (2022). Occupational matching and cities. *American Economic Journal: Macroeconomics*, 14(3):82–132.

Puga, D. (2010). The magnitude and causes of agglomeration economies*. *Journal of Regional Science*, 50(1):203–219.

Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199. Themed Issue: Treatment Effect 1.

Thoresson, A. (2024). Employer concentration and wages for specialized workers. *American Economic Journal: Applied Economics*, 16(1):447–79.

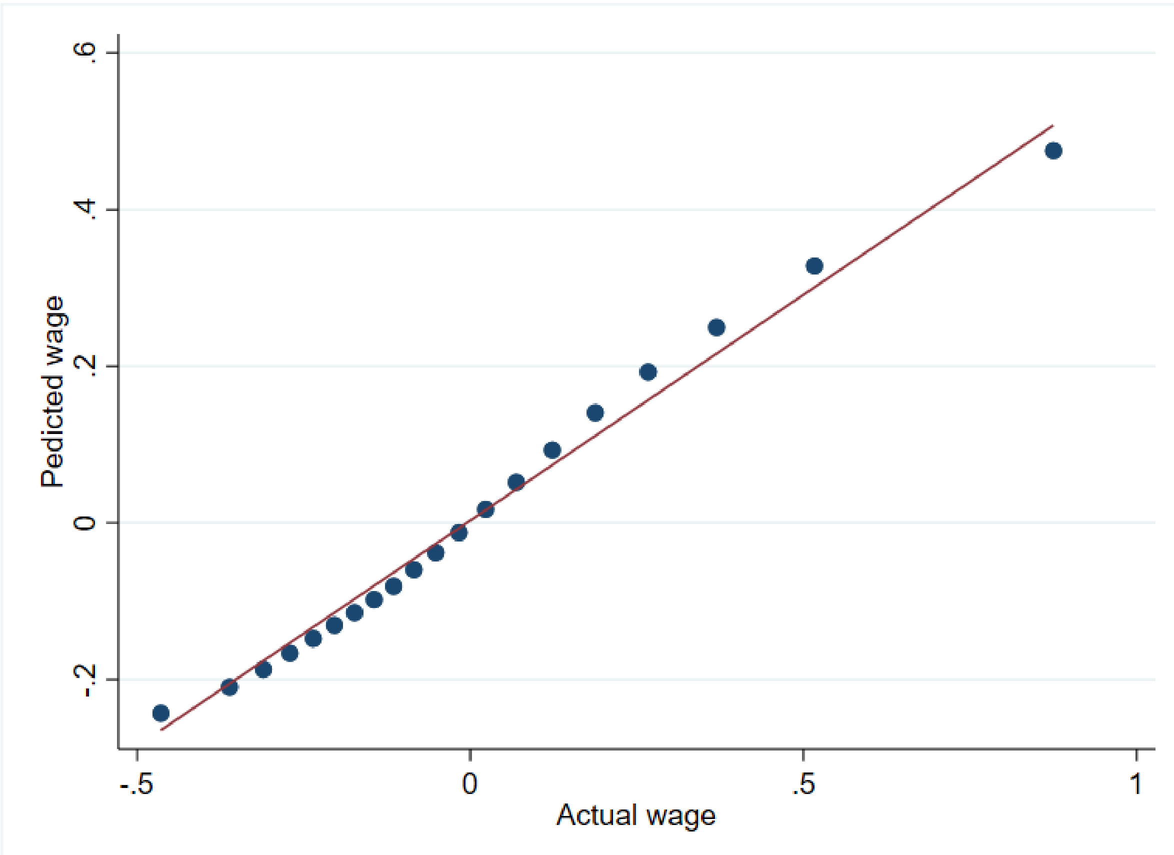
Wheeler, C. H. (2008). Local market scale and the pattern of job changes among young men. *Regional Science and Urban Economics*, 38(2):101–118.

Yagan, D. (2019). Employment hysteresis from the great recession. *Journal of Political Economy*, 127(5):2505–2558.

Appendix

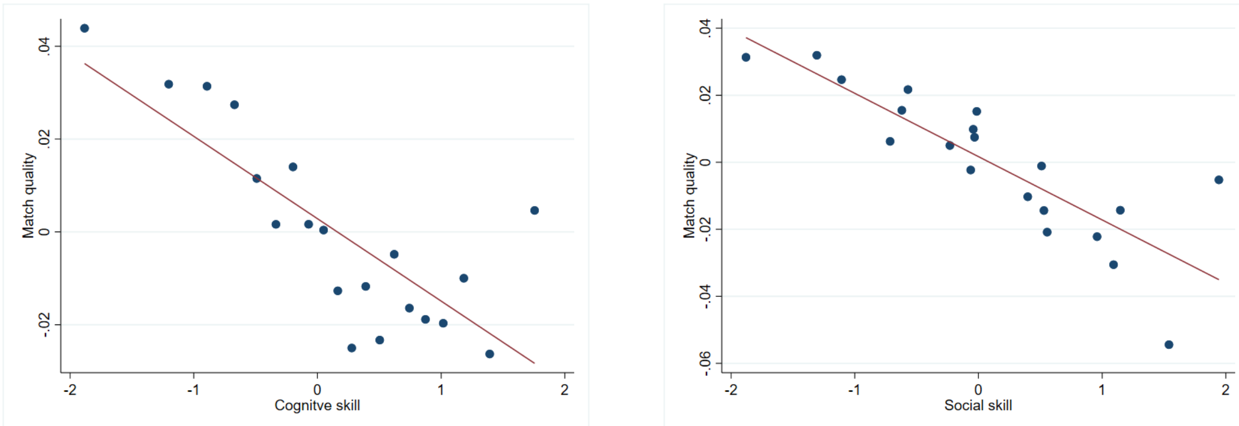
A Match quality measure

Figure A.1: Predicted wage with random forest vs actual wage



Notes: The figure shows predicted wages based on skills estimated with the random forest against actual wages. Wages are residualized on age and year-fixed effects. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. The individuals used for training the random forest are excluded from the sample.

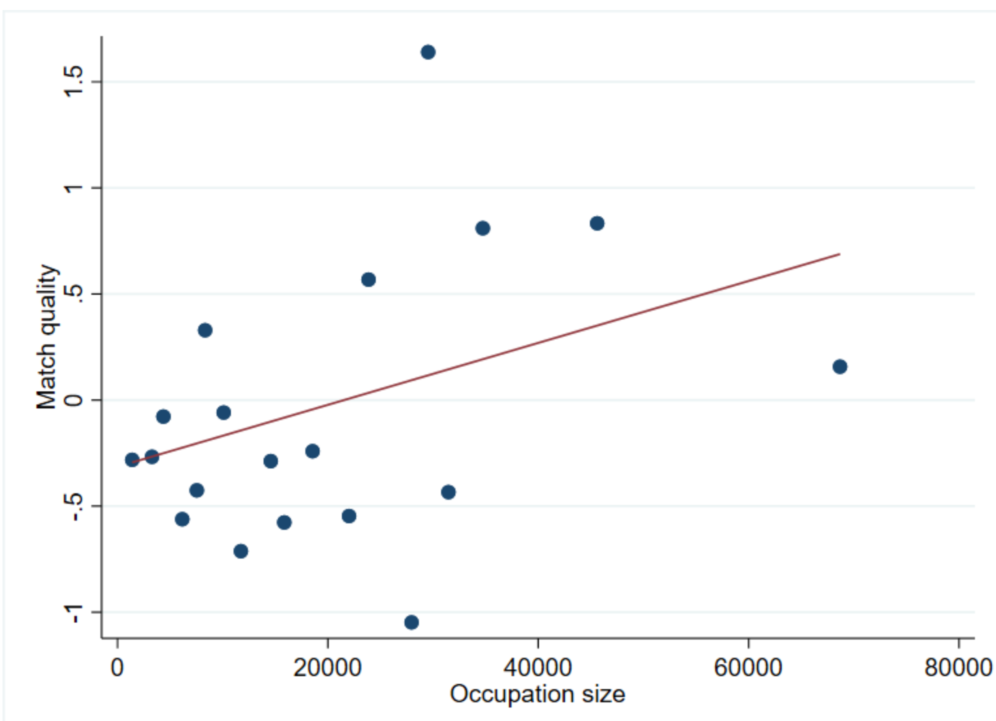
Figure A.2: Match quality and skills



(a) Cognitive skills (b) Social skills

Notes: The figure plot match quality against skills. Figure A shows the result for cognitive skills and Figure B shows the result for social skills. The match quality measure is standardized to have mean zero and standard deviation one. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013.

Figure A.3: Match quality and occupation size



Notes: The figure plot match quality against occupation size. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013.

Table A.1: Different match quality measures

	Dependent variable: Ln Wage						
VARIABLES	(1)	(2) O*Net	(3) Linear return	(4) Random forest probability	(5) Return vary with labor market	(6) Random forest trained on large labor market	Compared to optimal match
Log population size	0.039***	0.039***	0.036***	0.039***	0.010***	0.027***	0.030***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000
Match quality		0.010***	0.207***	0.028***	0.176***	0.167***	0.158***
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	6,005,629	5,820,642	6,005,629	6,005,629	6,005,629	6,005,629	6,005,62
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control skills	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table provides results estimating equation 1, with log wage as the dependent variable adding match quality as a control comparing the coefficient for local labor market size with and without controls for match quality. Column 1 shows the result without controls for match quality, and column 2-6 adds different match quality measures as controls. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Standard errors are adjusted for clusters at the individual level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

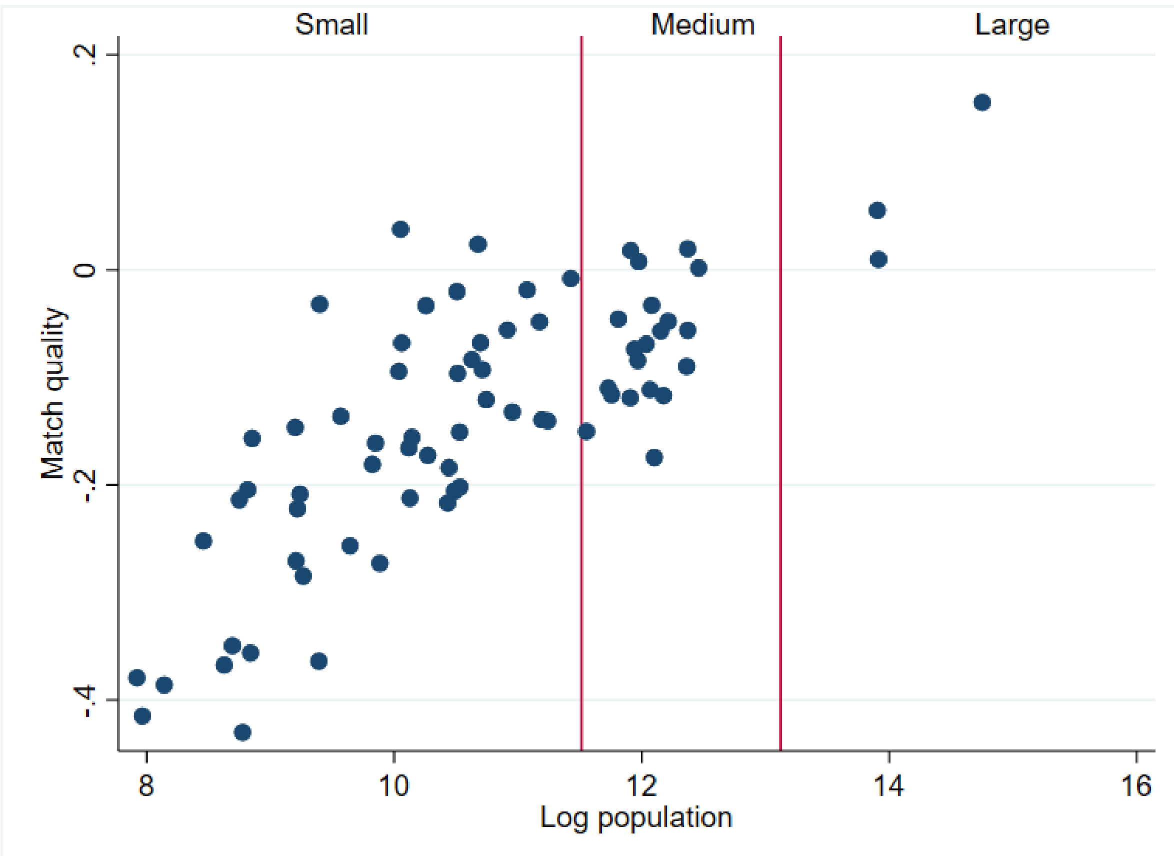
B Labor market definitions

Table B.1: Different labor market size definitions: Match quality

	(1)	(2)	(3)
VARIABLES	Match quality	Match quality	Match quality
Large city		0.285***	
		(0.003)	
Mid-sized city		0.090***	
		(0.003)	
LN, Populaton density	0.054***		
	(0.001)		
Large regions			0.226***
			(0.004)
Medium-sized regions			0.056***
			(0.004)
Observations	6,005,629	6,005,629	6,005,629
R-squared	0.021	0.023	0.019
Year FE	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Skill controls	Yes	Yes	Yes

Notes: This table provides results of the estimating equation 1, estimating the difference in match quality for individuals who live in labor markets of different sizes. In column one, labor market sizes are defined according Swedish Association of Local Authorities and Regions definition of municipality types, where municipalities have been divided into the three groups: large cities and municipalities near large cities, medium-sized towns and municipalities near medium-sized towns and smaller towns/urban areas and rural municipalities with smaller towns/urban areas and rural municipalities being the omitted category. In column 2, labor market size is defined as Ln of labor market density in the local labor market. Column 3 uses a categorical definition of labor market size, where larger labor markets are defined as labor markets with more than 500,000 inhabitants, medium-sized labor markets with between 100,000-500,000 inhabitants and small labor markets with less than 100,000 inhabitants, with small labor markets being the omitted category. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Standard errors are adjusted for 659,206 clusters at the individual level. *** p<0.01, ** p<0.05, * p<0.1

Figure B.1: Match quality and labor market size



Notes: The figure shows match quality against population size in the labor market. The vertical line shows the limit for small, medium, and large labor markets. The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013

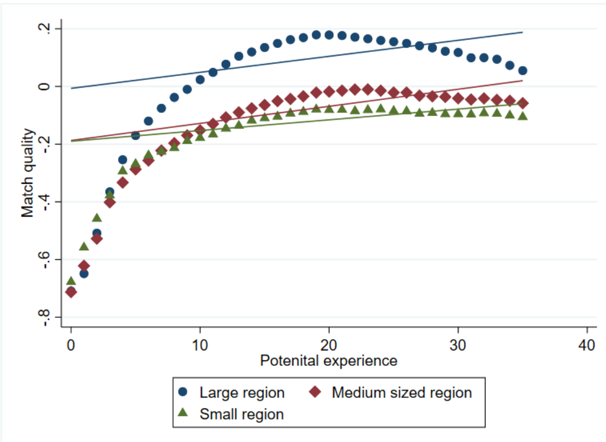
Table B.2: Different labor market size definitions: Wage

VARIABLES	(1) Ln wage	(2) Ln wage	(3) Ln wage
Large city		0.146***	
		(0.001)	
Mid-sized city		0.029***	
		(0.001)	
LN, Populaton density	0.031***		
	(0.000)		
Large regions			0.123***
			(0.001)
Medium-sized regions			0.019***
			(0.001)
Observations	6,005,629	6,005,629	6,005,629
R-squared	0.367	0.369	0.363
Year FE	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Skill controls	Yes	Yes	Yes

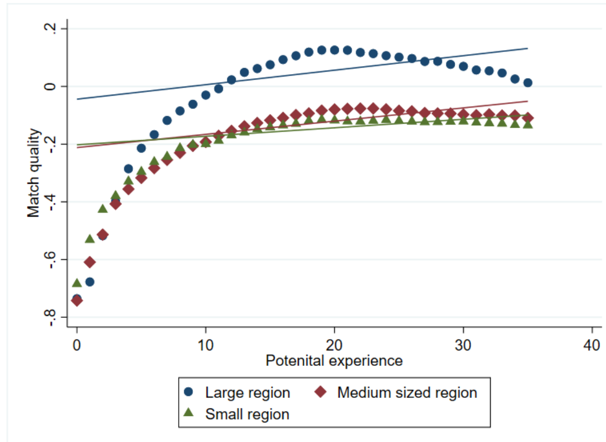
Notes: This table provides results of the estimating equation 1, with log wage as the outcome, estimating the difference in wages for individuals who live in labor markets of different sizes. In column one, labor market sizes are defined according Swedish Association of Local Authorities and Regions definition of municipality types, where municipalities have been divided into the three groups: large cities and municipalities near large cities, medium-sized towns and municipalities near medium-sized towns and smaller towns/urban areas and rural municipalities with smaller towns/urban areas and rural municipalities being the omitted category. In column 2, labor market size is defined as Ln of labor market density in the local labor market. Column 3 uses a categorical definition of labor market size, where larger labor markets are defined as labor markets with more than 500,000 inhabitants, medium-sized labor markets with between 100,000-500,000 inhabitants and small labor markets with less than 100,000 inhabitants, with small labor markets being the omitted category.The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Standard errors are adjusted for 659,206 clusters at the individual level. *** p<0.01, ** p<0.05, * p<0.1

C Life cycle pattern

Figure C.1: Match quality over the life-cycle



(a) All individuals



(b) Stayers

Notes: The figures plot match quality against potential experience separately for individuals living in small, medium, and large labor markets. Potential experience is constructed as age-(years of education + 6). The sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Figure (a) includes all individuals and figure (b) only includes individuals who always stay in the local labor market they were born in.

D Learning

Table D.1 shows estimates of regression 2, where experience is defined as years of living in a labor market since age 20.

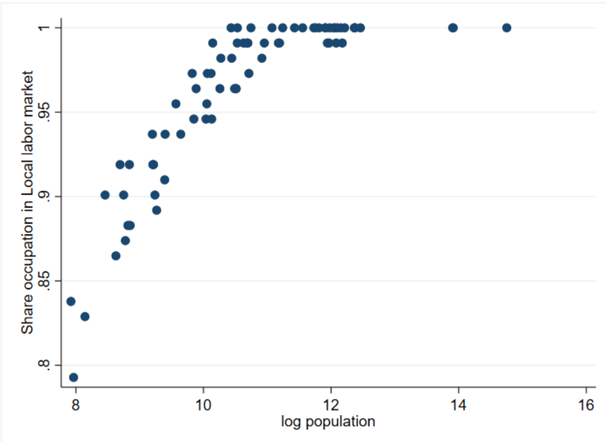
Table D.1: Learning

	Dependent variable: Match quality		
	(1)	(2)	(3)
VARIABLES	All	High school or less	University
Large regions	-0.010	0.056	-0.107***
	(0.028)	(0.040)	(0.038)
Medium-sized regions	-0.026	0.010	-0.113***
	(0.025)	(0.034)	(0.036)
Experience large	0.068***	0.039***	0.053***
	(0.006)	(0.009)	(0.008)
(Experience *large*)^2	-0.002***	-0.001*	-0.001**
	(0.000)	(0.001)	(0.001)
Experience medium	0.035***	0.022***	0.022***
	(0.005)	(0.007)	(0.007)
(Experience *medium*)^2	-0.001***	-0.001**	-0.001*
	(0.000)	(0.000)	(0.000
Experience	0.028***	0.015***	0.064***
	(0.002)	(0.002)	(0.004)
Experience^2	-0.000***	-0.000	-0.001***
	(0.000)	(0.000)	(0.000)
Large region#experience large	-0.037***	-0.013	-0.042***
	(0.006)	(0.009)	(0.008)
Large region#(experience *large*)^2	0.002***	0.001	0.002***
	(0.000)	(0.001)	(0.001)
Medium region#experience large	0.003	0.015	-0.006
	(0.006)	(0.009)	(0.008)
Medium region#(experience *large*)^2	-0.001	-0.001*	-0.000
	(0.000)	(0.001)	(0.001)
Large region#experience medium	0.003	-0.002	0.002
	(0.005)	(0.008)	(0.007)
Large region#(experience *medium*)^2	-0.000	0.000	-0.001
	(0.000)	(0.001)	(0.001
Medium region#experience medium	-0.025***	-0.019***	-0.029**
	(0.005)	(0.007)	(0.007)
Medium region#(experience *medium*)^2	0.001***	0.001**	0.001**
	(0.000)	(0.000)	(0.000)
Large region#experience	0.017***	0.004	0.034***
	(0.004)	(0.005)	(0.006)
Large region#*experience*^{2}	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)
Medium region#experience	0.005	0.003	0.022***
	(0.003)	(0.004)	(0.005
Medium region#*expeirence*^{2}	0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)
Individual FE	Yes	Yes	Yes
Observations	1,351,753	679,803	671,950

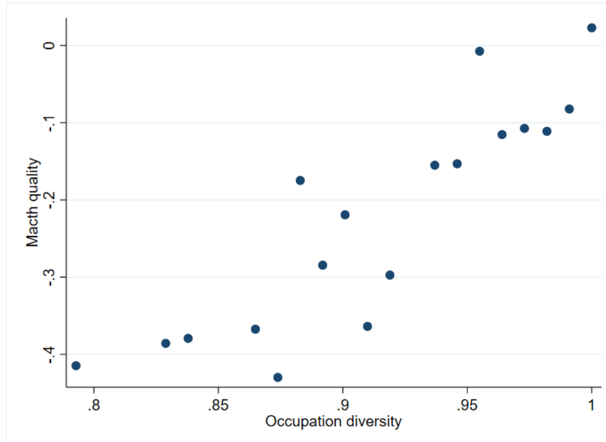
Notes: This table provides result estimating equation 2. The sample includes males born between 1970 and 1976, with outcomes for the years 1996-2013. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1

E Occupation diversity

Figure E.1: Occupation diversity



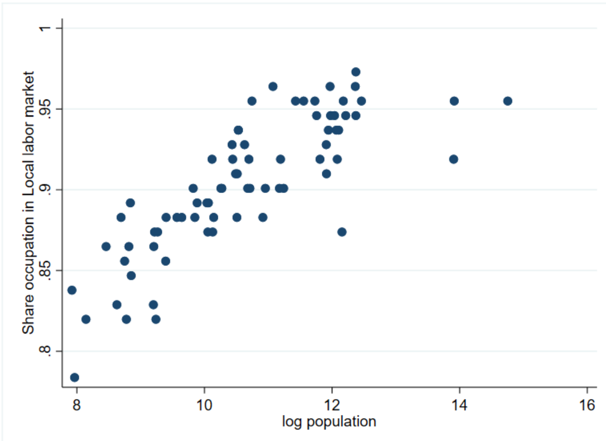
(a) Labor market size



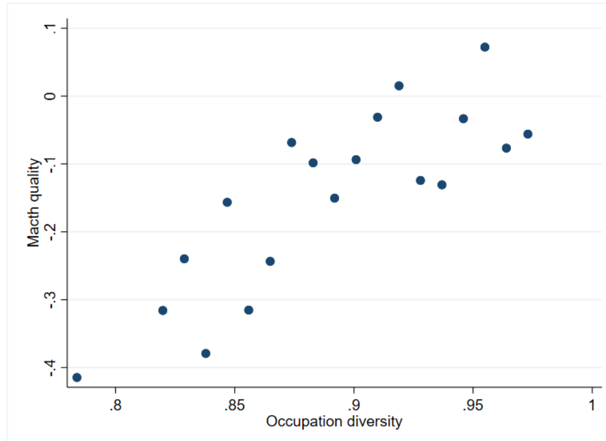
(b) Match quality

Notes: The figures plot occupation diversity, where occupation diversity is defined as the share of all possible occupations at the three-digit level that exists in the labor market. An occupation is defined as existing if somebody in the occupation sample works in the occupation. Both males and females are used to construct the occupation diversity variable. To measure the outcomes, the sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Figure (a) plots occupation diversity against labor market size, and figure (b) plots occupation diversity against match quality.

Figure E.2: Occupation diversity



(a) Labor market size



(b) Match quality

Notes: The figures plot occupation diversity, where occupation diversity is defined as the share of all possible occupations at the three-digit level that exists in the labor market. An occupation is defined as existing if somebody in the occupation sample works in the occupation. Both males and females are used to construct the occupation diversity variable. Occupation diversity is estimated by taking a random sample of 10,000 individuals from each labor market. To measure the outcomes, the sample includes males born between 1951 and 1976, with outcomes for the years 1996-2013. Figure (a) plots occupation diversity against labor market size, and figure (b) plots occupation diversity against match quality.

F Swedish population

Table F.1: Wage and labor market size females

VARIABLES	(1) Ln wage	(2) Ln wage	(3) Ln wage
log population size	0.033***	0.027***	0.013***
	(0.000)	(0.000)	(0.000)
Year FE	Yes	Yes	Yes
Age FE	Yes	Yes	Yes
Skill controls	No	No	No
Control education	No	Yes	No
Individual FE	No	No	Yes
Observations	14,005,036	14,005,036	14,005,036

Notes: This table provides results estimating equation 1, with log wage as the dependent variable, estimating the difference in log wages for individuals who live in labor markets of different sizes. The sample includes females born between 1951 and 1976, with outcomes for the years 1996-2013. In column 3, where individual fixed effects are included, the age dummies are normalized to be constant between 45 and 54 to avoid multicollinearity between age and year-fixed effects. Standard errors are adjusted for 659,206 clusters at the individual level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$