ISSN 1503-299X

WORKING PAPER SERIES

No. 1/2020

Assortative labor matching, city size, and the education level of workers*

Stefan Leknes
Research Department
Statistics Norway

Jørn Rattsø
Department of Economics
Norwegian University of Science and Technology

Hildegunn E. Stokke
Department of Economics
Norwegian University of Science and Technology

Department of Economics

Norwegian University of Science and Technology
N-7491 Trondheim, Norway
http://www.ntnu.edu/econ/working-papers

* We appreciate comments and discussions from colleagues, notably Magne Mogstad, Paolo Naticchioni, Kjell G. Salvanes and Bjarne Strøm. The project is funded by the Research Council of Norway (grant number 255509).

Assortative labor matching, city size, and the education level of workers*

Stefan Leknes[ξ], Jørn Rattsø[Ψ], and Hildegunn E. Stokke[Ψ]

Abstract

Recent research shows that thicker labor markets display better assortative matching. Our contribution addresses identification challenges and heterogeneity of effects, in particular with respect to education. Using a rich administrative worker-firm dataset for Norway, labor market size is shown to be of relevance for assortative matching mainly for the college educated. Among these, the pattern is most pronounced for workers of intermediate ages, with education related to business and administration, men, and service sector workers. Results are robust to instrumentation of population size using historical mines and sample adjustment to mitigate limited mobility bias.

Date: March 4, 2020

JEL codes: J24; J31; R23

Keywords: Assortative matching; Labor market size; Education; Register data; Historical mines

* We appreciate comments and discussions from colleagues, notably Magne Mogstad, Paolo Naticchioni, Kjell G. Salvanes and Bjarne Strøm. The project is funded by the Research Council of Norway (grant number 255509).
ξ Research Department, Statistics Norway. E-mail: sfl@ssb.no
Ψ Department of Economics, Norwegian University of Science and Technology.
E-mails: jorn.rattso@ntnu.no and hildegunn.e.stokke@ntnu.no

1. Introduction

Urban scholars have become increasingly aware of how skills create opportunities for workers to benefit from thicker labor markets. Many studies show that the urban wage premium is increasing with the level of education of workers. Bacolod et al. (2009) offer an overview understanding of allocation of skills across cities and the impact of agglomeration on wages. The observed wage premium is partly explained by the urban concentration of college-educated workers. Moretti (2004) analyzes the social return to education and finds that the number of college educated stimulates the productivity of a region. Baum-Snow and Pavan (2012) apply a structural model and find advantages for college-educated workers living in large cities. Carlsen et al. (2016) show that college-educated workers are positively selected into cities and they benefit more from working in cities even when sorting is accounted for. Combes and Gobillon (2015) survey methodological issues and recent studies relevant for sorting.

While the evidence of productivity effects from urban scale is convincing, the analyses of the sources of agglomeration economies are scarcer. One of the main channels of agglomeration effects is the advantage of large labor markets for the matching of workers and firms. A strand of this literature investigates the frequency of job change in labor markets of different sizes, and how job mobility corresponds to wage increases. An early contribution by Wheeler (2001) concludes that the advantage of large urban agglomerations results from enhanced worker productivity due to job search and matching. Wheeler (2006) and Yankow (2006) show that urban wage growth is related to turnover, and Finney and Kohlhase (2008) find that turnover is higher in urban areas. Wheeler (2008) and Bleakley and Lin (2012) find that younger workers in cities switch sectors and occupations more, and Leknes (2017) shows that the switching behavior in cities is increasing in education.

Other scholars have investigated the match between skills and jobs. Overeducation is less prevalent in urban settings according to Büchel and van Ham (2003). Boualam (2014) concludes that cities assist entrants to the labor market to find a job related to their education. Abel and Deitz (2015) present evidence of urban workers with graduate degrees being relatively more likely to have a job that both requires that level of education and is related to their college major. The demand side of the market has also been investigated. Strange et al. (2006) show that firms with demanding requirements regarding worker competences tend to locate in thick markets.

A novel approach entails the investigation of assortative matching between workers and firms. The method was innovated by Abowd, Karmarz and Margolis (1999), called the AKM-model. Among early studies related to urbanization, Andersson et al. (2007) estimate a positive relationship between labor market size and assortative matching using data for two US states. Melo and Graham (2014) correct for simultaneity between agglomeration and quality of match and support a positive relationship, but Mion and Naticchioni (2009) estimate a negative relationship for Italy. Figueiredo et al. (2014) concentrate on industrial clusters and find little evidence that the quality of matching increases with firms clustering within the same industry. Card et al. (2013, 2018) have developed the empirical methodology and shown that assortative matching affects the wage distribution. Dauth et al. (2018) study geographic wage disparities with this method and conclude that matching of workers to firms is an important explanatory factor. Wages are higher in large cities because they attract high-quality workers, but also because high-quality workers are likely to be better matched to high-quality firms. Assortative matching is stronger in large cities, in accordance with the standard understanding that large labor markets allow for more productive matches between workers and firms than small labor markets.

Using rich administrative data for Norway linking workers and firms, we investigate whether the overall positive relationship between city size and matching is driven by high-educated workers, similar to the skill-decomposition of the urban wage premium. Workers are separated by education level (primary, secondary and college). The positive relationship between city size and assortative matching is driven by college-educated workers. To address the concern of endogenous regional population size, we instrument population size with the number of historical mines. The mines started operating before the 19th century and are obsolete today, suggesting that the exclusion restriction holds. The difference between the OLS and IV estimates imply a negative bias of about 40%. The negative gap is robust to adjustment of the sample to deal with limited mobility bias.

To understand our results better, we dvelve deeper into the heterogeneity of effects. We start by analyzing whether there are specific education fields of the college educated that disproportionally drive the matching result. We find that high-educated within the field of business and administration are displaying higher assortative matching in cities. Next, the analysis is extended to address how the role of education varies across gender and industries. The scale coefficient is higher for college-educated men than women. Workers in services display better assortative matching in larger labor markets, while in manufacturing industries the match between worker and firm quality is independent of regional size. Conditional on industry, the gender differences follow the same general pattern. However, men have stronger relationship between regional size and matching quality, and even secondary educated workers take some advantage of larger labor markets. Finally, we explore the age gradient with respect to assortative matching in regions of different sizes. Independent of age group, college educated workers are better matched in more populous areas. In a separate analysis we look at the importance of turnover. The descriptive results indicate that rural-urban gap in job shifting corresponds to better assortative matching in cities. This result suggests that the superior job search opportunities in thicker labor markets may, at least partially, explain the results.

The econometric strategy and data are discussed in section 2. Section 3 reports results regarding the effect of city region population size for the strength of assortative matching for all workers using historical mines as instrument for regional population size. The bias compared to standard OLS is discussed. Section 4 estimates the relationship between regional size and matching quality separately for the education groups. Robustness of results are investigated in alternative model specifications in section 5. Further heterogeneity with respect to education fields, gender and industries, and age distribution are pursued in section 6. Section 7 looks at the role of turnover for assortative matching. Concluding remarks are given in section 8.

2. Data and econometric strategy

To study if there is a skill-heterogeneous relationship between assortative matching and city size, we employ the method innovated by Abowd, Kramarz and Margolis (1999) estimating two-way worker and plant fixed effects.[1] In this framework, workers and plant effects are identified by workers switching firms. It is therefore ideal to have a large longitudinal dataset covering the universe of workers. We use such a comprehensive dataset: employer-employee register data for the entire Norwegian labor force on hourly wages and worker characteristics during 2003–2010.

The individual level dataset is computed from three administrative registers: employment, tax, and education. The employment register links workers and plants and gives information on work contracts for all employees. It includes the duration of the contract, the type of contract[2], and the exact number of hours worked per week. We calculate the number of hours worked per year, which is combined with data on annual wage income from the tax register to give a measure of hourly wages for all employees. The education register covers the entire adult population and gives detailed information about workers' level and field of education. We also have information on the age, gender, immigrant status, occupation group, industry affiliation, plant affiliation, and home region of all individuals.

We concentrate on full-time workers aged 25–65 employed in the private sector.[3] The dataset includes about 575,000 workers every year during the period 2003–2010, giving a total of about 4.6 million worker-year observations in 54 industries, 350 occupation groups, 89 labor market regions, and about 115,000 plants. Workers can enter and leave the labor market during the eight-year period, and in total about 850,000 different workers are included.[4] Our main interest is to differentiate the degree of urban assortative matching based on skill levels.

---

[1] The terms *plant* and *firm* are used interchangeably throughout the paper.

[2] The employment register separates between three contract types: full-time contracts with at least 30 hours of work per week, part-time contracts with 20–29 hours of work per week, and part-time contracts with fewer than 20 hours of work per week.

[3] We exclude workers in the primary industries (agriculture, fishing, and forestry), as well as public sector workers. This gives a dataset of about 7.5 million worker-year observations. The tax register gives information on total annual earnings, rather than separate earnings for each work contract. Workers with more than two contracts during a year, as well as workers with one full-time and one part-time contract, are excluded. For workers with two full-time contracts, we allow for a maximum of three months of overlap between the contracts. We also exclude workers whose contract length is less than one month during a year. These restrictions reduce the dataset by about 0.5 million observations. Missing data on hours worked, annual earnings, level/field of education, occupation group, or industry affiliation, together with exclusion of workers that change education level after entering the labor market as full-time employees, further excludes approximately 2.2 million observations. To avoid extreme observations, we exclude the top and bottom 1% of the wage distribution.

[4] Workers that are observed in a single year are excluded, since individual fixed effects cannot be estimated.

We proceed with three subgroups of workers according to level of education: college (workers who have completed at least one year at college/university), secondary (workers who have completed at least one year of secondary education), and primary (workers with nothing more than compulsory schooling). About 19% of the workers have only a primary education, while workers with a secondary and college education account for 52% and 29% of the sample, respectively.

Panel A of Table 1 reports descriptive statistics of the individual level data, both aggregate and for three levels of education. The average hourly wage in the dataset is 270 NOK (log wage of 5.51) and wages of primary educated and secondary educated workers are, respectively, 39% and 26% below wages of the college educated. The average age is 42 years and decreases with the level of education, from 43 years among the primary educated to 40 years among college educated workers. Overall, 70% of workers are male, but the share is lower among the college educated. 12% of the workers are registered as immigrants. Separating between manufacturing industries and services, about 60% of primary- and secondary-educated workers are employed in services, increasing to 75% for college educated.

Table 1 about here

We estimate the following specification:

$$\ln w_{it} = \mu_i + \varphi_{J(i,t)} + \gamma_t + X_{it}\beta + \varepsilon_{it} \tag{1}$$

where w_{it} is the hourly wage income for worker *i* in year *t*. Worker and plant fixed effects are represented by μ_i and $\varphi_{J(i,t)}$, respectively. The vector of observable worker characteristics (X_{it}) includes education-specific cubic age profiles (quadratic and cubic age terms interacted with dummies for the three levels of education) and γ_t represents year dummies. β is a vector of parameters and ε_{it} is the error term. As seen from panel A of Table 1, the fixed effects are centered around zero overall. The mean value of worker fixed effects increases with the level of education, from -0.16 among the primary educated to 0.19 among college-educated workers. The differences in mean plant fixed effects are much smaller across education groups. The correlation between worker and plant fixed effects gives a measure of the strength of assortative matching. Under standard assumption of complementarity in production, we expect that there is an incentive for high-quality workers to match with high-quality firms. But overall, the correlation between worker and plant qualities is small. Across all 853,209 workers the correlation equals -0.11, and it is higher for college-educated workers (-0.04) than for workers with primary education (-0.27).

There are especially two estimation concerns related to the AKM method: limited mobility bias and the interpretation the plant effects. A plant's fixed effect is not identifiable without job switchers, and the estimates of the fixed effects might be noisy in small firms with few observations. The second issue relates to the sedentary nature of firms. As firms seldom change locations, their fixed effects will incorporate time-invariant traits (size, industry). Both these issues will only be of consequence if they are related to labor market scale. We use several methods to address these concerns in sensitivity tests.

To retrieve the strength of assortative matching within a region, we calculate the correlation between worker and plant fixed effects within each labor market $(CorrFE_r)$. The geographical units used in the analysis are based on information about commuting flows between municipalities. They are constructed by Statistics Norway, which divides Norway into 89 travel-to-work areas, denoted "economic regions". The economic regions conform to NUTS-4 regions, as defined by the European Union standard of regional levels. This level of aggregation captures functional regions understood as common labor markets.

To study the relationship between assortative matching and city size, we regress the correlation between worker and plant effects on regional population size (Pop_r): [5]

[5] Figures A1 and A2 in the appendix display results showing the relationship between population size and the fixed effects. As expected, the worker fixed effects are increasing in skills and the urban gradient is monotonically increasing with education level. This suggests that higher educated workers are better able to reap the benefits of thicker labor markets. The plant fixed effects are also increasing in education level, suggesting that higher educated workers are disproportionally located in high quality firms. The urban gradient for plant fixed effects is positive and quite similar across education groups.

$$CorrFE_r = \alpha_0 + \alpha_1 \log Pop_r + \varepsilon_r \tag{2}$$

Panel B of Table 1 presents descriptive statistics at the labor market region level. There are large differences in population size across regions with an average of about 50,000 inhabitants and a standard deviation of 75,000. Of the 89 regions, 36 have population below 20,000. The correlation between worker and plant fixed effects ranges from -0.52 to 0.05 and with an average of -0.20. The majority of negative correlations in city regions is consistent with Dauth et al. (2018). The variation is linked in complicated ways to worker and plant heterogeneity across regions.

Identification of the role of labor market size for the strength of assortative matching is challenging because workers and firms are sorted into urban areas motivated by superior labor matching opportunities. Although this may be an important productive advantage of cities, it bedevils interpretation of the population scale coefficient. An equally important challenge to identification is omitted variables; missing local variables that affect both matching opportunities and population size. To handle the possible endogeneity, we apply an instrument for current population size using information about the geographic distribution of historical mines introduced by Leknes (2015).

The mining industry was one of the first industries in Norway. In the same way as the locations of mineral resources are random, it can be argued that the geographical distribution of the mining industry is random. Also, all the historical mines were exhausted before the period of our study.[6] The discoveries of valuable mining resources incited economic activity that spurred local population growth, which is even traceable in the population patterns of today.

Norway has a long history of mining. The written source "Historia Norvegia" from 1170 mentions a silver mine in Oslo. After this period the mining industry gained momentum, and in the 18th century mining was one of the largest national industries. Today, however, the traditional mining industry is of marginal importance. The advantages of historical mining activity are obsolete, as all the historical mines where closed after the resources were

[6] We omit a region where there is mining activity in the vicinity of a historical mine.

exhausted. The instrument has therefore the same type of flare as the variation used in Bleakley and Lin (2012): they use natural features related to rivers, portage sites, and their importance in historical times for local economic activity and population growth. They argue that the natural advantages of these places can be considered forfeited today and find that they can still partially explain contemporaneous population patterns.

Data on historical mines are mainly collected from Thuesen (1979) and Carstens (2000), a detailed description of the data collection process can be found in Leknes (2015). We define historical mines as mines that opened sometime between the 12^{th} and 19^{th} century. As in the applications by Leknes (2015) and Carlsen et al. (2016), we use the number of historical mines in each region. The historical mining activity was reasonably spread out across the country with a mean number of 0.7 per region (s.e. = 1.28).

The fact that the historical mining activity has ceased sets this instrument apart from some of the other instruments in the urban economics literature, for instance historical population size. Using historical population size as instrument, identification hinges on different drivers of regional population growth historically compared to today. In this case, the mechanisms causing historical population growth is not explicitly stated and it is therefore more demanding to justify that they are not important today. We compare the results of the mining instrument with an alternative specification using early regional population size - beginning of the 19^{th} century population.

3. Assortative matching: IV estimation and bias

Our contribution adds to the limited literature about assortative matching and labor market size by addressing identification challenges and heterogeneity of effects, in particular with respect to the role of education.

The endogenous labor market size is a concern in the literature. We estimate the relationship between strength of assortative matching and city region size using an instrument for population size based on historical mining. The first stage estimate is documented in the first column of Table A1 in the appendix and shows a strong relationship between number of historical mines opened before the 19th century and current regional population size. The first column of Table 2 reports the second stage IV estimates and compares with OLS. The OLS estimated effect of population size equals 0.034 and is statistically significant at 1% level. IV-estimation with historical mines as instrument gives a coefficient of 0.057, significant at 10% level. This implies that the OLS estimate has a negative bias of about 40%. The interpretation of the result is that a doubling of the city region population size increases the correlation between worker and plant effect by 5.7 percentage points.

Table 2 about here

Our full sample results suggest that the degree of assortative matching is higher in larger labor markets. The conclusion is consistent with the recent contribution of Dauth et al. (2018) using a dataset of all private sector workers in Germany 1985–2014. They estimate an OLS coefficient for the 2002–2008 period of 0.061. This lies somewhat above our estimate of 0.034, which might reflect the larger scale of local German labor markets. Dauth et al. (2018) instrument regional population size with regional population in 1952. Identification then hinges on different drivers of regional population growth half a century ago compared to today. Using this instrument, they estimate a coefficient of 0.071 for the 2002–2008 period. It follows that they find less negative bias in the OLS estimate, around 13%. The difference in bias compared to our results may reflect that their historical population instrument is fairly recent and may offer less reliable exogenous variation.

We investigate the alternative identification strategy using early regional population size as instrument – beginning of the 19th century population, the year 1801. This is a well-known strategy in urban economics that originate from Ciccone and Hall (1996). However, it should be noticed that a shortcoming of historical population size instruments is that the mechanisms causing historical population growth is unknown and it is therefore hard to justify its validity. The first stage is documented in column 2 of Table A1 in the appendix and has a much higher R^2 than the mining instrument. The second stage is shown in the first column of Table A2 in the appendix and implies a coefficient of 0.06. We conclude that historical population back in 1801 as instrument gives a similar estimate as historical mines for all workers.

The interpretation of the negative OLS bias is not obvious. In a discussion of the endogeneity of labor market scale, Combes et al. (2008, 2010) emphasize two-way causality. Larger labor markets have been found to be more productive and are therefore more likely to attract firms and workers, which in turn will increase the labor market scale. Reverse causality suggests an upward bias of the estimates. But bias can also be related to heterogeneity. There may be unobserved traits of the region that have a positive relationship to both population size and degree of assortative matching. In this case, the OLS estimate is downward biased, consistent with our results. Industry composition may be a channel of heterogeneity. Small regions with specialized high-return industries may exhibit high assortative matching and attenuate the scale effect. Inspection of the 89 regions shows quite a few regions with small population that are dominated by a single industry (typically resource based, many of them linked to waterfalls and electricity production in fjords and valleys) and have good matches. Another factor that may dilute the scale effect is the structure of the education market. The larger universities are in urban areas and provide specialized educations that match well with specific jobs/occupations. Specialization of jobs have been found to be higher in cities (Duranton and Jayet, 2011). Mismatch may then occur when persons graduating with these educations settle for poor matches in cities waiting for better job opportunities to arise. Inspection of data for the regions again indicate that this may be important for some larger regional 'college cities'. More broadly, there are some large city regions with heterogenous private industries and many public sector jobs (not included in the analysis) that have less correlation between worker and firm quality.

4. Assortative matching heterogeneity: Education level of workers

Our main hypothesis is that the relationship between strength of matching and labor market size is positive and increasing in the formal skill level. In arguments relating back to Marshall, it is expected that large labor markets offer more opportunities for matching for both firms and workers. The number of potential matches increases with labor market size, and the incentive to search for a better match is stronger when the potential gains are larger. Several papers suggest that workers with higher skills have higher wage returns from thicker labor markets, which may also stem from better matching.

In columns 2-4 of Table 2 in section 3, we separate between primary-, secondary- and college-educated workers. The result is striking, population size is of relevance for assortative matching mainly for the college educated. Among primary- and secondary-educated workers, there is no significant relationship between size of labor market region and correlation of worker and plant fixed effects. This is consistent with the scatter plots in Figure 1 showing the association between city size and strength of assortative matching for the full sample and different education groups. For the college educated, the strength of assortative matching is increasing with population size. The IV-estimate equals 0.094, and the negative bias in the OLS estimate is still about 40%. Doubling the size of a city region leads to an increase in assortative matching as measured by the correlation coefficient of worker and plant fixed effects by 9 percentage points for college-educated workers.

Figure 1 about here

Labor markets differ with respect to skill and consequently we expect the matching between workers and plants to reflect differences both at the supply and the demand side. Our result implies that the market for the college educated takes advantage of larger city regions to achieve better quality matches between workers and plants. The result is consistent with the agglomeration literature discussed in the introduction – the college educated are overrepresented in larger cities (Bacolod et al., 2009) and the agglomeration effect is larger for college-educated workers (Carlsen et al., 2016). The urban wage premium has been shown to be smaller for workers with lower education. Here we find lack of assortative matching for primary and secondary educated, and this may contribute to the explanation of a lower agglomeration effect for the low educated.

5. Robustness

The AKM-model applied to estimate assortative matching has challenges related to limited mobility bias and interpretation of the plant effects. The plant fixed effects are identified based on job switchers, and the estimates may be noisy in small plants and small regions. Since the plants seldom change locations, their fixed effects incorporate time-invariant characteristics, notably industry. We use several methods to address the robustness of our results, and both OLS and IV estimates are reported in Table 3. The results are comparable to Table 2.

To check for the limited mobility bias problem, we reduce the sample in three alternative ways. First, in panel A, the smallest regions (below 20,000 inhabitants) are excluded. This leaves 53 regions covering 90% of the national population. The results are similar to Table 2, but the city region size effect for college educated is somewhat higher. In Panel B, we exclude all plants with less than five workers. This reduces the number of workers from 853,209 to 703,105 and the number of plants from 115,220 to 37,208. Excluding the smallest plants leads to higher correlation between worker and plant fixed effects. At the regional level, the average correlation increases from -0.20 to -0.06, and about 1/3 of the regions now have positive correlation between the fixed effects. The conclusions still hold, although the coefficient sizes for the college educated are somewhat smaller. To address the possibility of noisy estimates with limited mobility, we exclude in Panel C the top and bottom percentile of the fixed effects distributions. The results are robust to trimming the fixed effects.

Table 3 about here

As the plant fixed effects also include time-invariant characteristics, a concern is that the estimated relationship with urban scale reflects industry differences. In Panel D of Table 3, we purge the plant FE of industry influences by regressing them on 2-digit industry fixed effects and use the residual plant fixed effects in the correlation with worker FE. The results are very similar to the baseline. In all robustness checks, the negative bias in the OLS estimation is of similar magnitude as in Table 2.

6. Additional heterogeneity

6.1. Heterogeneity across education field for college educated

The literature on the college wage premium has moved on to investigate differences across fields of study and link this to geography, notably Cunningham et al. (2016). Given the registers with detailed education information for all workers, we can analyze how assortative matching varies across education fields. We concentrate on four large groups that cover about 85% of all college educated – humanities and art, social sciences and law, business and administration, and natural sciences. Table 4 reports the results for both OLS and IV estimation. The main conclusion is that stronger assortative matching with larger labor market regions primarily is observed for workers with an education in business and administration. The result is in line with the finding of Liu (2017) that workers with education related to Economics and Business display the highest urban wage premiums. There is no relationship between city region population size and strength of assortative matching for workers with college education in humanities and art and social sciences. City region size seems not to be relevant. There is a positive relationship for workers educated in natural sciences in the OLS estimation, but not in the IV model.

Table 4 about here

6.2. Heterogeneity with respect to gender and industry

Gender differences at the labor market have been studied in the context of urbanization and regional size. Hirsch et al. (2013) deal with regional differences in the gender pay gap, while Phimister (2005) analyzes the urban wage premium by gender. The variation in wages is linked to different gender intensities across industries. In our dataset, males represent a large share of workers in traditional manufacturing industries, construction, business services, information technology and finance. Females are more intensively employed in lighter manufacturing and services such as hotel/restaurants and cleaning. The importance of gender and industries consequently is investigated together.

When we study all workers, males and females have about the same relationship between city region size and strength of assortative matching. The first column of Table 5 shows that the estimated IV coefficient is about 0.07 for both genders, although significant only at 10% level for female workers. When separate regressions are run for the three levels of education, the same picture emerges as in Table 2, where only the college educated take advantage of larger labor markets. The quantitative effect of city region size on assortative matching is higher for college educated men than for college educated women (0.109 vs. 0.077), but both estimates are significant at 1% level.

Table 5 about here

Since the allocation of labor across industries is quite different between male and female workers, we pursue the assortative matching combining education, gender and industry. We separate between manufacturing industries (including construction) and services. The most striking result shown in Table 6 is that there is no effect of larger labor markets on assortative matching in manufacturing. This is true for all estimates – all education groups and when they are separated by gender (columns 1–3, panels A–C). In services, the effect of population size for the strength of assortative matching is statistically significant for both secondary- and college-educated men, but the estimated coefficient is much higher for the college educated (0.145 vs. 0.064). Female workers in services with college education have better matches in larger labor markets (coefficient of 0.086 significant at 1% level), while there is no significant effect for the two lowest education groups.

Table 6 about here

6.3. Heterogeneity with respect to age

Following the literature on job search, we expect workers early in their career to have imperfect information about their own preferences and abilities. They are therefore expected to experience worse matches early in their career relative to later. Because of data censoring (first observations are in 2003), we have limited knowledge about individual work histories. However, we can infer that younger workers have come shorter in the search process for better job matches. In Table 7, we split the workforce into four age categories: 25-34, 35-44, 45-54 and 55-65. We obtain the same conclusion as from the first part of the paper, positive urban assortative matching is only traceable for college-educated workers. Among these, in line with our hypothesis, the youngest age group has the lowest score.

Table 7 about here

7. The importance of job turnover

If worker characteristics and sector affiliation determine the opportunity to search for jobs in cities, this may explain why certain worker groups are better matched to firms. Leknes (2017) finds that workers in cities are more likely to switch jobs and that these turnover propensities are increasing in education. To explore this issue, we estimate a specification for job switching similar to equation (1), but without firm fixed effects:

$$S_{it} = \mu_i + \gamma_t + X_{it}\beta + \varepsilon_{it} \tag{3}$$

where S_{it} is an indicator variable equal to unity if the worker has a different firm identifier the subsequent year. Worker fixed effects are represented by μ_i. The vector of observable worker characteristics (X_{it}) includes education-specific cubic age profiles (quadratic and cubic age terms interacted with dummies for the three levels of education) and γ_t represents year dummies. β is a vector of parameters and ε_{it} is the error term.

To retrieve the propensity to switch jobs within each region, we calculate the mean worker fixed effect within each local labor market $(meanFE_r)$.[7] Next, we estimate a specification similar to equation (2), where we regress the mean worker effects on regional population size (Pop_r):

$$meanFE_r = \alpha_0 + \alpha_1 \log Pop_r + \varepsilon_r \tag{4}$$

As can be seen from Figure 2, college-educated workers are the only ones that have a significantly positive relationship between population size and regional probability of switching job.[8] Conducting the same analyses for the other splits in sample, we find descriptive evidence of rural-urban gaps in job switching that corresponds to better assortative matching in cities (see Tables A3 and A4 in the appendix). This is the case for college educated, within the field of business and administration, and in the service sector. The results suggest that turnover and assortative matching is linked. Women display a

[7] In the calculation of region means, each worker is only counted once and is allocated to the local labor market where he/she is first observed in the panel.

[8] Almost none of the IV-estimates are significant when using historical mines as instrument. The figures therefore display OLS estimates with robust standard errors.

somewhat different pattern. Although the propensity to switch jobs in cities for college-educated women are not very different from that of college-educated men, women with lower education levels switch more. The result is at odds with the search hypothesis. The pattern seems to be somewhat driven by occupation composition – women tend to disproportionally hold positions within culture, sales, administration and services and cleaning, which have higher turnover probabilities in cities. Women in the private sector may be a selected group, which confuses interpretation of results.

Figure 2 about here

8. Concluding remarks

The analysis adds to the limited literature about assortative matching and labor market size by addressing identification challenges and heterogeneity of effects, in particular with respect to the role of education.

Using rich administrative data for Norway linking workers and firms, we study if there is a skill-heterogeneous relationship between assortative matching and city size. Workers are separated by education level – primary, secondary and college. The method innovated by Abowd, Kramarz and Margolis (1999), the AKM-model, is employed to estimate two-way worker and plant fixed effects. The analysis addresses the concern of endogenous city region size. We instrument population size using historical mines before the 19th century that are obsolete today. The analysis shows that the overall positive relationship between city size and matching is driven by college-educated workers. Furthermore, we analyze whether there are specific education fields of the college educated that disproportionally link to the matching result. We find that high-educated workers within the field of business and administration are displaying higher assortative matching in cities. The analysis is extended to address how the role of education varies across gender and industries. The scale coefficient is higher for college-educated men than women. Also, workers in services display better assortative matching in larger labor markets, while in manufacturing industries, the match between worker and firm quality is independent of city region size. We also find that matching in cities is related to the age of workers. Independent of age group, college-educated workers are the only ones that display positive effects of population scale on assortative matching. However, the urban scale coefficients are higher for workers in the middle of the age distribution. Consistent with the search literature, matching is not as good early in the career as later. In an extension, descriptive evidence is provided suggesting that the patterns identified are related to turnover intensities. This suggests that the superior conditions for job search in cities may lead to better matches.

Our results indicate important variation in assortative matching between workers and firms dependent on education level, gender and industries. Methodological issues remain, in particular estimating the quality of firms. The plant fixed effects incorporate time-invariant regional characteristics that may be related to labor market size. We investigate the robustness of the results in alternative specifications. More broadly, it is of interest to develop characterizations of firms with more data about their structure and performance.

References

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

Abowd, J., F. Kramarz and D. Margolis (1999), High-wage workers and high-wage firms, *Econometrica* 67, 2, 251-233.

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

Bacolod, M., B. Blum and W. Strange (2009), Skills in the city, *Journal of Urban Economics* 65, 136-153.

Baum-Snow, N. and R. Pavan (2012), Understanding the city size wag gap, *Review of Economic Studies* 79, 1, 88-127.

Bleakley, H. and J. Lin (2012), Portage and path dependence, *Quarterly Journal of Economics* 127, 2, 587-644.

Boualam, B. (2014), Does culture affect local productivity and urban amenities? *Regional Science and Urban Economics* 46, 12-17.

Büchel, F. and M. van Ham (2003), Overeducation, regional labor markets, and spatial flexibility, *Journal of Urban Economics* 53, 3, 482-493.

Card, D., J. Henning and P. Kline (2013), Workplace heterogeneity and the rise of West German wage inequality, *Quarterly Journal of Economics* 128, 3, 967-1015.

Card, D., A. Cardoso, J. Heining and P. Kline (2018), Firms and labor market inequality: Evidence and some theory, *Journal of Labor Economics* 36, 51, S13-S70.

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

Carstens, H. (2000), …Bygger i Berge, Tapir Akademiske Forlag.

Ciccone, A. and R. Hall (1996), Productivity and the density of economic activity, *American Economic Review* 86, 1, 54-70.

Combes, P.P., G. Duranton and L. Gobillon (2008), Spatial wage disparities: sorting matters! *Journal of Urban Economics* 63, 723-742.

Combes, P.P., G. Duranton, L. Gobillon and S. Roux (2010), Estimating agglomeration economies with history, geology, and worker effects. In: Glaeser, E. (Ed.), *Agglomeration Economics*, The University of Chicago Press, pp. 15-66.

Combes, P.P. and L. Gobillon (2015), The empirics of agglomeration economies. In: Duranton, G., V. Henderson and W. Strange (Eds.), *Handbook of Urban and Regional Economics* vol. V. Elsevier-North Holland, Amsterdam, pp. 247-348.

Cunningham, C., M. Patton and R. Reed (2016), Heterogeneous returns to knowledge exchange: Evidence from the urban wage premium, *Journal of Economic Behavior & Organization* 126, 120-139.

Dauth, W., S. Findeisen, E. Moretti and J. Suedekum (2018), Matching in cities, NBER Working Paper No. 25227.

Duranton, G. and H. Jayet (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.

Figueiredo, O., P. Guimaraes and D. Woolward (2014), Firm-worker matching in industrial clusters, *Journal of Economic Geography* 14, 1-19.

Finney, M. and J. Kohlhase (2008), The effect of urbanization on labor turnover, *Journal of Regional Science* 48, 2, 311-328.

Hirsch, B., M. König and J. Möller (2013), Is there a gap in the gap? Regional differences in the gender pay gap, *Scottish Journal of Political Economy* 60, 4, 412-439.

Leknes, S. (2015), The more the merrier? Evidence on quality of life and population size using historical mines, *Regional Science and Urban Economics* 54, 1-17.

Leknes, S. (2017), Churning in thick labor markets: evidence of heterogeneous responses along the skill and experience gradients. Discussion Papers no. 866, Statistics Norway.

Liu, S. (2017), Agglomeration, urban wage premiums, and college majors, *Journal of Regional Science*, 57, 4, 611-630.

Melo P. and D. Graham (2014), Testing for labour pooling as a source of agglomeration economies: Evidence for labour markets in England and Wales, *Papers in Regional Science* 93, 1, 31-52

Mion, G. and P. Natacchioni (2009), The spatial sorting and matching of skills and firms, *Canadian Journal of Economics* 42, 28-55.

Moretti, E. (2004), Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data, *Journal of Econometrics* 121, 175-212.

Phimister, E. (2005), Urban effects on participation and wages: Are there gender differences? *Journal of Urban Economics* 58, 3, 513-536.

Strange, W., W. Heijazi, and J. Tang (2006), The uncertain city: competitive instability, skills, innovation, and the strategy of agglomeration, *Journal of Urban Economics* 59, 3, 331-351.

Thuesen, G. (1979), Den første dokumenterte bergverksdrift i Norge, Volund, pp. 7-60.

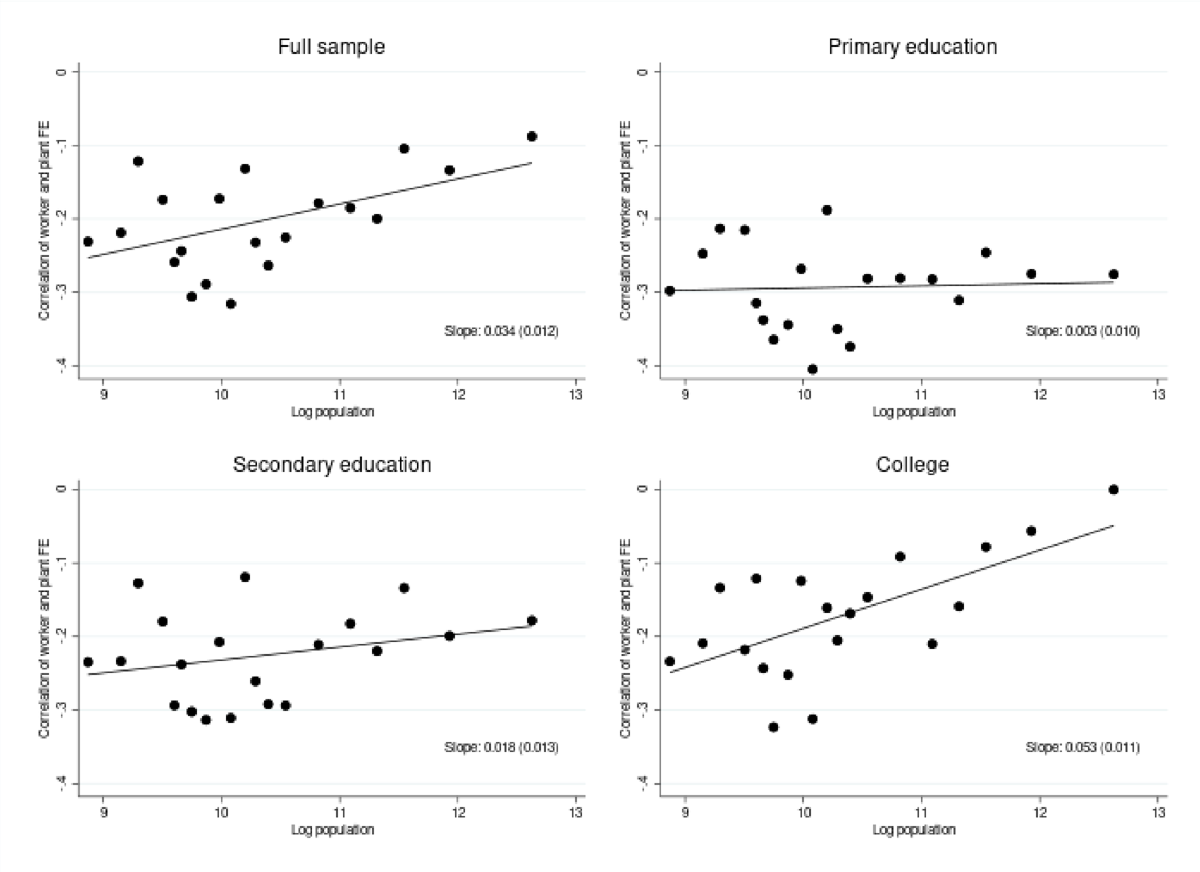
Wheeler, C. (2001), Search, sorting and urban agglomeration, *Journal of Labor Economics* 19, 879-899.

Wheeler, C. (2006), Cities and the growth of wages among young workers: Evidence from the NLSY, *Journal of Urban Economics* 60, 162-184.

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

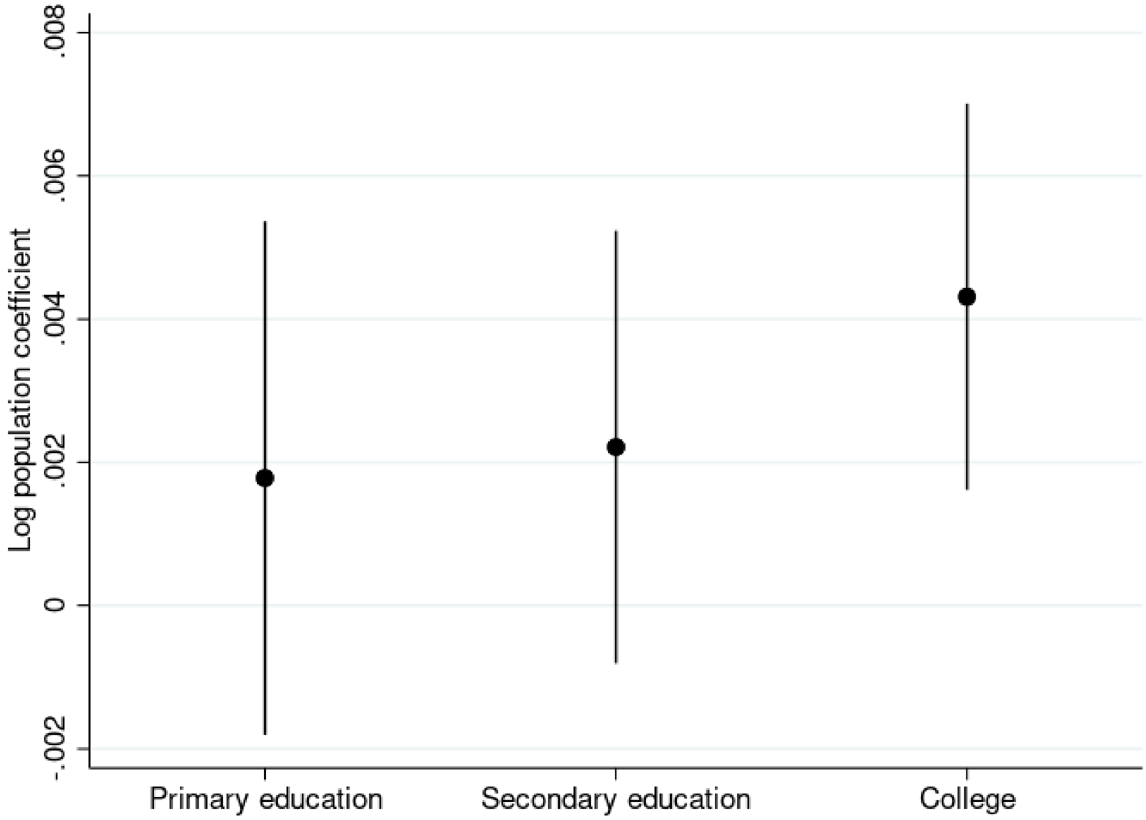
Yankow, J. (2006), Why do cities pay more? An empirical examination of some competing theories of the urban wage premium, *Journal of Urban Economics* 60, 139-161.

Figure 1. City size and strength of assortative matching. Heterogeneity across education groups.



Notes: The vertical axis shows correlation between worker and plant fixed effects at the regional level (N=89). In top-left figure, the correlation is calculated based on fixed effects for 853,209 individuals allocated across 89 labor market regions. In the other figures, the correlations are calculated based on fixed effects for three subgroups of workers defined by the level of education: 164,807 workers with primary education, 442,048 workers with secondary education, and 246,354 college-educated workers. The fixed effects follow from individual level AKM estimations during 2003–2010 of the log hourly wage on worker effects, plant effects, education-specific cubic age profiles, and year dummies. The regional population level is measured in 2003. The figures are binned scatter plots, where regions are grouped into 20 percentiles based on population size. The line in each figure is given by bivariate regressions and the slope coefficient and its robust standard error are provided within the figure.

Figure 2. City size and regional probability of turnover. Heterogeneity across education groups.



Notes: The figure shows the relationship between the regional mean worker fixed effects from a job switching regression and population at the regional level (N=89). The mean worker fixed effects are calculated based on subgroups of workers defined by the level of education: 164,807 workers with primary education, 442,048 workers with secondary education, and 246,354 college-educated workers. The fixed effects follow from a linear probability model at the individual level during 2003–2010. The dependent variable is an indicator of different firm ID the subsequent year, and the regressions control for worker effects, education-specific cubic age profiles, and year dummies. The regional population level is measured in 2003. The figure provides point estimates and 95 percent confidence intervals with robust standard errors.

Table 1. Descriptive statistics

	All	Primary	Secondary	College
Panel A: Worker level (mean values)				
Log hourly wage	5.51	5.33	5.46	5.72
Hourly wage	270.2	221.3	252.6	334.4
Age	42.0	43.1	42.6	40.3
Male	0.70	0.71	0.73	0.65
Immigrant	0.12	0.15	0.10	0.15
Manufacturing	0.36	0.39	0.42	0.25
Services	0.64	0.61	0.58	0.75
Worker FE	0.00	-0.16	-0.05	0.19
Plant FE	0.00	-0.02	-0.01	0.02
Corr (worker FE, plant FE)	-0.108	-0.272	-0.171	-0.04
No. of workers	853,209	164,807	442,048	246,354
Share of workers	1.00	0.19	0.52	0.29
No. of plants	115,220	54,172	88,228	49,018

	Mean	St.dev.	Min	Max
Panel B: Regional level				
Population 2003	51,149	74,858	5,595	517,401
No of historical mines	0.7	1.28	0	6
Historical population 1801	9,928	7,577	353	49,209
Corr (worker FE, plant FE) region	-0.204	0.116	-0.52	0.053
No. of regions	89			

Notes: The values for hourly wage and age refer to the average value across the period 2003-2010.

Table 2. City size and strength of assortative matching: Heterogeneity across education groups

	Dependent variable: Correlation of worker and plant FE			
	All	Primary	Secondary	College
	(1)	(2)	(3)	(4)
Panel A: OLS				
Log population	0.034***	0.003	0.018	0.053***
	(0.012)	(0.01)	(0.013)	(0.011)
R^2	0.08	0.001	0.018	0.178
Panel B: IV-2SLS				
Log population	0.057*	0.013	0.024	0.094***
	(0.03)	(0.025)	(0.038)	(0.028)
First stage F	18.04	18.04	18.04	18.04

Notes: The dependent variable is the correlation between worker and plant fixed effects at the regional level (N=89 in panel A and N=88 in panel B). In column (1), the correlation is calculated based on fixed effects for 853,209 individuals allocated across 89 labor market regions. In columns (2)–(4), the correlations are calculated based on fixed effects for three subgroups of workers defined by the level of education: 164,807 workers with primary education, 442,048 workers with secondary education, and 246,354 college-educated workers. The fixed effects follow from individual level AKM estimations during 2003–2010 of the log hourly wage on worker effects, plant effects, education-specific cubic age profiles, and year dummies. The regional population level is measured in 2003. In panel B, the instrument for log population is the number of historical mines opened before the 19th century. The first stage estimation is given in column (1) of Table A1 in the appendix. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Table 3. Robustness checks

	Dependent variable: Correlation of worker and plant FE			
	All	Primary	Secondary	College
	(1)	(2)	(3)	(4)
Panel A: Excluding the smallest regions (pop < 20,000)				
OLS: Log population	0.054***	0.02	0.035*	0.069***
	(0.017)	(0.015)	(0.019)	(0.016)
R^2	0.175	0.028	0.069	0.243
IV: Log population	0.07	0.025	0.019	0.122**
	(0.052)	(0.04)	(0.062)	(0.052)
Panel B: Excluding small plants (less than 5 workers)				
OLS: Log population	0.029**	0.008	0.015	0.031**
	(0.012)	(0.01)	(0.014)	(0.013)
R^2	0.048	0.000	0.011	0.046
IV: Log population	0.051**	0.028	0.026	0.054**
	(0.025)	(0.026)	(0.027)	(0.025)
Panel C: Excluding outliers in the FE distributions				
OLS: Log population	0.03***	0.006	0.017*	0.046***
	(0.009)	(0.007)	(0.01)	(0.01)
R^2	0.101	0.005	0.031	0.168
IV: Log population	0.049**	0.017	0.022	0.073***
	(0.019)	(0.017)	(0.02)	(0.024)
Panel D: Dependent variable is the correlation of worker and residual plant FE				
OLS: Log population	0.03***	0.004	0.017*	0.05***
	(0.009)	(0.009)	(0.01)	(0.009)
R^2	0.09	0.002	0.028	0.199
IV: Log population	0.063**	0.026	0.039	0.101***
	(0.029)	(0.025)	(0.034)	(0.028)

Notes: The table presents robustness checks on the regressions in Table 2, where the dependent variable is the correlation between worker and plant fixed effects at the regional level (N=89 with OLS and N=88 with IV). In panel A, we exclude regions with less than 20,000 inhabitants in 2003, which leaves 53 regions covering 90% of the national population (N=53 with OLS and N=52 with IV). In panel B, plants with less than five full-time workers are excluded. This reduces the number of workers from 853,209 to 703,105 and the number of plants from 115,220 to 37,208. In panel C, we exclude workers that are in the top or bottom 1% of the distribution of worker and/or plant fixed effects. This reduces the number of workers from 853,209 to 822,206 and the number of plants from 115,220 to 107,032. In panel D, the dependent variable is the correlation between worker and residual plant fixed effects at the regional level, where the residual plant effects are the residuals from a regression of the plant effects on 2-digit industry effects. In all panels, the instrument for log population is the number of historical mines opened before the 19th century. The first stage estimation is given in column (3) and column (1) of Table A1 in the appendix for panel A and for all other panels, respectively. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Table 4. City size and strength of assortative matching: College educated in different fields of study

	Dependent variable: Correlation of worker and plant FE			
	Humanities and arts	Social sciences and law	Business and administration	Natural sciences
	(1)	(2)	(3)	(4)
Panel A: OLS				
Log population	-0.006	0.013	0.069***	0.028**
	(0.027)	(0.029)	(0.012)	(0.014)
R^2	0.001	0.003	0.216	0.032
Panel B: IV-2SLS				
Log population	0.07	0.044	0.084***	0.04
	(0.047)	(0.046)	(0.023)	(0.028)

Notes: The dependent variable is the correlation between worker and plant fixed effects at the regional level for the four largest subgroups of college educated workers defined by field of study: *Humanities and arts* (26,183 workers), *Social sciences and law* (22,118 workers), *Business and administration* (71,217 workers), and *Natural sciences* (91,944 workers). The four subgroups cover 86% of all college-educated workers. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Table 5. City size and strength of assortative matching: Education and gender

	Dependent variable: Correlation of worker and plant FE			
	All	Primary	Secondary	College
	(1)	(2)	(3)	(4)
Panel A: MEN (IV-2SLS)				
Log population	0.07***	0.02	0.04	0.109***
	(0.026)	(0.025)	(0.028)	(0.034)
Panel B: WOMEN (IV-2SLS)				
Log population	0.068*	0.026	0.059	0.077***
	(0.04)	(0.04)	(0.05)	(0.026)

Notes: The dependent variable is the correlation between worker and plant fixed effects at the regional level (N=88). The fixed effects follow from individual level AKM estimations during 2003–2010 of the log hourly wage on worker effects, plant effects, education-specific cubic age profiles, and year dummies. The regional population level in 2003 is instrumented with the number of historical mines opened before the 19th century. The first stage estimation is given in column (1) of Table A1 in the appendix. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Table 6. City size and strength of assortative matching: Education, gender and industry

	Dependent variable: Correlation of worker and plant FE					
Sector	MANUFACTURING			SERVICES		
Education group	Primary	Secondary	College	Primary	Secondary	College
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: ALL (IV-2SLS)						
Log population	-0.003	-0.008	0.015	0.036	0.052	0.124***
	(0.028)	(0.029)	(0.026)	(0.028)	(0.04)	(0.029)
Panel B: MEN (IV-2SLS)						
Log population	0.002	-0.002	0.033	0.042	0.064**	0.145***
	(0.031)	(0.03)	(0.025)	(0.027)	(0.029)	(0.038)
Panel C: WOMEN (IV-2SLS)						
Log population	0.014	0.021	0.012	0.038	0.07	0.086***
	(0.047)	(0.045)	(0.053)	(0.04)	(0.048)	(0.025)

Notes: The dependent variable is the correlation between worker and plant fixed effects at the regional level (N=88). The fixed effects follow from individual level AKM estimations during 2003–2010 of the log hourly wage on worker effects, plant effects, education-specific cubic age profiles, and year dummies. The regional population level in 2003 is instrumented with the number of historical mines opened before the 19th century. The first stage estimation is given in column (1) of Table A1 in the appendix. Robust standard errors in parentheses. *** indicates significance at the 1 percent level.

Table 7. City size and strength of assortative matching, by age

	Dependent variable: Correlation of worker and plant FE			
	All	Primary	Secondary	College
	(1)	(2)	(3)	(4)
Panel A: Age 25-34 (IV-2SLS)				
Log population	0.034	-0.002	-0.008	0.049**
	(0.026)	(0.025)	(0.036)	(0.023)
Panel B: Age 35-44 (IV-2SLS)				
Log population	0.073*	0.033	0.024	0.122***
	(0.037)	(0.033)	(0.046)	(0.038)
Panel C: Age 45-54 (IV-2SLS)				
Log population	0.055*	-0.015	0.031	0.113**
	(0.030)	(0.031)	(0.034)	(0.047)
Panel D: Age 55-65 (IV-2SLS)				
Log population	0.061*	-0.006	0.034	0.078*
	(0.032)	(0.036)	(0.040)	(0.046)

Notes: The dependent variable is the correlation between worker and plant fixed effects at the regional level (N=88). The fixed effects follow from individual level AKM estimations during 2003–2010 of the log hourly wage on worker effects, plant effects, education-specific cubic age profiles, and year dummies. The regional population level in 2003 is instrumented with the number of historical mines opened before the 19th century. The first stage estimation is given in column (1) of Table A1 in the appendix. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Appendix: Additional tables and figures

Table A1. First stage IV estimation

	Log population (1)	Log population (2)	Log population (3)
Historical mines	0.288***		0.185***
	(0.068)		(0.067)
Log population 1801		0.884***	
		(0.127)	
Constant	10.099***	2.393**	10.713***
	(0.106)	(1.143)	(0.119)
Observations	88	89	52
R^2	0.148	0.502	0.124

Notes: The dependent variable is regional population size in 2003 (log form). In columns (1) and (3), the instrument is the number of historical mines opened before the 19th century, while the instrument in column (2) is historical population size (measured in the year 1801). In column (3), we drop regions with less than 20,000 inhabitants in 2003. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Table A2. Second stage IV-estimation with historical population size as instrument

	Dependent variable: Correlation of worker and plant FE			
	All (1)	Primary (2)	Secondary (3)	College (4)
Log population	0.06***	0.03	0.049**	0.073***
	(0.02)	(0.019)	(0.024)	(0.018)
First stage F	48.66	48.66	48.66	48.66

Notes: The dependent variable is the correlation between worker and plant fixed effects at the regional level (N=89). The regional population level in 2003 is instrumented with historical population size (1801 census). The first stage estimation is given in column (2) of Table A1. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Table A3. City size and job switching probabilities: College educated in different fields of study

	Dependent variable: Mean worker fixed effects from job switching regression			
	Humanities and arts (1)	Social sciences and law (2)	Business and administration (3)	Natural sciences (4)
Log population	0.002	-0.007	0.009***	-0.000
	(0.003)	(0.006)	(0.002)	(0.002)
R^2	0.002	0.022	0.155	0.000

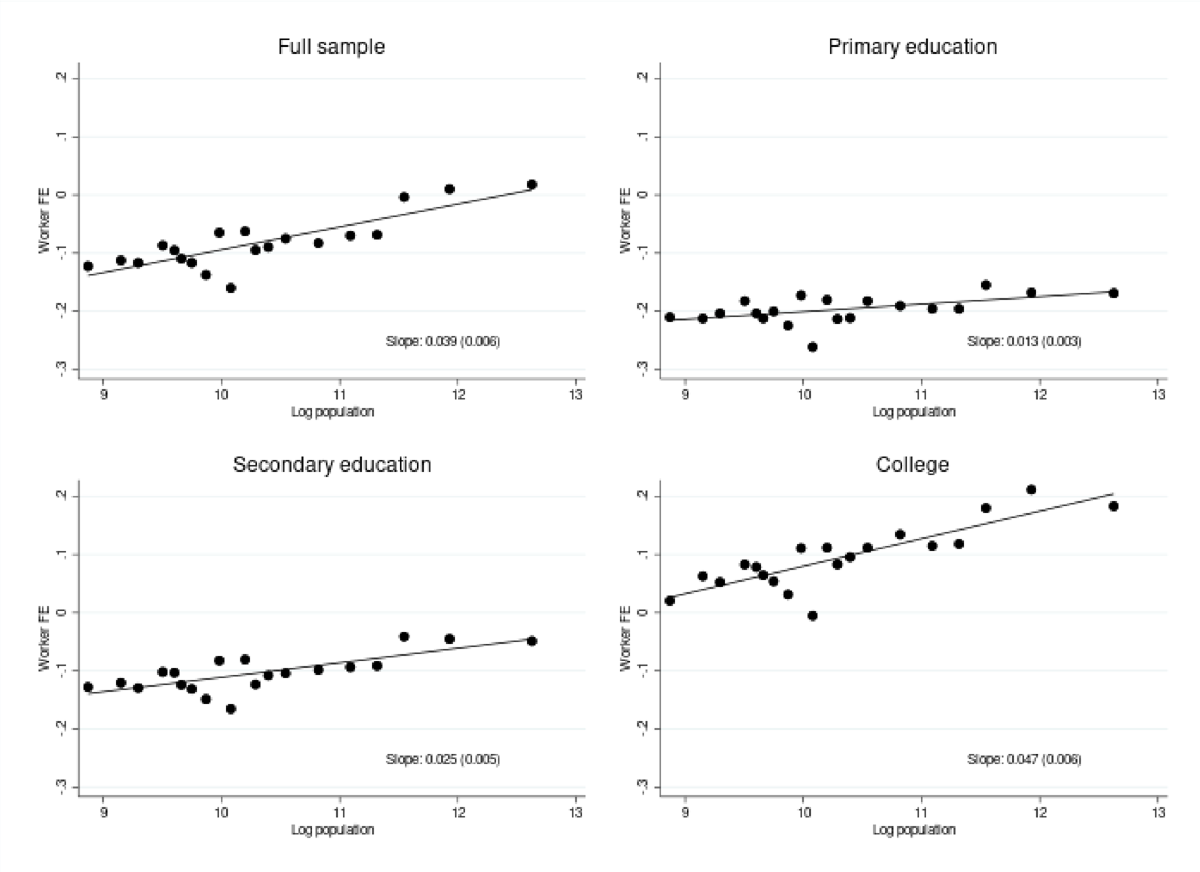
Notes: The table shows the relationship between the regional mean worker fixed effects from a job switching regression and population at the regional level (N=89). The mean worker fixed effects are calculated for the four largest subgroups of college educated workers defined by field of study: *Humanities and arts* (26,183 workers), *Social sciences and law* (22,118 workers), *Business and administration* (71,217 workers), and *Natural sciences* (91,944 workers). The worker fixed effects follow from a linear probability model at the individual level during 2003–2010. The dependent variable is an indicator of different firm ID the subsequent year, and the regressions control for worker effects, education-specific cubic age profiles, and year dummies. The regional population level is measured in 2003. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Table A4. Heterogeneity in urban job switching behavior, by education

	Dependent variable: Mean worker fixed effects from job switching regression		
	Primary (1)	Secondary (2)	College (3)
Panel A: Men			
Log population	0.001	0.001	0.004***
	(0.002)	(0.002)	(0.001)
R^2	0.002	0.003	0.076
Panel B: Women			
Log population	0.010***	0.011***	0.005**
	(0.003)	(0.002)	(0.002)
R^2	0.169	0.291	0.072
Panel C: Manufacturing			
Log population	-0.005*	-0.002	0.001
	(0.003)	(0.002)	(0.002)
R^2	0.035	0.009	0.002
Panel D: Services			
Log population	0.005***	0.005***	0.006***
	(0.002)	(0.002)	(0.001)
R^2	0.086	0.078	0.146
Panel E: Services for men			
Log population	0.003	0.002	0.005***
	(0.002)	(0.002)	(0.002)
R^2	0.018	0.009	0.092
Panel F: Services for women			
Log population	0.012***	0.012***	0.006**
	(0.003)	(0.002)	(0.002)
R^2	0.153	0.298	0.079

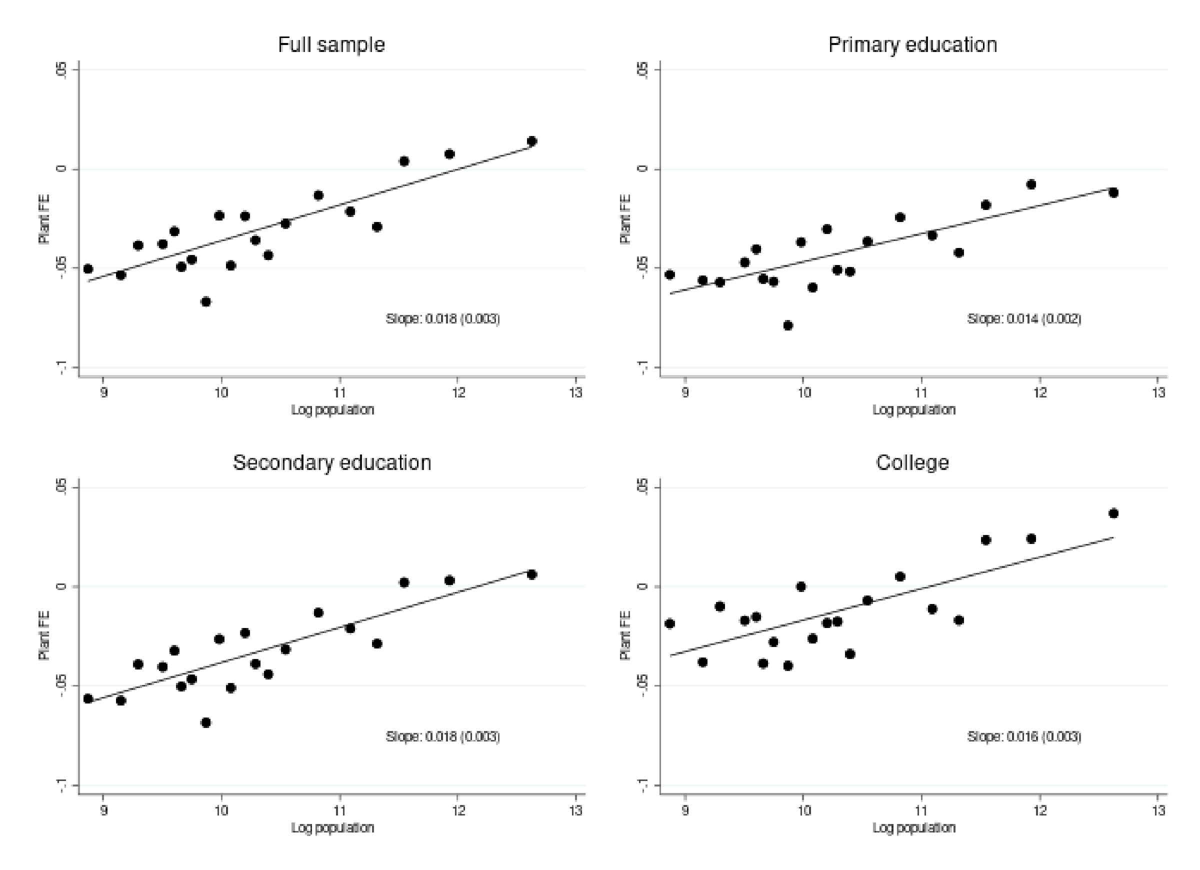
Notes: The table shows the relationship between the regional mean worker fixed effects from a job switching regression and population at the regional level (N=89). The mean worker fixed effects are calculated based on subgroups of workers defined by the level of education: 164,807 workers with primary education, 442,048 workers with secondary education, and 246,354 college-educated workers. The fixed effects follow from a linear probability model at the individual level during 2003–2010. The dependent variable is an indicator of different firm ID the subsequent year, and the regressions control for worker effects, education-specific cubic age profiles, and year dummies. The regional population level is measured in 2003. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1 percent, 5 percent and 10 percent level, respectively.

Figure A1. City size and worker fixed effects. Heterogeneity across education groups.



Notes: The vertical axis shows mean worker fixed effects at the regional level (N=89). In top-left figure, the mean is calculated based on fixed effects for 853,209 individuals allocated across 89 labor market regions. In the other figures, the means are calculated based on fixed effects for three subgroups of workers defined by the level of education: 164,807 workers with primary education, 442,048 workers with secondary education, and 246,354 college-educated workers. The fixed effects follow from individual level AKM estimations during 2003–2010 of the log hourly wage on worker effects, plant effects, education-specific cubic age profiles, and year dummies. They are centered around zero for the full sample. The regional population level is measured in 2003. The figures are binned scatter plots, where regions are grouped into 20 percentiles based on population size. The line in each figure is given by bivariate regressions and the slope coefficient and its robust standard error is provided within the figure.

Figure A2. City size and plant fixed effects. Heterogeneity across education groups.



Notes: The vertical axis shows mean plant fixed effects at the regional level (N=89). In top-left figure, the mean is calculated based on fixed effects for 853,209 individuals allocated across 89 labor market regions. In the other figures, the means are calculated based on fixed effects for three subgroups of workers defined by the level of education: 164,807 workers with primary education, 442,048 workers with secondary education, and 246,354 college-educated workers. The fixed effects follow from individual level AKM estimations during 2003–2010 of the log hourly wage on worker effects, plant effects, education-specific cubic age profiles, and year dummies. The regional population level is measured in 2003. The figures are binned scatter plots, where regions are grouped into 20 percentiles based on population size. The line in each figure is given by bivariate regressions and the slope coefficient and its robust standard error is provided within the figure.