Discussion Paper No. 17-055

Local Labor Market Size and Qualification Mismatch

Francesco Berlingieri

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 17-055

Local Labor Market Size and Qualification Mismatch

Francesco Berlingieri

Local labor market size and qualification mismatch*

Francesco Berlingieri†

ZEW Centre for European Economic Research, Mannheim

November 6, 2017

Abstract

This paper investigates the effect of the size of the local labor market on skill mismatch. Using survey data for Germany, I find that workers in large cities are both less likely to be overqualified for their job and to work in a different field than the one they are trained for. Different empirical strategies are employed to account for the potential sorting of talented workers into more urbanized areas. Results on individuals never moving from the place of childhood and fixed-effects estimates obtaining identification through regional migrants suggest that sorting does not fully explain the existing differences in qualification mismatch across areas. This provides evidence of the existence of agglomeration economies through better matches. However, lower qualification mismatch in larger cities is found to explain only a small part of the urban wage premium.

JEL-classification: I21, J24, J31, R23

Keywords: agglomeration, labor matching, qualification mismatch, urban wage premium

*I am indebted to Melanie Arntz, Wolfgang Dauth, Jorge De la Roca, Christina Gathmann, Henry Overman, Ulrich Zierahn for their valuable remarks and suggestions on earlier versions of the paper. The paper also benefitted from helpful comments at the workshop "New Research in the Economics of Educational and Skills Mismatch" (Aberdeen, 2014), the workshop "Perspectives on (Un-)Employment" (IAB Nuremberg, 2014), the 6th EGIT workshop (Dusseldorf, 2015), the 5th European meeting of the Urban Economics Association (Lisbon, 2015), the 2015 conference of the German Economic Association, the CEDEFOP/IZA Workshop on Skills and Skill Mismatch (Thessaloniki, 2015) and the 2015 Barcelona workshop on regional and urban economics. I thank the staff at DIW for their support with the remote access data from the German Socio-Economic Panel (SOEP). I gratefully acknowledge financial support from the Germany's Federal Ministry of Education and Research framework program "Economics of Science" (BMBF, research grant 01PW11019). The usual disclaimer applies.

†ZEW Mannheim, Labour Markets, Human Resources and Social Policy Research Department, P.O. Box 103443, D–68034 Mannheim, email: `berlingieri@zew.de`.

1 Introduction

There is ample evidence that workers earn higher wages in larger labor markets. For instance, Glaeser and Mare (2001) show that average wages in metropolitan areas with a big city (i.e. a city with more than 500.000 inhabitants) are about 33% higher than outside these areas. From an individual perspective, the higher cost of living in cities might explain why not all workers are willing to move to larger cities. However, the urban wage premium must reflect higher productivity in larger cities to explain why firms do not relocate to less urbanized areas. Duranton and Puga (2004) distinguish three mechanisms behind the higher productivity in larger cities: the sharing of facilities and risks, faster learning and knowledge diffusion and better matches between firms and workers. While the importance of better matches as a source of agglomeration economy is stressed from a theoretical perspective, there is little evidence of its empirical relevance (Puga, 2010). A major explanation for this is the paucity of data allowing to measure match quality in a comprehensive way. Previous studies have attempted to measure it indirectly through the share of occupational and industry changes (Bleakley and Lin, 2012) or through assortative matching in terms of worker and firm quality (Andersson et al., 2007; Dauth et al., 2016) and found evidence of better matches in more urbanized areas. The focus in this paper is on a direct measure of job match quality, namely the match between the formal qualifications earned by workers and the job requirements. This type of match has been found to be strongly related to wages and firm productivity (Leuven and Oosterbeek, 2011; Kampelmann and Rycx, 2012). More specifically, I look at the match between actual and required qualifications both in terms of level (vertical match) and in terms of content (horizontal match), since there is reason to expect both types of match to be better in thicker labor markets.

The question whether workers in more densely populated areas are less exposed to educational mismatch is also interesting in and of itself and relevant for the labor economics literature on skill mismatch. Does it actually pay off for individuals to move to larger cities in terms of better job matches and future career prospects? Previous studies have already investigated the impact of various regional labor market characteristics including the size of regional labor markets, regional unemployment rates and mobility restrictions on overqualification (Büchel and van Ham, 2003; Jauhiainen, 2011). However, these studies aimed at analyzing several determinants of overqualification and not at establishing a clear - and possibly causal - link between the size of the local labor market and qualification mismatch. OLS regressions with standard control variables might lead to biased estimates in this context. On the one hand, since cities generally have a higher share of university graduates, one could suspect that they could be more attractive for individuals with higher unobserved ability. In fact, several papers have stressed the importance of addressing the spatial sorting of workers by individual skills in order to estimate the urban wage premium (Glaeser and Mare, 2001; Combes et al., 2008). On the other hand, the skill mismatch literature has found a positive correlation between measures of individual ability and overqualification (Leuven and Oosterbeek, 2011). Since more talented individuals could be both more likely to live in large cities and to have a better job match, the sorting of workers across areas could lead to an overestimation of the effect of city size on the job match. To address this potential bias, I estimate linear regressions of qualification mismatch on regional employment density including an extensive set of control variables that correlate with individual ability, such as information on parental background, school grades and personality traits.[1] I then follow three main empirical strategies to sequentially deal with the major empirical concerns and test whether the baseline estimates are robust for different sub-groups of individuals.[2] Firstly, by restricting the sample to individuals who have remained in the region where they grew up (non-movers) biases from the direct migration of more talented workers into cities can be addressed. Secondly, by estimating a fixed effects model on our panel of workers and obtaining identification through individuals migrating form one region to another, I can get rid of unobserved time-invariant heterogeneity (such as individual ability). Thirdly, I deal with potential reverse causality by instrumenting current employment density with historical population data from the 19th century.

The obtained estimates of employment density on overqualification are fairly similar across the different specifications. An increase of 10% in the regional employment density is associated with a decrease of 1-1.5% in the probability of being overqualified. On the contrary, most of the estimates of employment density on the horizontal mismatch measure are smaller and not statistically significant. Finally, I investigate the contribution of

[1] I use regional employment density to measure the labor market size following previous studies (for a review of the literature on agglomeration economies, see Combes and Gobillon, 2015 and Heuermann et al., 2010). The results do not change qualitatively when using population density or dummy variables for urban areas.

[2] This procedure is common in the urban wage premium literature, because of the difficulty of finding exogenous sources affecting the mobility of individuals across regions.

better qualification matches in explaining the wage premium in thicker labor markets. By including our mismatch measures to an OLS regression of log hourly wages on employment density (and other control variables), overqualification is found to explain only 6% of the impact of regional employment density on hourly wages, while the contribution of horizontal mismatch appears to be insignificant.

Two other recent studies analyze the effect of population or employment density on job mismatch for the US (Abel and Deitz, 2015) and France (Boualam, 2014).[3] Abel and Deitz (2015) find evidence of a moderate effect of population size and employment density on measures of vertical and horizontal mismatch for US college graduates. They also find that mismatch accounts for 5-8% of the urban wage premium. Boualam (2014) investigates the impact of employment density on a measure of horizontal match based on the distribution of workers' fields of study within an occupation for French labor market entrants. While this measure of match quality is found to increase with employment density, it does not seem to explain differences in wages between thick and thin labor markets. The present paper provides at least three contributions. First, while the cited papers make use of cross-sectional data, I employ panel data that enables me to estimate fixed effects regressions to eliminate the time-invariant unobserved ability bias as in previous studies on the urban wage premium (Glaeser and Mare, 2001; Combes et al., 2008). Second, the survey data I use (i.e. the German Socio-Economic Panel) has extensive information on workers characteristics and biographies that might be very important to account for in the analysis to avoid potential omitted variable biases, such as detailed parental background information, high-school final grades and information on personality traits. Third, the data contains direct questions on the qualifications required by the job, allowing me to construct vertical and horizontal qualification mismatch variables based on workers' self-assessments. Measures based on worker's self-assessment are typically preferred over measures which infer the required qualification from the data at hand, since the latter not only depend on demand forces but also on qualifications supplied (Hartog, 2000).

The rest of the paper is organized as follows. Section 2 describes the data and presents descriptive evidence of the link between employment density and qualification mismatch. Section 3 contains the main results on the impact of employment density on overqualifi-

[3] Andini et al. (2013) also analyze the impact of population density on different measures of job matching, including the appropriateness of the educational qualification for the job. However, their coefficients are not statistically significant for the educational match, as well as for most of the other measures of matching.

cation and horizontal mismatch. While in Section 4 I attempt to disentangle the effect of labor market size from that of other characteristics of denser regions (such as specialization and skill structure) on the mismatch incidence, in Section 5 I investigate the contribution of qualification mismatch to the wage differential across regions. Finally, Section 6 concludes.

2 Data and Descriptive Statistics

2.1 Data Source and Key Variables

The sample used is drawn from the German Socio–Economic Panel (GSOEP), a panel data set for the years 1984-2012 consisting of about 20,000 individuals living in Germany (for details, see Kroh, 2012). I restrict the sample to males surveyed in the years 2000 to 2011 to avoid concerns about possible selection biases into labor force participation for women. The sample is further restricted to dependent workers employed full-time. The 12 GSOEP waves include 8,288 male adults aged between 16 and 64 with a university degree or a completed training qualification who are employed at least twice in the time framework considered. I end up with an unbalanced panel of 5,625 individuals (35,363 observations), for whom I have information on all variables relevant for our analysis.

I employ the regional employment density at the level of labor market regions as a measure of labor market size.[4] This is calculated by the number of employed individuals per square kilometer. Labor market regions are defined by the Federal Office for Building and Regional Planning to differentiate areas in Germany based on their economic interlinkages and of commuting patterns. This classification specifies 258 labor market regions with an average of 16 regions in each of the 16 federal states.[5] Information on employment density at the level of regional labor markets is gathered from administrative data sources (i.e. the INKAR database) and merged to the individual place of residence in the GSOEP data. [6]

[4] Similar results are found though when using population density or dummy variables for urban areas.

[5] Results using a different classification of 150 labor market regions show baseline estimates that are of similar magnitude, as shown in Table A.4.

[6] Ideally, I would consider the workplace location, because agglomeration economies are expected to arise where the production process takes place. Unfortunately, this information is not available in the GSOEP. Nevertheless, I do not expect this to affect much the results, since few individuals commute outside of regional labor markets. Moreover, the GSOEP data includes information on commuting distances, so that I can test whether the results are robust to excluding long-distance commuters.

I employ two measures for qualification mismatch: vertical mismatch (i.e. overqualification) and horizontal mismatch. Overqualification is measured based on workers' self–assessment about the educational requirements of the job. More precisely, the following question is asked in the GSOEP questionnaire: "What type of education or training is usually necessary for this type of work?" I consider an individual to be overqualified if he reports that his job requires a lower degree than the he possesses.One drawback of this measure, which is widespread in the literature on overeducation, is the reliance on the subjective individual self–assessment. For instance, according to Hartog (2000), respondents might have a tendency to upgrade the status of their position. However, since I employ this measure as an outcome variable, subjectivity would be an issue only if workers in small and large labor markets systematically differ in the way they answer such a question. Several authors have claimed that overqualification measures based on self-assessments are preferable to measures based on the distribution of educational qualifications within occupations – i.e. "realized matches" (Leuven and Oosterbeek, 2011). Measurement error might be more severe for measures based on realized matches, because they ignore any variation in required education within occupations. Moreover, realized matches do not reflect only job requirements but are already the outcome of the interplay of supply and demand (Hartog, 2000). I also rely on workers' self–assessment to compute the horizontal mismatch measure similarly to previous studies (see, e.g., Robst, 2007). The question asked in the GSOEP is: "Is this position the same as the profession for which you were educated or trained?". Since the only possible answers are yes or no, I construct a dummy that is equal to 1 if individuals answer negatively to this question.

Hourly wages are measured through the self-reported monthly gross income divided by monthly working hours. I calculate real wages based on the CPI deflator using 2010 as the base year. In order to ensure that outliers are not driving the main results I trim wages excluding the 1st and the 99th percentile (individuals receiving a hourly wage lower than EUR 4 or higher than EUR 75) and I employ the standard logarithmic form for the wage regressions. The richness of the data allows me to include in the analysis an extensive set of control variables. I consider demographic characteristics, educational and parental background information, job characteristics and some geographic characteristics, such as previous regional mobility. Further information on school grades, personality traits and risk preferences is available only for certain waves or for given individuals depending on the time of their first participation to the survey. I thus add this characteristics only in a separate analysis. In Section 4 I also add further regional information at the local labor market level which is gathered from the INKAR database.

2.2 Descriptive Results

Table 1 presents the mean and standard deviation for the variables included in the analysis. The overqualification incidence is about 15% in the sample, while the incidence of horizontal mismatch amounts to 30%. Employment density ranges from 16 employed individuals per square km in Salzwedel (Sachsen-Anhalt) to 1,889 in Berlin.[7] Most individuals have a vocational degree as their highest qualification (68%), while the rest of the sample has a tertiary degree either from a standard university or a university of applied science (*Fachhochschule, FH*). I further include information on the school leaving qualification, which is a further important control variable in Germany because of the tracking system of secondary education. Individuals have on average 21 years of work experience, most of which (13 years on average) gained with their current employer. Most of the individuals in the sample (61%) never left the city where they grew up.

Figure 1 shows the existence of a negative relationship between employment density and qualification mismatch as measured through the subjective assessment of the qualification level required by the job (vertical mismatch or overqualification) and the relatedness of the job to the worker's field of education or training (horizontal mismatch). The unit of observation in both graphs is the labor market region, meaning that the information on the individual match is aggregated at the regional level. The slope of the fitted regression line is -0.014 for vertical mismatch and -0.022 for horizontal mismatch and the coefficients are statistically significant at standard levels for both regressions.

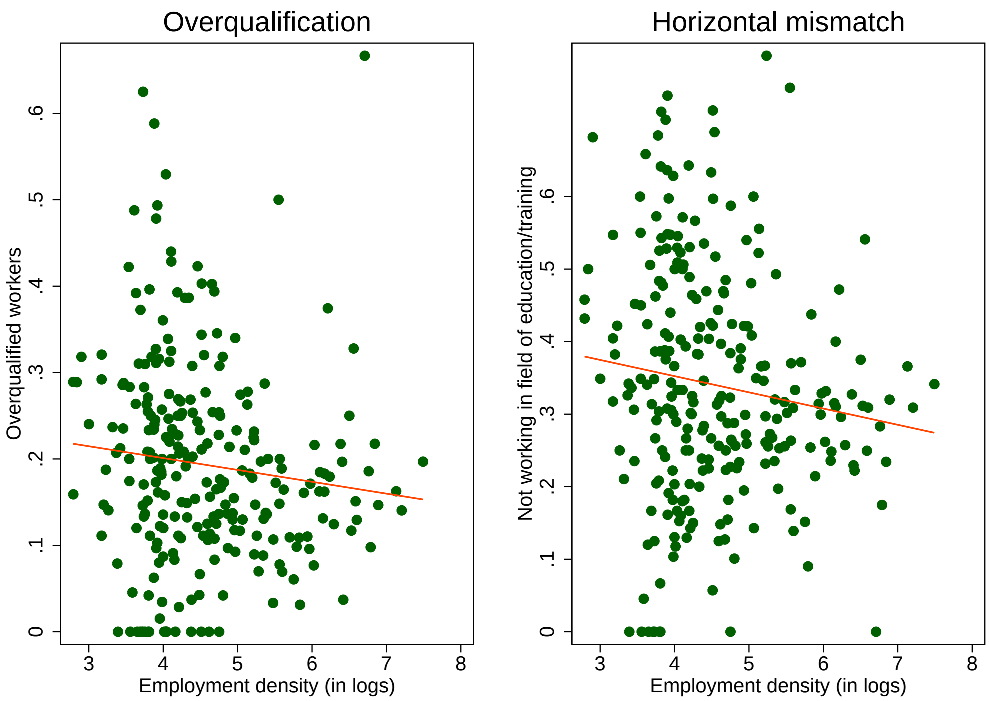
[7] Figure A.2 shows the differences in employment density across the 258 German labor market regions in 2010 (darker colors depict a higher employment density).

Table 1: Summary Statistics

	Mean	Std. Dev.	Min.	Max.
Dependent variables and other main variables				
Overqualified	0.19	0.39	0	1
Horizontal mismatch	0.34	0.47	0	1
Hourly wage (log)	2.76	0.44	1.59	3.91
Employment density (log)	5.03	1.02	2.75	7.56
Main control variables				
University degree	0.21	0.41	0	1
FH degree	0.11	0.31	0	1
Vocational degree	0.68	0.47	0	1
Migration background	0.08	0.27	0	1
Married or living with partner	0.82	0.38	0	1
Actual work experience	21.1	10.5	0	48
Has children	0.40	0.49	0	1
School leaving qualification				
University access (Abitur)	0.27	0.44	0	1
FH access (Fachhochschulreife)	0.08	0.27	0	1
Realschulabschluss	0.35	0.48	0	1
Hauptschule or no degree	0.30	0.46	0	1
Parental background				
Father: higher educ.	0.14	0.34	0	1
Mother: higher educ.	0.06	0.25	0	1
Mother non-employed (age 15)	0.24	0.43	0	1
Geographic characteristics				
Lives in city of childhood	0.61	0.49	0	1
Macro-region				
Centre	0.33	0.47	0	1
North	0.13	0.34	0	1
South	0.27	0.44	0	1
East	0.27	0.44	0	1
Job characteristics				
Public sector	0.25	0.43	0	1
Firm tenure	12.8	10.5	0	47
Industry				
Agriculture	0.01	0.12	0	1
Energy	0.02	0.13	0	1
Mining	0.01	0.07	0	1
Manufacturing	0.23	0.42	0	1
Construction	0.19	0.39	0	1
Trade	0.10	0.30	0	1
Transport	0.07	0.25	0	1
Bank & Insurance	0.05	0.23	0	1
Services	0.32	0.47	0	1

Note: The summary statistics are based on the baseline sample of 35,363 observations (5,625 individuals). Main control variables include year fixed effects, as well as a squared term for work experience. Job characteristics also include firm tenure squared.

Figure 1: Employment Density and Qualification Mismatch



3 Impact of Agglomeration on Qualification Mismatch

3.1 Baseline Regressions

Having seen that there is a negative relationship between employment density and qualification mismatch, I first test whether the results change when I include an extensive set of control variables. I thus estimate a the following simple linear probability model[8]:

$$Pr(mismatch_{ijt} = 1) = \alpha + \beta\ empdensity_{jt} + \gamma \mathbf{X_{ijt}} + \epsilon_{ijt} \tag{1}$$

where *mismatch* is a dummy variable that takes value 1 in case of a qualification mismatch for individual i in year t, *empdensity* denotes the employment density of the region of residence j in year t and $\mathbf{X_{ijt}}$ is a vector of covariates that differs across specifications. Panel A of Table 2 shows the results for the overqualification dummy, and Panel B those for horizontal mismatch. Column (1) reports results for a regression with the inclusion of the

[8]Average marginal effects estimates of a probit model lead to results that are very similar to the linear probability model estimates.

main control variables only (i.e. highest educational qualification, migration background, marital status, having children in household, actual experience, experience squared, year dummies). The remaining five columns show results by gradually including dummies for the school leaving qualification, parental background characteristics (i.e. father and mother education, whether the mother was employed when the individual was 15 years old), geographic characteristics (macro-region dummies and whether individual still lives in place of childhood), job characteristics (i.e. tenure, public sector, industry dummies) and occupation fixed effects in column (6).

Table 2: Impact of Employment Density on Qualification Mismatch

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Overqualification			
Empl. density (log.)	-0.031***	-0.026***	-0.025***	-0.017***	-0.019***	-0.012***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Main controls	Yes	Yes	Yes	Yes	Yes	Yes
School degree	No	Yes	Yes	Yes	Yes	Yes
Parental background	No	No	Yes	Yes	Yes	Yes
Geographic charact.	No	No	No	Yes	Yes	Yes
Job charact.	No	No	No	No	Yes	Yes
Occupation FE	No	No	No	No	No	Yes
Observations	35,363	35,363	35,363	35,363	35,363	35,363
R-squared	0.021	0.049	0.051	0.060	0.090	0.207
			Panel B: Horizontal mismatch			
Empl. density (log.)	-0.030***	-0.027***	-0.025***	-0.012*	-0.015**	-0.012**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Main controls	Yes	Yes	Yes	Yes	Yes	Yes
School degree	No	Yes	Yes	Yes	Yes	Yes
Parental background	No	No	Yes	Yes	Yes	Yes
Geographic charact.	No	No	No	Yes	Yes	Yes
Job charact.	No	No	No	No	Yes	Yes
Occupation FE	No	No	No	No	No	Yes
Observations	35,363	35,363	35,363	35,363	35,363	35,363
R-squared	0.041	0.055	0.057	0.069	0.099	0.184

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variables. Standard errors are clustered at the individual level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).

Column (1) in Panel A shows the existence of a negative relationship between regional employment density and the probability of being overqualified for the job when standard control variables are included. The coefficient is equal to -0.031 and is significant at the 1%

significance level. A 10% increase in employment density would imply a decrease of about 0.3 percentage points in the overqualification probability, which is equal to a decrease of about 1.6% (given that the overqualification rate in our sample is 19%). The employment density coefficient decreases to -0.018 when school degree dummies, parental background information, geographic characteristics and job characteristics are included. The inclusion of occupation fixed effects (ISCO 1-digit) in column (6) leads to a smaller coefficient (-0.012), but is still statistically significant. While the ISCO classification at the 1-digit level is relatively broad, its inclusion together with the information about educational qualifications might partially capture vertical qualification mismatch. Therefore, in the overqualification regressions it is better not to control for occupation fixed effects, since these might be bad control variables.

Panel B shows that regional employment density appears to have a negative impact on horizontal mismatch as well, i.e. whether one works in the same field of one's education or training. The coefficient in column (1) is equal to -0.030 and is statistically significant. A 10% increase in employment density would imply a decrease of about 1% in horizontal mismatch (since the incidence of horizontal mismatch is about 30%). The coefficient decreases to -0.016 when school degree dummies, parental background information, geographic controls and job characteristics are included. In particular, a large part of the correlation between density and horizontal mismatch can be explained by differences between West and East Germany, as East Germany is characterized on average by both a lower employment density and a higher incidence of horizontal mismatch. When occupation fixed effects are included, the coefficient drops (in absolute value) further to 0.012 but is still significant at the 5% significance level. Since no information on the field or orientation of the highest qualification obtained is included, there are less arguments against the inclusion of occupation fixed effects in the case of horizontal mismatch. Workers in denser regions appear thus to have a better job match in terms of their actual qualifications.

3.2 Controlling for School Grades, Personality Traits and Risk Preferences

The GSOEP data contains further individual information which might be important to control for when analyzing the effect of employment density on skill mismatch. First, high-school grades might proxy individual ability and motivation and thus reduce potential biases from the sorting of talented individuals into larger cities. Second, personality traits and risk preferences might differ on average across regional areas and are likely to affect the job match, as well as the individual assessment of the match. Since these characteristics are available only for a relative small sample of individuals, I exclude these from the baseline regressions and present separate results for a sub-sample of 2,141 individuals, for whom I have information on all relevant characteristics.

Table A.1 presents results of a linear probability model of qualification mismatch by gradually adding mathematics and German grades from workers' final school reports, standard measures of the big five personality traits (extraversion, consciousness, agreeableness, neuroticism and openness to experience) and a subjective measure of risk preference.[9] Columns (1) and (5) of the table present the results of the same model of column (5) in Table 2, where all baseline control variables are included except for occupation fixed effects. The employment density coefficient is slightly larger in absolute terms for overqualification compared to the baseline sample and is statistically significant. Conversely, the estimate is not significant for horizontal mismatch, because of the smaller sample size. For both variables, however, the estimates remain very similar when school grades, personality traits and risk preference are included. While some characteristics matter for the qualification mismatch measures, they appear to be almost irrelevant for the impact of employment density on the match.[10]

3.3 Addressing the Omitted Ability Bias

Two empirical strategies are employed to address the potential overestimation of the results due to omitted ability bias stemming from the sorting of talented individuals into larger cities. Similarly to Boualam (2014) I first investigate whether the results are different for the sub-samples of individuals ever moving from a district to another (movers)

[9]Mathematics and German are the only compulsory courses for the high school diploma in most federal states in Germany. The grades are measured using the 6 points scale typical for the German system, where 1 is the best grade and 6 the worst. The big 5 personality traits are indexes in the range of 1 to 21, which are computed basing on a larger set of personality items contained in the survey following Gerlitz and Schupp (2005). The measure of risk preference is a index ranging from 1 to 10 based on an individual statement. Since I have information for both the big five and risk preference only for specific years, I compute the individual average of all observed values.

[10]In particular, it appears that individuals with better math and German school grades have a lower likelihood of mismatch. Among the big 5 personality traits neuroticism in particular appears to be positively associated with both mismatch variables.

and the ones who stay in the place where they grew up (non-movers). Focusing on non-movers helps me to avoid biases from the direct migration of more talented workers to cities. Column (1) of Table 3 reports the results for the same linear regression estimated in the last column of Table 2 (with the inclusion of all control variables apart from occupation fixed effects). The same model is then estimated on the sub-sample of individuals never moving from the place of childhood who represent about 61% of the sample.[11] In the case of overqualification, the coefficient for the sub-sample of non-movers is very similar to the one for all individuals. In the case of horizontal mismatch, the coefficient is even slightly larger.

Table 3: Impact of Employment Density on Qualification Mismatch: Addressing Spatial Sorting

	Pooled OLS		Fixed effects			
	all	non-movers	all		restricted sample	
	(1)	(2)	(3)	(4)	(5)	(6)
			Overqualification			
Empl. density (log.)	-0.018***	-0.017***	-0.017*		-0.020*	
	(0.004)	(0.006)	(0.009)		(0.011)	
Ave. empl. density (log)				-0.014		-0.017*
				(0.009)		(0.010)
			Horizontal mismatch			
Empl. density (log.)	-0.016***	-0.019**	-0.005		-0.005	
	(0.006)	(0.008)	(0.009)		(0.010)	
Ave. empl. density (log)				-0.002		-0.002
				(0.009)		(0.009)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No	No	No
Observations	35,363	21,532	35,363	35,363	35,137	35,137

Note: Standard errors are clustered at the individual level in the cross-sectional regressions and at the individual level in the panel regressions; *** p<0.01, ** p<0.05, * p<0.1; Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies). [a] Only control variables that change over time are included.

The results for the sub-sample of non-movers might still be biased if talented individuals are more likely to grow up in cities because of inherited abilities by parents and grandparents who moved to large agglomerations (Glaeser and Mare, 2001; Bosquet and

[11]This question is constructed based on a survey question asking if the individual still lives in the city or regional area where most of the childhood was spent.

Overman, 2016). Thus, in a second step, I exploit the panel structure of the data in order to control for individual fixed effects. On the one hand, this enables me to address the problem that unobserved individual ability might lead to an overestimation of the results. On the other hand, the identification will be achieved through individuals migrating from one district to another and individuals moving to a different region are likely to do so because they find a better job match (Gould, 2007). Therefore, the identification strategy will rely on the assumption that the reason to change region will not differ for the same individual whether he moves to a larger or a smaller region. Only 486 individuals in our sample change their labor market region of residence. 260 of these change both region and job. Since the identification will hinge upon those changing their region of residence and only job switchers can change the match status, I also estimate a regression excluding the spells in which individuals change region but not the job.

For simplicity, I estimate the following linear fixed effects model that gets rid of the time-constant unobserved individual heterogeneity:

$$Pr(\ddot{overqual}_{ijt} = 1) = \beta \ \ddot{empdensity}_{jt} + \gamma \ddot{\mathbf{X}}_{\mathbf{1,ijt}} + \ddot{\epsilon}_{ijt} \tag{2}$$

where the "double dot" denotes that the variables are time demeaned, *overqual* is a dummy variable denoting if individual i in year t is overqualified for the job, *empdensity* denotes the employment density of the region of residence and the vector $\mathbf{X}_{\mathbf{1,ijt}}$ includes all time variant control variables, excluding occupation fixed effects. These are part of the demographic characteristics, job and geographic characteristics. In a separate regression I also use the average regional employment density over the period considered (2000-2011), so that change in density across years are not taken into account. I do this because, unlike wages, the mismatch measures are dummies that are typically constant if the worker does not change job and it is unlikely that they respond quickly to small changes in the size of the labor market. Column (4) of Table 3 presents the results of the fixed effects model. The coefficient for overqualification (-0.017) turns out to be similar to the estimate of the baseline model, but it is less precise and statistically significant only at the 90% confidence level. In column (5) I not allow regional employment density to vary over time and the coefficient turns out to be smaller (in absolute value) and not anymore statistically significant. When I exclude the spells of those changing region but not job, the size of the effect increases slightly and both the time-varying and the average employment density coefficients become statistically significant at the 90% confidence level. Differently from overqualification, the impact of employment density on horizontal mismatch appears to vanish in the fixed effects model.

Thus the fixed effects estimate turns out to be similar to the baseline LPM regression for overqualification, but almost zero for horizontal mismatch. As said, while the fixed effects estimation gets rid of the omitted ability bias, it relies on the assumption that the individual reasons to change region do not differ systematically depending on the move to a bigger or smaller region. If the same individual moving to a larger city because of a better job match then returns to his place of childhood at the cost of a worse match (e.g. to take care of the parents), this will of course affect the results.

3.4 Addressing Potential Reverse Causality

Another potential bias arising in the estimation of agglomeration economies is reverse causality. In fact, higher local wages or areas with better matching could attract workers and increase local labor supply and thus employment density (Combes and Gobillon, 2015). Following several studies in the literature I use historical population values as an instrument for current density (Ciccone and Hall, 1996; Combes et al., 2008; Andini et al., 2013). Specifically, I employ population data from the 1880 census of all cities and towns with more than 10,000 inhabitants.[12] The cities and towns from 1880 are then matched to actual regional labor markets. In regional labor markets, for which no city is listed in the 1880 census, a population of 10,000 is imputed. As it will be shown, the instruments are relevant because of inertia of the local population. However, because of the profound changes that affected Germany over the past century, I am confident that the instrument is exogenous to the current match. The country experienced a structural shift from agriculture to manufacturing and services, dramatic political and administrative changes, two world wars that strongly reshaped its borders, as well as the separation and reunification of East and West Germany.

Column (1) of Table 4 shows the results of a two-stage least square regression on overqualification, where the log of employment density is instrumented by the log of the (city) population in 1880. The coefficient of the 1880 population in the first-stage

[12] The data was collected from the Statistical Yearbooks of the German Empire (Statistisches Jahrbuch für das Deutsche Reich).

Table 4: Impact of Employment Density on Qualification Mismatch: IV Regressions

	Overqualification		Horizontal mismatch	
	2SLS	FE IV	2SLS	FE IV
	(1)	(2)	(3)	(4)
Empl. density (log.)	-0.028***	-0.018*	-0.023***	0.006
	(0.006)	(0.009)	(0.008)	(0.010)
	First-stage equation			
1880 Population (log.)	0.679***	0.710***	0.679***	0.710***
	(0.008)	(0.003)	(0.008)	(0.003)
Control variables	Yes	Yes[a]	Yes	Yes[a]
Occupation FE	No	No	No	No
Observations	35,363	35,363	35,363	35,363

Note: Standard errors are clustered at the individual level in the cross-sectional regressions and at the individual level in the panel regressions; *** p<0.01, ** p<0.05, * p<0.1; Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies). [a] Only control variables that change over time are included.

equation is large and significant and the F-statistics on the excluded instruments is equal to 6648. The coefficient of employment density in the overqualification regression is negative, statistically significant and even larger in absolute value than in the the baseline regression. Column (3) shows the results of the same model for horizontal mismatch. Also in this case the coefficient is negative, statistically significant and larger than the baseline estimate. Column (2) and (4) report the results for fixed-effects models, where employment density is instrumented with historic population. Again the results are similar to the fixed-effects equation. The employment density is negative and significant (at the 90% confidence level) in the overqualification regression and not significantly different from zero in the horizontal mismatch regression (similar to the baseline model).

3.5 Further Robustness Checks

I carry out several robustness checks to ensure that the main results are not driven by outliers and specific regions. Given that regions in East Germany have on average a higher mismatch and a lower employment density, it is important to test if the results are driven by differences between East and West Germany. Estimating the baseline regression only for West Germany gives very similar estimates for overqualification, while the horizontal mismatch estimate is slightly smaller (in absolute values) and not statistically significant (see Table A.2). Very similar results are also found when leaving out Berlin, which is the largest agglomeration. I then test if the results are comparable for younger and older workers and split the sample into those younger and older than 45. The results do not change for overqualification, while the density estimate for individuals younger than 45 in the horizontal mismatch regression is similar but not statistically significant. This shows again the robustness of the results for overqualification, while the outcomes for horizontal mismatch are much less precise and not always statistically significant.

While one would ideally use the density of the place of work, in this paper only the place of residence is observed. I am confident that the vast majority of individuals live and work in the same labor market region, as the definition of the regional level that I employ is based on commuting patterns. However, not observing the place of work can lead to a certain degree of measurement error. This is likely to lead to an attenuation bias, as for any type of measurement error in the explanatory variable. In the present setting, there are also reasons to believe that the employment density in larger cities is underestimated, since individuals are more likely to commute to denser regions. The attenuation bias could thus get even larger. While there is no information on the place of work, there is information on commuting distances for a sub-sample of the population. To minimize the error, I drop those individuals who commute long distances and are thus likely to work in a different region from the one of residence. However, one needs to bear in mind that individuals in less dense areas are more likely to commute, even within the regional labor market, and commuters are more likely to have a better match. So, if commuters within the regional labor market are dropped, one could actually induce a further attenuation bias. When excluding individuals commuting more than 50 km we get results that are about 10% larger in the overqualification equation and very similar in the horizontal mismatch equation (see Table A.3). When dropping all individuals commuting more than 25 km the employment density coefficient gets smaller than the baseline both for horizontal mismatch and overqualification, suggesting that the attenuation bias gets larger when restricting the sample to individuals commuting short distances. These results suggest that the main estimates suffer from a certain degree of attenuation bias. Nevertheless, it is very difficult to estimate the bias precisely with the data at hand.

In this paper I use employment density at the regional labor market level (258 regions) to measure agglomeration. However, the main results are very similar when employing population density as an agglomeration measure (see Table A.4). The baseline results are also qualitatively very similar if I employ a broader definition of local labor markets (150 regions) or the finer administrative definition of 402 districts. Finally, even a binary variable distinguishing urban from rural districts leads to negative and significant results.[13] The likelihood to be overqualified is about 20% lower in urban regions, while horizontal mismatch is about 11% lower.

3.6 Heterogeneous Effects by Education Level

So far I have estimated the impact of employment density on qualification mismatch without distinguishing among individuals with a different highest qualification. However, denser regions have typically a higher share of high-skilled individuals and if the qualification mismatch measures differ across individuals with different levels of education this is likely to lead to biased results. Moreover, there is evidence that urban wage premium increases over the wage distribution (Matano and Naticchioni, 2012) and is larger for white-collar workers (Gould, 2007). It is thus interesting to analyze whether the effect of agglomeration on better matches differs between tertiary graduates and individuals with a vocational degree. To analyze heterogeneous effects by qualification level I first add an interaction term to the baseline regression and then estimate separate regressions for individuals with a tertiary education degree and for those who have a vocational degree as their highest qualification.

Better educated individuals are more likely to be overqualified in our sample (see Figure A.1 for more details). Conversely, the incidence of horizontal mismatch differs substantially across qualifications and is much higher for individuals with a vocational degree compared to university graduates. Even if I controlled for the highest degree obtained in the baseline specifications, it is very important to make sure that the results obtained are not biased from the different composition of qualified labor across regions.

To better address these compositional issues I first add an interaction term between employment density and the highest degree obtained in the baseline model for the whole sample considered. Columns (1) and (2) of Table 5 show the results of this estimation

[13] The urban variable is based on the definition of the Federal Office for Building and Regional Planning.

without and with the inclusion of occupation fixed effects. For simplicity, I include only the interaction between employment density and vocational qualification, so that I can interpret the employment density coefficient as the effect for tertiary graduates. The regressions for overqualification (panel A) show estimates for tertiary graduates that are similar but slightly higher than the baseline estimates for the full sample. Conversely, the estimations for horizontal mismatch show a zero effect of density on horizontal mismatch for tertiary graduates. Thus, if there is a significant impact of agglomeration on horizontal mismatch, this seems to be only present for individuals with a vocational degree.

Table 5: Impact of Employment Density on Qualification Mismatch by Qualification Level

	All degrees		Vocational degree		Tertiary degree	
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Overqualification			
Empl. density (log.)	-0.025***	-0.025***	-0.020***	-0.012***	-0.020**	-0.014*
	(0.010)	(0.009)	(0.005)	(0.005)	(0.010)	(0.008)
Vocational degree	-0.291***	-0.466***				
	(0.061)	(0.055)				
Vocational degree ×	0.008	0.019*				
Empl. density	(0.011)	(0.010)				
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	Yes	No	Yes	No	Yes
Observations	35,363	35,363	24,277	24,277	11,290	11,290
			Panel B: Horizontal mismatch			
Empl. density (log.)	-0.012	-0.011	-0.019**	-0.015**	-0.011	-0.009
	(0.010)	(0.009)	(0.008)	(0.007)	(0.010)	(0.010)
Vocational degree	0.119*	0.073				
	(0.066)	(0.065)				
Vocational degree ×	-0.006	-0.003				
Empl. density	(0.012)	(0.012)				
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	Yes	No	Yes	No	Yes
Observations	35,363	35,363	24,277	24,140	11,223	11,223

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variables. Standard errors are clustered at the individual level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).

The results of the separate regressions on the sub-sample of tertiary graduates and on that of individuals with a vocational degree as their highest qualification are fairly similar to the model with interaction terms. However, as regards overqualification, the estimates are now very similar for the two education groups. As regards horizontal mismatch, the coefficient for individuals with a vocational degree is larger than the baseline estimates and statistically significant at standard confidence levels. For tertiary graduates, again there does not seem to be any difference in the horizontal mismatch between smaller and larger cities. This could also point to the fact that this type of measure is not always a "real" job mismatch for university graduates. In fact, even if graduates that are horizontally mismatched earn on average less than better matched graduates, some individuals in highly remunerated jobs also report being mismatched with respect to their field of study.

4 Determinants of the Qualification Mismatch Differential across Regions

So far I have established that thick labor markets reduce the probability of workers being overqualified for their job. In this section I investigate the channels that contribute to the mismatch differential across cities. More precisely, I am interested in highlighting those characteristics of denser regions (apart from a pure market size effect) that contribute to better average job matches. I did not include those characteristics in the previous chapters, because I consider these to be outcomes or intrinsic characteristics of larger cities. However, from a theoretical perspective it is very important to try to disentangle agglomeration economies and localization economies, as well as to separate the agglomeration economies due to better matches from those due to knowledge spillovers.

Denser regions typically have a higher proportion of high-skilled individuals. On the one hand, one would like to exclude the effects of skills from agglomeration economies, as far as this represents a pure composition effect (Combes and Gobillon, 2015). High-skilled individuals might be over-represented in cities, because they value city amenities more or because of historical migration of high-skilled individuals (with their skills being partly transmitted to their children). On the other hand, people could become more skilled through living in cities, tanks to stronger learning effects in denser regions. Faster learning and knowledge diffusion is indeed one of the main mechanisms of agglomeration economies (De La Roca and Puga, 2017). In our setting, it is interesting to analyze the qualification mismatch differential across regions while keeping the regional skill composition fixed. Column (2) in Table 6 shows the results of our baseline regression augmented with the regional share of tertiary educated individuals in the workforce. This variable has often been used in the literature to account for knowledge spillovers (Moretti, 2004). A higher share of high-skilled workers is associated with a lower risk of overqualification (but the coefficient is not statistically significant), probably because of a greater availability of high-skilled jobs. Controlling for skill composition, the employment density coefficient drops (in absolute value) to -0.012 but remains statistically significant at the 90% confidence level. Conversely, the high-skill share coefficient is positive and not statistically significant in the horizontal mismatch equation, so that the employment density estimate increases slightly in absolute value.

Previous studies have often included an index of industrial concentration or diversity in order to isolate the effect of agglomeration economies from urban specialization (localization economies) (see, for instance, D'Costa and Overman, 2014). I thus include an index of industrial concentration of the region (Herfindahl-Hirschman-Index, HHI) based on the regional share of employment in 7 major industries in column (3) of Table 6. The coefficient of the industrial concentration turns out to be negative in both equations but is not statistically significant. The effect of employment density on qualification mismatch appears thus to be robust to the inclusion of this variable.

Firm size has been found to be an important determinant of the urban wage gap in Germany (Lehmer and Möller, 2010). In column (4) I also add two dummies for firm size to see if the lower mismatch incidence in denser regions can be partly explained by the presence of larger firms. The firm size coefficients turn out not to be significant in the overqualification equation and the main results are not affected. On the contrary, horizontal mismatch turns out to be more common in larger firms. Since firms are on average larger in thicker labor markets, the employment density coefficient increases (in absolute value) in the horizontal mismatch equation.

Finally, column (5) presents the results of a regression in which all discussed determinants are included. The coefficient of the variables included do not lose their magnitude suggesting that they affect the two mismatch variables through different channels. The coefficient of horizontal mismatch does not change substantially compared to the baseline equation, suggesting that it can be interpreted as a pure labor market size effect. Conver-

Table 6: Determinants of the Qualification Mismatch Differential

	(1)	(2)	(3)	(4)	(5)
			Panel A: Overqualification		
Empl. density (log.)	-0.019***	-0.012*	-0.018***	-0.019***	-0.011*
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)
High-skilled share		-0.005*			-0.005**
		(0.002)			(0.002)
HHI industry			-0.324		-0.386
			(0.268)		(0.270)
Large firm (>200 empl.)				0.005	0.006
				(0.010)	(0.010)
Small firm (<=20 empl.)				-0.008	-0.007
				(0.012)	(0.012)
			Panel B: Horizontal mismatch		
Empl. density (log.)	-0.015**	-0.018**	-0.014**	-0.018***	-0.019**
	(0.006)	(0.008)	(0.006)	(0.006)	(0.008)
High-skilled share		0.002			0.001
		(0.003)			(0.003)
HHI industry			-0.439		-0.442
			(0.353)		(0.353)
Large firm (>200 empl.)				0.046***	0.046***
				(0.012)	(0.012)
Small firm (<=20 empl.)				-0.086***	-0.086***
				(0.016)	(0.016)
Control variables	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No	No
Observations	35,363	35,363	35,363	35,363	35,363

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variables. Standard errors are clustered at the individual level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).

sely, the overqualification effect drops by almost one half when the regional high-skilled share is included. Part of the agglomeration effect found in the baseline is thus associated with a different skill composition in denser regions.

5 Qualification Mismatch and the Urban Wage Premium

I will now investigate the importance of qualification mismatch as a mechanism of agglomeration economies. More precisely, I analyze what portion of the effect of regional employment density on earnings can be explained by better job matches with respect to workers' qualifications. To do so, I first estimate an OLS regression with hourly wages as the dependent variable, the regional employment density as the variable of interest and the full set of control variables presented in the previous sections.I then add to this regression our measures of qualification mismatch and look at how the coefficient for employment density is affected. Since I found a relatively large effect of employment density on overqualification and the overeducation literature documents a strong negative relationship between overqualification and wages, I expect that a large part of the effect of regional employment density on wages can be explained by a lower probability of being overqualified.

Table 7 shows the results of these regressions, where both employment density and hourly wages are expressed in logarithmic form. The coefficient of employment density in column (1) is equal to 0.049 and is significant at the 1% significance level. Doubling the number of employed workers per square kilometer is associated with an increase in wages of about 5%. This result appears to be in line with previous studies which estimate the magnitude of urban agglomeration economies as ranging from 0.02 to 0.07 (Combes and Gobillon, 2015). In column (2) and (3) I add our measure of overqualification and horizontal mismatch separately. Consistent with the literature, vertical and horizontal mismatch are both associated with lower wages. As expected the coefficient of employment density decreases in both specifications. However, the decrease is relatively small. According to the estimations, overqualification accounts for about 6% of the urban wage premium while horizontal mismatch accounts for less then 2%. In column (4) both mismatch measures are added and it becomes clear that they are positively correlated, since their coefficients decrease significantly. Their impact on the wage premium does not appear to add up, suggesting that they explain slightly more than 6% of the effect of employment density on wages. In column (5) I also add an interaction term between employment density and the mismatch variables. The results show that the regional difference in wages are significantly smaller for overqualified workers. Doubling employment density is associated with about 3% higher wages for the overqualified compared to 5% higher wages for those that are well-matched according to their qualifications.

Table 7: Impact of Employment Density and Mismatch on Log Hourly Wages

	(1)	(2)	(3)	(4)	(5)
Empl. density (log.)	0.049***	0.046***	0.048***	0.046***	0.049***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
Overqualification		-0.156***		-0.154***	-0.062
		(0.009)		(0.009)	(0.048)
Horizontal mis.			-0.053***	-0.004	-0.002
			(0.008)	(0.008)	(0.040)
Overqualification ×					-0.019*
Empl. density					(0.010)
Horizontal mis. ×					-0.000
Empl. density					(0.008)
Control variables	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No	No
Observations	34,204	34,204	34,204	34,204	34,204
R-squared	0.512	0.530	0.515	0.530	0.530

Note: Standard errors are clustered at the individual level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).

The wage results are consistent with previous studies using different measures of skill mismatch, which found that it only accounts for about 5-8% (Abel and Deitz, 2015) of the urban wage premium or has almost no contribution at all (Boualam, 2014). Overqualification seems to be the most important channel here, while horizontal mismatch does not seem to add much to this. Furthermore, knowing that less talented individuals are more likely to be overqualified (Leuven and Oosterbeek, 2011), part of this explained effect might actually denote unobserved ability. Indeed, due to spatial sorting, controlling for ability is expected to lead to a decrease in the coefficient of the urban wage premium and the overqualification dummy might to some extent proxy unobserved ability.

6 Conclusion

The aim of this paper is to measure the effect of local labor market size on vertical and horizontal qualification mismatch. Estimating a linear probability model with an extensive set of control variables, I find that German male workers who live in denser regions are less likely to be overqualified and to work in a different field than that of their education or training. The impact on overqualification is robust to the inclusion of an extensive set of control variables (including school grades, personality traits and risk preference) and is relatively large. An increase of 10% in the regional employment density is associated with a decrease of 1-1.5% in the overqualification incidence. The impact of horizontal mismatch is slightly smaller and less precise (not statistically significant in some specifications). I then follow three empirical strategies to deal with the fact that talented workers might sort into larger cities. First, by restricting the sample to individuals that remain in the area they grew up in, I get a smaller but still sizable estimate of the effect of employment density on overqualification and a similar estimate for horizontal mismatch. Second, by exploiting the panel structure of the data and accounting for individual fixed effects, I get a coefficient for overqualification that is less precise, but very similar in size to the baseline regressions. Conversely, the estimate in the horizontal mismatch regression is not statistically different from zero. Third, I address potential reverse causality by instrumenting current employment density with historical population data. The IV regression estimates are very similar to the baseline and fixed effects regression results.

While differences in the regional skill composition account for a large portion of the match differential across regions, most of the impact found seems to be attributable to a pure labor market size effect. Finally, I investigate the extent to which lower qualification mismatch in large agglomerations contributes to the urban wage premium. I find that overqualification explains about 6% of the impact of regional employment density on hourly wages, while the contribution of horizontal mismatch appears to be insignificant.

References

Abel, J. R. and Deitz, R. (2015). Agglomeration and Job Matching among College Graduates. *Regional Science and Urban Economics*, 51:14–24.

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.

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.

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.

Bosquet, C. and Overman, H. G. (2016). Why does Birthplace Matter so Much? Sorting, Learning and Geography. SERC Discussion Papers, 190. Spatial Economics Research Centre, London, UK.

Boualam, B. (2014). Getting a First Job: Quality of the Labor Matching in French Cities. Mimeo, University of Geneva.

Büchel, F. and van Ham, M. (2003). Overeducation, Regional Labor Markets, and Spatial Flexibility. *Journal of Urban Economics*, 53(3):482–493.

Ciccone, A. and Hall, R. E. (1996). Productivity and the Density of Economic Activity. *American Economic Review*, 86(1):54–70.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial Wage Disparities: Sorting Matters! *Journal of Urban Economics*, 63(2):723–742.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. In Duranton, G., Henderson, V., and Strange, W., editors, *Handbook of Regional and Urban Economics*, volume 5, chapter 5, pages 247–341. Elsevier.

Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2016). Spatial Wage Disparities–Workers, Firms, and Assortative Matching. Mimeo, Heinrich Heine University Düsseldorf.

D'Costa, S. and Overman, H. G. (2014). The Urban Wage Growth Premium: Sorting or Learning? . *Regional Science and Urban Economics*, 48(0):168 – 179.

De La Roca, J. and Puga, D. (2017). Learning by Working in Big Cities. *Review of Economic Studies*, 84(1):106–142.

Duranton, G. and Puga, D. (2004). Micro-foundations of Urban Agglomeration Economies. In Henderson, J. V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 48, pages 2063–2117. Elsevier.

Gerlitz, J.-Y. and Schupp, J. (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. DIW Research Notes 4, DIW Berlin.

Glaeser, E. L. and Mare, D. C. (2001). Cities and Skills. *Journal of Labor Economics*, 19(2):316–42.

Gould, E. D. (2007). Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium. *Review of Economic Studies*, 74(2):477–506.

Hartog, J. (2000). Over-Education and Earnings: Where Are We, Where Should We Go? *Economics of Education Review*, 19(2):131–147.

Heuermann, D., Halfdanarson, B., and Suedekum, J. (2010). Human Capital Externalities and the Urban Wage Premium: Two Literatures and their Interrelations. *Urban Studies*, 47(4):749–767.

Jauhiainen, S. (2011). Overeducation in the Finnish Regional Labour Markets. *Papers in Regional Science*, 90(3):573–588.

Kampelmann, S. and Rycx, F. (2012). The Impact of Educational Mismatch on Firm Productivity: Evidence from Linked Panel Data. *Economics of Education Review*, 31(6):918–931.

Kroh, M. (2012). Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2011). DIW Data Documentation 66, DIW Berlin.

Lehmer, F. and Möller, J. (2010). Interrelations between the Urban Wage Premium and Firm-size Wage Differentials: a Microdata Cohort Analysis for Germany. *Annals of Regional Science*, 45(1):31–53.

Leuven, E. and Oosterbeek, H. (2011). Overeducation and Mismatch in the Labor Market. In Hanushek, A. E., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 4, pages 283–326. Elsevier.

Matano, A. and Naticchioni, P. (2012). Wage Distribution and the Spatial Sorting of Workers. *Journal of Economic Geography*, 12:379–408.

Moretti, E. (2004). Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-sectional Data. *Journal of Econometrics*, 121(1):175–212.

Puga, D. (2010). The Magnitude and Causes of Agglomeration Economies. *Journal of Regional Science*, 50(1):203–219.

Robst, J. (2007). Education and Job Match: The Relatedness of College Major and Work. *Economics of Education Review*, 26(4):397–407.

A Further Tables and Figures

Figure A.1: Qualification Mismatch by Highest Degree

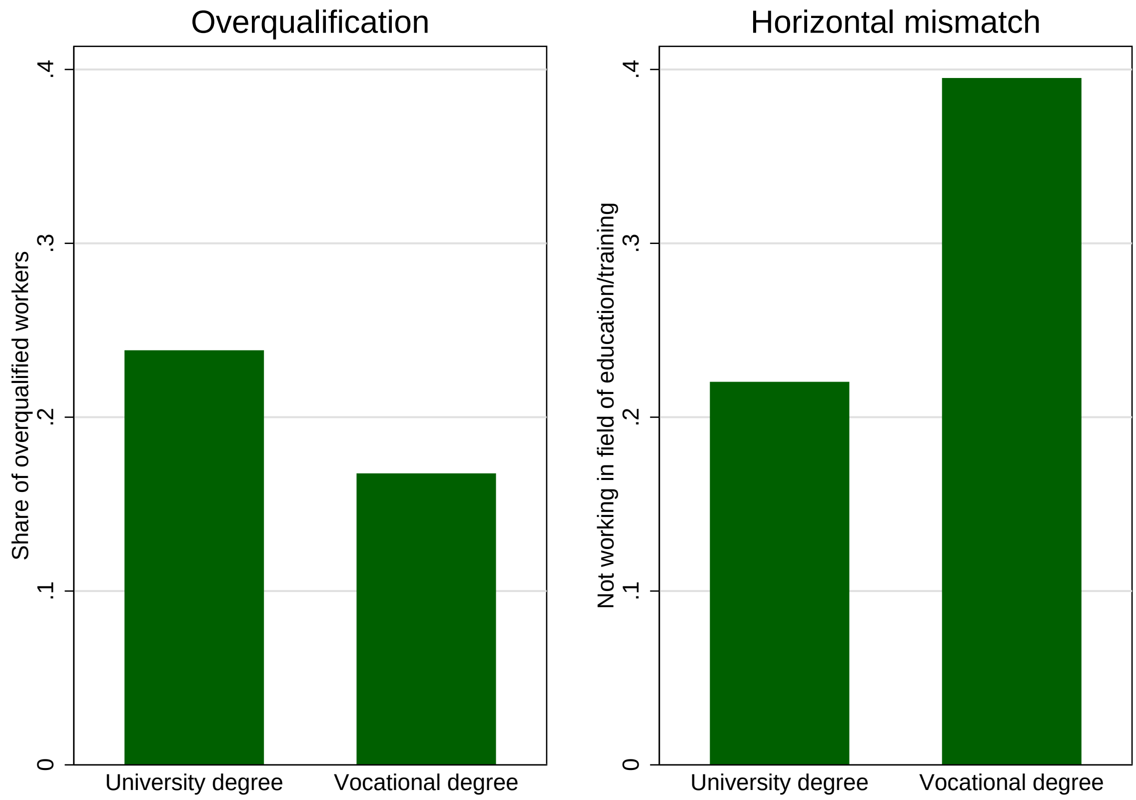


Figure A.2: Employment Density in German Labor Market Regions (in 2010)

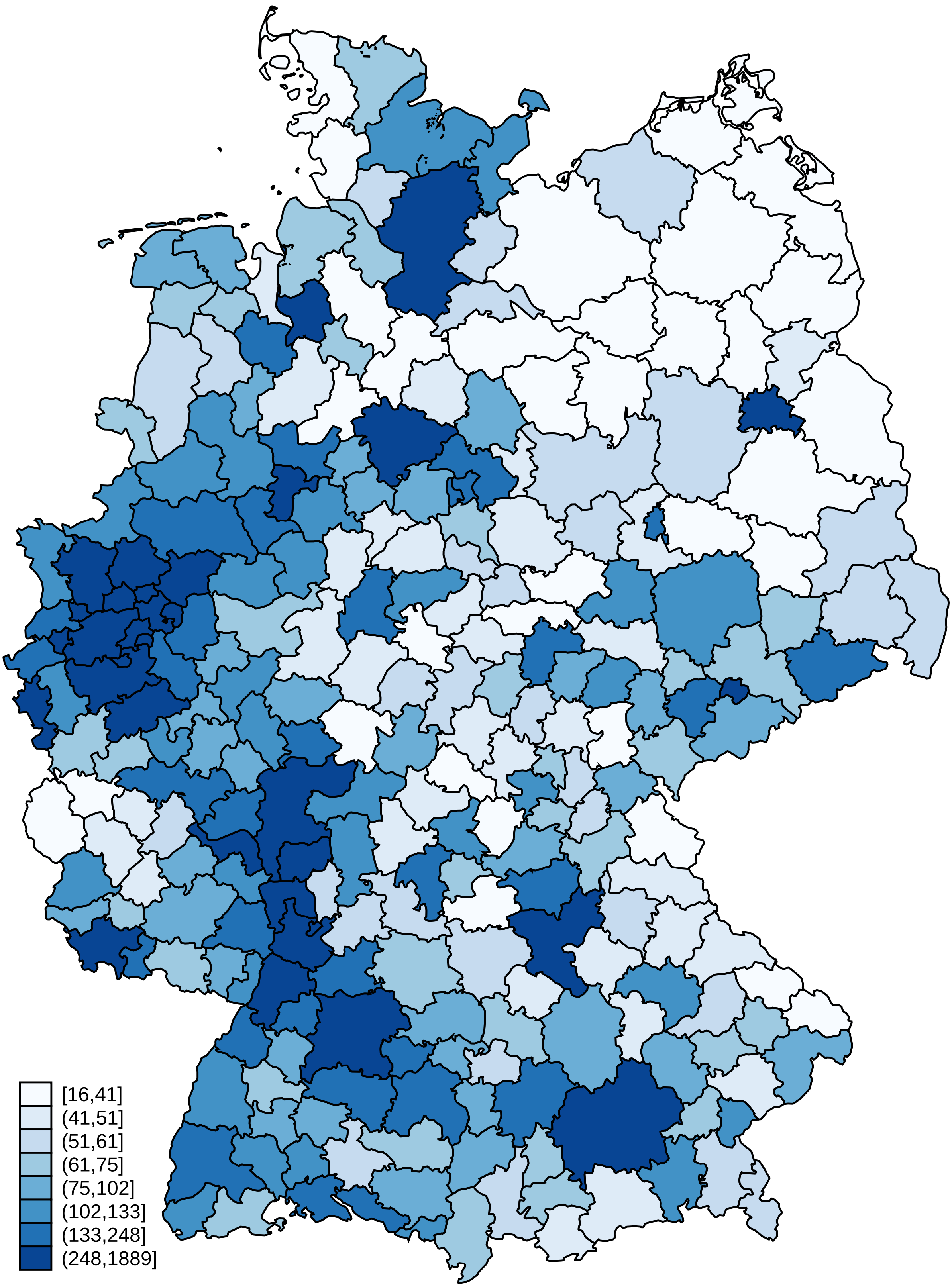


Table A.1: Impact of Employment Density on Overqualification: Further Controls

	Overqualification				Horizontal mismatch			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Empl. density (log.)	-0.020***	-0.022***	-0.020**	-0.020**	-0.012	-0.013	-0.013	-0.013
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
School grade: Math		-0.013*	-0.013*	-0.013*		-0.006	-0.004	-0.004
		(0.008)	(0.008)	(0.008)		(0.010)	(0.010)	(0.010)
School grade: German		-0.019**	-0.017*	-0.017*		-0.018	-0.019	-0.018
		(0.010)	(0.009)	(0.009)		(0.012)	(0.012)	(0.012)
Extraversion			0.002	0.002			0.003	0.003
			(0.002)	(0.003)			(0.003)	(0.003)
Consciousness			0.003	0.003			-0.003	-0.003
			(0.003)	(0.003)			(0.004)	(0.004)
Agreableness			0.003	0.003			0.007*	0.007**
			(0.003)	(0.003)			(0.004)	(0.004)
Neuroticism			0.005**	0.005**			0.008***	0.008***
			(0.002)	(0.002)			(0.003)	(0.003)
Openness to experience			-0.005*	-0.005*			-0.001	-0.001
			(0.002)	(0.003)			(0.003)	(0.003)
Risk aversion				-0.000				0.005
				(0.005)				(0.006)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No	No	No	No	No
Observations	14,814	14,814	14,814	14,814	14,814	14,814	14,814	14,814
R-squared	0.092	0.095	0.098	0.098	0.094	0.095	0.099	0.99

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variables. Standard errors are clustered at regional labor market level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).

Table A.2: Excluding Specific Regions or Age Groups

	Excluding regions		Excluding age groups	
	only west	excl. Berlin	45 or younger	over 45
	(1)	(2)	(3)	(4)
		Overqualification		
Empl. density (log.)	-0.019**	-0.018***	-0.018***	-0.020***
	(-0.007)	(0.006)	(0.006)	(0.007)
		Horizontal mismatch		
Empl. density (log.)	-0.013	-0.011	-0.013*	-0.016*
	(0.008)	(0.008)	(0.008)	(0.009)
Control variables	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No
Observations	25,780	34,215	18,449	16,914

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variable. Standard errors are clustered at the individual level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).

Table A.3: Excluding Long-Distance Commuters

	Commuters excluded			
	None	>50 km	>30 km	> 20 km
	(1)	(2)	(3)	(4)
		Overqualification		
Empl. density (log.)	-0.020***	-0.022***	-0.021***	-0.018***
	(0.005)	(0.005)	(0.006)	(0.006)
		Horizontal mismatch		
Empl. density (log.)	-0.016**	-0.017**	-0.014**	-0.012*
	(0.006)	(0.007)	(0.007)	(0.007)
Control variables	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No
Observations	33,313	30,406	26,947	22,688

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variables. Standard errors are clustered at the individual level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).

Table A.4: Other Agglomeration Measures

	258 regions	150 regions	402 districts	
	(1)	(2)	(3)	(4)
	Overqualification			
Pop. Density (log.)	-0.019***			
	(0.005)			
Emp. Density (log.)		-0.020***	-0.014***	
		(0.006)	(0.003)	
Urban region				-0.040***
				(0.010)
	Horizontal mismatch			
Pop. Density (log.)	-0.016**			
	(0.007)			
Emp. Density (log.)		-0.023***	-0.015***	
		(0.008)	(0.008)	
Urban region				-0.034***
				(0.013)
Control variables	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No
Observations	35,363	35,363	35,363	35,363

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variables. Standard errors are clustered at the individual level; *** p<0.01, ** p<0.05, * p<0.1. Control variables included are the main control variables (highest degree, migration background, marital status and children, experience, experience squared and year fixed effects), school degree, parental background (higher education of mother/father and working status of mother), geographic characteristics (macro-regions and whether living in city of childhood) and job characteristics (tenure, tenure squared, industry and public sector dummies).