Local labour market size and qualification mismatch*

Francesco Berlingieri†

ZEW Centre for European Economic Research, Mannheim

March 18, 2015

Preliminary draft - Please do not cite or circulate

Abstract

This paper investigates the effect of the size of the local labour market on skill mismatch. Using survey data for Germany, we find that male 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, better job matching in larger cities seems to explain only a small part of the urban wage premium.

JEL-classification: I21, J24, J31, R23

Keywords: agglomeration, qualification mismatch, urban wage premium

*This paper was written as part of the project "Adequate employment of graduates: An economic analysis of job match quality" (ADÄQUAT) supported within the programme "Economics of Science" by the German Ministry of Education and Research (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 a large amount of evidence that workers earn higher wages in larger labour markets. For instance, descriptive estimates using a comparable definition of urban areas range from about 33% in the US (Glaeser and Mare, 2001) to about 14% in Germany (Lehmer and Möller, 2010). From an individual perspective the higher costs of living in cities might explain why not all workers are willing to move to large agglomerations. However, the urban wage premium must reflect a higher productivity in larger cities to explain why firms do not relocate to less urbanized areas. Duranton and Puga (2004) distinguish between three mechanisms behind the higher productivity of larger cities (i.e. agglomeration economies): the sharing of facilities and risks, faster learning and knowledge diffusion and better matches between firms and workers. While the importance of the latter source of agglomeration economy is stressed from a theoretical side, there is little evidence about its empirical relevance (Puga, 2010). This is also because of the difficulties of measuring the 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) finding some evidence of better matches in more urbanized areas. The focus in this paper is on the match between the formal qualifications earned by workers and the job requirements. We look at the match between possessed and required qualifications both in terms of level (vertical match) and in terms of content (horizontal match), since there are reasons to expect both to be better in thicker labour markets.

The question whether workers in more urban areas are less exposed to educational mismatch is also interesting by itself and relevant for the labour economics literature on skill mismatch. Does it actually pay off for individuals to move to larger agglomerations in terms of better job matches and for their future careers? Previous studies have already investigated the impact of various regional labour market characteristics including the size of regional labour markets and regional unemployment rates, as well as of individual restrictions to mobility, on overqualification (Buchel 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 labour market and qualification mismatch. Simple regressions with standard control variables might lead to biased estimates in this context, because more talented individuals are both more likely to live in large cities and to have a better job match. Several papers have stressed the importance of addressing spatial sorting of workers by individual skills for estimating the urban wage premium (Glaeser and Mare, 2001; Combes et al., 2008). Since there are reasons to believe that the sorting of workers across areas could lead to an overestimation of the effect of larger agglomerations on the job match, we first try to mitigate the problem by estimating linear regressions including an extensive set of control variables, such as information on parental background, school grades and personality traits. We then corroborate these estimates by employing two empirical strategies that have been used in previous studies on the urban wage premium. One the one hand, by restricting the sample to individuals that remain in the region where they grew up (non-movers) we can avoid biases from the direct migration of more talented workers into cities. On the other hand, by estimating a fixed effects model on our panel of workers and obtaining identification through individuals migrating form one region to another, we can get rid of unobserved time-invariant heterogeneity (such as individual ability).

Similarly to previous studies (for a review of the literature on agglomeration economies see Combes and Gobillon, 2014 and Heuermann et al., 2010), we use the regional employment density to measure the labour market size. We obtain estimates of employment density on overqualification that 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 to be overqualified. On the contrary, most of the estimates of employment density on the horizontal mismatch measure are smaller and not statistically significant. Finally, we investigate the contribution of better qualification matches in explaining the wage premium in thicker labour markets. By including our mismatch measures to an OLS regression of log hourly wages on employment density (and other control variables), we find that overqualification explains only 8% 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).[1] Abel and Deitz (2015) find evidence of a moderate effect of population size and employment density on

[1] Also Andini et al. (2013) 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.

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 labour market entrants. While this measure of match quality is found to increase with employment density, it does not seem to explain difference in wages between thick and thin labour markets. The present paper has at least three different features. First, the survey data we use (i.e. the German Socio-Economic Panel) contains direct questions on the qualification required by the job, so that we can construct vertical and horizontal qualification mismatch variables based on workers' self-assessments. Second, the data has extensive information on individual characteristics and their biography 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 panel structure of the data enables us to estimate fixed effects regressions to get rid of the unobserved ability bias analogously to previous studies on the urban wage premium (Glaeser and Mare, 2001; D'Costa and Overman, 2014).

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 overqualification and horizontal mismatch. While in section 4 we attempt to disentangle the effect of labor market size from that of other characteristics of larger agglomerations (such as specialization and the skill structure) on the mismatch incidence, in section 5 we investigate the contribution of qualification mismatch on 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). We focus on males surveyed in the years 2000 to 2011.[2] The sample is further restricted to dependent workers employed full-time. The 12 GSOEP waves include 12,700 male adults aged between 20 and 65 with a university degree or a completed training that are employed in one of the 11 waves. For the baseline analysis, we select one observation per individual such that the time from graduation is minimized, but is at least 2 years. We end up with a sample of 4,281 individuals, for whom we have information on all variables relevant for our analysis.

The literature has used different measures for agglomeration or labour market size. Following other studies we employ the regional employment density measured at the level of *Raumordnungsregionen* (ROR).[3] This is calculated by the number of employed individuals per square kilometer. There are 96 ROR regions in Germany with an average of 6 regions for each of the 16 federal states.[4] ROR 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. Information on employment density, as well as on the unemployment rate, at the ROR level is gathered from administrative data sources (i.e. the INKAR database) and merged to GSOEP data. Ideally, we would consider the effect of workplace location, but unfortunately only the residence location is available in the GSOEP.

We 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 requirement 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?" We consider an individual to be overqualified if he reports that his job requires a lower degree than the one possessed.[5] The measure, which is widespread in the overeducation literature, has the drawback of relying on the subjective individual self–assessment. Nevertheless, several authors have claimed that the

[2]We restrict the analysis to male graduates, since female labour market participation in Germany is strongly influenced by child care and family responsibilities. The investigation of females therefore requires a different econometric approach that takes into account selection out of the labour market. The extension of the results to include women is on the agenda, but the first results appear to be relatively similar to the ones for males.

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

[4]We plan to augment the results by using more precise regional information. Preliminary results using two different classifications of local labour market regions (with 150 and 258 regions) show baseline estimates that are of similar magnitude, but more precise.

[5]Note that we do not distinguish between university and university of applied science (*Fachhochschule*) degrees, although the variable allows such a distinction.

measurement errors are probably less severe for this measure than for measures based on the distribution of educational qualification within occupations – i.e. "realized matches" on the qualification required by the job. This is because the latter is the result of demand and supply forces and it ignores variation in required schooling across jobs within an occupation (Leuven and Oosterbeek, 2011). Horizontal mismatch is also self-reported. The question asked in the GSOEP is: "does the job fit to your higher education or training?". Since the only possible answers are yes or no, we construct a dummy that is equal to 1 if individuals answer negatively to this question.

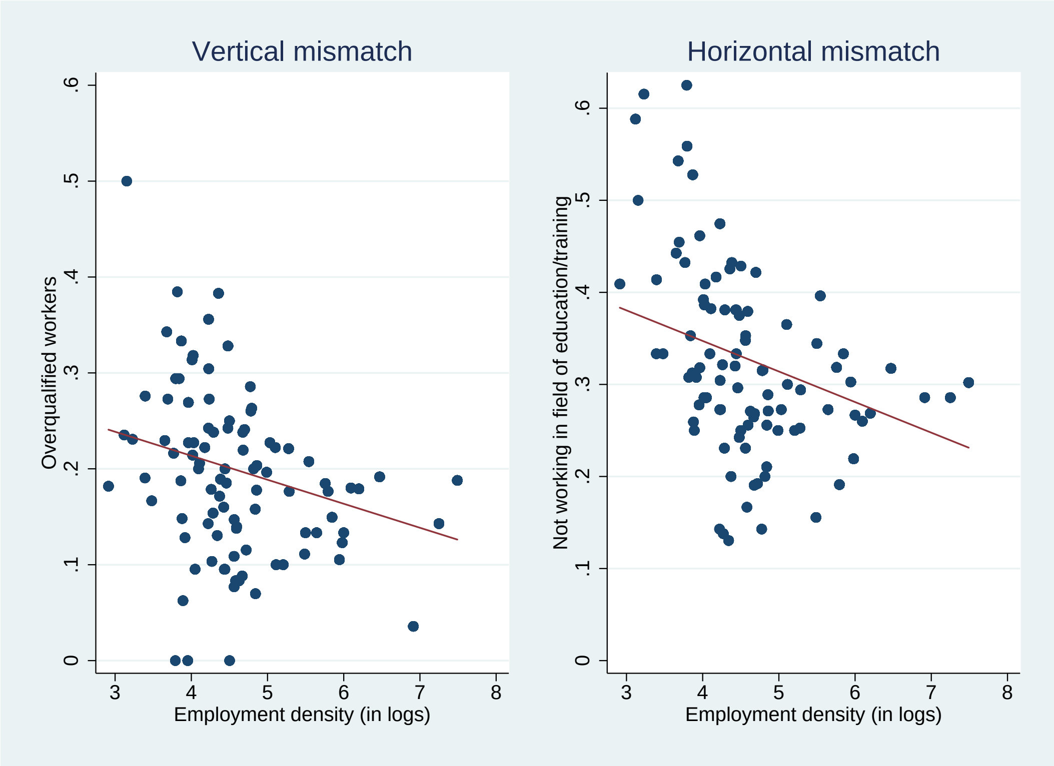
Hourly wages are measured through the self-reported monthly gross income divided by monthly working hours. We calculate real wages based on the CPI deflator using 2010 as the base year. In order to ensure that outliers are not driving the main results we trim wages excluding the 1st and the 99th percentile (individuals receiving a hourly wage lower than EUR 4 or higher than EUR 75) and we employ the standard logarithmic form for the wage regressions.

2.2 Descriptive results

Table A.1 presents the mean and standard deviation for the variables included in the analysis. The overqualification incidence is of about 15% in the sample, while the incidence of horizontal mismatch amounts to 30%. Figure A.2 shows the differences in employment density across the 96 German ROR regions. Darker colors depict a higher employment density, which ranges from 18 employed individuals per square km in Altmark (Sachsen-Anhalt) to 1871 in Berlin.

Figure 1 shows that 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 between the job and the field of education or training (horizontal mismatch). The unit of observation in both graphs is the ROR region, meaning that the information on the individual match is aggregated at the regional level. The slope of the fitted regression line is of -0.025 for vertical mismatch and -0.033 for horizontal mismatch and the coefficients are statistically significant at standard levels for both regressions.

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 of the region of residence and qualification mismatch, we first want to test whether the results change when we include an extensive set of control variables. We thus estimate a the following simple linear probability model[6]:

$$Pr(mismatch_{ij} = 1) = \alpha + \beta \ empdensity_j + \gamma \mathbf{X_{ij}} + \epsilon_{ij} \tag{1}$$

where *mismatch* is a dummy variable that takes value 1 in case of a qualification mismatch for individual i, *empdensity* denotes the employment density of the region of residence j and $\mathbf{X_{ij}}$ is a vector of covariates that differs across specifications. Panel A of Table 1 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 main control variables only (i.e. highest educational qualification, migration background, marital status,

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

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 at age 15), 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 1: Impact of employment density on qualification mismatch

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Overqualification			
Empl. Density (log.)	-0.027***	-0.022***	-0.021***	-0.019***	-0.021***	-0.015***
	(0.008)	(0.008)	(0.008)	(0.005)	(0.005)	(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	4,281	4,281	4,281	4,281	4,281	4,281
R-squared	0.013	0.030	0.034	0.040	0.048	0.168
			Panel B: Horizontal mismatch			
Empl. Density (log.)	-0.020**	-0.017*	-0.015*	-0.010	-0.013*	-0.010
	(0.008)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)
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	4,281	4,281	4,281	4,281	4,281	4,281
R-squared	0.053	0.071	0.073	0.082	0.090	0.181

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variable. Standard errors are clustered at ROR 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 dummy).

Column (1) in Panel A shows the existence of a negative relationship between regional employment density and the probability to be overqualified for the job when standard control variables are included. The coefficient is equal to -0.027 and is significant at the 1% significance level. A 10% increase in employment density would imply a decrease of about 0.27 percentage points in the overqualification probability, which is equal to a decrease of about 1.5% (given that the overqualification rate in our sample is 19%). The employment density coefficient decreases to -0.022 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 (-0.015), but is still statistically significant. While the ISCO classification at the 1-digit level is relatively broad, its inclusion together with the information about the educational qualification is likely to partially capture vertical qualification mismatch. For overqualification it seems thus better to avoid to control for occupation fixed effects.

Panel B shows that regional employment density appears to have a negative impact also on horizontal mismatch, i.e. whether one works in the same field of one's education or training. The coefficient in column one is equal to -0.020 and is statistically significant. A 10% increase in employment density would imply a decrease of about 0.7% in horizontal mismatch (since the incidence of horizontal mismatch is about 30%). The coefficient decreases slightly when school degree dummies and parental background information are included. It becomes though equal to -0.013 (and significant at the 10% significance level) with the inclusion of geographic controls and job characteristics. In particular, a large part of the correlation between density and horizontal mismatch can be explained by differences between West and East Germany, which is both characterized on average by 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 0.010 and becomes insignificant. 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. To sum up, larger cities seem to have a relatively large impact on overqualification, while the impact on horizontal mismatch appears to be smaller and not robust to the inclusion of our extensive list of control variables.[7]

[7] Note that the lack of an effect on horizontal mismatch might be due to our specific self-reported measure. The extension of the analyses to measures of horizontal and vertical mismatch based on the distribution of qualifications within occupations is on the research agenda.

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 agglomerations. 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, we exclude these from the baseline regressions and present separate results for a sub-sample of 2141 individuals, for whom we have information about all characteristics.

Table A.2 presents results of a linear probability model of qualification mismatch by gradually adding mathematics and German grades from the last school report, standard measures of the big five personality traits (openness, consciousness, agreeableness, neuroticism and openness to experience) and a subjective measure of risk preference.[8] Column (1) and (5) of the table present the results of the same model of column (5) in table 1, where all baseline control variables are included except for occupation fixed effects. The employment density coefficients are slightly larger in absolute terms for both overqualification and horizontal mismatch compared to the baseline sample and are statistically significant. These estimates remain very similar when school grades, personality traits and risk preference are included (if anything they get larger in absolute value). In facts, while some characteristics matter for the qualification mismatch measures, they appear to be almost irrelevant for the impact of employment density on the match.

3.3 Addressing the omitted ability bias

We employ two empirical strategies to address the potential overestimation of the results due to omitted ability bias stemming from sorting of talented individuals to larger

[8] 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 we have information for both the big five and risk preference only on specific years, we compute the individual average of all observed values.

agglomerations. In the following we focus only on overqualification as a measure of mismatch, since the results for horizontal mismatch are not robust to the inclusion of all control variables. Similarly to Boualam (2014) we first investigate whether we find different results for the sub-samples of individuals ever moving from a district to another (movers) and the ones staying in the place where they grew up (non-movers). Focusing on non-movers enables to avoid biases from the direct migration of more talented workers to cities. However, it might be the case that talented individuals are more likely to rise in cities, because of inherited abilities by parents and grandparents that moved to large agglomerations (Glaeser and Mare, 2001). Column (1) of table 2 reports the results for the same linear regression estimated in the last column of table 1 (with the inclusion of all control variables apart from occupation fixed effects). The same model is then estimated on the sub-sample of individuals that did not move from the place they grew up, who represent about 57% of the sample.[9] The coefficient for the sub-sample of non-movers appears to be slightly smaller than the one for all individuals, but still statistically significant. If anything, this result suggests that a moderate sorting of high-ability individuals into cities is taking place. Nevertheless, larger labour markets seem to allow for better job matches also for non-movers.

Table 2: Impact of employment density on overqualification: addressing spatial sorting

	Cross-section		Panel data		
	All	Non-movers	Pooled OLS	Fixed effects	
	(1)	(2)	(3)	(4)	(5)
Empl. density (log.)	-0.022***	-0.020***	-0.017***	-0.020**	-0.025**
	(0.005)	(0.007)	(0.005)	(0.010)	(0.011)
Control variables	Yes	Yes	Yes	Yes[a]	Yes[a]
Occupation FE	No	No	No	No	No
Observations	4,281	2,444	36,140	36,140	35,970

Note: Standard errors are clustered at ROR 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), 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 dummy). [a] Only control variables that change over time are included.

In a second step, we exploit the panel structure of the data in order to control for

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

individual fixed effects. On the one hand, this enables to get rid of 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 agglomeration compared to moves to a smaller region. For this analysis we use an unbalanced panel of the male individuals of the previous estimations, who are observed as being employed at least twice in the data. Only 419 individuals change the ROR region of residence in our sample. 249 of these change both region and job. Since the identification will hinge upon those changing the region of residence and only job switchers can change the match status, we also estimate a regression excluding the spells in which individuals change region but not the job. In the regression we use the average regional employment density, so that we do not allow this to change across years. We 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 labour market. For simplicity, we estimate the following linear fixed effects model that gets rid of the time-constant unobserved individual heterogeneity:

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

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 average employment density of the region of residence in the period considered (2001-2011) and the vector $\mathbf{X_{1,ijt}}$ includes all control variables that may change across years with the exclusion of occupation fixed effects. These are part of the demographic characteristics, job and geographic characteristics.Column (3) of table 2 reports the results of a pooled OLS estimation including all spells. The coefficient (-0.017) turns out to be very similar to the estimate of the baseline model. Column (5) presents the results of the fixed effects model. The coefficient turns out to be significant and even larger than the coefficient of the pooled OLS (column 3). In column (5) we exclude the spells of those changing region but not job. As expected, the estimate in absolute value increases even further and is equal to -0.025.

Thus the fixed effects estimate turns out to be even larger than in the baseline LPM regression. 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 will then return 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 our results. Further investigations are needed to test whether this assumption is plausible.[10]

3.4 Heterogeneous effects by qualification level

So far we have estimated the impact of employment density on qualification mismatch without distinguishing among individuals with a different highest qualification. However, larger agglomerations have typically a higher share of high-skilled individuals and if the qualification mismatch measures differ across individuals with different education this is likely to lead to biased results. Moreover, it would be 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 we first add an interaction term to the baseline regression and then estimate separate regressions for individuals with a tertiary education degree and for those have a vocational degree as highest qualification.

Figure A.1 shows the incidence of qualification mismatch by the highest degree obtained. We distinguish among vocational degrees, university degrees and universities of applied sciences (FH) degrees, since the latter two are considered separately for the required qualification reported that is relevant for the overqualification measure. The figure shows that the overqualification incidence is fairly similar across degrees. However, better educated individuals are on average slightly more likely to be overqualified. Differently, the incidence of horizontal mismatch differs a lot across qualifications and is much higher for individuals with vocational education compared to university graduates. Even if we

[10]We are planning to carry out a more in depth analysis distinguishing between moves to larger agglomeration and moves to smaller agglomerations, as well as the reasons of such moves. In a second step we also wish to estimate the longer run career impacts in terms of job match of moving to a large or small region, since moving to a larger city might be related to a persistent better job match.

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 labour across regions.

To better address these compositional issues we 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 A.3 show the results of this estimation without and with the inclusion of occupation fixed effects. For simplicity, we include only the interaction between employment density and vocational qualification, so that we can interpret the employment density coefficient as the impact 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 education.

We get similar results for the separate regressions on the sub-sample of tertiary graduates and on that of individuals with a vocational degree as highest qualification. Concerning overqualification, the estimates are larger for vocational graduates when occupation fixed effects are excluded. However, the estimate for tertiary graduate appears also to be statistically significant (at the 90% confidence level) despite the small sample and becomes larger than the one for vocational graduates with the inclusion of occupation fixed effects. Concerning horizontal mismatch, the coefficient for individuals with vocational education are sizable and larger than the baseline estimates, but remain statistically insignificant also because of the bigger standard errors. For tertiary graduates, again there does not seem to be any difference in horizontal mismatch between smaller and larger agglomerations. This could also point to the fact that this type of measure is not always a “real” job mismatch for university graduates. In facts, even if graduates that are horizontally mismatch earn on average less then matched graduates, some individuals in highly remunerated jobs also report to be mismatched with respect to their field of study.

4 Determinants of the qualification mismatch differential across regions

So far we have established that thick labour markets reduce the probability for workers to be overqualified for the job. We wish now to investigate the channels that contribute to the mismatch differential across cities. More precisely, we are interested in highlighting those characteristics of larger agglomerations (apart from a pure market size effect) that are contributing to better average job matches. We did not include those characteristics in the previous chapters, because we consider these to be outcomes or intrinsic characteristics of larger agglomerations. However, form 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.

Larger agglomerations have typically 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, 2014). High skilled individuals might be over-represented in cities, because they value city amenites more or because of historical migration of high-skilled individuals (transmitting part of the skills to their children). On the other hand, people could be made more skilled by cities, through stronger learning effects in larger cities. Faster learning and knowledge diffusion is indeed one of the main mechanisms of agglomeration economies. In our setting, it might be tempting to investigate the qualification mismatch differential across regions while keeping the regional skill composition fixed. Column (2) in Table 3 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 related with a lower risk of overqualification (but the coefficient is not statistically significant), probably mostly because of a larger availability of high skilled jobs. Controlling for the skill composition, the employment density coefficient drops (in absolute value) to -0.014 but remains statistically significant at standard confidence levels.

To isolate the effect of agglomeration economies from urban specialization (localization economies) we include the regional share of employment in 7 major industries, as well as

Table 3: Determinants of the qualification mismatch differential

	(1)	(2)	(3)	(4)	(5)	(6)
Empl. density (log.)	-0.020***	-0.013**	-0.024**	-0.021***	-0.020***	-0.019*
	(0.006)	(0.006)	(0.009)	(0.006)	(0.006)	(0.010)
High-skilled share		-0.004				-0.007
		(0.002)				(0.004)
HHI industry			-0.788			-1.081
			(0.683)			(0.760)
Innovative indust. share				-0.003**		-0.005**
				(0.002)		(0.002)
Large firm (>200 empl.)					-0.002	-0.000
					(0.014)	(0.014)
Small firm (<=20 empl.)					-0.008	-0.009
					(0.018)	(0.018)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Industry shares	No	No	Yes	No	No	Yes
Occupation FE	No	No	No	No	No	No
Observations	3,974	3,974	3,974	3,974	3,974	3,974
R-squared	0.049	0.049	0.049	0.050	0.049	0.052

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variable. Standard errors are clustered at ROR 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 dummy).

an index of industrial concentration of the region (Herfindahl-Hirschman-Index, HHI), in column (3) of Table 3. The coefficient of the industrial concentration turns out to be negative but not statistically significant. If anything, it seems that the incidence of overqualification is lower in regions that are more specialized in specific industries. As a result, the employment density coefficient increases (in absolute value) to -0.024. Another characteristic that might be related with both agglomeration economies and with the qualification mismatch incidence is the share of innovative industries in the region. Innovative industries, and especially high tech industries, have been found to have a particularly high multiplier effect in the local economy in terms of additional jobs in other sectors (Moretti, 2010). These industries might thus both lead to faster agglomeration economies and lead to better qualification matches since they have positive spillovers to many other sectors of the economy. In column (4) we add the share of innovative and research intensive industries as defined by the NIW/ISI/ZEW list (for details, see Gehrke et al., 2010). The coefficient turns out to be negative and significant, pointing towards a lower overqualification incidence in those regional areas that have innovative industries. Nevertheless the employment density coefficient if anything increases (in absolute value) slightly, revealing that in Germany the share of these industries is not higher on average in larger agglomerations.

Firm size has been found to be an important determinant of the urban wage gap in Germany (Lehmer and Möller, 2010). In column (5) we add also two dummies for firm size to see if the lower overqualification incidence in larger agglomerations can be partly explained by the presence of larger firms. The coefficient denoting firms with more than 200 employees and firms with less than 20 employees are both slightly negative (even if not statistically significant), so that overqualification seems to be especially relevant in middle-sized firms. Even if firms are on average larger in thicker labour markets, the employment density coefficient appears to be unaffected by the inclusion of firm size dummies. Finally, column (6) presents the results of a regression, where all discussed determinants are included. The coefficient of the variables included do not lose their magnitude suggesting that they affect overqualification through different channels. While the precision of the estimate of employment density decreases a bit, the magnitude remains very similar to the baseline estimate. Summing up, especially the share of high–skilled individuals explains a large portion of the overqualification differential across regions. However, when we also take into account other structural regional characteristics, we end up with with a sizable estimate of employment density that is comparable to the one of the baseline regression. This suggests that we can interpret the employment density coefficient as a “pure” labour market size effect.

5 Qualification mismatch and the urban wage premium

In this section we want to investigate the extent to which qualification mismatch contributes to the urban wage premium. More precisely, we wish to analyze what portion of the effect of regional employment density on earnings is explained by better job matches with respect to the qualification possessed. To do so we 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.

We then add to this regression our measures of qualification mismatch and look how the coefficient of employment density is affected. Since we found a relatively large effect of employment density on qualification mismatch (on overqualification in particular) and we know from the literature on overeducation that there is a strong negative relationship between overqualification and wages, we expect that a large part of the effect of regional employment density on wages will be explained by a lower probability to be overqualified.

Table 4: Impact of employment density and mismatch on log hourly wages

	(1)	(2)	(3)	(4)
Empl. density (log.)	0.049***	0.045***	0.048***	0.045***
	(0.010)	(0.009)	(0.009)	(0.009)
Overqualification		-0.177***		-0.156***
		(0.014)		(0.015)
Horizontal mis.			-0.098***	-0.043***
			(0.011)	(0.011)
Control variables	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	No
Observations	4,022	4,022	4,022	4,022
R-squared	0.430	0.455	0.439	0.456

Note: Standard errors are clustered at ROR 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 dummy).

Table 4 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 squared kilometer is associated with an increase in wages of about 5%. This result appears to be in line with previous studies estimating the magnitude of urban agglomeration economies that range from 0.02 to 0.07 (Combes and Gobillon, 2014). In column (2) and (3) we add separately our measure of overqualification and horizontal mismatch. Consistently with the literature, vertical and horizontal mismatch are both associated to lower wages. As expected the coefficient of employment density decreases in both specifications. However, the decrease is very small. According to the estimations, overqualification explains about 8% of the urban wage premium while horizontal mismatch less then 3%. 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, so that they explain slightly more than 8% of the effect of employment density on wages. These results are consistent with previous studies using different measures of skill mismatch, which found that it explains only about 5-8% (Abel and Deitz, 2015) of the urban wage premium or does not have almost any contribution (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 decrease the coefficient of the urban wage premium and the overqualification dummy might proxy to some extent unobserved ability.

6 Conclusion

This paper seeks to measure the effect of local labour market size on vertical and horizontal qualification mismatch. Estimating a linear probability model with an extensive set of control variables, we find that more densely populated regions are associated with a lower probability for German male workers to be overqualified and to work in a different field than the one of 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 insignificant when macro-region controls and occupation fixed effects are included. We then follow two 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 place of childhood we get a smaller but still sizeable estimate of employment density on overqualification. Second, by exploiting the panel structure of the data and accounting for individual fixed effects, we get a coefficient that is even slightly larger compared to the baseline regressions. When looking at the determinants of the match differential across regions a large portion can be related to the different regional skill composition, but overall nearly the whole impact found seems to be attributable to a pure labour market size effect. Finally, we investigate the extent to which lower qualification mismatch in large agglomerations contributes to the urban wage premium. We find that overqualification explains only 8% 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*, forthcoming.

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.

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

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

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. (2014). The Empirics of Agglomeration Economies. IZA Discussion Papers 8508, Institute for the Study of Labor (IZA).

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.

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.

Gehrke, B., Rammer, C., Frietsch, R., Neuhäusler, P., and Leidmann, M. (2010). Listen Wissens-und Technologieintensiver Güter und Wirtschaftszweige. *Zwischenbericht zu den NIW/ISI/ZEW-Listen*, 2011.

Gerlitz, J.-Y. and Schupp, J. (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Technical report, 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.

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.

Kroh, M. (2012). Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2011). Technical report, 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. *The Annals of Regional Science*, 45(1):31–53.

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

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.

Moretti, E. (2010). Local Multipliers. *American Economic Review*, 100(2):373–77.

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

A Further tables and figures

Figure A.1: Qualification mismatch by highest degree

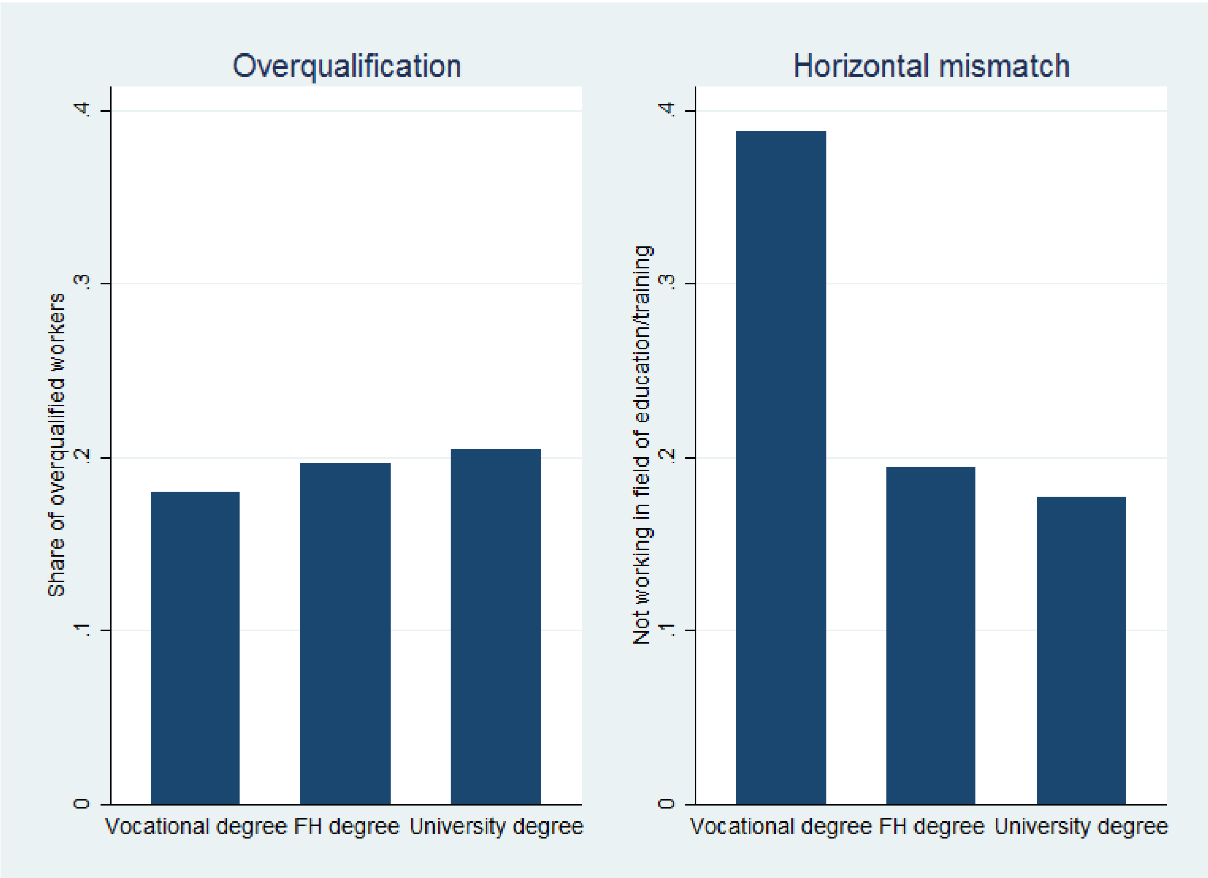


Figure A.2: Employment density of German regions (in 2010)

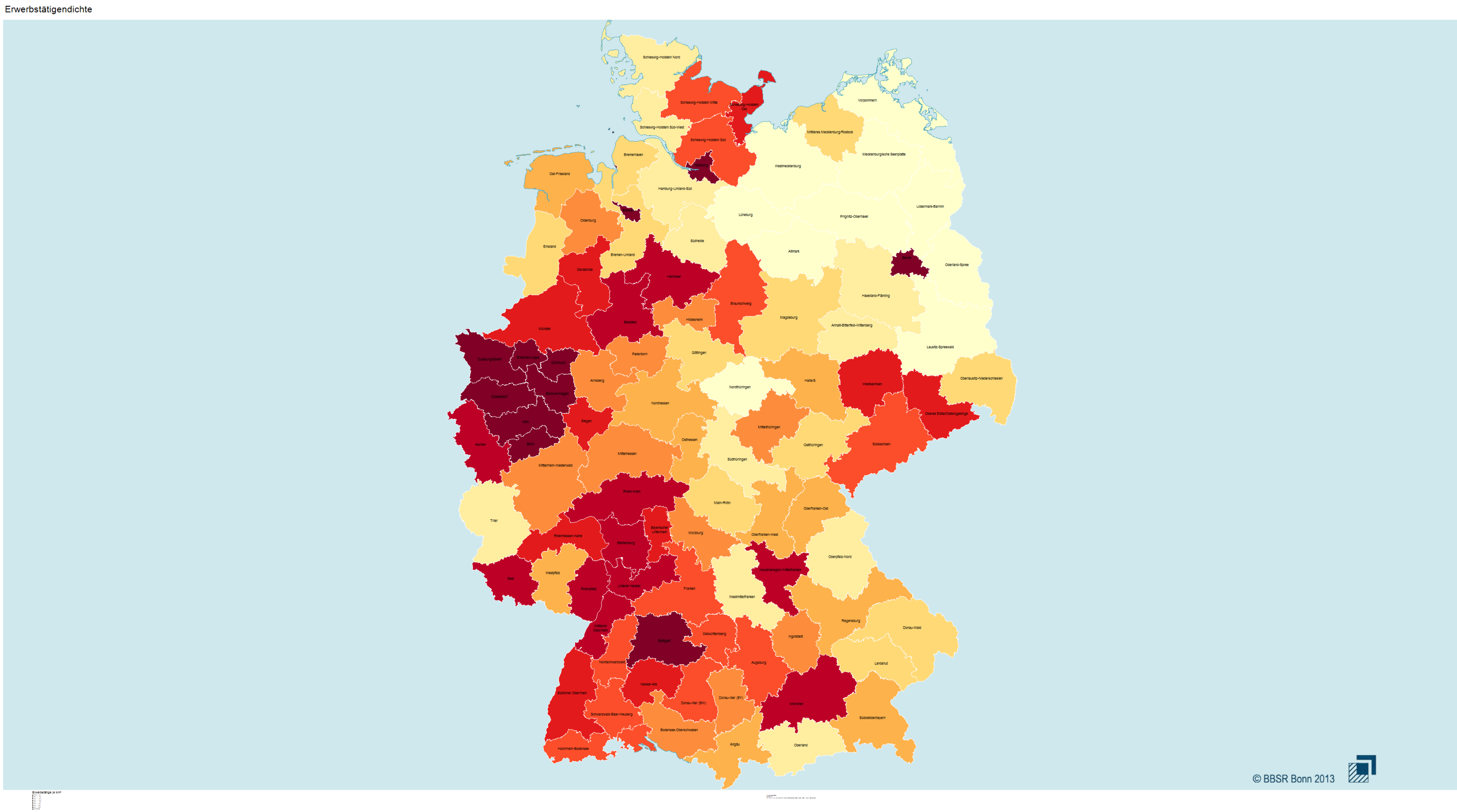


Table A.1: Summary statistics

	Mean	Std. Dev.	Min.	Max.
Dependent variables and other main variables				
Overqualified	0.19	0.39	0	1
Horizontal mismatch	0.31	0.46	0	1
Hourly wage (log)	2.75	0.48	1.35	4.29
Employment density (log)	5.03	0.99	2.89	7.53
Main control variables				
Migration background	0.08	0.27	0	1
FH degree	0.12	0.33	0	1
Vocational degree	0.63	0.48	0	1
Married or living with partner	0.82	0.38	0	1
Actual work experience	18.7	11.1	0	48
Has children	0.40	0.49	0	1
Survey year				
Surveyed in 2000, 2001 or 2002	0.51	0.50	0	1
Surveyed in 2003, 2004 or 2005	0.09	0.28	0	1
Surveyed in 2006, 2007 or 2008	0.14	0.35	0	1
Surveyed in 2009, 2010 or 2011	0.26	0.44	0	1
School leaving qualification				
University access (Abitur)	0.31	0.31	0	1
FH access (Fachhochschulreife)	0.10	0.31	0	1
Realschulabschluss	0.32	0.47	0	1
Parental background				
Father: higher educ.	0.15	0.36	0	1
Mother: higher educ.	0.07	0.25	0	1
Mother non employed (age 15)	0.39	0.49	0	1
Geographic characteristics				
Lives in city of childhood	0.57	0.50	0	1
Macro-region				
North	0.15	0.36	0	1
South	0.30	0.46	0	1
East	0.22	0.41	0	1
Centre	0.33	0.47	0	1
Job characteristics				
Public sector	0.26	0.44	0	1
Firm tenure	11.5	10.6	0	47

Note: The summary statistics are based on the baseline sample of 4281 observations. Main control variables include a squared term for work experience and tenure as well. Job characteristics also include 9 dummies for the industry or sector of the firm (agriculture, energy, mining, manufacturing, construction, trade, transport, bank/insurance and services) and firm tenure squared.

Table A.2: Impact of employment density on overqualification: further controls

	Overqualification				Horizontal mismatch			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Empl. density (log.)	-0.026***	-0.027***	-0.026***	-0.027***	-0.019***	-0.020***	-0.021***	-0.022***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
School grade: German		0.014	0.013	0.014		0.020**	0.017*	0.018*
		(0.010)	(0.009)	(0.010)		(0.010)	(0.010)	(0.010)
School grade: Math		0.021*	0.022*	0.021*		0.011	0.014	0.013
		(0.012)	(0.013)	(0.013)		(0.012)	(0.012)	(0.012)
Extraversion			0.003	0.003			0.003	0.003
			(0.003)	(0.003)			(0.004)	(0.003)
Consciousness			0.001	0.001			-0.005	-0.005
			(0.004)	(0.004)			(0.004)	(0.004)
Agreableness			0.003	0.004			0.004	0.005
			(0.004)	(0.004)			(0.004)	(0.004)
Neuroticism			0.006*	0.006**			0.005*	0.006**
			(0.003)	(0.003)			(0.003)	(0.003)
Openness to experience			-0.002	-0.003			0.002	0.001
			(0.003)	(0.003)			(0.004)	(0.004)
Risk preference				0.006				0.010*
				(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	2,141	2,141	2,141	2,141	2,141	2,141	2,141	2,141
R-squared	0.046	0.049	0.053	0.054	0.076	0.079	0.081	0.082

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variable. Standard errors are clustered at ROR 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 dummy).

Table A.3: 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.)	-0.024***	-0.025**	-0.025***	-0.015**	-0.017*	-0.019**
	(0.009)	(0.011)	(0.007)	(0.006)	(0.009)	(0.009)
Vocational degree ×	0.005	0.016				
Empl. density	(0.012)	(0.014)				
Vocational degree	-0.210***	-0.376***				
	(0.068)	(0.074)				
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	No	Yes	No	Yes	No
Observations	4,281	4,281	2,716	2,716	1,565	1,565
Panel B: Horizontal mismatch						
Empl. density (log.)	-0.000	0.002	-0.021	-0.014	-0.001	0.000
	(0.008)	(0.007)	(0.013)	(0.013)	(0.007)	(0.007)
Vocational degree ×	-0.022	-0.017				
Empl. density	(0.014)	(0.014)				
Vocational degree	0.210***	0.127*				
	(0.075)	(0.075)				
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	No	Yes	No	Yes	No
Observations	4,281	4,281	2,716	2,716	1,565	1,565

Note: The table shows the estimates of a linear probability model with skill mismatch measures as dependent variable. Standard errors are clustered at ROR 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 dummy).