Do Ethnic Networks Ameliorate Education–Occupation Mismatch?

Eric Schuss*,†

Abstract. The question to what extent ethnic networks affect occupational mismatch has so far been overlooked. This paper exploits supraregional variation in ethnic composition in Germany and shows that a one standard deviation increase in the share of the own ethnic group per zip code significantly reduces the years of overqualification for females, by 0.27 years. For males, neither the foreign share nor the ethnic share per residency area is found to significantly impact the extent of overqualification. Selection into residency groups and occupations and different endowments in language capital explain the more efficient benefit of ethnic networks accrued to females.

1. Introduction

The ethnic composition of areas can ameliorate or hamper the labor market integration of immigrants (e.g., Cutler and Glaeser, 1997; Borjas, 1998; Edin *et al.*, 2003). Considering this ambiguity, my paper analyzes the case of Germany and examines the question of whether ethnic networks causally induce education–occupation mismatch.[1]

In a recent paper, Dustmann *et al.* (2016a) provide evidence that ethnic networks in Germany increase initial wages, reduce employee turnover, and increase tenure duration of foreign-born immigrants. Empirical evidence also demonstrates that recently arrived migrants face a sharper trade-off between unemployment and education–occupation mismatch (Bentolila *et al.*, 2010). One source of this trade-off is the exclusion of recently arrived migrants and other migrant groups from obtaining social benefits. A lack of command of the host country's official

*I thank Jan Goebel and the Research Data Center of the German Socio-Economic Panel (SOEP, DIW Berlin) for making data access to spatial data of the German Socio-Economic Panel (SOEP) and Microm data possible. Furthermore, I am grateful to Rüdiger Budde and Sandra Schaffner from the Research Data Center Ruhr at the RWI Leibniz Institute for Economic Research Essen to allow me to work with Microm data at the RWI. I appreciate helpful comments and remarks by Melanie Arntz, Herbert Brücker, Franco Peracchi, Reinhold Schnabel, and two anonymous reviewers. I thank the Institute for Employment Research (IAB) and the University of Duisburg-Essen for support while writing substantial parts of this paper. In addition, I am grateful for discussions with conference participants at the 11th RGS Doctoral Conference 2018 in Essen, the 8th ifo Dresden Workshop on Labour Economics and Social Policy, the 32th Annual Conference of the European Society of Population Economics (ESPE), and internal seminars. All remaining errors are my own. Open access funding enabled and organized by Projekt DEAL.

[Correction added on 25 September 2020, after first online publication: Projekt DEAL funding statement has been added.]

†Federal Institute for Vocational Education and Training (BIBB), Robert-Schuman-Platz 3, Bonn, D-53175 Germany.
E-mail: Eric.Schuss@bibb.de

LABOUR 34 (4) 441–476 (2020) DOI: 10.1111/labr.12183
 JEL C26, J24, J61, R23

language, higher reliance on ethnic networks, and exclusion from social benefits merge for recently arrived migrants. Does a higher reliance on ethnic networks result in a higher risk of accepting a job that does not meet a migrant's skills and qualification? This question is economically relevant because overqualification can be interpreted as inefficient returns to education. If networks increase tenure durations — one major finding of Dustmann *et al.* (2016a) — then inefficient occupational matching would not only mean wasted productivity in the short term but also a long-term accumulation of deadweight losses if the migrant works permanently in jobs for which he or she is overqualified.

This paper makes several contributions. First, it tackles the novel question of whether ethnic networks affect education–occupation mismatch, which has barely been examined so far (to my knowledge, the exceptions are Charpin, 2014; Kalfa and Piracha, 2018). Concerning the definition of education–occupation mismatch, this paper only considers foreign-born immigrants in measuring overqualification, without including native-born Germans. This measure captures the education of a foreign-born worker in a particular occupation and how his or her education deviates from the average level of education of foreign-born workers who are employed in the same occupation. Thus, my definition deviates from the standard definition of overqualification in Poot and Stillman (2016), who pool natives and migrants together.[2] Second, I pay special interest to gender differences and have a particular focus on females when examining the effect of ethnic networks on education–occupation mismatch. Because of data availability and research focuses, previous studies have mainly considered males in the context of labor market integration. However, the gender gap in labor market participation and wages is larger for immigrants (for instance, see, Fendel and Jochimsen, 2017), which is one reason for a more detailed analysis of the employment potentials of female migrants. Third, research on the determinants of education–occupation mismatch is scarce in general, with only a few exceptions (for more details, see Section 2.2). In this context, the central econometric challenge is to uncover the endogeneity between ethnic networks and individual labor market outcomes and to cancel out reversed causality and negative selection into districts. I apply the instrumental variable (IV) approach of Bertrand *et al.* (2000), which tackles the issue of endogeneity by exploiting the supraregional variation in the ethnic composition of areas that is pre-determined by historical population numbers. I incorporate this approach and use aggregated numbers of ethnic population per zip code area to solve the problem of sorting into districts.[3] Fourth, research on the effects of immigration on employment and wages has recently highlighted some methodological drawbacks of previously approaches (see Dustmann *et al.*, 2016b; Jaeger *et al.*, 2018). Lessons from those discussions are applied to my IV approach, including the empirical analyses of the exogeneity of my instruments (see Section 6.2).

To study the education–occupation mismatch of migrants, individual survey data from the IAB-SOEP Migration Sample for 2013–2015 are linked with spatial data on the ethnic composition of German districts.[4] After finding evidence of sorting into zip codes, IV estimations demonstrate a significant negative marginal effect of the share of the own ethnic group per zip code (henceforth the ethnic share) on years of overqualification for females by 0.27 years. For males, neither the foreign share nor the ethnic share per zip code is found to significantly affect occupational mismatch. For females — a group often ignored in the context of segregation and networks — it can be concluded that networks are ethnically stratified (see Damm, 2014; Edin *et al.*, 2003; Glitz, 2014). Thus, for females, while residential segregation does not induce mismatch, the linkage to their own ethnic group decreases the extent of education–occupation mismatch. This gender difference is explained by different selection into occupations and different endowments of pre-migration skills, language capital, and the different reasons for migration by gender. Thus, the focus on gender differences allows examining the transmission channels of ethnic networks concerning the labor market performance of immigrants.

The structure of the paper is as follows. The next section provides a detailed literature review on the determinants of overqualification and the effects of ethnic networks. Section 3 gives a presentation of the different datasets, and Section 4 explains the IV approach as the econometric method. In Sections 5 and 6, I present the results of the baseline regressions and tackle econometric issues behind the identification strategy. Section 7 concludes with a critical discussion of the underlying identification strategy and presents the political implications of my empirical findings.

2. Literature

2.1. Ethnic networks and labor market outcomes

Based on theoretical models by Altonji and Card (1991), Card (2001), and Calvo-Armengol and Jackson (2004, 2007), a large number of empirical papers examine the effect of ethnic networks — indicated by the share of the own ethnic group in the area — on wages, employment, and the demand for welfare benefits. The endogeneity of residential choices is the main concern in prior research. Negative sorting into districts produces endogeneity, which arises from the fact that immigrants with low levels of qualification and low pre-migration skills select into deprived areas with weak economic conditions and with a large foreign-born population share. There are three main lessons from these previous analyses: Networks are 'ethnically stratified', network quality plays an important role in producing effect heterogeneity, and the network effect evolves dynamically over time and ethnic networks also affect human capital accumulation.

Bertrand *et al.* (2000) analyze whether higher ethnic concentrations in districts increase the demand for welfare benefits.[5] The authors define an ethnic network as the number of people living in a district who belong to the same language group relative to the whole population in the area.[6] To account for the endogeneity of residential choices, they assume that individual choices of the metropolitan area are given and pre-determined by historical population numbers and traditional migration patterns. Following this argument, the authors instrument the ethnic concentration per district by the ethnic share of the supraregional level of metropolitan areas. This instrumental variable (IV) approach justifies the assumption of exogeneity if the bias due to self-selection disappears at the aggregated metropolitan area level. They find that the interaction between mean welfare use of the own ethnic group and the group's population share per district strongly increases individual welfare participation. This finding highlights that networks are 'ethnically stratified'. In a recent paper, Dustmann *et al.* (2016a) confirm this ethnic stratification and, using German matched employer–employee data for the period 1980–2001, demonstrate that the share of the own ethnic group at a workplace reduces employee turnover and increases tenure duration and initial wages. Moreover, wage growth is found to be smaller if tenure relies on networks. These results are confirmed by Brown *et al.* (2016) with matched applicant–employee data from the United States.[7]

Deri (2005) and Andersson *et al.* (2014) adopt the IV approach of Bertrand *et al.* (2000). Andersson *et al.* (2014) replace mean welfare use by the mean employment rate of the own ethnic group as the indicator of network quality. Edin *et al.* (2003) find that low-skilled immigrants realize positive income effects from living in ethnic segregation, and these effects increase with average labor income and with the share of self-employment within ethnic milieus. On the contrary, living in an ethnic group of low quality and being a high-skilled immigrant leads to a negative effect. Thus, this literature strand underlines the important role of network quality in producing effect heterogeneity (see Borjas, 1995, 1998; Cutler and Glaeser, 1997; Damm, 2009, 2014). By exploiting residency assignments for political refugees in the United States, Beaman (2012) also emphasizes the role of network quality. She finds that competition among immigrants increases with the size of the ethnic community. While the number of network members with labor market experience acquired in the United States improves the labor market performance of recently arrived immigrants, an increase in the number of recently arrived immigrants worsens their labor market performance.

By distinguishing between the short run and long run, Battisti *et al.* (2018) emphasize the dynamic evolution of the network effect. By estimating a panel model with a large set of fixed effects and covariates, they find that immigrants living in districts with a large ethnic share find their first job faster. However, at the same time, they are less likely to invest in human capital. The positive network effect on employment fades away after five years, while the positive returns from human capital remain.

My paper tackles the endogeneity puzzle and transfers the IV approach of Bertrand *et al.* (2000) to European labor economics for the first time. I instrument the share of the own ethnic group at the zip code level by the respective numbers at the metropolitan level, approximated by zip code areas. In Section 4, I provide a detailed justification for this identification strategy. Moreover, I take the first two lessons from the literature into account: Network effects are ethnically stratified, and heterogeneity in effects arises from the quality of networks.

2.2. Theory and literature on education–occupation mismatch

The constrained portability of human capital acquired in the country of origin is the major source of education–occupation mismatch and overqualification for immigrants. Weak language skills, imperfect information on immigrants' abilities, and problems in receiving the recognition of vocational degrees acquired abroad restrict the portability of skills over borders (Brücker *et al.*, 2018; Chiswick, 1978; Mattoo *et al.*, 2008). The majority of studies focus on the question of whether education–occupation mismatch is more likely in a group of migrants compared to natives (Chiswick and Miller, 2010; Nieto *et al.*, 2015; Poot and Stillman, 2016; Pellizzari and Fichen, 2017). However, knowledge on the determinants of occupational mismatch is scarce. This gap is tackled by examining whether ethnic networks ameliorate occupational matching quality. As the previous section illustrated, Dustmann *et al.* (2016a) find that the share of the own ethnic group at a workplace reduces employee turnover and increases tenure duration and initial wages. Are higher wages induced by networks a sign of improved job matching quality?

Theoretical models outline that the decision to accept a job for which a person is overqualified is driven by alternatives to this job (outside options). Unemployment makes a person more likely to accept a job for which he or she is overqualified (Nielsen, 2011) because unemployed individuals are under pressure to subsist. If outside options of overqualification are weak, which is the case if a person is not eligible for unemployment benefits, the tendency to accept a job for which the person is overqualified increases. This leads to the expectation that weaker outside options in cases of overqualification induce mismatch (Bentolila *et al.*, 2010).

In many countries, recently arrived migrants are excluded from unemployment benefits and face a lower level of social benefits at first. In Germany, a person can only claim unemployment benefits if they have already had a twelve-month period of employment.[8]

This institutional setting means that the trade-off between occupational mismatch and unemployment is strengthened for recently arrived migrants. Both limited eligibility to outside options of employment and a higher reliance on ethnic networks apply to recently arrived migrants, suggesting the expected link between ethnic networks and the extent of overqualification. Per this hypothesis, Patel and Vella (2013) find that recently arrived immigrants often choose the same occupations as their compatriots. The following sections examine whether the relationship between ethnic networks and overqualification is causal.

In this context, it is essential to discuss theoretical transmission channels of this link. Besides language skills, residency status is one potential transmission channel. Eligibility for social benefits depends on residency status and country of origin. For instance, eligibility differs between migrants from states of the European Union (EU) and third countries. Particular groups, such as economic migrants, seasonal workers, asylum seekers, and tolerated migrants, are also excluded from unemployment assistance and are only eligible for social assistance. However, family members have access to economic resources and family support in the country of destination. Thus, for family members, unemployment instead of overqualification is less harmful and the use of ethnic networks is expected to be more efficient compared to asylum seekers, whose migration decisions are less prepared and who are likely to suffer under a greater degree of precariousness from uncertainty in asylum decisions (Chiswick *et al*., 2005; Cortes, 2004). Furthermore, the chance to receive the recognition of vocational degrees acquired abroad as a further transmission channel depends on residency status, country of origin, and other characteristics.

In this theoretical framework, the gender difference is of particular interest. Females and males differ in the endowment of language capital and select themselves into different occupations and residency status groups (details are provided in Section 3.1). For instance, females more often choose jobs in which language capital is more important. I exploit this gender difference in occupational selection to examine how the relationship between ethnic networks and overqualification depends on different endowments in language capital and selection patterns into occupations. The command of the destination country's language, the choice of occupation, occupational recognition, and residency status are the potential transmission channels that can explain this link. If this is the case, I suggest that males and females have different relationships between ethnic networks and occupational mismatch.

To the best of my knowledge, two paper examine the effect of ethnic networks on occupational mismatch. Kalfa and Piracha (2018) use the Households Income and Labour Dynamics in Australia (HILDA) data and apply a dynamic random-effects probit model. They tackle the problem of endogeneity in social capital and ethnic concentration by using lags of the explaining variables, finding that ethnic concentration significantly increases the incidence of overeducation. This applies particularly to females and is not significant for males. Apparently, the negative effects from ethnic networks dominate the positive effects, which contradicts other studies who find ethnic networks to be beneficial during job search (Bentolila *et al*., 2010; Nielsen, 2011).

Charpin (2014) does not find any significant relationship between ethnic networks and occupational mismatch when using French data from the Longitudinal Survey of the Integration of First-Time Arrivals (ELIPA) and the Training and Professional Qualification Survey (FQP). However, Charpin does not distinguish between the general foreign-born population and the population of certain ethnic groups. Furthermore, the problem of negative selection into districts is not tackled.

Note that Charpin (2014) considers the effects on occupational downgrading and does not have a focus on education–occupation mismatch. Although the terms downgrading and education–occupation mismatch display similarities (Nieto *et al.*, 2015), a few differences exist that justify the separation of literature on occupational mismatch and downgrading in my paper.[9]

3. Data and descriptive statistics

3.1. Survey data and descriptive statistics on occupational mismatch

The empirical analysis is based on individual survey data from the IAB-SOEP Migration Sample for 2013–2015, linked with spatial data on ethnic composition across zip codes in Germany. The IAB-SOEP Migration Sample offers unique data on households with a household head that has a migration history (Brücker *et al.*, 2014) and household members who are at least 16 years old (Wagner *et al.*, 2008).[10] In the sampling procedure, certain countries were given above-average consideration (e.g., Turkey, Southern Europe, [former] Yugoslavia, and countries of the 2004 EU enlargement). Representative statements can be made using weighting factors that take into account the composition of the population according to age, gender, and country of origin. I define education–occupation mismatch as cases in which an immigrant's level of education deviates from the required level of education in his or her occupation, considering the immigration population. To capture this definition in my dataset, I apply the approach by Poot and Stillman (2016) and calculate years of overqualification in the following way:

$$\text{Years of overqualification}_i = \text{Years}Ed_i - \text{Years}Ed_{Occ(i)} \qquad [1]$$

First, I capture individual education by individual years of schooling $YearsEd_i$. Second, the required level of education in one occupation, $YearsEd_{Occ(i)}$, is understood as the average level of education of sampled immigrants within the same occupation. For each occupation, I calculate the average schooling years of surveyed immigrants within the same occupation, an approach known as the realized matches procedure.[11] While Poot and Stillman (2016) include natives in their analysis, I deviate from the standard definition of overqualification and exclusively consider foreign-born persons in my calculations.

To calculate the average schooling years for a given occupation, it is essential to characterize occupations by diverse features. If they are only distinguished between a few categories (e.g., white-collar, blue-collar, and the public sector), this results in little variation in the required schooling years. However, it is desirable to generate a continuous variable and produce a large variation in the required level of education. Moreover, in a sample considering foreign-born immigrants, it is essential to make jobs with workers from different countries comparable. One concept that meets those requirements is the Standard International Occupational Prestige Scale (SIOPS). By using this classification, I distinguish between 66 different occupations in the survey. Thus, to capture the required years of schooling in each occupation, I calculate the average number of schooling years surveyed.[12]

To avoid any distortions, I exclude occupations with fewer than ten individuals employed. The calculations of average years of schooling for each occupation are based on an average of 76.3 observations. One drawback is that this measurement of occupational mismatch is driven by individual selection patterns into occupations and by the distribution of qualifications (Poot and Stillman, 2016). Moreover, it conflates both undereducation and overeducation into one variable. Nevertheless, this measurement is the least problematic choice (for a detailed discussion of different concepts of measuring occupational mismatch, see Hartog, 2000; Chiswick and Miller, 2008). In fact, this measurement has two further important advantages. First, to account for systematic differences in educational and occupational choices by gender, the average schooling years per occupation can be calculated separately for males and females. Second, the calculation of average schooling years can be stratified by the country of origin to allow for the different values of one schooling year across different education systems.

In the context of occupational mismatch, it is essential to restrict the empirical analysis to employed persons aged between 16 and 64 and to exclude self-employed migrants. Although a sample exclusively considering employed persons is selective, this approach ensures a focus on the leading research question. Furthermore, the sample is restricted to first-generation immigrants, so that a sample of 3,560 observations over 2,502 individuals remains. Although sample attrition is larger for migrants compared to natives in most surveys, this does not fully explain why only one third was surveyed several times in this case. Two reasons apply here. First, in 2015, the survey was refreshed to increase the number of observations and to study migrants born in countries of the EU enlargement in 2004 and 2007. This refresher sample makes up around 20 percent of my sample. Second, the restriction on employed migrants increases attrition due to the high turnover of recently arrived migrants.

Table 1 shows that the average years of overqualification are slightly larger for females than for males, while females display more years of schooling. Panel B presents binary information about the share of the sample that is mismatched. For instance, 40.5 percent of females and 41.9 percent of males are overqualified by at least six months. Additionally, 23.3 percent of females and 20.3 percent of males display overqualification of between −0.5 and + 0.5 years and are appropriately qualified for their occupation. In general, males are more likely to be undereducated than females. This suggests that males select into qualifications with low requirements or display lower educational attainment on average.

Previous research consistently identifies the duration of residence as a major determinant of overqualification. This follows the expected trade-off between unemployment and education–occupation mismatch for recently arrived migrants. Figure 1 displays the negative association between occupational mismatch and years of residence. Years of overqualification fall below 0.5 years when considering migrants who have lived in Germany for five years. Furthermore, family members and asylum seekers are expected to suffer more from occupational mismatch compared to labor migrants (Chiswick *et al.*, 2005; Cortes, 2004). In Figure A1 in the Appendix A, left-hand kernel densities confirm this expectation, while right-hand graphs demonstrate that German language skills particularly influence the left tail of the distribution.

Table 2 offers a summary of the covariables used in the analysis. Besides basic socioeconomic characteristics, the IAB-SOEP Migration Sample offers retrospective information about the residency status and language skills of migrants on arrival and provides information about occupational choices and employment participation in the year before migration to Germany. Such pre-migration characteristics are often summarized under unobserved heterogeneities in migration economics (see Borjas and Bratsberg, 1996; Chiswick and Miller, 2005; Dustmann and van Soest, 2002). Furthermore, the previous literature underlines that residency status is a further major determinant of occupational choice (Amuedo-Dorantes and Bansak, 2011; Fasani, 2015). These pre-migration characteristics affect both the choice of ethnic networks and the extent of occupational mismatch.

Additionally, I control for the country of origin and for the occupational sector to capture systematic differences in the extent of occupational mismatch across ethnic groups

Table 1. Education–occupation mismatch by gender

	Females	Males	Mean Diff.
Panel A: Metric measurement of education–occupation mismatch			
Years of overqualification	0.064 (1.875)	−0.019 (1.999)	0.084
Years of schooling (*YearsEd*$_i$)	10.5 (2.0)	10.3 (2.1)	0.2**
Panel B: Binary measurement of education–occupation mismatch (in %)			
±0.5 years			
Overqualified	40.5	41.9	−1.4
Underqualified	36.2	37.8	−1.6
Appropriately qualified	23.3	20.3	3.0**
±1.0 Years			
Overqualified	35.8	36.8	−1.0
Underqualified	24.3	27.9	−3.7**
Appropriately qualified	40.0	35.3	4.7***
±1.5 Years			
Overqualified	21.5	22.7	−1.2
Underqualified	21.5	25.8	−4.3***
Appropriately qualified	57.0	51.5	5.5***
±2.0 Years			
Overqualified	17.2	17.3	−0.1
Underqualified	11.1	13.1	−2.0*
Appropriately qualified	71.7	69.6	2.0
Observations	1,702	1,858	3,560
Individuals	1,312	1,192	2,504

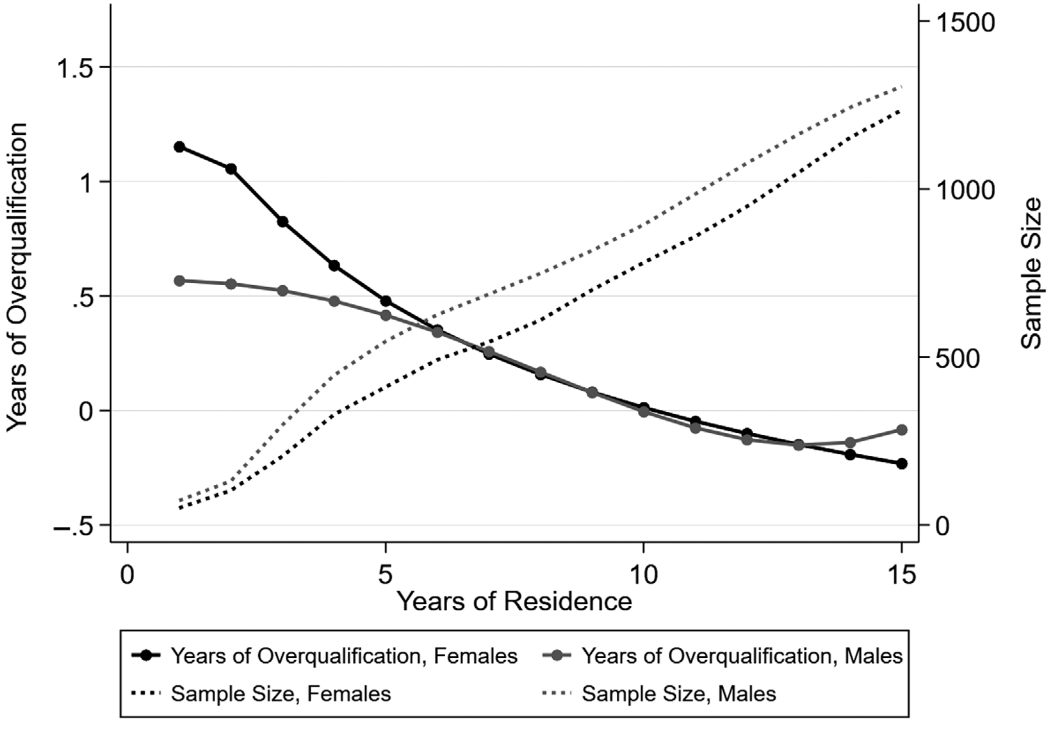
Source: IAB-SOEP Migration Sample, own calculations.
Notes: *p < 10%, **p < 5%, ***p < 1%; standard deviations in parentheses.
Two-sided t-tests show whether mean differences significantly differ by gender.

and job sectors. Distinct differences by gender legitimate separated econometric analyses. First, females are more likely to display appropriate German language skills — currently and up to the point of time at arrival — and to have attended a German language course. Second, men are more often employed in the country of origin before migration and are differently distributed across the occupational sectors compared to women. Women are overrepresented in the sector of white-collar jobs and more likely to have migrated to Germany as family members (38.6 percent). In the group of men, only 17.9 percent arrived in Germany as family members, while 11.2 percent were asylum seekers.

3.2. Spatial data on the ethnic composition of zip codes

Information about the exact residency of the surveyed individuals allows me to link survey data with Microm Raster data. This dataset includes spatial information about ethnic compositions and economic conditions across 1-km^2-cells, zip codes, and municipalities in Germany. My indicator of ethnic networks accounts for important facts drawn from the literature. Prior research demonstrates that networks are ethnically stratified. To consider this, I distinguish between the share of the overall foreign-born population FS_{jt} and the

Figure 1. The Link between Education–Occupation Mismatch and Years of Residence. *Source:* IAB-SOEP Migration Sample, own calculations. *Note:* The left-hand vertical axis displays years of overqualification depending on the duration of residence. The right-hand vertical axis gives the number of observations depending on the duration of residence. The two curves result from fractional polynomial regression of years of overqualification on years of residence. Note that each data point in this figure is based on at least 50 observations. There are 51 observations for females after one year of residence and 73 observations for males after one year of residence.



share of the own ethnic group k in area j in survey year t, henceforth the ethnic share $\mathrm{ES}_{\mathrm{jkt}}$ (Schaffner and Treude, 2014):

$$FS_{jt} = \text{Share of the foreign} - \text{born population in area } j \text{ (residential segregation)} = \frac{\text{Immigrant}_{jt}}{\text{Population}_{jt}} \quad [2]$$

$$ES_{jkt} = \text{Ethnic share in area } j \text{ for the ethnic group } k \text{ (ethnicnetwork)} = \frac{\text{Immigrant}_{jkt}}{\text{Population}_{jt}} \quad [3]$$

On which level of geography are networks formed? Empirical economics shows that the ethnic networks of migrants are developed at the zip code level due to norms and the information channel. Because ethnic networks are built upon visible features, such as skin, look, and common values (such as religion and traditions [norm channel]), networks are constructed on a coarser geographical unit than neighborhoods (Glitz, 2014; Dustmann *et al.*, 2016a). Moreover, in Germany, natives and immigrants are assigned to authorities and administrations based on zip codes. Thus, the information channel also predicts that networks are constructed at the zip code level (Schaffner and Treude, 2014).[13]

In Microm Raster data, ethnic affiliation is identified by the first names and surnames of household heads (Budde and Eilers, 2014; Microm Consumer Marketing, 2014). By using international registers of names, and by combining collected names in different zip codes with the linguistic ancestry of names, foreign and ethnic shares can be captured (Budde and Eilers, 2014; Microm Consumer Marketing, 2014; Schaffner and Treude, 2014).[14]

Panel A of Table 3 displays an average foreign share of 10.2 percent for females and males, whereas the sample also includes migrants living in zip codes with surpassing shares of up to 39 percent. The share of the own ethnic group is 1.2 percent for females and slightly larger for males. The Turkish community and migrants from the Balkans are the two largest communities in Germany. Spatial data provide large variation in both foreign shares and ethnic shares; however, both variables are highly right-skewed. To produce normally distributed errors and reliable t-tests, logarithmized shares are used.

The measurement of ethnic networks by spatial indicators suffers from some drawbacks. First, these concepts capture only the potential for social contacts. Thus, effective contact with networks remains unobservable (Bertrand *et al.*, 2000; Devillanova, 2008). Second, these concepts do not distinguish between weak, strong, and regular ties to the members of a network (Card, 2001; Green *et al.*, 1999; Patacchini and Zenou, 2012).[15] Third, Microm data do not include information about network quality, such as the share of highly educated immigrants per area (see Andersson *et al.*, 2014; Klaesson et al., 2019). Battu *et al.* (2011) and Damm (2014) capture quality and the effective benefit by survey-based indicators of job searching channels. A similar indicator of networks is captured in the IAB-SOEP Migration Sample. Immigrants state whether they found their first job in Germany via personal networks, advertisement channels (job advertisements in newspapers or on the web), or institutional channels (job centers or job agencies in Germany or the country of origin) — a categorization based on Battu *et al.* (2011). However, as Damm (2014) argues, the choice of such job searching channels is highly selective and does not allow for causal inference. I focus on geographic indicators despite the stated drawbacks. The econometric method is outlined in the next section to help understand how I attempted to identify causal inference.

4. Identification strategy

The first step in estimating the effect of ethnic networks on education–occupation mismatch is to apply ordinary least squares (OLS):

$$y_{ijkt} = \beta_0 + \beta_1 \log ES_{jkt} + \beta_2 \log FS_{jt} + X'_{it}\beta_3 + M'_i\beta_4 + \gamma_j + \delta_k + \alpha_i + \varepsilon_{ijkt} \quad [4]$$

The main parameter of interest β_1 gives the impact of the logarithmized ethnic share $logES_{jkt}$ on years of overqualification y_{ijkt} of immigrant i from ethnic group k. The

Table 2. Descriptive statistics

	Females	Males	Mean Diff.
Socioeconomic & educational variables			
Age	41.3	41.5	−0.2
Years of residence	11.4	11.5	−0.1
Age at immigration	29.9	29.9	0.0
Relationship (in %)	61.3	62.5	−1.2
High education, ISCED 5–6 (in %)	30.6	25.8	4.8***
Middle education, ISCED 3–4 (in %)	48.1	50.5	−2.4
Low education, ISCED 1–2 (in %)	21.2	23.7	−2.5*
Integration & pre-migration variables (in %)			
German citizenship	33.3	30.1	3.2**
Command of the German language	76.6	66.9	9.7***
Language course in Germany	56.3	45.6	10.7***
At arrival			
Command of the German language	22.2	17.7	4.5***
Support at migration	42.0	42.2	−0.2
Employed in the source country	68.3	78.8	−10.5***
Employee without managerial function	53.5	58.3	−4.8***
Employee with managerial functions	11.6	13.7	−2.1*
Self-employed	3.2	6.6	−3.4***
Relationship	27.7	27.8	−0.1
Residency status (in %):			
At arrival			
Family member	38.6	17.9	20.7***
Asylum seeker	4.1	11.2	−7.1***
Ethnic Germans	18.8	16.5	2.3*
Job searcher	14.5	21.7	−7.2***
With job commitment	10.2	22.7	−12.5***
Other status groups	13.9	9.5	4.4***
Currently			
Perm. right of residency, EU	39.0	37.4	1.6
Blue Card	0.8	1.5	−0.7**
Temporary right of residency	9.0	10.8	−1.8*
Tolerance permit and visas	1.4	1.3	0.1
Country of origin (in %)			
EU	46.9	44.2	2.7
Countries of EU enlargement in 2004	18.9	14.5	4.4***
Turkey	4.5	7.7	−3.2***
Arabic-speaking countries	1.6	3.1	−1.5***
Guest-worker Countries	13.2	20.6	−7.4***
Russia and (former) USSR	50.4	36.8	13.6***
Asia	3.2	2.1	1.1**
Sub-Saharan Africa	1.7	1.9	−0.2
Occupational sectors (in %)			
Blue collar			
Unskilled or semiskilled	35.7	39.6	−3.9**
Skilled or Technician Foreman, etc.	1.4	11.9	−10.5***
Foreman etc.	0.1	1.3	−1.2***
White collar			
Without apprenticeship	20.2	10.8	9.4***
With apprenticeship or foreman	8.0	5.5	2.5***
Qualified, high qualified, or executive function	26.1	22.7	3.4**
Public sector (in %)	*17.0*	*6.0*	11.0***
Self-employed (in %)	*3.2*	*6.6*	−3.4***

Table 2. Continued

	Females	Males	Mean Diff.
Observations	1,702	1,858	3,560
Individuals	1,312	1,192	2,504

Italics indicates the significance levels for the variables public sector and self-employed.
Source: IAB-SOEP Migration Sample, own illustration.
Notes: *p < 10%, **p < 5%, ***p < 1%; two-sided t-tests show whether mean differences differ significantly by gender.

application of random-effects panel estimation is feasible because of having panel data across three years.[16] The parameter β_2 estimates the impact of the share of the overall foreign-born population on the outcome. Additionally, I control for two sets of covariates. The matrix X_{it} includes socioeconomic and migration-specific variables, such as current residency status and German language skills (see Table 2). The set M_i controls for time-invariant pre-migration characteristics and initial conditions, such as German language skills at arrival, residency status at arrival, and employment status in the country of origin in the year before emigration. To control for the different economic conditions across zip codes and other pull factors, I add area dummies γ_{j}. Furthermore, by including dummies for countries of origin δ_k, I adjust for systematic differentials in the labor market attachment (of women) and exposures to discrimination across different ethnic groups.

Because residency choices are selective and driven by socioeconomic characteristics and unobservables (e.g., Blotevogel and Jeschke, 2001; Edin *et al.*, 2003; Aslund and Skans, 2009, 2010; Schönwälder and Söhn, 2009), OLS analysis results in a downward bias of β_1 and β_2; that is, the values are biased toward zero. To tackle this issue, I follow the IV strategy developed by Bertrand *et al.* (2000). They and other economists (Andersson *et al.*, 2014; Deri, 2005) assume that sorting into districts is only endogenous at the level of neighborhoods and boroughs, which they define as standardized areas of at least 100,000 inhabitants (so-called Public Use Microdata Areas [PUMAs]). Concurrently, the choice of the metropolitan area — a geographical unit that represents the superior extended city — is exogenous to the individual's labor market performance.

If we follow this argument, endogeneity can be canceled out by instrumenting the share of the foreign-born population and the ethnic share per PUMA by the respective share at the level of metropolitan areas. This strategy is supported by another approach applied by Altonji and Card (1991), Card (2001), Cascio and Lewis (2012), and Patacchini and Zenou (2012). They use the lagged ethnic composition of areas (and interactions with national inflows by countries of origin) to instrument the current population composition. For the instruments to be valid, lagged population numbers must be unrelated to (unobserved) factors determining current employment activity, apart from their effect through the current ethnic population in districts. This can be achieved by exploiting the fact that immigrants tend to settle in communities established by earlier immigrants of the same ethnic origin. Evidence provided by Card (2001), Cascio and Lewis (2012), and Patacchini and Zenou (2012) shows that the metropolitan area an immigrant chooses is pre-determined by the residency choices of his or her ancestors of the same ethnicity. This results in a stable presence of cultural and sports clubs, shops, supermarkets, and culinary offers in certain metropolitan areas over decades that reflect the norms and traditions of certain countries of origin. The establishment of certain ethnic groups in metropolitan areas is empirically

Table 3. Descriptive statistics on the ethnic composition in German zip codes

	Females				Males			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Panel A: Shares at the zip code level								
Foreign share (in %)	10.245	6.431	1.002	39.336	10.221	6.251	0.920	39.336
Ethnic share (in %)	1.206	1.522	0.000	14.359	1.330	1.898	0.000	27.577
Different countries of origin (in %)								
Italy	1.113	0.963	0.000	6.066	1.114	0.951	0.000	5.456
Turkey	3.748	3.892	0.000	27.577	3.715	3.796	0.000	27.577
Greece	0.725	0.594	0.000	3.720	0.727	0.626	0.000	4.803
Spain, Portugal	0.288	0.250	0.000	2.419	0.295	0.280	0.000	2.419
Latin America								
The Balkans	1.349	1.230	0.000	7.103	1.324	1.186	0.000	7.103
Eastern Europe	0.986	0.762	0.000	6.147	1.003	0.694	0.000	6.147
Sub-Saharan Africa	0.184	0.229	0.000	1.719	0.185	0.234	0.000	2.046
Islam	0.347	0.425	0.000	2.690	0.352	0.433	0.000	2.690
Asia	0.123	0.136	0.000	1.633	0.122	0.135	0.000	1.633
Other Western	1.038	0.430	0.130	2.907	1.023	0.436	0.130	2.609
Ethnic German	0.318	0.323	0.000	2.150	0.336	0.334	0.000	1.794
Panel B: Instruments at the level of zip code areas								
Foreign share (in %)	8.809	4.033	2.091	20.069	8.709	4.019	2.091	20.069
Ethnic share (in %)	0.945	0.945	0.030	8.883	1.003	1.057	0.022	8.883
Panel C: Instrument at the federal state level								
Language course density (per 100,000 inhabitants)	12.489	4.884	2.205	32.855	12.364	4.470	2.205	32.855
Observations	1,702				1,858			
Individuals	1,312				1,192			

Source: Microm Raster data; Federal Statistical Office; Federal Office for Migration and Refugees; own illustration.
Notes: SD denotes standard deviations, and min and max provides minima and maxima of variables. The Balkans includes Albania, Bulgaria, Hungary, Romania, and the former Yugoslavia. East Europe is the Czech Republic, Poland, Slovakia, and the (former) Soviet Union. Islam covers Northern Africa, Iran, the Middle East, Pakistan, and Muslims from Southeast Asia. Asia includes China, India, Japan, Korea, Sri Lanka, Thailand, and Vietnam. Other Western countries are Benelux, France, Great Britain, Northern Europe, Australia, Canada, New Zealand, and the United States.

observable for Germany. For instance, as a consequence of bilateral agreements on recruiting guest-worker with Turkey in 1961 and Yugoslavia in 1968, immigrants from those countries are still the two largest ethnic communities in certain areas (Federal Office for Migration and Refugees, 2005). If the residency choice is pre-determined by traditional migration patterns, it seems unreliable to call recent immigrants' choices exogenous. However, previous papers suppose that historical population numbers do not directly affect individuals' current labor market performance (Card, 2001; Patacchini and Zenou, 2012).

Despite these arguments that support my identification strategy, Jaeger *et al.* (2018) recently uncovered one drawback of the underlying identification assumption. They point out that the slow speed of adjustment to previous demand shocks and the high correlation between lagged and current ethnic composition results in the conflation of short- and long-run responses to immigration shocks. Following their argument, adjustment to previous shocks produces a direct correlation of the instrument with the current labor market performance of individuals. This drawback also applies to my IV approach, which should be kept in mind throughout the paper.

The German equivalent of the metropolitan area is the extended zip code area, which is identified by the first two digits of the five-digit zip code number. In Germany, there are 99 different zip code areas. Zip code areas are on a coarser level of local labor market regions. Thus, only commuting a large distance would make it possible for personal networks to work on such a level (Kosfeld and Werner, 2012; Schaffner and Treude, 2014). Descriptive statistics on the ethnic composition at this aggregated level are displayed in Panel B of Table 3. I argue that the application of this IV approach demonstrates a negative relationship between the ethnic share and years of overqualification ($\beta_1 < 0$). This approach enables consideration of the whole population of migrants and does not restrict the analysis to particular migrant groups, which is required when residential assignments for certain migrant groups are exploited. The second advantage is that migrants are not required to stay in the destination country for a minimum duration, which is required when instrumenting the current residency by the initial residency in exploiting residential assignments (Boeri *et al.*, 2012; Damm, 2009, 2014; Edin *et al.*, 2003; Danzer and Yaman, 2016). To take the discussions by Jaeger *et al.* (2018) into account, I provide an empirical analysis of whether the exclusion restriction holds. In Section 6.2, I examine whether economic differences between zip codes are canceled out at the aggregated level of zip code areas and whether negative selection is still observable at this level.

Additionally, in Panel C of Table 3, I present the supply of language courses across federal states as a second instrument for the foreign share per zip code. The supply of language courses is randomly distributed across Germany with large variations between and within federal states. The supply depends on the number of course providers but does not follow indicators such as the foreign share and the unemployment rate of foreign-born people, which may predict the actual demand for language courses (Federal Office for Migration and Refugees, 2009). The non-existence of local authorities that may estimate the local demand for language courses supports my argument, whereas the language density per federal state is exogenous regarding the individual's labor market performance. To calculate this variable, I merged spatial data on the number of language courses from the German Federal Office for Migration and Refugees (BAMF) to population numbers per federal state from the German Federal Statistical Office.

5. Main empirical results

5.1. Baseline results

A closer look at the sample composition by gender in Tables 1 and 2 indicates that females and males select into different occupations and legal status. Thus, Table 4 reports empirical results of estimating equation [4] with random effects separately by gender. Years of overqualification are regressed on the logarithmized ethnic share $logES_{jkt}$ and the logarithmized foreign share $logFS_{jt}$. Each model controls for the covariates given in Table 2, including dummies for countries of origin δ_k, residency status, and occupational sectors.

If region dummies are included in Model 4 of Panel A, OLS estimation results in a small negative effect of ethnic share on years of overqualification for females, which is insignificant. On the contrary, if the ethnic share and the foreign share per zip code are instrumented by the respective share at the zip code area level, the coefficient of the logarithmized ethnic share shows a significant impact on years of overqualification. IV results in Models 2–4 of Panel A show that an increase in the ethnic share significantly reduces the years of overqualification for females. The significant effect of the foreign share in Model 1 becomes insignificant after controlling for the share of the own ethnic group. After including area dummies in Model 4, a one standard deviation increase in the ethnic share significantly reduces years of overqualification by almost four months (or 0.27 years). This effect is highly significant and more than eight times larger than the respective OLS estimation in Model 4. This confirms the expected downward bias of β_1 in OLS estimations and demonstrates the supposed negative association between ethnic networks and years of overqualification.

The first stage, reported in the lower panel of Table 4, illustrates that both instruments are highly relevant, with the lowest F-statistics of foreign shares of 12.0 after including dummies for areas. The results are different for males — neither the foreign share nor the ethnic share results in significant effects. Also, squared terms of the foreign share and the ethnic share do not identify a significant U-shaped impact on education–occupation mismatch. Using the mode instead of the mean when calculating average years of schooling for occupations (see Kiker *et al.*, 1997; Piracha and Vadean, 2013) and modeling shares in absolute terms, without logarithmizing, leave the results unchanged.[17] OLS estimations for males demonstrate the importance of distinguishing between the general foreign-born population and the own ethnic group. The significant effect of the foreign share in Model 1 becomes insignificant after controlling for the ethnic share in Models 3 and 4.[18] Instead of a continuous measurement of occupational mismatch, I adopt the approach of Kalfa and Piracha (2018) and model a binary variable to capture the probability of overqualification. This tackles the disadvantage that undereducation and overeducation are conflated into one variable in equation [1]. For females, a one standard deviation increase in the ethnic share decreases the probability of being overqualified by at least 1.5 and 2 years by 7.1 and 5.7 percentage points, respectively (IV estimates in Table A2). Again, no significant effects can be detected for males.

My results contradict Kalfa and Piracha (2018), who found that ethnic concentration increases the probability of overqualification, particularly for females. While there are many differences in their study, such as the different juridical migration systems in Australia and Germany and the different compositions of the foreign-born populations, three main transmission channels explain both a positive effect from ethnic networks and the different results by gender. First, a closer look at the sample composition, illustrated in Table 2, demonstrates that female immigrants are more likely to choose white-collar jobs or work in the public sector. Thus, differential findings by gender are partly driven by selection into the labor market. Such occupational selection can be explained by observable characteristics. In white-collar jobs and the public sector, command of the host country's official language and (partially) German citizenship is essential conditions (Gathmann and Keller, 2017).[19] This argument is supported by prior research, which emphasizes larger gains from language skills and citizenship for females regarding their labor market performance (Yao and van Ours, 2015; Gathmann and Keller, 2017). Table 2 illustrates that females are more likely to display good initial and current German language skills and to be naturalized. Thus, language is one potential transmission channel from ethnic networks to education–occupation mismatch and drives selection into different occupations.

Second, besides the differential selection into the labor market, the recognition of vocational degrees acquired abroad plays a relevant role in immigrants' labor market outcomes (Brücker *et al.*, 2018; Tani, 2017). On the one hand, based on Bertrand *et al.* (2000), it can be expected that stronger links to ethnic networks and more limited contact with natives

Table 4. The effects of ethnic networks and residential segregation on education–occupation mismatch by gender

	Model 1		Model 2		Model 3		Model 4	
	OLS	**2SLS IV**	**OLS**	**2SLS IV**	**OLS**	**2SLS IV**	**OLS**	**2SLS IV**
Panel A: Females								
Networks & Segregation								
(Log) ethnic share			−0.0202	−0.1533***	−0.0172	−0.1502***	−0.0254	−0.2128***
			(0.0135)	(0.0552)	(0.0127)	(0.0586)	(0.0192)	(0.0715)
(Log) foreign share	−0.0471	−0.2318*			−0.0254	−0.0141	−0.0110	−0.0800
	(0.0441)	(0.1200)			(0.0452)	(0.1240)	(0.0576)	(0.3430)
F-statistic of 1st Stage:								
(Log) ethnic share (per zip code area)				32.84***		17.57***		27.50***
(Log) foreign share (per zip code area)		78.14***				50.12***		12.02***
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies							Yes	Yes
R^2	0.2383	0.2335	0.2394	0.2285	0.2392	0.2284	0.3857	0.3912
Observations	1,702	1,702	1,702	1,702	1,702	1,702	1,702	1,702
Individuals	1,312	1,312	1,312	1,312	1,312	1,312	1,312	1,312
Panel B: Males								
Networks & segregation								
(Log) ethnic share			0.0122	−0.1132	0.0039	−0.0899	0.0015	−0.0193
			(0.0077)	(0.3542)	(0.0053)	(0.1265)	(0.0026)	(0.0728)
(Log) foreign share	0.1510**	0.1178			0.1438	0.1645	0.0551	0.2360
	(0.0661)	(0.1495)			(0.0704)	(0.1536)	(0.0510)	(0.4054)
F-statistic of 1st Stage								
(Log) ethnic share (per zip code area)				0.26		3.80**		62.77***
(Log) foreign share (per zip code area)		27.40***				53.39***		31.63***
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies							Yes	Yes
R^2	0.2088	0.2100	0.2100	0.1881	0.2091	0.2074	0.3552	0.3860
Observations	1,858	1,858	1,858	1,858	1,858	1,858	1,858	1,858
Individuals	1,192	1,192	1,192	1,192	1,192	1,192	1,192	1,192

Source: IAB-SOEP Migration Sample, own illustration.
Note: $*p < 10\%$, $**p < 5\%$, $***p < 1\%$; heteroskedastic robust standard errors in parentheses.
Table 4 presents OLS and 2SLS IV estimations of equation [4] with random effects separately for females and males. Years of overqualification y_{ijkt} are regressed on the log. ethnic share $logES_{jkt}$ and on the log. foreign share $logFS_{jt}$. Additional controls are gradually added and include dummies for countries of origin, occupations, regions, and covariates given in Table 2.

will prevent immigrants from accessing information about the German system of recognition of vocational degrees. On the other hand, the opposite effect could also be conceivable, as ethnic networks could also help when collecting information (e.g., sharing information about which documents are required by the administration). In my sample, 39.4 percent of immigrants with a vocational degree applied for recognition. Further analyses not shown in this paper display that, conditional on the set of covariates in Table 2, exposure to ethnic networks and the effort to apply for recognition are negatively correlated.[20] However, if the estimations are repeated with additional control on occupational recognition, the effect strength of the ethnic networks does not change significantly. Males do significantly apply more often for recognition. Thus, occupational recognition is a further factor that explains differential results by gender. Third, residency status is an additional transmission channel that explains the gender difference and displays one further determinant that drives occupational decisions and the tendency to connect to ethnic networks. For instance, family migrants, the largest group of residency status in the female sample, face better outside options of being occupationally mismatched and face less pressure to quickly find a job after migration. Thus, selection into different groups of residency status also explains why females benefit more efficiently from ethnic networks.

When discussing transmission channels, it should be borne in mind that selective out-migration decisions affect the sample composition and the distribution of qualification (for an overview of the basic theory, see Borjas and Bratsberg, 1996). First, this kind of selection affects the measurement of occupational mismatch (Poot and Stillman, 2016), and second, outmigration patterns differ by gender (Bijwaard, 2010) and ethnic groups.

One further issue claims that network effects are driven by females from traditional countries that emphasize conservative roles for women, with generally low labor market participation. This is indicated by a lower share of employment in the source country of 68.3 percent for females in Table 2. However, the validity of this claim would result in a downward bias of β_1. Thus, the presented results are lower bounds. Nevertheless, this claim justifies a closer examination of heterogeneities for females as tackled in the next section.

5.2. Heterogeneities

Bentolila *et al.* (2010) suppose that the exclusion from social benefits and less attractive outside options of employment cause a stronger association between ethnic networks and education–occupation mismatch. Eligibility for social benefits depends on the country of origin and residency status. Since females and males choose different residency statuses, examining heterogeneous effects is helpful to explain gender differences.[21] For female non-EU citizens, a larger ethnic share significantly reduces the years of overqualification, as shown in Table 5, while no significant effect can be detected for female EU citizens and female immigrants from countries of the EU enlargement in 2004. The insignificance in Models 2 and 3 can be explained by the insufficient relevance of the ethnic share per zip code area in the first stage. This is because the largest ethnic communities are of a non-Union nature (Turkey and the Balkans).

Concerning residency status, ethnic Germans are a particular group in the context of ethnic networks (Glitz, 2012; Schaffner and Treude, 2014). For instance, only this group faced a sustained residential assignment between 1989 and 2009 in Germany (Federal Office for Migration and Refugees, 2007). By excluding ethnic Germans in Model 4, the impact of ethnic share is strengthened. Less restricted access to German citizenship and ingrained roots in Germany and the German language explain this finding.

Table 5. Heterogeneous effects for females regarding origin and residency status at arrival

	Country of Origin			Residency Status at Arrival		
	(1) Non-EU Citizens	**(2) EU Citizens**	**(3) Countries of EU Enlargement in 2004**	**(4) Ethnic Germans Excluded**	**(5) Family Members**	**(6) West Germany**
Networks & Segregation						
(Log) ethnic share	−0.1537**	−0.0753	0.0057	−0.2359***	−0.1761*	−0.2751***
	(0.0746)	(0.0655)	(0.0727)	(0.0749)	(0.0940)	(0.0838)
(Log) foreign share	−0.0233	−0.1886	−0.1079	0.0852	−0.0044	−0.0544
	(0.1979)	(0.1724)	(0.2482)	(0.1688)	(0.2103)	(0.3103)
F-statistic of 1st Stage						
(Log) ethnic share (per zip code area)	35.90***	2.49*	1.24	19.17***	6.70***	63.21***
(Log) foreign share (per zip code area)	22.04***	15.10***	23.90***	27.61***	18.11***	12.06***
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.1915	0.2736	0.2930	0.2690	0.3475	0.3933
Observations	904	798	322	1,382	657	1,542
Individuals	609	583	241	991	447	1,082

Source: IAB-SOEP Migration Sample, own illustration.
Note: $^*p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; heteroskedastic robust standard errors in parentheses.
Table 5 presents two-stage least squares (2SLS) IV estimations of equation [4] with random effects for females, where years of overqualification are regressed on the log. ethnic share $logES_{jkt}$ and on the log. foreign share $logFS_{jt}$. Each model controls on dummies for countries of origin, occupations, regions γ_j, and includes the covariates given in Table 2. Models 1–3 exclusively consider certain countries of origin. Models 4–5 stratify on residency status at arrival. Model 6 excludes migrants who live in federal states in East Germany (the former German Democratic Republic [GDR]).

For family migrants, economic resources and family support are often available in the country of destination (Chiswick *et al.*, 2005; Cortes, 2004). Empirical evidence shows a significant relationship between ethnic networks and education–occupation mismatch. However, the effect is lower compared to the results in Table 4. This confirms the expectation expressed at the end of Section 5.1 that the exclusive consideration of females from traditional and conservative countries underestimates the effect of ethnic networks.

Model 6 has a special focus on federal states in West Germany because they contain a larger foreign-born community compared to East Germany (the former German Democratic Republic [GDR]). IV regressions show that the impact of ethnic share on years of overqualification is raised to 0.35 years from a one standard deviation increase. Results not presented in this paper do not show any significant effect for East Germany.

6. Further examinations

6.1. Robustness checks

Before Section 6.2 takes a detailed look at the exclusion restriction of the underlying IV approach, Section 6.1 tackles a set of economic and methodological claims considering female immigrants.[22] First, the calculation of average schooling years $YearsEd_{Occ(i)}$ in equation [1] may be biased because different factors between women and men drive educational choices (see, for instance, Reuben *et al.*, 2015; Rapoport and Thibout, 2018). Second, the calculation of required schooling years per occupation can be critical as the value of one schooling year depends on the underlying educational system and on the country in which the school was attended. To tackle those claims, average schooling years for women and men are calculated separately in Model 1 of Table 6. In Model 2, computing average schooling years per prestige score is stratified on the country of origin groups. Both alternative specifications show significant effects and result in only slightly decreased impacts of ethnic networks.

The third claim supposes that zip code areas as the instrumental level are chosen arbitrarily. To illustrate why this level of geography is the correct level, I instrument the foreign share and ethnic share per zip code by alternative instruments. First, I instrument the ethnic composition of zip codes by the respective variables at the zip code department level (Model 3 of Table 6). The zip code department corresponds to the first digit of the five-digit zip code. Thus, there are ten different zip code departments in Germany. Alternatively, I apply the language course density per 100,000 inhabitants at the federal state level as the instrument for the foreign share across zip codes (Model 4). In both specifications, the effect of ethnic share is only weakly significant. The low relevance in the first stage demonstrates that federal states and zip code departments are too coarse and cannot produce sufficient variation in ethnic composition. Such coarse levels lose information about the variation in ethnic share within each zip code department because both rural and urban structures characterize German departments.[23] Econometric theory highlights that low variation in the instrument is a drawback in IV settings and potentially creates imprecise estimates (Imbens and Angrist, 1994). This can be seen in Table 6 by the broad range of coefficients of – 0.19 and – 0.34, which strengthens the assumption that the zip code area level creates enough variation in ethnic composition and ensures a high relevance of the instruments in the first stage.

Table 6. IV Estimations: robustness checks for females

	Years of Overqualification Stratified by···		Alternative Instruments				
	(1) Gender	(2) Countries	(3)	(4)	(5) Unemployment Rate	(6) Age at migration ≥ 18	(7) Arrival Cohort Trends
Networks & Segregation							
(Log) ethnic share	−0.2094***	−0.1998**	−0.3371*	−0.1935*	−0.2185***	−0.2060***	−0.2132***
	(0.0694)	(0.0865)	(0.1948)	(0.1107)	(0.0711)	(0.0726)	(0.0712)
(Log) foreign share	−0.1348	−0.0622	−1.8846	−0.2575	−0.2032	−0.2158	0.0334
	(0.3393)	(0.3264)	(2.7284)	(0.7137)	(0.5055)	(0.3400)	(0.3920)
Unemployment Rate					0.0387		
					0.0385		
F-statistic of 1st Stage							
(Log) ethnic Share (per zip code area)	23.75***	22.65***		11.68***	29.08***	30.43***	32.59***
(Log) foreign share (per zip code area)	11.70***	11.02***			4.38***	10.98***	10.21***
Language Course Density (per Federal State)				7.94***			
(Log) ethnic share (per zip code Department)			25.57***				
(Log) foreign share (per zip code Department)			1.24				
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.3960	0.3641	0.2681	0.3881	0.3898	0.3868	0.4148
Observations	1,702	1,702	1,702	1,702	1,702	1,658	1,702
Individuals	1,312	1,312	1,312	1,312	1,312	1,165	1,312

Source: IAB-SOEP Migration Sample, own illustration.
Note: $*p < 10\%$, $**p < 5\%$, $***p < 1\%$; heteroskedastic robust standard errors in parentheses.
Table 6 presents two-stage least squares (2SLS) IV estimations of equation [4] with random effects for females. In each model, years of overqualification are regressed on the log. ethnic share $logES_{jkt}$ and on the log. foreign share $logFS_{jt}$, and controls on dummies for countries of origin, occupations, regions, and the covariates given in Table 2. Models 1 and 2 use alternative indicators of occupational mismatch. In Model 1, the calculation of average schooling years per prestige is stratified by gender. In Model 2, the calculation of average schooling years is separately computed with respect to country of origin groups. IV estimations 3 and 4 use two different kinds of instruments. In Model 3, the foreign and the ethnic share per zip code are instrumented by the respective shares at the level of zip code departments (first digit of the zip code number). In Model 4, the foreign share across zip codes is instrumented by language course density across federal states. In addition, Model 5 controls on unemployment rate across zip codes and Model 6 excludes individuals which arrived under the age of 18 in Germany. Finally, specification 7 includes dummies for arrival cohorts.

Models 5–7 show the robustness of my findings to some minor issues. Model 5 controls for unemployment rates across zip codes to make sure that the effect of the ethnic share can be attributed to the ethnic composition of an area. This check rules out the possibility that the social deprivation of areas drives the network effect. Furthermore, I exclude childhood immigrants in Model 6 and add dummies for immigration years in Model 7.

6.2. A critical inspection of the exclusion restriction

Besides the assumption of relevance, I assumed that the residential choice of the zip code area is random. Here, I rely on the theoretical justifications made by Card (2001), Cascio and Lewis (2012), and Patacchini and Zenou (2012). Although this IV approach has already been transferred from Bertrand *et al.* (2000) by several such as Deri (2005) and Andersson *et al.* (2014), empirical validation of the exogeneity assumption has so far been avoided. Moreover, Jaeger *et al.* (2018) underlined channels that compromise the exogeneity of my instrument (see Section 4).

To fill this research gap, I undertake spatial analyses by splitting my sample into groups. Migrants living in an area with a larger ethnic share than a certain critical value are assigned to the treatment group, while those living in an area with a lower ethnic share than this limit, are assigned to the control group. I define this critical value twice: First, this limit is equivalent to the lower quartile of the ethnic share and the foreign share. Second, it equals the median of the shares. In the empirical analysis, negative selection into zip codes is detected by downward biases of β_1, so that immigrants move to districts with poor economic conditions. Table 7 displays how treatment and control regions differ concerning observable indicators of the economic structure of areas. This comparison is made for the ethnic share at the level of zip codes on the left-hand side and its instrument at the level of zip code areas on the right-hand side. First, Table 7 demonstrates a negative selection into zip codes. Zip codes with a larger ethnic share than the lower quartile of the ethnic share (treatment group) differ significantly concerning unemployment rates, purchasing power per capita, and share of the population aged above 60. The ethnic share is positively correlated with the unemployment rate and the purchasing power per capita and negatively associated with the share of the population above the age of 60. However, at the level of zip code areas, those differences between treatment and control individuals disappear. Unemployment rates and the share of over-60s also no longer differ significantly. Furthermore, numbers for each specific wave show that the trends are congruent in each variable. However, a significant difference is detected for purchasing power per capita, which is puzzling in general and contradicts economic expectations, as the ethnic share and the foreign share are expected to be larger in socially deprived areas. However, higher purchasing power in the treatment group would result in an upward bias of the effect of ethnic share on years of overqualification.

Nevertheless, choices of zip code area can still be driven by endogenous decisions and (unobservable) individual characteristics, just as treatment and control regions may differ concerning the unobservable characteristics of the regions (see Jaeger *et al.*, 2018). This should be kept in mind and requires caution when deriving policy implications from my paper. As a final step, Table 8 demonstrates that immigrants living in a treatment region display significantly fewer years of overqualification (0.8 years), while residential segregation indicated by a large foreign share does not significantly affect years of overqualification. However, no significant effect of ethnic networks is found when the critical value is defined as the median (0.7 percent for the ethnic share). This illustrates one interesting point: While ethnic networks must have a large enough size to generate a substantial effect

Table 7. Inspection of the balancing property for females concerning the lower quartile of ethnic shares

	At the level of zip codes			At the level of zip code areas		
	Treatment	**Control**	**Mean Diff.**	**Treatment**	**Control**	**Mean Diff.**
Overall sample						
Unemployment rate (in %)	8.102	6.893	−1.209***	7.738	7.985	0.248
Purchasing Power per Capita (in)	21,688.6	21,275.5	−413.1*	21,814.6	20,898.0	−916.6***
Share of Over-60s (in %)	3.399	3.611	0.212***	3.436	3.500	0.064
Specific waves						
Unemployment rate (in %)						
2013 ($N = 810$)	8.373	6.804	−1.569***	7.911	8.155	0.244
2014 ($N = 573$)	8.034	7.245	−0.789*	7.725	8.039	0.314
2015 ($N = 319$)	7.579	6.306	−1.273**	7.347	7.262	−0.085
Purchasing Power per Capita (in)						
2013 ($N = 810$)	21,087.2	21,029.4	−57.7	21,251.0	20,512.6	−738.4**
2014 ($N = 573$)	22,051.6	21,438.5	−613.1*	22,136.5	21,292.7	−843.8***
2015 ($N = 319$)	22,508.2	21,690.7	−817.5	22,647.27	21,034.3	−1, 613.4**
Share of Over-60s (in %)						
2013 ($N = 810$)	3.429	3.627	0.198**	3.455	3.557	0.102
2014 ($N = 573$)	3.452	3.703	0.250**	3.508	3.551	0.043
2015 ($N = 319$)	3.242	3.329	0.087	3.281	3.163	−0.119
Observations	1,276	426	1,702	1,276	426	1,702
Individuals	984	328	1,312	984	328	1,312

Source: IAB-SOEP Migration Sample; Federal Office for Migration and Refugees; own illustration.
Notes: *p < 10%, **p < 5%, ***p < 1%.
Table 7 shows differences in unemployment rate, purchasing power per capita, and share of the population over the age of 60 for treatment individuals and control individuals for the full sample and specific waves. This is done twice: first, concerning the geographical level of zip codes, and second, at the level of zip code areas. Migrants living in a zip code (area) with a higher ethnic share than the lower quartile of the ethnic share (0.4 percent) are assigned to the treatment group, whereas those living in an area with a lower ethnic share than the lower quartile are assigned to the control group.

on labor market outcomes, this size does not need to be larger than the average. This is consistent with the finding in Section 5.2, whereas no effect of ethnic networks could be detected for East Germany. Presumably, an ethnic share larger than average is both an indicator of a large ethnic community and social deprivation.

7. Conclusion

Do ethnic networks causally affect occupational mismatch? By focusing on this unexplored research question for the case of Germany, I extend the current state of research in several ways: First, I tackle the scarcity of research on the determinants of education–occupation mismatch by focusing on the question of whether ethnic networks affect overqualification. Second, I particularly consider gender differences and use this to examine the transmission channels that link ethnic networks to occupational mismatch. Third, I exploit supraregional variation in the ethnic composition of German zip codes and apply the IV approach of Bertrand *et al.* (2000) to novel data from Germany. Furthermore, regarding recent findings by Jaeger *et al.* (2018) that challenge the reliability of the IV approach, I conduct a critical inspection of this method. Empirical analyses provide hints to the fact that negative selection is indeed partly

Table 8. Regression results for females with binary information about residential segregation and ethnic networks

	Lower quartile		Median	
	(1) OLS	**(2) 2SLS IV**	**(3) OLS**	**(4) 2SLS IV**
Networks & Segregation				
Large ethnic share	−0.0554**	−0.8445*	0.0008	0.5168
	(0.0237)	(0.4468)	(0.0361)	(0.8942)
Large foreign share	0.0631	5.1820	−0.0022	−1.4452
	(0.0904)	(3.3247)	(0.0198)	(1.500)
F-statistic of 1st Stage				
Large ethnic share (per zip code area)		122.90***		3.00**
Large foreign share (per zip code area)		3.11**		1.15
Additional controls	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes
R^2	0.3841	0.2254	0.3829	0.3468
Observations	1,702	1,702	1,702	1,702
Individuals	1,312	1,312	1,312	1,312

Source: IAB-SOEP Migration Sample, own illustration.
Notes: *p < 10%, **p < 5%, ***p < 1%; heteroskedastic robust standard errors in parentheses.
Table 8 displays ordinary least squares (OLS) and two-stage least squares (2SLS) IV estimations of equation [4] with random effects for females. In each model, the years of overqualification are regressed on dummies, which indicate whether an individual i lives in a zip code (area) with a large ethnic share and a large foreign share. Estimations also control for dummies for countries of origin, occupations, regions, and the covariates provided in Table 2. In Models 1 and 2, a large ethnic and foreign share is indicated if an individual i lives in a zip code (area) with an ethnic share and a foreign share above the lower quartile, whereas in Models 3 and 4, these dummies are equal to 1 if the shares are at least its median.

canceled out at the aggregated level of zip code areas. However, closer inspection uncovered reasons that compromise the exclusion restriction.[24]

The applied analysis demonstrates a negative marginal effect of the ethnic share per residency area on overqualification for females, by about 0.27 years. For males, neither the general foreign share nor the ethnic share per residency area is found to affect occupational mismatch. Empirically, females are more likely to choose occupations in which command of the host country's official language and higher levels of education are required. This explains why women benefit more from ethnic networks. Females in my sample have access to better economic resources due to better initial conditions on arrival, higher levels of education, and family support. This initial position expresses better outside options for occupational mismatch and facilitates a more efficient input of networks. Furthermore, I show that if ethnic Germans are excluded from the sample or if only federal states in West Germany are considered, the link between ethnic networks and education–occupation mismatch becomes tighter and stronger for females. Furthermore, the link becomes weaker if only family members are considered. This underlines that my findings are not merely driven by females from traditional countries in which women generally have a low level of labor market participation.[25]

The reliability of my results is highlighted by agreeing with prior empirical studies that apply different methods (Dustmann *et al.*, 2016b). The linkage to the own ethnic group decreases the extent of occupational mismatch for females, a finding that emphasizes and confirms that networks are ethnically stratified (Borjas, 1998; Damm, 2014; Edin *et al.*, 2003, etc.) and that networks can be beneficial in terms of the labor market integration of immigrants, by reducing information asymmetries between worker and employer.

Appendix A

Figure A1. The Distribution of Years of Overqualification Regarding Residency Status at Arrival and German Language Skills. *Source*: IAB-SOEP Migration Sample, own calculations. *Note:* The four graphs display the density of years of overqualification regarding residency status at arrival and German language skills. The left-hand graph of Panel (a) provides the density for female migrants that arrived either as a family member, ethnic German or asylum seeker in Germany. The right-hand graph of Panel (a) illustrates the density of years of overqualification, depending on whether the individual has German language skills or not. Panel (b) illustrates the respective statistics for males.

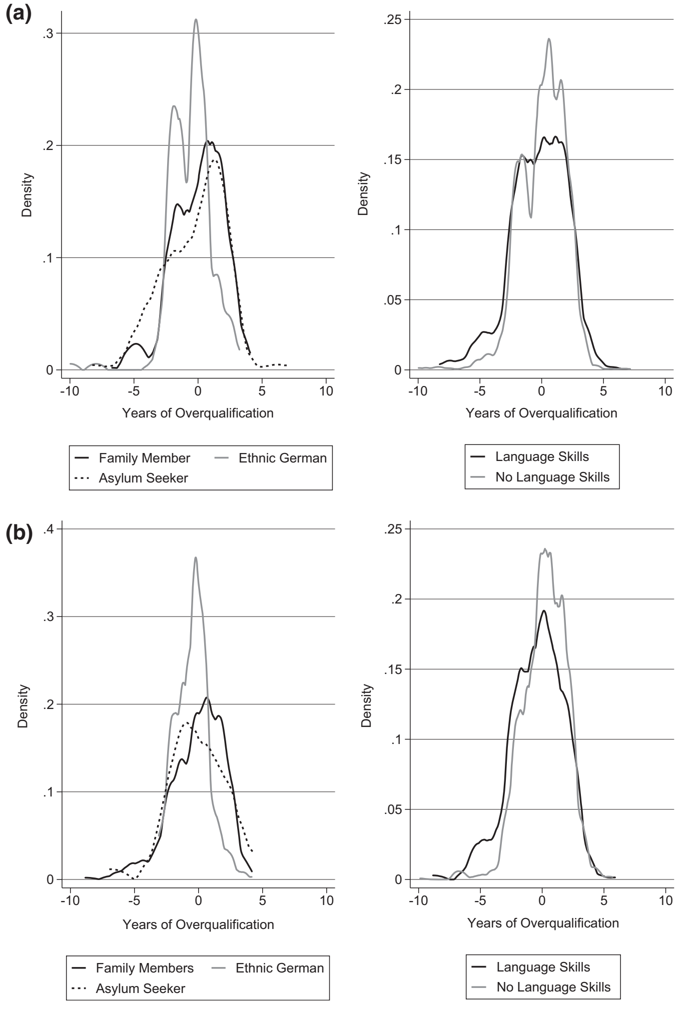


Table A1. The pooled impact of the ethnic share and the foreign share on years of overqualification

	(1) OLS, No Region FE		(2) 2SLS IV, No Region FE		(3) OLS, Region FE		(4) 2SLS IV, Region FE	
	Coeff.	**Rob. SE**	**Coeff.**	**Rob. SE**	**Coeff.**	**Rob. SE**	**Coeff.**	**Rob. SE**
Networks & segregation								
(Log) ethnic share	0.0009	0.0056	−0.0800*	0.0428	−0.0046*	0.0026	−0.1084*	0.0585
(Log) foreign share	0.0497	0.0442	0.1176	0.0799	0.0110	0.0352	−0.3216	0.2456
Socioeconomic & educational variables								
Age	−0.0121***	0.0045	−0.0125***	0.0045	−0.0100**	0.0046	−0.0098**	0.0047
Years of residence	−0.0181***	0.0063	−0.0253***	0.0070	−0.0215***	0.0006	−0.0278***	0.0069
Relationship	−0.0693**	0.0342	−0.0946	0.0636	−0.0672*	0.0384	−0.1178**	0.0512
High education	0.8786***	0.2158	1.3803***	0.1180	0.8667***	0.2284	0.9786***	0.1895
Middle education	0.8550***	0.2113	1.3683***	0.1049	0.8416***	0.2232	0.9862***	0.1939
Residency status at arrival (reference group: ethnic Germans)								
Family member	0.6227	0.6034	0.7871	0.6082	0.1961	0.1245	0.3104**	0.1530
Asylum seeker	0.6673	0.6216	0.8789	0.6264	0.3292*	0.1985	0.5211**	0.2273
Job searcher	0.5005	0.6090	0.6772	0.6130	0.0429	0.1435	0.1643	0.1705
With job commitment	0.6210	0.6085	0.7824	0.6134	0.1869	0.1493	0.2719	0.1719
Other status groups	0.5127	0.6099	0.6309	0.6138	0.0684	0.1575	0.1578	0.1707
Countries of origin								
Member of EU	0.0388	0.1200	0.0692	0.1216	0.1324	0.1223	0.1844	0.1305
Countries of EU Enlargement in 2004	0.7638***	0.1499	0.6765***	0.1539	0.6090***	0.1554	0.5193***	0.1646
Turkey	−1.2874***	0.2536	−0.9844***	0.2584	−1.4104***	0.2480	−1.0682***	0.2865
Afghanistan, Iran, Iraq, Syria, and Lebanon	0.09768	0.3273	0.1098	0.3237	0.1115	0.3311	0.1059	0.3314
Guest-worker countries	0.0016	0.1374	0.0501	0.1342	0.1336	0.1355	0.1441	0.1355
Russia, (Foreign) USSR	−0.7806***	0.1104	−0.7434***	0.1155	−0.7220***	0.1143	−0.6322***	0.1264
F-statistic of 1st stage								
(Log) ethnic share (per zip code area)			121.92***				27.37***	
(Log) foreign share (per zip code area)			647.22***				22.32***	
Additional control variables								
Integration & pre-migration variables	Yes		Yes		Yes		Yes	
Current Residency Status	Yes		Yes		Yes		Yes	

Table A1. Continued

	(1) OLS, No Region FE		(2) 2SLS IV, No Region FE		(3) OLS, Region FE		(4) 2SLS IV, Region FE	
	Coeff.	**Rob. SE**	**Coeff.**	**Rob. SE**	**Coeff.**	**Rob. SE**	**Coeff.**	**Rob. SE**
Occupational Sector	Yes		Yes		Yes		Yes	
Region Dummies					Yes		Yes	
R^2	0.2119		0.2215		0.3101		0.3007	
Observations	3,560		3,560		3,560		3,560	
Individuals	2,504		2,504		2,504		2,504	

Source: IAB-SOEP Migration Sample, own illustration.
Notes: $*p < 10\%$, $**p < 5\%$, $***p < 1\%$; standard errors are heteroskedastic robust.
Table A1 presents ordinary least squares (OLS) and two-stage least squares (2SLS) IV estimations of equation [4] with random effects and heteroskedastic robust standard errors SE for the overall sample. In each estimation, the years of overqualification are regressed on the log. ethnic share $logES_{jkt}$ and on the log. foreign share $logFS_{jt}$ and on the covariates given in Table 2.

Table A2. The effects of ethnic networks and residential segregation on the probability of overqualification

	(1) +1.5 years of overqualification		**(2) +2.0 years of overqualification**	
	OLS	**2SLS IV**	**OLS**	**2SLS IV**
Panel A: Females				
Networks & segregation				
(Log) ethnic share	−0.0147*	−0.0566***	−0.0137*	−0.0452***
	(0.0089)	(0.0164)	(0.0079)	(0.0152)
(Log) foreign share	0.0157	0.0144	0.0217	0.0008
	(0.0280)	(0.1020)	(0.0285)	(0.0980)
F-statistic of 1st stage				
(Log) ethnic share (per zip code area)		76.27***		75.19***
(Log) foreign share (per zip code area)		22.13***		21.30***
Additional controls	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes
R^2	0.2998	0.2959	0.2727	0.2728
Observations	1,702	1,702	1,702	1,702
Individuals	1,312	1,312	1,312	1,312
Panel B: Males				
Networks & segregation				
(Log) ethnic share	0.0003	−0.0153	0.0026	−0.0053
	(0.0013)	(0.0195)	(0.0028)	(0.0153)
(Log) foreign share	0.0207	−0.0865	−0.0317	0.1840
	(0.0143)	(0.0844)	(0.0441)	(0.1267)
F-statistic of 1st stage				
(Log) ethnic share (per zip code area)		40.38***		24.79***
(Log) foreign share (per zip code area)		20.29***		55.18***
Additional controls	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes
R^2	0.2292	0.2273	0.1960	0.2555
Observations	1,858	1,858	1,858	1,858
Individuals	1,192	1,192	1,192	1,192

Source: IAB-SOEP Migration Sample, own illustration.
Notes: *$p < 10\%$, **$p < 5\%$, ***$p < 1\%$; heteroskedastic robust standard errors in parentheses.
Table A2 presents ordinary least squares (OLS) and two-stage least squares (2SLS) IV estimations of equation [4] with random effects, whereas the continuous variable years of overeducation is replaced by a binary variable. In Model 1, the dependent variable y_{ijkt} equals 1 if individual *i* is overqualified by at least 1.5 years, and in Model 2, overqualification is indicated if years of overeducation is at least 2 years. In both models, this binary outcome is regressed on the log. ethnic share $\log FS_{jt}$ and on the log. ethnic share $\log ES_{jkt}$, and it is controlled on countries of origin, area dummies, occupational dummies, and covariates given in Table 2.

Table A3. Heterogeneous effects for males regarding origin and residency status at arrival

	Country of Origin			Residency status at arrival		
	(1) Non-EU citizens	**(2) EU citizens**	**(3) Countries of EU enlargement in 2004**	**(4) Ethnic Germans excluded**	**(5) Family members**	**(6) West Germany**
Networks & Segregations						
(Log) ethnic share	−0.0246	0.0256	−0.1626	−0.0397	−0.1689	0.0169
	(0.0925)	(0.1066)	(0.2384)	(0.0830)	(0.4402)	(0.0844)
(Log) foreign share	0.0674	−0.2570	0.1435	0.0542	0.0245	−0.2676
	(0.2786)	(0.1842)	(0.3633)	(0.1176)	(0.8080)	(0.1681)
F-statistic of 1st stage						
(Log) ethnic share (per zip code area)	13.05***	22.08***	56.93***	50.79***	4.41**	16.09***
(Log) foreign share (per zip code area)	3.16*	15.03***	111.29***	17.08***	0.31	14.01***
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.2305	0.1838	0.2473	0.2362	0.3278	0.3618
Observations	1,036	822	270	1,552	333	1,718
Individuals	665	527	172	995	227	1,100

Source: IAB-SOEP Migration Sample, own illustration.
Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; heteroskedastic robust standard errors in parentheses.

Table A4. IV estimations: robustness checks for males

	Years of Overqualification Stratified by···		Alternative Instru-ments				
	(1) Gender	**(2) Countries**	**(3)**	**(4)**	**(5) Unemployment rate**	**(6) Age at migration ≥ 18**	**(7) Arrival cohort trends**
Networks & segregation							
(Log) ethnic share	−0.0654	0.1065	−0.0051	−0.4526	−0.0083	−0.0129	0.0926
	(0.0865)	(0.0946)	(0.0791)	(1.9878)	(0.1360)	(0.0749)	(0.0756)
(Log) foreign share	0.4545	0.6534	0.6429	0.4786	0.6417	0.5344	0.3564
	(0.6486)	(0.6974)	(0.7006)	(3.2862)	(0.484)	(0.5781)	(0.5235)
Unemployment Rate					0.0288		
					(0.0295)		
F-statistic of 1st Stage							
(Log) ethnic share (per zip code area)	58.56***	40.56***		10.89***	4.82***	23.64***	68.56***
(Log) foreign share (per zip code area)	32.64***	27.64***			5.81***	57.54***	35.56
Language Course Density (per Federal State)				0.62			
(Log) ethnic share (per zip code department)			5.75***				
(Log) foreign share (per zip code department)			3.85**				
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.3765	0.3956	0.3355	0.2477	0.3459	0.3891	
Observations	1,858	1,858	1,858	1,858	1,858	1,799	1,858
Individuals	1,192	1,192	1,192	1,192	1,192	1,134	1,192

Source: IAB-SOEP Migration Sample, own illustration.
Note: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; heteroskedastic robust standard errors in parentheses.
Table A4 presents two-stage least squares (2SLS) IV estimations of equation [4] with random effects for females. In each model, years of overqualification are regressed on the log. ethnic share $logES_{jkt}$ and on the log. foreign share $logFS_{jt}$, and controls on dummies for countries of origin, occupations, regions, and the covariates given in Table 2. The Models 1 and 2 use alternative indicators of occupational mismatch. In Model 1, the calculation of average schooling years per prestige is stratified by gender. In Model 2, the calculation of average schooling years is separately computed with respect to country of origin groups. The IV estimations 3 and 4 use two different kinds of instruments. In Model 3, the foreign and the ethnic share per zip code are instrumented by the respective shares at the level of zip code departments (first digit of the zip code number). In Model 4, the foreign share across zip codes is instrumented by language course density across federal states. In addition, Model 5 controls on unemployment rate across zip codes and Model 6 excludes individuals which arrived under the age of 18 in Germany. Finally, Model 7 includes dummies for arrival cohorts.

Notes

[1]In previous literature, an education–occupation mismatch occurs in the case of overqualification, which means that a worker's education level is higher than that required by his or her occupation (Nieto *et al.,* 2015; Poot and Stillman, 2016).

[2]There are several reasons for this approach. The German dual vocational training system makes it hard to compare education years acquired in Germany and education years acquired abroad (Deissinger, 2015). Furthermore, as I study the effects of ethnic networks on years of overqualification, the regressions can only include immigrants. However, if the calculation of years of overqualification is based on natives and immigrants but the estimations only consider immigrants, the distribution of overqualification years in the estimations will be strongly left-skewed.

[3]The use of spatial shares to capture ethnic networks is in accordance with Borjas (1995, 1992) and is also incorporated by Bertrand *et al.* (2000): 'An ethnic network is determined by geographical proximity and furthermore are based on ethnic similarity'. Sociologists define these terms in a similar way: 'Residential Segregation may be understood as the systematic and uneven presence of racial minorities in city areas that are separate and apart from the places where the native population lives' (Nelson, 2013; Iceland *et al.*, 2002). Thus, I capture ethnic networks of immigrants by the share of their own ethnic group per zip code. Although the main interest lies on the effect of the ethnic share on occupational mismatch, distinguishing between the ethnic share and the share of the foreign-born population is important, as the latter is understood as an indicator of residential segregation. Although prior research offers further proxies of segregation and networks, such as the dissimilarity index (Cutler and Glaeser, 1997), the isolation index (Glitz, 2014), the assimilation index (Elsner *et al.,* 2018), and others (for an overview, see Massey and Denton, 1988; Aslund and Skans, 2009), the use of simple shares per area is a reliable approach that is in line with Edin *et al.* (2003), Bauer *et al.* (2005), Deri (2005), Schaffner and Treude (2014), and Dustmann et al. (2016a).

[4]Throughout the article, I exclusively consider foreign-born immigrants. Thus, migration background is captured by country of birth.

[5]Theory predicts that the information channel supplies information about eligibility rules for accessing social benefits, health care, and facilitating visits of other national authorities (Ioannides and Loury, 2004). Besides the positive effects of information disposal, Bertrand *et al.* (2000) illustrate the hazard of networks: Unilateral network knowledge hampers the collection of information about alternative options for accessing healthcare services and job opportunities.

[6]The authors also compare this ratio to the analogous proportion of the entire United States.

[7]Edin *et al.* (2003) and Damm (2014) confirm the ethnic stratification of networks with quasi-experimental evidence. They exploit the introduction of (compulsory) residential assignments to particular migrant groups and instrument the current residence by the initial and assigned residence of immigrants (also see Aslund and Skans, 2010; Damm, 2009; Foged and Peri, 2015). Similar quasi-experimental studies can be found for the Netherlands (Beckers and Borghans, 2011), Italy (Boeri *et al.,* 2012), and Germany (Danzer and Yaman, 2016). A further strand of IV methods is based on work by Bayer and Ross (2006), who estimate ethnic shares by all possible combinations of covariates. They find that similar individuals live in areas with similar ethnic shares (see also Bayer *et al.,* 2008; Bauer *et al.,* 2011; Schaffner and Treude, 2014).

[8]If the person paid contributions within an employment less than twelve months, he or she has claims on obtaining unemployment assistance for a maximum duration of six months.

[9]In accordance with Nieto *et al.* (2015), Dustmann *et al.* (2016b) describe downgrading as first, cases in which immigrants work in jobs for which they are overqualified, and second, cases in which immigrants receive lower returns to education than natives. In general, however, occupational downgrading considers cases in which the prestige of an immigrant's occupation is reduced by migration (Akresh, 2006, 2008). Thus, downgrading is interpreted as occupational mobility over time and may be because immigrants are often endowed with low skills or skills that are underutilized (Mattoo *et al.,* 2008). On the contrary, the term education–occupation mismatch defines a reference level of required education and does not perform a comparison of occupations over time. Instead, it

compares a person's education to the required education level in the individual's occupation (see Section 3.1). Thus, occupational mismatch arises if a given level of human capital is underutilized.

[10]Thus, migrants who partner with a native spouse and who are also not household head are not part of the data set. In my sample, males (70.1 percent) are more likely to be household head than females (60.9 percent), which should be noted considering sample selection by gender. Although I do not observe binational marriages in my restricted sample, different shares by gender who partner with a native spouse may influence the sample procedure of the data set when focusing the entire data set without restrictions.

[11]Chiswick and Miller (2008) discuss the advantages and drawbacks of the three major approaches to capturing the required level of education. Next to the realized matches procedure, the reference level of education can be defined by public authorities or by subjective feeling in which the worker assesses whether he or she has more or less education than is required.

[12]The International Socio-Economic Index of Occupational Status (ISEI) is an alternative to the SIOPS. However, using the ISEI classification does not change my empirical results in a noteworthy way. Both classifications are based on the International Standard Classification of Occupations (ISCO) by the International Labour Organisation (ILO). However, the application of this classification is not feasible because it distinguishes between a much larger number of occupations. As a consequence, the average of schooling years would be based on too few observations in my sample.

[13]Since the 1st of July 1993, German zip codes have been five-digit numbers, with approximately 30,000 different zip codes throughout Germany (for details, see Budde and Eilers, 2014).

[14]This concept suffers from biases because only the ethnic affiliation of household heads is collected and differences between first-generation immigrants and persons born in Germany cannot be regarded.

[15]Elsner *et al.* (2018) state that weak ties are more important than close ties because friends of friends overlap less frequently in their social contacts. This makes it possible to obtain information from outside their own network (see also Granovetter, 1983; Green *et al.,* 1999; Calvo-Armengol and Jackson, 2004; Patacchini and Zenou, 2012).

[16] Although only one third of my sample were surveyed several times, the random-effects approach is preferred to account for time-invariant unobservable heterogeneity a_i for these individuals. Controlling for the unobservable determinants of labor market performance is particularly important for immigrants because of many unobservable factors such as discrimination, the motivation to obtain guest-country-specific human capital, the actual willingness to stay in Germany, and the local supply of language training. Note that the results of the following section do not change significantly when pooled OLS is applied.

[17]Note that the application of cross-sectional IV estimation does not significantly change the results. The cross-sectional estimate of (log.) ethnic share for males remains insignificant (−0.0238), and the analogue coefficient for females remains significant (−0.2252), when I control for the covariates listed in Table 2.

[18]In a pooled estimation for females and males, a weak significant effect of ethnic share with negative sign dominates (Table A1 in the Appendix). Furthermore, Table A1 displays coefficients of socioeconomic, educational, and migration-specific variables. The results confirm the negative effect of years of residency and age from literature (Chiswick *et al.,* 2005) and show that the country of origin is an important factor. Here, evidence by Saleheen and Shadforth (2006), Drinkwater *et al.* (2009), Kahanec *et al.* (2009), and Elsner (2013) is confirmed: Migrants from countries of the EU enlargement in 2004 suffer significantly from education–occupation mismatch. The opposite can be concluded about immigrants from Russia and the former USSR, which includes a majority of ethnic Germans, and about immigrants from Turkey.

[19]In this context, one drawback of my measure of capturing occupational mismatch is identified. Years of overqualification are also determined by the educational choices of individuals, which are endogenous. Despite this drawback, applying this indicator is currently the best choice.

[20]This result is available upon request from the author.

[21] Although I exclusively focus on female immigrants in this section, I also provide heterogeneous estimates for males (see Table A3 in the Appendix).

[22] See Table A4 in the Appendix for the results considering male immigrants.

[23] For instance, the fifth department includes the metropolis Cologne with a high foreign share of 17.5 percent, and rural areas such as the Sauerland.

[24] Basically, testing the exclusion restriction is not directly feasible. Nevertheless, recent research has developed novel test procedures (see van Kippersluis and Rietveld, 2018; Clarke and Matta, 2018).

[25] Note that occupational mismatch is not harmful on its own. Migrants are often willing to work in a job that does not perfectly fit with their experience or qualifications as long as real wages are higher than in the country of origin (Akresh, 2008). Thus, education–occupation mismatch does not have to result in a worse standard of living; however, it is one of the indicators that future research should use to evaluate the labor market integration of immigrants.

References

Akresh I. R. (2006) 'Occupational Mobility Among Legal Immigrants to the United States', *International Migration Review* 40(4): 854–884.

Akresh I. R. (2008) 'Occupational Trajectories of Legal US Immigrants: Downgrading and Recovery', *Population and Development Review* 34(3): 435–456.

Altonji J. G. and Card D. (1991) 'The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives' in Abowd J. and Freeman R. B. (eds.) *Immigration, Trade and the Labor Market*. Chicago: University of Chicago Press: pp. 201–234.

Amuedo-Dorantes C. and Bansak C. (2011) 'The Impact of Amnesty on Labor Market Outcomes: A Panel Study Using the Legalized Population Survey', *Industrial Relations* 50(3): 443–471.

Andersson F., Burgess S. and Lane J. (2014) 'Do as the Neighbors Do: Examining the Effect of Residential Neighborhoods on Labor Market Outcomes', *Journal of Labor Research* 35(4): 373–392.

Aslund O. and Skans O. N. (2009) 'How to Measure Segregation Conditional on the Distribution of Covariates', *Journal of Population Economics* 22: 971–981.

Aslund O. and Skans O. N. (2010) 'Will I See You atWork? EthnicWorkplace Segregation in Sweden, 1985–2002', *Industrial & Labor Relations Review* 63(3): 471–493.

Battisti M., Peri G. and Romiti A. (2018) 'Dynamic Effects of Co-Ethnic Networks on Immigrants' Economic Success', NBER Working Paper Series No: 22389.

Battu H., Seaman P. and Zenou Y. (2011) 'Job Contact Networks and the Ethnic Minorities', *Labour Economics* 18: 48–56.

Bauer T., Epstein G. S. and Gang I. N. (2005) 'Enclaves, Language, and the Location Choice of Migrants', *Journal of Population Economics* 18: 649–662.

Bauer T. K., Fertig M. and Vorell M. (2011) 'Neighborhood Effects and Individual Unemployment', IZA Discussion Paper, No. 6040.

Bayer P. and Ross S. L. (2006) 'Identifying Individual and Group Effects in the Presence of Sorting: A Neighborhood Effects Application', NBER Working Paper Series, No. 12211.

Bayer P., Ross S. L. and Topa G. (2008) 'Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes', *Journal of Political Economy* 116(6): 1150–1196.

Beaman L. A. (2012) 'Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S', *Review of Economic Studies* 79(1): 128–161.

Beckers P. and Borghans L. (2011) 'Segregation in Neighbourhoods and Labour Market Outcomes of Immigrants: Evidence from Random Assignment in the Netherlands', UNUMERIT Working Paper Series, No. 2011–068.

Bentolila S., Michelacci C. and Suarez J. (2010) ‘Social Contacts and Occupational Choice’, *Economica* 77(305): 20–45.

Bertrand M., Luttmer E. F. P. and Mullainathan S. (2000) ‘Network Effects and Welfare Cultures’, *The Quarterly Journal of Economics* 115(3): 1019–1055.

Bijwaard G. E. (2010) ‘Immigrant Migration Dynamics Model for The Netherlands’, *Journal of Population Economics* 23: 1213–1247.

Blotevogel H. H. and Jeschke M. (2001) ‘Determinanten der Stadt-Umland-Wanderungen im Raum Duisburg’, Universität Duisburg, Institut für Geographie, Diskussionspaper, 4/2001.

Boeri T., de Philippis M., Patacchini E. and Pellizzari M. (2012) ‘Moving to Segregation: Evidence from 8 Italian Cities’, IZA Discussion Paper, No. 6834.

Borjas G. J. (1992) ‘Ethnic Capital and Intergenerational Mobility’, *The Quarterly Journal of Economics* 107(1): 123–150.

Borjas G. J. (1995) ‘Ethnicity, Neighborhoods, and Human-Capital Externalities’, *American Economic Review* 85(3): 365–390.

Borjas G. J. (1998) ‘To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation’, *Journal of Urban Economics* 44: 228–253.

Borjas G. J. and Bratsberg B. (1996) ‘Who Leaves? The Outmigration of the Foreign-Born’, *The Review of Economics and Statistics* 78(1): 165–176.

Brown M., Setren E. and Topa G. (2016) ‘Do Informal Referrals Lead to Better Matches? Evidence from a Firm’s Employee Referral System’, *Journal of Labor Economics* 34(1): 161–209.

Brücker H., Glitz A., Lerche A. and Romiti A. (2018) ‘Occupational Recognition and Immigrant Labor Market Outcomes’, IZA Discussion Paper, No. 12030.

Brücker H., Kroh M., Bartsch S., Goebel J., Kühne S., Liebau E., Trübswetter P., Tucci I. and Schupp J. (2014). ‘The New IAB-SOEP Migration Sample: An Introduction into the Methodology and the Content’, SOEP Survey Papers 216, Series C. DIW Berlin.

Budde R. and Eilers L. (2014) *Sozioökonomische Daten auf Rasterebene – Datenbeschreibung der Microm-Rasterdaten*. RWI Essen: RWI Materialien: p. 77.

Calvo-Armengol A. and Jackson M. O. (2004) ‘The Effects of Social Networks on Employment and Inequality’, *American Economic Review* 94(3): 426–454.

Calvo-Armengol A. and Jackson M. O. (2007) ‘Networks in Labor Markets: Wage and Employment Dynamics and Inequality’, *Journal of Economic Theory* 132(1): 27–46.

Card D. (2001) ‘Immigrant Inows, Native Outows, and the Local Market Impacts of Higher Immigration’, *Journal of Labor Economics* 19(1): 22–64.

Cascio E. U. and Lewis E. G. (2012) ‘Cracks in the Melting Pot: Immigration, School Choice, and Segregation’, *American Economic Journal: Economic Policy* 4(3): 91–117.

Charpin A. (2014) ‘Occupational Downgrading of Immigrants in France’, Economies and Finances, No. dumas-01105391.

Chiswick B. R. (1978) ‘The Effect of Americanization on the Earnings of Foreign-born Men’, *Journal of Political Economy* 86(5): 897–921.

Chiswick B. R., Lee Y. L. and Miller P. W. (2005) ‘A Longitudinal Analysis of Immigrant Occupational Mobility: A Test of the Immigrant Assimilation Hypothesis’, *International Migration Review* 39(2): 332–353.

Chiswick B. R. and Miller P. W. (2005) ‘Do Enclaves Matter in Immigrant Adjustment?’, *City and Community* 4(1): 5–35.

Chiswick B. R. and Miller P. W. (2008) ‘Why is the Payoff to Schooling Smaller for Immigrants?’, *Labour Economics* 15(6): 1317–1340.

Chiswick B. R. and Miller P. W. (2010) ‘The Effects of Educational-Occupational Mismatch on Immigrant Earnings in Australia, with International Comparisons’, *International Migration Review* 44(4): 869–898.

Clarke D. and Matta B. (2018) ‘Practical Considerations for Questionable IVs’, *Stata Journal* 18(3): 663–691.

Cortes K. E. (2004) 'Are Refugees Different from Economic Immigrants? Some Empirical Evidence on the Heterogeneity of Immigrant Groups in the United States', *Review of Economics and Statistics* 86(2): 465–480.

Cutler D. M. and Glaeser E. L. (1997) 'Are Ghetto Good or Bad?', *The Quarterly Journal of Economics* 112(3): 827–872.

Damm A. P. (2009) 'Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence', *Journal of Labor Economics* 27(2): 281–314.

Damm A. P. (2014) 'Neighborhood Quality and Labor Market Outcomes: Evidence from Quasi-Random Neighborhood Assignment of Immigrants', *Journal of Urban Economics* 79: 139–166.

Danzer A. M. and Yaman F. (2016) 'Ethnic Concentration and Language Fluency of Immigrants: Evidence from the Guest-Worker Placement in Germany', *Journal of Economic Behavior & Organization* 131(Part: A):151–165.

Deissinger T. (2015) 'The German Dual Vocational Education and Training System as 'Good Practice'?', *Local Economy* 30(5): 557–567.

Deri C. (2005) 'Social Networks and Health Service Utilization', *Journal of Health Economics* 24: 1076–1107.

Devillanova C. (2008) 'Social Networks, Information and Health Care Utilization: Evidence from Undocumented Immigrants in Milan', *Journal of Health Economics* 27: 265–286.

Drinkwater S., Eade J. and Garapich M. (2009) 'Poles Apart? EU Enlargement and the Labour Market Outcomes of Immigrants in the United Kingdom', *International Migration* 47(1): 161–190.

Dustmann C., Glitz A., Schönberg U. and Brücker H. (2016a) 'Referral-based Job Search Networks', *Review of Economic Studies* 83: 514–546.

Dustmann C., Schönberg U. and Stuhler J. (2016b) 'The Impact of Immigration: Why Do Studies Reach Such Different Results?', *Journal of Economic Perspectives* 30(4): 31–56.

Dustmann C. and van Soest A. (2002) 'Language and the Earnings of Immigrants', *Industrial and Labor Relations Review* 55(3): 473–492.

Edin P.-A., Fredriksson P. and Aslund O. (2003) 'Ethnic Enclaves and the Economic Success of Immigrants – Evidence From a Natural Experiment', *The Quarterly Journal of Economics* 118 (1): 329–357.

Elsner B. (2013) 'Does Emigration Benefit the Stayers? Evidence from EU Enlargement', *Journal of Population Economics* 26(2): 531–553.

Elsner B., Narciso G. and Thijssen J. J. J. (2018) 'Migrant Networks and the Spread of Information', *Oxford Bulletin of Economics and Statistics* 80(3): 659–688.

Fasani F. (2015) 'Understanding the Role of Immigrants Legal Status: Evidence from Policy Experiments', *CESifo Economic Studies* 61(3): 722–763.

Federal Office for Migration and Refugees (2005). 'Der Einuss von Zuwanderung auf die deutsche Gesellschaft. Deutscher Beitrag zur Pilotforschungsstudie "The Impact of Immigration on Europe's Societies" im Rahmen des Europäischen Migrationsnetzwerks', Bundesamt fur Migration und Flüchtlinge (BAMF), Nüurnberg, Forschungsbericht 1.

Federal Office for Migration and Refugees (2007). 'Zuwanderung und Integration von (Spät-)Aussiedlern – Ermittlung und Bewertung der Auswirkungen des Wohnortzuweisungsgesetzes', Bundesamt für Migration und Flüchtlinge (BAMF), Nürnberg, Forschungsbericht 3.

Federal Office for Migration and Refugees (2009). 'Konzept "Trägerzulassungsverfahren" Bundesamt für Migration und Flüchtlinge (BAMF), Nürnberg'.

Fendel T. and Jochimsen B. (2017) 'Child Care Reforms and Labor Participation of Migrant and Native Mothers', *IAB Discussion Paper* No: 9/2017.

Foged M. and Peri G. (2015) 'Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data', *American Economic Journal: Applied Economics* 8(2): 1–34.

Gathmann C. and Keller N. (2017) 'Access to Citizenship and the Economic Assimilation of Immigrants', *The Economic Journal* 128: 3141–3181.

Glitz A. (2012) 'The Labor Market Impact of Immigration: A Quasi-Experiment Exploiting Immigrant Location Rules in Germany', *Journal of Labor Economics* 30(1): 175–213.

Glitz A. (2014) 'Ethnic Segregation in Germany', *Labour Economics* 29: 28–40.

Granovetter M. S. (1983) 'The Strength of Weak Ties: A Network Theory Revisited', *Sociological Theory* 1: 201–233.

Green G. P., Tigges L. M. and Diaz D. (1999) 'Racial and Ethnic Differences in Job-Search Strategies in Atlanta, Boston, and Los Angeles', *Social Science Quarterly* 80(2): 263–278.

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

Iceland J., Weinberg D. H. and Steinmartz E. (2000) 'Racial and Ethnic Residential Segregation in the United States: 1980-2000', Census 2000 Special Reports, Series CENSR-3, U. S. Census Bureau Washington.

Imbens G. W. and Angrist J. D. (1994) 'Identification and Estimation of Local Average Treatment Effects', *Econometrica* 62(2): 467–475.

Ioannides Y. M. and Loury L. D. (2004) 'Job Information Networks, Neighborhood Effects, and Inequality', *Journal of Economic Literature* 42(4): 1056–1093.

Jaeger D. A., Ruist J. and Stuhler J. (2018) 'Shift-Share Instruments and the Impact of Immigration', CReAM Discussion Paper Series, CPD 02/18.

Kahanec M., Zaiceva A. and Zimmermann K. F. (2009) 'Lessons from Migration after EU Enlargement' in Kahanec M. and Zimmermann K. F. (eds.) *EU Labor Markets After Post-Enlargement Migration, Chapter 3*. Berlin: Springer: pp. 3–45.

Kalfa E. and Piracha M. (2018) 'Social Networks and the Labour Market Mismatch', *Journal of Population Economics* 31(3): 877–914.

Kiker B. F., Santos M. C. and de Oliveira M. M. (1997) 'Overeducation and Undereducation: Evidence for Portugal', *Economics of Education Review* 16(2): 111–125.

Klaesson J., Öner Ö. and Pennerstorfer D. (2019). 'Getting the First Job – Size and Quality of Ethnic Enclaves for Refugee Labor Market Entry', Economics Working Papers, Department of Economics, Johannes Kepler University Linz, No. 2019–07.

Kosfeld R. and Werner A. (2012) 'Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007–2011', *Raumforschung und Raumordnung* 70(1): 49–64.

Massey D. S. and Denton N. A. (1988) 'The Dimensions of Residential Segregation', *Social Forces* 67(2): 281–315.

Mattoo A., Neagu I. C. and Özden C. (2008) 'Brain Waste? Educated Immigrants in the US Labor Market?', *Journal of Development Economics* 87: 255–269.

Microm Consumer Marketing (2014) *Microm Datenhandbuch: Arbeitsgrundlagen für microm MARKET & GEO*. Neuss: Microm Micromarketing-Systeme und Consult GmbH.

Nelson K. A. (2013) 'Does Residential Segregation Help or Hurt? Exploring Differences in the Relationship between Segregation and Health among U.S. Hispanics by Nativity and Ethnic Subgroup', *Social Science Journal* 50: 646–657.

Nielsen C. P. (2011) 'Immigrant Over-Education: Evidence from Denmark', *Journal of Population Economics* 24(2): 499–520.

Nieto S., Matano A. and Ramos R. (2015) 'Educational Mismatches in the EU: Immigrants vs Natives', *International Journal of Manpower* 36(4): 540–561.

Patacchini E. and Zenou Y. (2012) 'Ethnic Networks and Employment Outcomes', *Regional Science and Urban Economics* 42: 938–949.

Patel K. and Vella F. (2013) 'Immigrant Networks and Their Implications for Occupational Choice and Wages', *The Review of Economics and Statistics* 95(4): 1249–1277.

Pellizzari M. and Fichen A. (2017) 'A New Measure of Skill Mismatch: Theory and Evidence from PIAAC', *IZA Journal of Labor Economics* 6(1): 1–30.

Piracha M. and Vadean F. (2013) 'Migrant Educational Mismatch and the Labour Market' in Constant A. F. and Zimmermann K. F. (eds.) *International Handbook on the Economics of Migration, Chapter 9*. Northampton: Edward Elgar Publishing: pp. 176–192.

Poot J. and Stillman S. (2016) 'Skill Composition of Immigration Flows and the Measurement of Education-Occupation Mismatch', *IZA Journal of Migration* 5(18): 1–28.

Rapoport B. and Thibout C. (2018) ‘Why Do Boys and Girls Make Different Educational Choices? The Inuence of Expected Earnings and Test Scores’, *Economics of Education Review* 62: 205–229.

Reuben E., Wiswall M. and Zafar B. (2015) ‘Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender’, *Economic Journal* 127 (604): 2153–2186.

Saleheen J. and Shadforth C. (2006) ‘The Economic Characteristics of Immigrants and Their Impact on Supply’, *Bank of England Quarterly Bulletin* Q4: 373–385.

Schaffner S. and Treude B. (2014) ‘The Effect of Ethnic Clustering on Migrant Integration in Germany’, Ruhr Economic Papers, No. 536.

Schönwälder K. and Söhn J. (2009) ‘Immigrant Settlement Structures in Germany: General Patterns and Urban Levels of Concentration of Major Groups’, *Urban Studies* 46(7): 1439–1460.

Tani M. (2017) ‘Local Signals and the Returns to Foreign Education’, *Economics of Education Review* 61: 174–190.

van Kippersluis H. and Rietveld C. A. (2018) ‘Beyond Plausibly Exogenous’, *The Econometrics Journal* 21: 316–331.

Wagner G. G., Göbel J., Krause P., Pischner R. and Sieber I. (2008) ‘Das Soziooekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender)’, *AStA Wirtschafts- und Sozialstatistisches Archiv* 2(4): 301–328.

Yao Y. and van Ours J. C. (2015) ‘Language Skills and Labor Market Performance of Immigrants in the Netherlands’, *Labour Economics* 34: 76–85.