RUHR

ECONOMIC PAPERS

Sandra Schaffner
Barbara Treude

The Effect of Ethnic Clustering on Migrant Integration in Germany

rwi

#536

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #536

Responsible Editor: Jochen Kluve

ISSN 1864-4872 (online) – ISBN 978-3-86788-613-0

Ruhr Economic Papers #536

Sandra Schaffner and Barbara Treude

The Effect of Ethnic Clustering on Migrant Integration in Germany

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788613
ISSN 1864-4872 (online)
ISBN 978-3-86788-613-0

Sandra Schaffner and Barbara Treude[1]

The Effect of Ethnic Clustering on Migrant Integration in Germany

Abstract

Since ethnic clustering is common in Germany, a better understanding of its effects on the integration of immigrants could be important for integration policies, especially in the light of rising immigration and a skilled worker shortage. Yet, both economic theory and empirical research for other countries cannot give a clear-cut answer to whether clustering is beneficial or detrimental for immigrants' integration. In this paper, the effect of residential clustering on the labour market outcome of first-generation immigrants in Germany is analysed empirically. It, thus, contributes to the literature by extending it to Germany on which hardly any research has been conducted. For the analysis, two measures for labour market integration are used: the employment probability and wage levels. In order to control for the endogeneity of the location decision, a two-step strategy is used, combining a control function and an instrumental variable (IV) approach. The results suggest a negative enclave effect on both employment and wages, that is even larger when sorting is taken into account.

JEL Classification: J61, J64, J31, R23

Keywords: Ethnic enclaves; residential clustering; labour market integration; migrants; wage differentials

December 2014

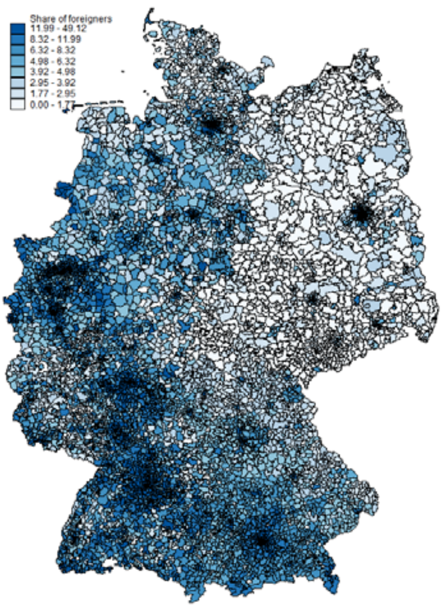
1 Sandra Schaffner, RWI; Barbara Treude, RWI and UCL, London. *– We thank Thomas K. Bauer for helpful comments and suggestions.* *– All correspondence to: Sandra Schaffner, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, e-mail: sandra.schaffner@rwi-essen.de*

1 Introduction

When migrating to another country, fellow immigrants are often essential during the first time after arrival. They provide the new immigrants with valuable information and help them get along in the foreign country. Yet, this "warm embrace" (Borjas, 2000, p. 93) comes at a potential cost: less integration. In recent years, the public debate on immigrant integration in Germany has been centred around this issue, with "parallel society" (*Parallelgesellschaft*) becoming the catch word.

Indeed, ethnic clustering is observable in Germany. Figure 1 shows the share of foreigners in Germany on the post code level for the year 2010. In general, there are far less immigrants in East Germany, which is likely due to the worse labour market situation there (Aretz, 2013; Brücker and Trübswetter, 2007). Immigrants in Germany also tend to live in the large cities, in particular around Berlin, Hamburg, Munich, Frankfurt and Stuttgart as well as in the Ruhr Area. These cities do not only offer good labour market prospects but also allow immigrants to consume ethnic goods, such as religious services and traditional food. Overall, ethnic enclaves hence do seem to matter for Germany.

Figure 1: Immigrant Share in Germany (2010)



Data source: microm. The share is presented on the post code level. The shapefile is provided by Esri Geoinformatik GmbH Germany.

At the same time, empirical research has shown that immigrants fare worse in the German labour market: Aldashev et al. (2012, p. 504) show that even those who completed their education in Germany have on average 19% lower wages than natives. Lehmer and Ludsteck (2014) support this finding of lower immigrant wages, and further provide evidence that economic assimilation differs substantially across ethnic groups. It is above average for many European immigrants and those from other advanced countries – in fact resulting in above-native wage levels –, while wages remain much lower for immigrants from Greece, Turkey, Vietnam and the Middle East.

Thus, why do we see immigrants with identical human capital obtained in Germany faring worse than Germans? Why do some ethnic groups do significantly better in the labour market than others? In the light of these persisting differentials, ethnic enclaves could be an important factor to understanding. Immigrants living in enclaves might adapt less to the German culture and learn less German. Both can severely hinder the transferability of their foreign human capital. In addition, spatial mismatch of enclaves might hinder highly skilled immigrants to work in high pay positions. Hence, a negative enclave effect could (partially) explain wage differentials between natives and immigrants. However, enclaves might also have a positive net effect, actually helping immigrants overcome labour market discrimination.

However, economic theory cannot give a clear-cut answer to whether clustering is beneficial or detrimental for immigrants' integration overall. Nor does the empirical literature provide a convincing answer. Yet, surprisingly little empirical research has dealt with enclave effects for Germany. This paper, thus, contributes to the literature by extending it to Germany. For the analysis, a new and recent dataset is used. Thereby, two measures for labour market integration are used: the employment probability and wage levels. The analysis suggests a negative enclave effect on both employment and wages, that is even larger when taking sorting into account.

The remainder of this paper is structured as follows: Section 2 reviews the relevant literature, providing an overview both of theoretical channels and empirical evidence of the enclave effect. In section 3, the dataset used is discussed and summary statistics presented. Section 4 describes the identification strategy. Section 5 presents the results of the empirical analysis, including a comprehensive sensitivity analysis, and section 6 concludes.

2 Literature Review

From economic theory, the effect of ethnic enclaves on immigrants' labour market integration is unclear. This is due to the many, often opposing channels of how living in a neighbourhood with many immigrants (from the same ethnic background) can affect the labour market attainment of immigrants. The net effect of all of these channels will be referred to as *enclave effect* or, equivalently, *clustering* *effect* in the following. An enclave thereby is defined as an area where there is a "concentration of ethnic minority individuals" (Clark and Drinkwater, 2002, p. 7).

If ethnic enclaves are seen as a network, they work based on information and norms (Bertrand et al., 2000). On the one hand, the network provides them with information, such as information on job offers (Bayer et al., 2008; Beaman, 2012; Goel and Lang, 2009). Therefore, for an immigrant, being part of that network can help overcome discrimination otherwise faced. However, the network could also provide newly-arrived immigrants with information on other sources of income, such as the welfare system (Bertrand et al., 2000). On the other hand, the network can also affect labour market outcomes via the channel of social norms. Damm (2009, p. 284) differentiates between direct influences, e.g. work ethics or attitudes towards self-employment, and indirect ones, e.g. by impacting other factors such as marriage and educational attainment. The network channel of the enclave effect is thereby driven by the quality of the enclave and is in itself not unambiguous. This is closely interlinked with the idea of human capital externalities within a neighbourhood (cf. Acemoglu and Angrist, 2000), which also depend on the quality of the enclave and, thus, may vary by skill-level of the immigrants.

Besides, living in an enclave can also lower the incentives to invest in country-specific human capital of the host country as it is not necessary in the enclave. This channel has extensively been investigated for language skills, e.g. by Lazear (1999), Chiswick and Miller (1996), Chiswick and Miller (2001) and Warman (2007). As host country-specific human capital is important for most jobs, individuals living in an enclave will thus have less chances to integrate in the labour market as their human capital is less transferable to the host country's labour market without host country-specific knowledge (Warman, 2007).

Yet, if enclaves are sufficiently large, they might constitute a labour market of their own (Portes and Jensen, 1989; Portes and Shafer, 2007), providing ethnic goods (Waldfogel, 2003) and thus inducing more enclave inhabitants to open up businesses. This creates job opportunities for migrants who live in enclaves (cf. Portes, 1987), even though the effects of immigrant self-employment on earnings are unclear (Logan et al., 2003; Portes and Zhou, 1996). In addition, regarding larger ethnic businesses, chances of finding employment are independent from the host country's human capital, which increases the likelihood to find a job (McManus, 1990) and returns higher wages due to better transferability of human capital (Wilson and Portes, 1980; Xie and Gough, 2011; Zhou and Logan, 1989). On the other hand, Borjas (2000) and Tu (2010) argue that enclave labour markets generally pay lower wages due to monopsy power.

Furthermore, the spatial mismatch theory by Ihlanfeldt and Sjoquist (1989); Kain (1968); Zenou (2013) is based on a twofold argument why residential segregation will be detrimental to the labour market success of ethnic minorities. On the supply side, there is a mismatch between residential areas of immigrants and workplaces resulting in higher commuting costs and a lack of information on jobs. On the demand side, prejudices are fostered leading to increased labour market discrimination.

In addition to the most prominent channels of the network, Cutler et al. (2008, p. 761) stress the importance of below average quality of local institutions. They argue that enclaves are often located in older residential areas that are further away from the centres of city development. Thus, inhabitants of enclaves are quite likely to be served by more constrained, incumbent central city institutions. Another channel outlines that immigrants' reservation wages decrease if they can consume ethnic goods of the enclave, such as traditional food or religious services (Chiswick and Miller, 1995; Damm, 2009). However, as Edin et al. (2003, p. 335) point out, this cannot really be considered a causal channel of living in an enclave as it merely reveals immigrants' preferences.

Overall, the theoretical channels of enclave effects do not allow for the prediction of the net enclave effect. Thus, this is ultimately an empirical question. However, for both employment and wages, the empirical results (mainly for the US) are mixed. On the one hand, enclaves seem to have the power to provide an efficient network, increasing the probability to be employed (Aguilera, 2002, 2003; Lancee, 2010; Sanders et al., 2002; Tu, 2010). Due to enclave labour markets, the employment probability can be increased by a higher rate of self-employment (Borjas, 1986; Le, 2000). However, Razin and Langlois (1996) and Clark and Drinkwater (2002) find a negative effect on self-employment and Pedace and Rohn (2008) on employment in general. No convincing relationship is found by Damm (2009) and Cutler et al. (2008). Regarding wages, empirical results are equally mixed. While a few studies find positive wage effects (Cutler et al., 2008; Damm, 2009; Edin et al., 2003), other find negative ones (Borjas, 2000; Warman, 2007) or an insignificant relationship (Chowdhury and Pedace, 2007; Pedace and Rohn, 2008; Xie and Gough, 2011). Goel and Lang (2009) show that social networks can help immigrants find jobs faster but also at a lower wage.

One reason why the results are mixed could be that enclave effects differ among immigrant groups. This could be different ethnic groups (Hou and Picot, 2003), different skill level groups (Sousa, 2013; Warman, 2007) or gender differences (Aguilera, 2005; Gilbertson, 1995; Zhou and Logan, 1989). Another reason could be differences in the enclave quality, e.g. in the duration that its inhabitants have been in the host country (Beaman, 2012), the average human capital (Cutler et al., 2008; de Graaf and Flap, 1988) or average income levels (Edin et al., 2003). In addition, there is no certain time frame in which the clustering effect is assumed to take place. Tu (2010) finds that the clustering effect diminishes over time, while Edin et al. (2003), however, argue that there are cumulative enclave effects. Samples that consist of immigrants with different years since migration could thus also explain the opposing results found in the literature.

There has been surprisingly little research on Germany. Most closely related to this paper is a study by Kanas et al. (2011) who find no enclave effects in Germany regarding both occupational status and earnings, arguing that possibly a sufficient threshold of the immigrant share is not reached. However, they use the ethnic share on the level of federal states (*Länder*) which is likely to be too large to allow for observing any enclave effects as this is certainly beyond the scope of a personal network. In addition, there is some literature on ethnic German migrants to Germany after World War II (*Aussiedler*), and clustering of these immigrants is generally found to have a positive effect on their labour market outcomes (Bauer and Zimmermann, 1997; Dietz, 1999). However, since these immigrants are ethnic Germans, their integration into the German labour market is likely to differ considerably from the one of later immigration waves, such as the guest workers.

As for other countries, the language channel has been analysed for Germany. Dustmann (1994), Dustmann and Van Soest (2001) and Dustmann and Van Soest (2002) show that German skills positively impact the earnings position of migrants, arguing that ethnic language skills differentials could be explained by ethnic differentials in the clustering behaviour. Similarly, Danzer and Yaman (2010) find a negative relationship between ethnic concentration and language fluency which Kalter (2006) supports for second generation Turkish immigrants.

Overall, the effect of ethnic enclaves remains unclear both from a theoretical and empirical perspective. This is particularly true for Germany, on which there are hardly any studies. The present paper thus contributes to the existing literature by extending the enclave literature to Germany. In contrast to Kanas et al. (2011), the ethnic share is observed on the post code level which allows for a more precise estimation of the enclave effect. Post code areas, especially in cities where most immigrants live, can be considered sufficiently small to proxy for a social network. With respect to the institutions channels, this definition further makes sense as inhabitants are often allocated to institution based on their residential post code.

3 Data

To investigate the effects of ethnical clustering on the labour market outcome of migrants, data from the German Socio-Economic Panel (SOEP)[1] are employed, which is an annual representative household survey of individuals aged 16 and above and which started in 1984 (Schupp, 2009). For the present paper, the regional version with post code level regional information is used. Data for

[1] The data used in this paper was extracted using the Add-On package PanelWhiz for Stata®. PanelWhiz (`http://www.PanelWhiz.eu`) was written by Dr. John P. Haisken-DeNew (`john@PanelWhiz.eu`). See Hahn and Haisken-DeNew (2013) and Haisken-DeNew and Hahn (2010) for details. The PanelWhiz generated DO file to retrieve the data used here is available from me upon request. Any data or computational errors in this paper are my own.

the years 2007 to 2012 is extracted and the sample is restricted to those between 17 and 64, further excluding those in education, military service, on parental leave and working in sheltered workshops. In addition, immigrants with African background have been dropped from the sample, as they are too few to estimate reliable results. Based on these restrictions, the sample consists of 7593 observations of first generation immigrants which corresponds to 2397 persons. Since there are only few second-generation migrants in the sample, the analysis focusses on first-generation migrants.

The SOEP is matched with microm[2] raster data which provide information on the share of immigrants and their ethnic background in each 1-km^2-cell of Germany (see Budde and Eilers, 2014, for a detailed data description). The data are available for the years 2005, 2009 and 2010, and they are obtained by the analysis of pre- and surnames of the household's head. This allows for an efficient approximation of the share of foreigners in a small area. As the SOEP sample is not stratified for regional representation, a computation of the immigrants' share using the SOEP data is very likely to be biased. By merging the microm data, a reliable share of immigrants can be obtained.

Furthermore, the microm data provide separate variables for different ethnic groups. This allows to identify the effect of living in an enclave with people from the same or a similar cultural background. Arguably, cross-group effects are the lower the farther the distance between the two groups' cultures. Therefore, using the general share of immigrants in the regression is likely to underestimate the true effect. Based on the microm variables, the observed immigrants are grouped into 11 ethnic backgrounds[3], thus incorporating the possibility of cross-group effects between similar cultures (cf. e.g. Chowdhury and Pedace, 2007).

In addition, data obtained from the The Federal Institute for Research on Building, Urban Affairs and Spatial Development (Bundesinstitut für Bau-, Stadt- und Raumforschung) (BBSR) on the small-scale regional unemployment rate for the years 2009 to 2011 have been used. The data are obtained on the county level (*Kreis*) and aggregated to the local labour market area (LLMA) level, following definitions by Kosfeld and Werner (2012)[4]. Both raster data and data on the LLMA level were mapped to the post code level using a table obtained by intersecting different regional layers[5]. For the

[2] Microm Micromarketing-Systeme und Consult GmbH is a German marketing and consultancy agency.

[3] The groups are: African, Asian, Balkan, Greek, Middle East, Italian, Eastern European (Russia), Spanish/Portuguese/Latin Amerika (Spain), Aussielder, Turkish, other/Western. A detailed list of countries in the sample that were considered for each ethnic group can be found in the appendix, table A1.

[4] The necessary shapefile is provided by the Research Data Centre (RDC) Ruhr at Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI) based on 2013 county level definitions of the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie) (BKG).

[5] The table is provided by the RDC Ruhr at RWI. The intersection is based on shapefiles from 2013 obtained from the BKG for administrative German regions, from the European Environment Agency (EEA) for grid definitions and from Esri Geoinformatik GmbH Germany for 2004 post code regions. County areas that have changed within the observed time frame have been converted to 2013 boundaries.

aggregation, regional units have been weighted with the population size. Since both the microm and the Indicators and Maps for the Spatial and Urbal Development in Germany and Europe (Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa) (INKAR) data are not available during the whole observed time period, missing data have been linearly inter- and extrapolated.

Table 1: Summary Statistics of the Ethnic Measures

Variable	**Mean**	**SD**	**Min**	**Max**	**N**
Ethnicity dummies					
Italy	0.070	0.255	0	1	7593
Turkey	0.162	0.368	0	1	7593
Greece	0.016	0.126	0	1	7593
Spain	0.037	0.188	0	1	7593
The Balkans	0.148	0.355	0	1	7593
Russia	0.120	0.325	0	1	7593
Middle East	0.036	0.185	0	1	7593
Asia	0.037	0.190	0	1	7593
Western	0.120	0.325	0	1	7593
Aussiedler	0.254	0.436	0	1	7593
Share of foreigners*	9.016	4.765	1	33	7593
Ethnic share*	1.433	1.884	0	27	7593

*In percent.

Observations have been weighted using population weights. Shares pertain to post code area shares for observations in the sample. Data source: SOEP and microm.

Table 1 shows summary statistics of the ethnic groups in the sample. For the analysis, all observations have been weighted using the population weights provided with the SOEP (cf. Haisken-DeNew and Frick, 2005). In the sample, the largest ethnic groups constitute of immigrants from Turkey, the Balkans and Eastern Europe with a Russian cultural background, which is in line with the whole immigrant population in Germany (Statistisches Bundesamt, 2014a). Apart from Africans, who were so few in the sample that they were excluded, the smallest immigrant groups are from Greece, the Middle East and Asia. Given these ethnic groups, for post codes in the sample, the average share of foreigners of the same ethnicity in the post code area, the ethnic share, is 1.4% with a standard deviation of 1.88. The mean share of all foreigners in a post code area is 9.0% with a standard deviation of 4.77.

Table 2: Summary Statistics of the Controls

Variable	**Mean**	**SD**	**Min**	**Max**	**N**
Employment:					
Employed	0.697	0.459	0	1	7593
Gross wage (euro/ month)	2114.810	1427.965	100	9773	4811
ln(wage)	7.384	0.833	5	9	4811
Hours p/w	35.353	12.388	2	50	4811
Tenure	9.032	8.514	0	45	4811
SME	0.591	0.492	0	1	4811
Partner's net wage p/m	757.376	1100.436	0	16000	7593
Personal:					
Female	0.543	0.498	0	1	7593
Age	44.460	11.399	17	64	7593
Disabled	0.083	0.276	0	1	7593
Married	0.724	0.447	0	1	7593
Children in HH	0.709	1.001	0	9	7593
Single parent	0.031	0.174	0	1	7593
YSM	24.077	11.681	1	62	7593
Education:					
Low	0.270	0.444	0	1	7593
Medium	0.500	0.500	0	1	7593
High	0.231	0.421	0	1	7593
Region:					
Unemployment LLMA (%)	6.463	2.415	1	20	7593
Mean hourly wage LLMA (euro)	18.662	2.201	10	26	4811
West	0.975	0.155	0	1	7593
German:					
Oral German	3.952	0.921	1	5	4752
Written German	3.524	1.151	1	5	4752

Observations have been weighted using population weights. Data source: SOEP.

Table 2 shows the weighted summary statistics of the dependent and control variables for the first generation migrants in the sample. Most immigrants have stayed in Germany for a long period of time with the mean years since migration being 24 years. 70% of the persons observed are employed, ranging from marginal employment to full-time employment. The mean gross monthly wage in the sample is 2115 euro, which is lower than the average for Germans in the SOEP that is 2536 euros (cf.

table A3 in the appendix).

The average immigrant in the sample lives in an LLMA with an unemployment rate of 6.5% which is lower than the German average unemployment rates in the observed years that ranged from 7.6% to 10.1% (Statistisches Bundesamt, 2014b). This in line with the average native in the SOEP living in an LLMA with an unemployment rate of 7.9% (cf. table A3 in the appendix). This might indicate sorting of immigrants into regions that exhibit relatively favourable labour market conditions. LLMA gross hourly wages for the average immigrant amount to 18.7 euros which is more than the average gross wage of 15.5 euros in 2010 in Germany (Statistisches Bundesamt, 2013) and also more than the mean for natives in the SOEP which is 17.7 euros, further hinting at the sorting of immigrants. The majority of the observed persons (98%) lives in West Germany.

4 Empirical Strategy

The clustering effect on immigrants' labour market integration can be identified by comparing the labour market outcomes of individuals of the same ethnic background who live in neighbourhoods that differ with respect to the share of migrants with their own background. Labour market integration of immigrants is measured in two ways, employment probability and wage levels. Thus, the analysis is analogously conducted in two steps. However, due to non-random sorting of immigrants into enclaves based on unobservable characteristics of both the individual itself and the neighbourhood, the estimates may still be biased.

4.1 Baseline Models

Two baseline specifications are estimated for each step. Initially, the clustering effect is assumed to be the same for all ethnic groups. In the following step, this assumption is loosened and the effect is allowed to vary for different groups.

First, a linear probability model of being employed for individual i of the ethnic group g living in neighbourhood j located in LLMA k in year t is estimated as follows:

$$e_{ijkgt} = \alpha + C_{jgt}\beta_1 + U_{kt}\beta_2 + \bar{X}_{ijkgt}\beta_3 + \delta_t + u_{ijkgt}, \tag{1}$$

where e denotes employment which is one if someone is employed and zero otherwise. For the estimation, a normal distribution is assumed. α is a constant, C_{jgt} the share of immigrants with the same ethnic background g living in j in the year t. U_{kt} measures the unemployment rate in LLMA k. $\bar{X}_{ijkgt}$ is a vector of controls and δ_t denotes year fixed effects (FE) to account for exogenous shocks, such as overall changes of the German economy within a particular year. u is the error term.

Arguably, labour market outcomes depend on those nearby. The unemployed will commute within a certain area if necessary and consequently unemployment probabilities for similar individuals should equalise within the commuting area. If immigrants are assumed to move to regions where the general employment probability is high, i.e. cluster there and less elsewhere, the estimate of β_1 is likely to be upward biased. Therefore, the average unemployment rate in the respective LLMA, U_{kt}, is also included in the analysis. Including the average unemployment rate thus controls for potential reverse causality as well as region-specific labour market differences, and it is therefore similar to including LLMA FE. In contrast to the latter, including the average unemployment rate has the advantage of being robust to non-random sorting of immigrants into LLMAs. This is true as the LLMA unemployment rate includes the observations of both natives and immigrants. In addition, LLMA FE do not allow for LLMA-specific differences that change over time. Even though there could be other LLMA-specific characteristics that influence employment, which could be captured by including FE, it is argued that average unemployment is by far the most important LLMA specific factor. In addition, Chowdhury and Pedace (2007, p. 243) argue that including area FE could result in a collinearity problem with the ethnic share if the latter does not change substantially over time. Hence, including the LLMA unemployment rate is preferred over LLMA FE.

The estimates of equation 1 might still be biased if immigrants from certain ethnic groups tend to have worse labour market outcomes, e.g. for reasons of discrimination or cultural differences, and the same groups exhibit a considerably different clustering behaviour than other groups. Therefore, ethnic group FE, κ_g, are added to the regression.

In the second step, for those employed, the effect of ethnical clustering on the wages is analysed, extending Mincer's wage equation (Mincer, 1974) as follows:

$$\ln w_{ijkgt} = \alpha + C_{jgt}\beta_1 + W_{kt}\beta_2 + \bar{X}_{ijkgt}\beta_2 + \delta_t + u_{ijkgt}, \tag{2}$$

where $\ln w$ denotes log gross monthly wages and the rest is specified as above. W_{kt} denotes the average wage of both natives and immigrants in LLMA k. The model is estimated using ordinary least squares (OLS) and standard errors are clustered on the LLMA level, as explained above.

4.2 Endogeneity of the Location Decision

As individuals are likely to sort non-randomly into specific neighbourhoods based on unobserved individual and neighbourhood characteristics, the regression might suffer from an omitted variable bias. Thereby, the endogeneity problem due to omitted variables is twofold. First, there might be a selection of a specific group of immigrants, e.g. those that are less prone to learn the domestic language, into enclave neighbourhoods. Thus, unobserved individual characteristics may be correlated with both the labour market outcome and the ethnic share, also leading to biased estimates. Second, immigrants might particularly be drawn to neighbourhoods where unobserved characteristics are particularly favourable. The error can be divided into the idiosyncratic error ε_{ijkgt} and individual unobservables ψ_i and neighbourhood unobservables τ_j. A bias arises if $cov(C_i, \psi_i) \neq 0$ or $cov(C_i, \tau_j) \neq 0$, i.e. there is sorting into neighbourhoods on individual characteristics or neighbourhood characteristics are correlated with the ethnic share and labour market outcomes.

In order to control for individual sorting, a cell-based instrumental variable (IV) approach is used, as proposed in Bayer and Ross (2006) and Bauer et al. (2011). All possible combinations of major control characteristics, such as age and education constitute one cell. For each cell the ethnic shares are averaged, excluding the individual for which the average shall be used as an IV. These cell averages $\bar{C}_{jgt}$ are used as an instrument for the real clustering measure of regressions 1 and 2. This assumes that similar individuals "live" in similar neighbourhoods, eliminating the impact of individual sorting behaviour due to unobservables. The model is estimated using two-stage least-squares (2SLS).

Since the cell-based IV approach is sensitive to outliers in a relatively small sample, the group formation cannot be based on all control variables. The age variable is grouped into three categories. Additional controls used for the cell formation include the ethnic group, a children dummy and a dummy indicating a high educational level. Besides, cells are formed for each single year. Like this, the average cell contains 47 observations, which is reasonably large in order to argue that the approach is not extremely sensitive to outliers. Additionally, all cells with less than five observations are dropped.

These cell-averages constitute a valid instrument for two reasons:

1. Exogeneity: The cell averages are exogenous to individual labour market outcomes as individuals with the same observable characteristics exhibit the same ethnic share, leaving out individual unobservable characteristics. Since the individual's own share is not used for the calculation of cell averages, own unobservables are not even partly considered.

2. Relevance: The instrument is sufficiently correlated with the instrumented variable, ethnic share, if one assumption is made: Immigrants with similar observable characteristics exhibit similar individual preferences in the decision where to live, i.e. in the selection process. That is, arguably, a reasonable assumption. Characteristics, such as age, whether one has children and educational attainment, are major determinants of the residential location decision because individual returns to the same neighbourhood characteristics differ for those with different demographic characteristics (Bayer et al., 2007).

Besides selection on individual characteristics, unobserved neighbourhood characteristics may be correlated with the ethnic share and labour market outcomes. This may be the case for two reasons. First, there could be a direct link, e.g. an efficient aliens' or unemployment authority could a) attract further immigrants leading to a high ethnic share and b) decrease the unemployment among immigrants. However, this channel is of less concern for the present paper as they constitute one of the channels how the enclave influences labour market outcomes. The second channel could arise if individual preferences regarding unobservable neighbourhood characteristics τ_j change with unobservable abilities. This implicitly influences the chosen ethnic share. For example, if only the abled valued professional childcare facilities, all the abled immigrants would move to an area where there is a good provision of childcare and thus form an enclave. Immigrants in this enclave would then be much more likely to exhibit positive labour market outcomes as their unobserved abilities are more favourable but not because they live in an enclave.

To deal with this selection bias, first, a measure for the neighbourhood's unobserved characteristics has to be found. To this end, a hedonic flat price regression[6], as shown below, is run where h denotes flats and j a specific neighbourhood:

$$\ln P_{hjt} = \xi + \bar{H}_{hjt}\phi + \bar{N}_{jt}\zeta + \delta_t + \lambda_{hjt}. \tag{3}$$

The vector $\bar{H}_{hjt}$ includes flat-specific characteristics, such as living area, age and state of the flat, as well as other facilities such as an elevator, a cellar room, a balcony and a fitted kitchen. In addition, the neighbourhood-specific characteristics of the main regression are included in vector $\bar{N}_{jt}$, in particular the share of foreigners, average wage levels and average unemployment. δ_t denotes year FE.

The residuals λ_{hjt} of each flat h are averaged over the respective neighbourhood j. They should capture all neighbourhood-specific characteristics that cannot be observed – at least with the present data –, such as childcare facilities, parks and other amenities. Their average for each neighbourhood j in year t is included as an additional regressor in regressions 1 and 2 and serves as a proxy variable for the unobserved neighbourhood characteristics. For the flat price regression, data obtained from the internet portal *ImmobilienScout24.de* is used, covering almost 510,000 flats for rent between 2007 and 2012. The estimation results of this regression can be found in the appendix, table A4.

Once a proxy variable has been found, it is possible to control for the selection bias 2. First, including the proxy as an additional regressor, allows for assessing the first channel. If there is a significant estimate, it can be inferred that the unobservable neighbourhood characteristics influence labour market outcomes, which could also impact the ethnic share by immigrant sorting. However, there might still be a bias due to the correlation between individual labour-market relevant unobservables and individual preferences for the neighbourhood unobservables. For analysing this second channel, the same cell-based IV strategy described for solving selection bias 1 applies to the proxy of

[6]Due to bad coverage of housing offers for sale in cities, where most immigrants live in Germany, rather than taking the residuals from a flat price regression as Bayer and Ross (2006) and Bauer et al. (2011) suggest, flats offered for rent were used.

unobserved neighbourhood characteristics $\bar{\lambda}_{jt}$. Again, for each year, observations are sorted into cells based on their observable characteristics, i.e. age, ethnic group, a children indicator and a high education indicator. For each individual i, the average proxy of these cells is taken as an instrument for the proxy of individual i's neighbourhood. The instrument is exogenous to individual labour market outcomes as the cells are averaged without the observation's own value, thus leaving out any variation due to unobserved individual preferences. If assumed, again, that the location decision is similar for those with similar characteristics, the instrument is also relevant and correlated with $\bar{\lambda}_{jt}$.

In order to account for both types of sorting behaviour, a two-step approach is used as proposed by Bayer and Ross (2006) and Bauer et al. (2011), combining a control function and an instrumental variable (IV) approach. The two-step approach yields the following model for estimating labour markets outcomes y, which are either the employment probability or log wages:

$$y_{ijkgt} = \alpha + \hat{C}_{jgt}\beta_1 + Y_{kt}\beta_2 + \bar{X}_{ijkgt}\beta_3 + \delta_t + \hat{\bar{\lambda}}_{jt}\beta_4 + \varepsilon_{ijkgt}, \tag{4}$$

where $\hat{C}_{jgt}$ and $\hat{\bar{\lambda}}_{jt}$ are instrumented variables.

5 Results

5.1 Results of the Employment Regression

Table 3 shows the main results of the employment regression. Column (1) and (2) present results of the OLS estimation using first the total share of foreigners in the neighbourhood and second the ethnic share, i.e. foreigners of the same ethnicity, as the clustering measure. The total share of foreigners is slightly negative correlated to the employment status, whereas the own ethnic share is at the 1% level. This means the higher the share of immigrants (with the own ethnic background) the lower the employment probability. This is in line with theoretical predictions, arguing that a network is based on common norms and a common language. This is, of course, more likely to occur with persons with a similar ethnic background[7].

Also the estimates of the control variables exhibit the expected sign: The employment probability increases with age but at a diminishing rate, due to increasing labour market experience. An additional year spent in Germany is related to a higher employment probability which, intuitively, captures the assimilation process. Disability is associated with a lower employment probability whereas the employment probability increases with higher education. Women *per se* are not less likely to work but do so when they are either married and/or have children. Married men, on the other hand, are more

[7] In addition to the results shown here, a quadratic specification of the ethnic share has been tested. The results suggest that a linear relationship is appropriate.

Table 3: Employment Regression – Main Results

	OLS						IV			
	(1)		(2)		(3)		(4)		(5)	
Share of foreigners	-0.004*	(0.002)								
Ethnic share			-0.026***	(0.006)	-0.013*	(0.008)	-0.050***	(0.010)	-0.049***	(0.011)
Age	0.063***	(0.006)	0.065***	(0.006)	0.069***	(0.006)	0.063***	(0.007)	0.060***	(0.007)
Age sqaured	-0.001***	(0.000)	-0.001***	(0.000)	-0.001***	(0.000)	-0.001***	(0.000)	-0.001***	(0.000)
YSM	0.003***	(0.001)	0.004***	(0.001)	0.004***	(0.001)	0.005***	(0.001)	0.004***	(0.001)
Disabled	-0.293***	(0.050)	-0.292***	(0.049)	-0.290***	(0.049)	-0.292***	(0.046)	-0.308***	(0.043)
Education – low as base										
Medium	0.126***	(0.035)	0.111***	(0.033)	0.085***	(0.030)	0.100***	(0.034)	0.101***	(0.035)
High	0.146***	(0.032)	0.125***	(0.032)	0.114***	(0.030)	0.114***	(0.038)	0.122***	(0.040)
Female	0.053	(0.037)	0.053	(0.037)	0.022	(0.038)	0.050	(0.040)	0.080*	(0.045)
Married	0.062	(0.040)	0.074*	(0.040)	0.064	(0.038)	0.087**	(0.042)	0.096**	(0.046)
Married × female	-0.153***	(0.044)	-0.155***	(0.044)	-0.125***	(0.044)	-0.150***	(0.048)	-0.173***	(0.053)
Children in HH	-0.004	(0.016)	0.001	(0.016)	0.007	(0.016)	0.015	(0.016)	0.015	(0.018)
female × Children in HH	-0.089***	(0.020)	-0.090***	(0.019)	-0.083***	(0.020)	-0.098***	(0.020)	-0.107***	(0.021)
Single parent	-0.093	(0.069)	-0.098	(0.070)	-0.105	(0.073)	-0.112	(0.073)	-0.108	(0.078)
Partner's net wage	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
Unemployment LLMA	-0.014***	(0.004)	-0.012***	(0.004)	-0.010**	(0.004)	-0.009**	(0.004)	-0.007	(0.005)
West	0.141	(0.105)	0.157	(0.108)	0.158	(0.105)	0.217**	(0.107)	0.210	(0.148)
Ethnic share - Turkish as base										
Italy × ethnic share					-0.024	(0.054)				
Greece × ethnic share					0.066*	(0.034)				
Spain × ethnic share					0.136	(0.204)				
Balkans × ethnic share					0.029	(0.023)				
Russia × ethnic share					-0.026	(0.038)				
Middle East × ethnic share					-0.041	(0.073)				
Asia × ethnic share					-0.968	(0.615)				
Western × ethnic share					0.077**	(0.032)				
Aussiedler × ethnic share					0.025	(0.038)				
Lambda									0.012	(0.048)
Constant	-0.561***	(0.160)	-0.627***	(0.164)	-0.834***	(0.161)	-0.633***	(0.175)	-0.599***	(0.202)
Year dummies	Yes		Yes		Yes		Yes		Yes	
Ethno dummies	No		No		Yes		No		No	
R^2	0.195		0.203		0.227		0.204		0.217	
N	7593		7593		7593		6982		5328	
F-statistic (1st stage)							97.69		79.20	

Standard errors in parentheses. Standard errors are clustered on the LLMA level. Estimates are weighted using population weights.
IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

likely to work which indicates the predominance of the male breadwinner role model among immigrants in Germany. Single parents do not exhibit significantly different employment patterns which could be due to the opposing incentives, i.e. greater need for money and greater need for childcare time, that cancel each other out. The partner's net wage seems to be economically insignificant when immigrants decide upon their labour market availability. The unemployment rate in the LLMA is associated with a lower probability to be employed for the individual which is suggestive evidence for the assumed arbitrage effects.

Previous research has suggested that enclave effects might differ between ethnic groups (Hou and Picot, 2003). Therefore, in column (3), the ethnic share is interacted with the different ethnic groups where the largest immigrant group, the Turkish, are the base category. However, the main coefficient stays negative. In particular, the estimate of the ethnic share is halved to -0.013. While there is a negative and insignificant correlation for most ethnic groups, it is found to be positive and quite large for Western immigrants. This could be explained by their cultural background that is close to the German culture and the different reasons for migration compared to other groups. Thus, they can benefit from the network effects while many of the disadvantages like increased discrimination, different work ethics and higher language barriers do not apply to them to the same extent.

As explained before, the results may still suffer from an endogeneity bias. Column (4) deals with selection bias 1, i.e. selection based on individual unobservable characteristics. To that end, the ethnic share is instrumented with cell averages of observations having similar observable characteristics. The instrument is valid with a F-statistic of 54.67. While the other estimates remain quite stable, the estimate of the ethnic share almost doubles from -0.026 to -0.050. An increase of one percentage point in the ethnic share is associated with an employment probability lowered by 5 percentage points. As the average ethnic share is only around 1.3 in the sample, this might not seem economically significant at first. An increase in the ethnic share of one standard deviation results in a lower probability to be employed of ($1.884 \times -0.050 = -0.094$) 9.4 percentage points. Given that the mean employment in the sample is 73.5%, the found negative impact seems relevant in size. After accounting for selection bias 1, living among same-ethnic immigrants seems to impact the employment probability more severely. This direction of the bias is also observed by Cutler et al. (2008). Since the selection mechanism cannot be observed, any reason for the observed bias is speculative. However, it is possible that those with more favourable unobserved labour market attributes also have a higher preference for the ethnic goods of the enclave which would lead to positive selection into the enclave and results in the observed upward bias of the OLS estimate.

In addition, the estimates might still suffer from selection bias 2, i.e. selection on unobservable neighbourhood characteristics. Column (5) shows the estimation results, when the proxy λ for the unobserved neighbourhood characteristics, the residuals from a house price regression, is included. All estimates remain roughly the same and the estimate of the proxy is not significantly different from zero. Thus, unobserved neighbourhood characteristics do not appear to influence labour market outcomes directly. This would be possible if they captured amenities like closeness to large firms or good transportation. Thus, the first channel of selection bias 2 seems to be no problem for the current analysis. Unfortunately, it is not feasible to test for the second channel. This could be done by instrumenting for the proxy, but the instrument is clearly weak (F-statistic of 3.88). Yet, for immigrants it is likely that individual unobservables are far more important than unobservable neighbourhood characteristics when deciding where to live, in particular whether to live in an enclave or not. Since one of the channels of the selection bias could be ruled out by a insignificant estimate of the proxy and the other seems to be of little relevance, selection bias 2 is thus assumed to be rather small and not considered further.

Instrumenting the interacted ethnic share of column (3) has also been tested. Yet, due to the weakness of the instrument for some ethnic groups, the results are not reliable. Since the OLS estimation in column (3), however, does not suggest large differences among most immigrant groups, the issue seems less pressing.

5.2 Results of the Wage Regression

The results of the wage regression are displayed in table 4. Columns (1) to (3) present the OLS results. As in the case of employment, ethnic networks seem to be the driving factor rather than the overall share of foreigners with an estimate of -0.028 that is significant at the 1% level. Living in an enclave is therefore associated with lower gross monthly wages. Similar to the results from the employment regression, there also seem to be only small differences among ethnic groups. A notable exception are immigrants from the Middle East and Greece. The first exhibit a positive estimate for the ethnic share which is significantly different to Turkish immigrants. Possibly, these immigrants face stronger discrimination on the labour market and thus benefit more from the information within the enclave network (for Australia, see Booth et al., 2012). Likewise, due to the large differences in the cultures, it is possible that there are more possibilities for a competitive enclave labour market, resulting in higher wages. Although the coefficient is less pronounced for Greek immigrants it is positively correlated on the 1% level. This is in line with the findings for employment. However, as there are only about 85 (Middle East) and 171 (Greece) observations each, the results may not be representative.

The IV estimates are presented in columns (4) and (5). Column (4) shows the results of the IV estimation with regard to selection bias 1. The OLS estimate seems slightly upwards biased, with the estimate dropping from -0.028 to -0.036, while the other estimates remain fairly stable. Including the

Table 4: Wage Regression – Main Results

	OLS						IV			
	(1)		(2)		(3)		(4)		(5)	
Share of foreigners	-0.003	(0.003)								
Ethnic share			-0.028***	(0.007)	-0.020*	(0.011)	-0.036***	(0.012)	-0.034***	(0.012)
Age	0.043***	(0.009)	0.046***	(0.009)	0.049***	(0.009)	0.039***	(0.008)	0.035***	(0.010)
Age sqaured	-0.001***	(0.000)	-0.001***	(0.000)	-0.001***	(0.000)	-0.000***	(0.000)	-0.000***	(0.000)
YSM	0.005***	(0.002)	0.006***	(0.002)	0.006***	(0.002)	0.005***	(0.002)	0.005***	(0.002)
Female	-0.085*	(0.046)	-0.086*	(0.047)	-0.104**	(0.047)	-0.109**	(0.045)	-0.117**	(0.055)
Disabled	-0.031	(0.051)	-0.026	(0.050)	-0.022	(0.047)	-0.024	(0.051)	-0.053	(0.066)
Education – low as base										
Medium	0.081**	(0.031)	0.073**	(0.031)	0.044	(0.031)	0.080**	(0.032)	0.090***	(0.031)
High	0.351***	(0.041)	0.333***	(0.040)	0.310***	(0.035)	0.347***	(0.037)	0.368***	(0.040)
Married	0.091*	(0.047)	0.106**	(0.048)	0.097**	(0.046)	0.074	(0.052)	0.085	(0.060)
Married × female	-0.120**	(0.056)	-0.126**	(0.056)	-0.107*	(0.055)	-0.093	(0.058)	-0.074	(0.072)
Hours	0.046***	(0.002)	0.046***	(0.002)	0.046***	(0.002)	0.046***	(0.002)	0.046***	(0.002)
Tenure	0.015***	(0.003)	0.015***	(0.003)	0.015***	(0.002)	0.016***	(0.002)	0.017***	(0.002)
SME	-0.218***	(0.027)	-0.216***	(0.026)	-0.199***	(0.023)	-0.210***	(0.028)	-0.209***	(0.027)
Mean wage LLMA	0.024***	(0.008)	0.024***	(0.007)	0.027***	(0.008)	0.027***	(0.008)	0.024***	(0.007)
West	0.452***	(0.074)	0.455***	(0.073)	0.412***	(0.089)	0.407***	(0.080)	0.417***	(0.068)
Ethnic share - Turkish as base										
Italy × ethnic share					-0.054	(0.033)				
Greece × ethnic share					0.101***	(0.034)				
Spain × ethnic share					-0.076	(0.124)				
Balkans × ethnic share					0.028	(0.019)				
Russia × ethnic share					0.020	(0.049)				
Middle East × ethnic share					0.239*	(0.135)				
Asia × ethnic share					0.439	(1.299)				
Western × ethnic share					0.040	(0.090)				
Aussiedler × ethnic share					0.049	(0.051)				
Lambda									0.088***	(0.032)
Constant	3.806***	(0.240)	3.750***	(0.242)	3.651***	(0.271)	3.931***	(0.240)	4.047***	(0.250)
Year dummies	Yes		Yes		Yes		Yes		Yes	
Ethno dummies	No		No		Yes		No		No	
R^2	0.733		0.736		0.745		0.739		0.740	
N	4811		4811		4811		4468		3413	
F-statistic (1st stage)							57.04		47.84	

Standard errors in parentheses. Standard errors are clustered on the LLMA level. Estimates are weighted using population weights.
IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

proxy for the unobserved neighbourhood characteristics slightly increases the estimate of the ethnic share, as seen from column (5). The estimate for the proxy itself is quite large compared to the other estimates with 0.088 and significant at the 1% level. Thus, certain neighbourhood unobservables, such as closeness to certain firms, seem to be connected to higher wages. However, as mentioned earlier, it is likely that for immigrants selection within an LLMA is rather based on individual characteristics. This is supported by the small change of the ethnic share estimate. Unfortunately, due to the weakness of the instrument (F-statistic of 1.73), this cannot be tested however, which is why selection bias 2 is again not further considered albeit appearing more important for wages. Also similar to the employment regression, the interacted IV estimation is not feasible. However, again, this does not seem to be a severe problem, as there seem to be only little differences between ethnic groups in the first place – especially considering the low number of observations from Greek immigrants and those from the Middle East. Overall, it can be summarised that an increase in the ethnic share of one percentage point, leads to a decrease in wages of almost 3%. A standard deviation increase in the ethnic share thus leads to lower wages of about ($1.884 \times -0.034 = -0.070$) 7.0%, which seems to be relevant in size, and at the mean wage of 2059 Euro translates to a loss of 122 Euros per month.

5.3 Robustness Checks

To test the reliability of the above results, several tests were performed to check the robustness of the empirical specification. First, a different measure for clustering is used. While the absolute share of one ethnic group in the same neighbourhood is used in the baseline specification, following Borjas (2000), the share relative to the overall share of that ethnic group in Germany is also used. The relative share C^r_{jg} of group g in neighbourhood j is defined as:

$$C^r_{jg} = \frac{N_{gj}/N_j}{N_g/N}, \tag{5}$$

where N is the total number of persons. This relative share is less sensitive to outliers and thus helps detect clustering of small ethnic groups in Germany. Given that networks of small minorities can be especially strong because of the small number of its members, this makes sense. However, less value is given to the network size. Even large networks will have a low relative share if the ethnic group is also comparably large in Germany.

Tables 5 and 6 show how the clustering estimates change with different measures. First, the relative clustering measure is used, i.e. the share of immigrants in a neighbourhood relative to the overall share of this group in Germany. With the relative clustering measure, for both employment and wages, OLS estimates drop by roughly two thirds. However, since the unit is no longer the same, the two estimates are not comparable in size. In spite of that, two main findings remain valid. First,

Table 5: Employment Regression – Robustness: Measures

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Ethnic share	-0.027***	-0.050***				
	(0.007)	(0.010)				
Relative clustering			-0.010**	-0.033***		
			(0.004)	(0.010)		
Ethnic share lagged					-0.028***	-0.051***
					(0.007)	(0.011)
R^2	0.212	0.204	0.205	0.178	0.222	0.213
N	6982	6982	6982	6982	6234	6234
F-statistic (1st stage)		97.69		50.31		88.64

Standard errors in parentheses. Standard errors are clustered on the LLMA level.Estimates are weighted using population weights. IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

there is a negative correlation between economic outcomes and clustering (column (3)) and there is a positive bias of the OLS estimate (column (4)). Furthermore, to account for the potential reflection problem, the lagged ethnic share is used instead of the current ethnic share as van Ham et al. (2011, p. 6) suggest. The lagged share cannot be simultaneously determined with the individual labour market outcome, rendering the simultaneity problem irrelevant. Using the lagged ethnic share hardly alters the results at all. This indicates that the analysis does not suffer from a reflection problem.

The literature has further suggested that the enclave effect is not the same for all ethnic subsamples. In particular, two differentiations are considered here. First, it seems reasonable to assume that clustering effects differ for men and women (Aguilera, 2005; Gilbertson, 1995; Zhou and Logan, 1989). Especially for those with a cultural background where women do not usually participate in the labour market, enclaves might affect the employment probability for women more negatively, or less positively, than for men. Second, it has been suggested that the clustering effect might differ across skill levels (Sousa, 2013; Warman, 2007).

Table 6: Wage Regression – Robustness: Measures

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Ethnic share	-0.028***	-0.036***				
	(0.007)	(0.012)				
Relative clustering			-0.011***	-0.039***		
			(0.004)	(0.013)		
Ethnic share lagged					-0.029***	-0.034***
					(0.007)	(0.012)
R^2	0.739	0.739	0.738	0.727	0.741	0.741
N	4468	4468	4468	4468	3977	3977
F-statistic (1st stage)		57.04		29.01		55.04

Standard errors in parentheses. Standard errors are clustered on the LLMA level.

Estimates are weighted using population weights.

IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7 and 8 explore possible differences between men and women. Columns (1) and (2) show the estimates of the male subsample and columns (3) and (4) the estimates of the female subsample. Women seem to be more negatively affected by living in an enclave both with respect to employment and wages which is in line with theoretical predictions and several empirical findings (Gilbertson, 1995; Zhou and Logan, 1989). It can thus be concluded that overall, by living in an enclave, women are not affected differently from men. Furthermore, selections seems to play a greater role for women.

Table 7 also shows the results for the employment regression with different skill groups. Interestingly, medium educated immigrants seem least negatively affected by the enclave regarding employment. This is seen by the subsample results, both OLS and IV ones, and also confirmed by the interacted estimates. Several reasons could explain this. High-skilled workers possibly suffer from spatial mismatch the strongest because they could be overqualified for certain jobs. This might not apply to medium-skilled workers to the same extent. Besides, in contrast to low-skilled workers, they can build on previous knowledge to benefit from human capital externalities in the enclave. In addition, medium-skilled immigrants might be particularly attractive for the enclave labour market. While they are better educated than low-skilled ones, they are also less expensive than high-skilled immigrants. Taken together, these reasons could explain why the employment effect of the enclave

Table 7: Employment Regression – Subgroups

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	OLS	IV	OLS	IV
	Male		Female			
Ethnic share	-0.024***	-0.035***	-0.030***	-0.065***		
	(0.008)	(0.012)	(0.009)	(0.015)		
R^2	0.262	0.260	0.167	0.154		
N	3149	3149	3833	3833		
F-statistic (1st stage)		65.84		59.51		
	Low Skilled		Medium Skilled		High Skilled	
Ethnic share	-0.033***	-0.070***	-0.017*	-0.038***	-0.030**	-0.029
	(0.006)	(0.021)	(0.010)	(0.012)	(0.012)	(0.029)
R^2	0.276	0.251	0.161	0.155	0.245	0.245
N	1948	1948	3587	3587	1447	1447
F-statistic (1st stage)		45.13		68.95		16.30

Standard errors in parentheses. Standard errors are clustered on the LLMA level.

Estimates are weighted using population weights.

IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

is less negative for medium-skilled immigrants. Regarding wages, living in an enclave reduces wages for high skilled even to a greater extent than for low skilled, as can be seen from table 8. Another concern is the large number of *Aussiedler* in the sample. They differ from other immigrants in many aspects. In particular, many of them speak good German and also familiarise easily with the German cultural and societal system. Therefore, their integration process is entirely different to that of other immigrants. The main analysis with interacted ethnic shares has already suggested that there is no significant difference between *Aussiedler* and other immigrants. Likewise, the results remain stable for the employment regression when excluding *Aussiedler*, as displayed in panel (a) of table 9. However, for wages, excluding *Aussiedler* from the regression alters the results considerably, as can be seen from panel (b). The estimated enclave effect becomes smaller and insignificant in the IV specification, which might be due to the lower number of observations. Likewise, the estimate for *Aussiedler* only is larger but not significant. Taken together, this could indicate that the negative enclave effect is driven by the *Aussiedler* group to a large extent. Yet, this is somewhat contradicted by the fact that, in the main regression, an interaction of the ethnic share with an *Aussiedler* dummy has not yielded

Table 8: Wage Regression – Subgroups

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
	Male		Female			
Ethnic share	-0.029***	-0.029*	-0.032***	-0.056***		
	(0.009)	(0.016)	(0.008)	(0.019)		
R^2	0.585	0.585	0.717	0.715		
N	2244	2244	2224	2224		
F-statistic (1st stage)		40.80		28.50		
	Low Skilled		Medium Skilled		High Skilled	
Ethnic share	-0.024***	-0.025	-0.023**	-0.034**	-0.048**	-0.081**
	(0.009)	(0.033)	(0.010)	(0.014)	(0.021)	(0.037)
R^2	0.765	0.765	0.730	0.729	0.741	0.739
N	1016	1016	2488	2488	964	964
F-statistic (1st stage)		29.60		41.54		6.19

Standard errors in parentheses. Standard errors are clustered on the LLMA level.

Estimates are weighted using population weights.

IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

a significant estimate. However, all in all, a negative enclave effect on wages can still be confirmed.

In the preceding analysis, the neighbourhood is represented by post code areas. We argue that these are sufficiently small to allow for a representation of an actual social network of neighbours. However, it is possible that the network is in fact larger. Therefore, a weighted sum of the neighbouring post code areas is included in the regression as another regressor. Table 10 shows the results. For employment, the neighbouring ethnic shares do not seem to matter. While the ethnic share estimate slightly decreases, the estimate for the neighbours remains insignificant. In addition, the R^2 suggests that no additional information is gained by including the neighbours.

Table 9: Robustness: *without Aussiedler*

	Employment		Wages	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Ethnic share	-0.022***	-0.043***	-0.019**	-0.011
	(0.007)	(0.013)	(0.008)	(0.015)
Year dummies	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R^2	0.208	0.201	0.740	0.739
N	4762	4762	2817	2817
F-statistic (1st stage)		78.15		43.89

Standard errors in parentheses. Standard errors are clustered on the LLMA level.
Estimates are weighted using population weights.
IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Robustness: Neighbouring Post Codes

	(1)		(2)		(3)		(4)	
	employed		employed		wage		wage	
Ethnic share	-0.027***	(0.007)	-0.024**	(0.012)	-0.028***	(0.007)	-0.018	(0.011)
Neighb. ethnic share			-0.005	(0.017)			-0.015	(0.012)
R^2	0.212		0.212		0.739		0.739	
N	6982		6982		4468		4468	

Standard errors in parentheses. Standard errors are clustered on the LLMA level.
Estimates are weighted using population weights.
IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

For the wage regression, the single estimate is split up between the own share and the neighbouring ones, though both turn insignificant. This hints at a strong correlation between the neighbouring post code areas that do, however, not increase the model's fit. It seems rather likely that due to the high correlation some of the correlation of the own post code area is now captured by the neighbouring ones. In return, it can be argued that they themselves seem to be of little importance. Overall, it

Table 11: Placebo Regression – Results

	(1) employed		(2) employed		(3) wage		(4) wage	
Share of foreigners	-0.002**	(0.001)			0.000	(0.002)		
Ethnic share			-0.000	(0.002)			0.000	(0.005)
R^2	0.208		0.208		0.665		0.665	
N	65045		65045		46341		46341	

Standard errors in parentheses. Standard errors are clustered on the LLMA level.

Estimates are weighted using population weights.

IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

seems, therefore, appropriate to use post code areas as approximations of relevant neighbourhoods.

Albeit the previous evidence for enclave effects, it might still be possible that the estimates capture other neighbourhood effects that are completely unrelated to the ethnic network. To explore that, a placebo regression is run. To that end, Germans are randomly assigned to an ethnic group and, based on these, placebo regressions are estimated, results of which can be found in table 11. For both employment and wages, the OLS estimates are insignificant and very small in size. Also the IV estimates do not change these results. The estimates of the other covariates are, however, as expected and significant, indicating the reliability of the results. Thus, it can be concluded that the previous findings on enclave effects really capture ethnic network effects rather than general neighbourhood effects.

5.4 Insights into Channels of the Enclave Effect

Apart from evaluating the overall effect of living in an enclave, the channels through which the enclave effect works are of particular interest. If insight could be gained as to what are the most important channels, this could be valuable for policy-making. Therefore, in the following, two channels are explored. First, the spatial mismatch hypothesis is looked at, second the language channel is evaluated.

Table 12 shows the estimation results when the job type is included as an additional regressor in the wage estimation for gaining insights into the spatial mismatch hypothesis. As can be seen, high-skilled white collar workers have significantly higher wages than low-skilled blue collar job workers, which is as expected. At the same time, the estimate of the ethnic share is decreased compared to the estimation without job types. This indicates that some of the enclave effect is captured by the job type variables. This could hint at spatial mismatch. In the light of spatial mismatch, immigrants in an enclave live far off the good job opportunities and are forced to work in occupations that pay less, typically low-skilled and blue collar work. Thus, table 12 provides suggestive evidence of the spatial mismatch hypothesis.

Another prominent channel of the clustering effect is via language. Especially against the background of the complexity of the German language and its predominance in the German business world, lower language skills seem to be a huge disadvantage regarding both employment chances and wage levels (Dustmann, 1994; Dustmann and Van Soest, 2001, 2002). Table 13 shows the results of regressions with language skills as the dependent variable, where skills are self-measured on a scale from one to five, five being best. For oral German skills, a negative correlation between the ethnic share and German skills significant at the 1% level is found. The estimate becomes larger in size when accounting for individual sorting (F-statistic of 50.58). This would imply that those with rather good language learning abilities are drawn into enclaves, while the opposite is expected intuitively.

However, since the IV cells are also based on ethnic groups, it is possible that it captures part of a measurement error that is common to a particular ethnic group. This is supported by the interacted OLS estimates of column (3). The language skills of almost all immigrants groups are significantly higher than those of Turkish immigrants. Notably, Greek immigrants seem to have language skills that are superior to any other group albeit Greek and German are not extraordinarily similar. This could hint at a systematic higher estimation of language skills by Greek people. Also the clustering effect on language skills potentially differs between ethnic groups. While there is literally no correlation for most, a strong negative correlation is found for Eastern European immigrants. However, overall, the results do not draw a clear picture and might suffer from a measurement error due to the self-reported language skills. Unfortunately, the interacted IV estimation is, again, unfeasible due to weak instruments. The results for written German skills are quite similar and can be found in the appendix, table A5.

Overall, the evidence for negative clustering effects on employment and wages seem both economically relevant and reliable. The OLS estimate is found to be upward biased due to individual sorting, while there is only little evidence for the selection on neighbourhood characteristics. No convincing evidence could be found for either ethnic group or gender differences in the enclave effect. Regarding skill levels, medium-skilled immigrants seem to suffer least from living in an enclave with respect to their employment probability, while no difference is found between low- and high-skilled immigrants. With respect to wages, no such differentials can be seen. Furthermore, questions regarding the channels of the enclave effect remain to a large extent unanswered.

Table 12: Wage Regression – Robustness: Job Type

	(1) OLS		(2) IV		(3) OLS	
Ethnic share	-0.022***	(0.007)	-0.029***	(0.011)	-0.019*	(0.010)
high-skilled white collar job	0.223***	(0.051)	0.222***	(0.050)	0.208***	(0.047)
low-skilled white collar job	0.021	(0.032)	0.021	(0.031)	0.016	(0.031)
high-skilled blue collar job	0.037	(0.034)	0.039	(0.034)	0.025	(0.034)
Trained for job	0.134***	(0.033)	0.131***	(0.033)	0.120***	(0.032)
Turkish as base						
Italy × ethnic share					-0.045	(0.030)
Greece × ethnic share					0.207***	(0.079)
Spain × ethnic share					-0.069	(0.076)
Balkans × ethnic share					0.032	(0.019)
Russia × ethnic share					0.028	(0.043)
Middle East × ethnic share					0.190***	(0.063)
Asia × ethnic share					1.726*	(0.893)
Western × ethnic share					0.059	(0.081)
Aussiedler × ethnic share					0.044	(0.048)
Year dummies	Yes		Yes		Yes	
Ethno dummies	No		No		Yes	
Controls	Yes		Yes		Yes	
R^2	0.759		0.759		0.765	
N	4424		4424		4424	
F-statistic (1st stage)			48.35			

Standard errors in parentheses. Standard errors are clustered on the LLMA level.

Estimates are weighted using population weights.

IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusion

Since ethnic clustering is common in Germany, a better understanding of its effects on the integration of immigrants could be important for integration policies, especially in the light of rising immigration and a skilled worker shortage. Yet, only little empirical research has been conducted for Germany. In

Table 13: Regression of Oral German Skills

	(1) OLS		(2) IV		(3) OLS	
Ethnic share	-0.083***	(0.020)	-0.159***	(0.026)	-0.031	(0.024)
Age	-0.065***	(0.012)	-0.059***	(0.014)	-0.049***	(0.012)
Age sqaured	0.000**	(0.000)	0.000	(0.000)	0.000	(0.000)
YSM	0.044***	(0.003)	0.047***	(0.003)	0.047***	(0.003)
Female	0.122**	(0.051)	0.099*	(0.051)	0.082*	(0.048)
Disabled	-0.149*	(0.075)	-0.146**	(0.072)	-0.134**	(0.060)
Married	-0.278***	(0.063)	-0.252***	(0.066)	-0.241***	(0.058)
Medium	0.415***	(0.067)	0.371***	(0.068)	0.290***	(0.053)
High	0.766***	(0.087)	0.699***	(0.089)	0.552***	(0.078)
Children in HH	-0.003	(0.029)	0.020	(0.028)	0.004	(0.026)
Single parent	-0.060	(0.127)	-0.062	(0.131)	-0.010	(0.120)
West	0.025	(0.160)	0.089	(0.160)	-0.007	(0.115)
Turkish as base						
Italy × ethnic share					-0.041	(0.112)
Greece × ethnic share					0.069	(0.080)
Spain × ethnic share					-0.357	(0.300)
Balkans × ethnic share					0.008	(0.045)
Russia × ethnic share					-0.224**	(0.087)
Middle East × ethnic share					0.220	(0.205)
Asia × ethnic share					-1.337	(1.561)
Western × ethnic share					0.138	(0.109)
Aussiedler × ethnic share					-0.142	(0.151)
Ethno dummies	No		No		Yes	
Constant	4.924***	(0.286)	4.838***	(0.305)	4.253***	(0.328)
Year dummies	Yes		Yes		Yes	
R^2	0.322		0.295		0.381	
N	4310		4310		4310	
F-statistic (1st stage)			90.73			

Standard errors in parentheses. Standard errors are clustered on the LLMA level.
Estimates are weighted using population weights.
IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

addition, the results from both economic theory and empirical research for other countries do not draw a clear picture. In this paper, the effect of residential clustering on the labour market outcome of first-generation immigrants in Germany was analysed empirically. It, thus, contributes to the literature by extending it to Germany. In contrast to the most closely related study by Kanas et al. (2011), who use data on the level of federal states (*Länder*), the ethnic share was observed on the postcode level which allows for a sensible approximation of actual neighbourhoods and social networks. This was achieved by combining the SOEP with small-scale raster data from microm on the ethnic share.

For the analysis, two measures for labour market integration were used: the employment probability and the wage level. Yet, the selection of immigrants into enclaves is likely to be non-random, e.g. based on motivation and language learning abilities. In particular, two selection biases could be distinguished. First, the selection of unobserved individual characteristics, and, second, the selection on unobserved neighbourhood characteristics. Both could result in an endogeneity bias. In order to control for this, an IV strategy was combined with a control function approach.

The analysis suggests a negative enclave effect on both employment and wages, that is economically relevant and robust to a variety of specifications. A standard deviation increase in the ethnic share relates to a lower employment probability of 10.0 percentage points. In addition, it is associated with a wage loss of 7.0% which, at the mean, is equivalent to 122 euros per month. There seem to be only little differences between different immigrant groups. In particular, the enclave effect on employment is positive for Western immigrants and for immigrants from Greece. It has been argued, this could be due to lower costs of living in an enclave because of the similarity of cultures and their good social integration. For immigrants from the Middle East and Greece, the clustering effect on wages is notable in size and positive. Yet, for both groups there are only few observations so that these results have to be taken with a grain of salt.

The estimates of the IV strategy confirm the presence of selection bias due to individual sorting, which is found to be upwards, but little evidence is found for the a correlation between the ethnic share and unobserved neighborhood characteristics. Possibly, those with more favourable unobserved labour market attributes also have a higher preference for living in the enclave which could lead to positive selection into the enclave and results in the observed upward bias of the OLS estimate.

To test the reliability of these results, several robustness checks were performed. The results remained stable. In addition, two channels of the enclave effects were analysed in more detail. If insight could be gained as to what are the most important channels, this could be valuable for policy-making. First, the spatial mismatch hypothesis was looked at, second the language channel was evaluated. There is suggestive evidence for spatial mismatch. The language channel could not be confirmed convincingly as the analysis is likely to suffer from a measurement bias.

Overall, enclaves could thus indeed be one reason why immigrants persistently earn less than natives, even when they have completed their education in Germany. In addition, the enclave effect can also partly explain ethnic group differentials in wage assimilation. While enclaves can help immigrants get on in the new country, in the long-run – the average immigrant in the sample has been in the country for almost 25 years – labour market effects seem to be detrimental for immigrants in Germany. Hence, for most immigrants in Germany, the "warm embrace" of the enclave is but cold comfort when immigrants face the labour market.

References

Acemoglu, D. and Angrist, J. (2000). How large are human-capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, 15:9.

Aguilera, M. B. (2002). The impact of social capital on labor force participation: Evidence from the 2000 social capital benchmark survey. *Social Science Quarterly*, 83(3):853–874.

Aguilera, M. B. (2003). The impact of the worker: How social capital and human capital influence the job tenure of formerly undocumented mexican immigrants. *Sociological Inquiry*, 73(1):52–83.

Aguilera, M. B. (2005). The impact of social capital on the earnings of Puerto Rican migrants. *Sociological Quarterly*, 46(4):569–592.

Aldashev, A., Gernandt, J., and Thomsen, S. L. (2012). The immigrant-native wage gap in germany. *Jahrbücher für Nationalökonomie*, (05):490–517.

Aretz, B. (2013). Gender differences in german wage mobility. IZA Discussion Paper 7158, Institute for the Study of Labor (IZA).

Bauer, T. K., Fertig, M., and Vorell, M. (2011). Neighborhood effects and individual unemployment. IZA Discussion Paper 6040, Institute for the Study of Labor (IZA).

Bauer, T. K. and Zimmermann, K. F. (1997). Unemployment and wages of ethnic Germans. *The Quarterly Review of Economics and Finance*, 37, Supplement 1:361–377.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. Working Paper 13236, National Bureau of Economic Research.

Bayer, P., Ross, S., 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.

Bayer, P. and Ross, S. L. (2006). Identifying individual and group effects in the presence of sorting: A neighborhood effects application. Working Paper 12211, National Bureau of Economic Research.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S. *The Review of Economic Studies*, 79(1):128–161.

Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.

Booth, A. L., Leigh, A., and Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment: Discrimination across minority groups. *Oxford Bulletin of Economics and Statistics*, 74(4):547–573.

Borjas, G. J. (1986). The self-employment experience of immigrants. *Journal of Human Resources*, 21(4):485–506.

Borjas, G. J. (2000). Ethnic enclaves and assimilation. *Swedish Economic Policy Review*, 7(2):89–122.

Brücker, H. and Trübswetter, P. (2007). Do the best go West? An analysis of the self-selection of employed East-West migrants in Germany. *Empirica*, 34(4):371–395.

Budde, R. and Eilers, L. (2014). *Sozioökonomische Daten auf Rasterebene - Datenbeschreibung der microm-Rasterdaten*, volume 77 of *RWI Materialien*. RWI, Essen.

Chiswick, B. R. and Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics*, 13(2):246–288.

Chiswick, B. R. and Miller, P. W. (1996). Ethnic networks and language proficiency among immigrants. *Journal of Population Economics*, 9(1):19–35.

Chiswick, B. R. and Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, 38(3):391–409.

Chowdhury, M. and Pedace, R. (2007). Ethnic enclaves and labor markets: An analysis of immigrant outcomes in California. *Contemporary Economic Policy*, 25(2):238–249.

Clark, K. and Drinkwater, S. (2002). Enclaves, neighbourhood effects and employment outcomes: Ethnic minorities in England and Wales. *Journal of Population Economics*, 15(1):5–29.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (2008). When are ghettos bad? lessons from immigrant segregation in the united states. *Journal of Urban Economics*, 63(3):759–774.

Damm, A. P. (2009). Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence. *Journal of Labor Economics*, 27(2):281–314.

Danzer, A. M. and Yaman, F. (2010). Ethnic concentration and language fluency of immigrants in germany. IZA Discussion Paper 4742, Institute for the Study of Labor (IZA).

de Graaf, N. D. and Flap, H. D. (1988). "With a little help from my friends": Social resources as an explanation of occupational status and income in West Germany, The Netherlands, and the United States. *Social Forces*, 67(2):452.

Dietz, B. (1999). Ethnic german immigration from eastern europe and the former soviet union to germany: the effects of migrant networks. IZA Discussion Paper 68, Institute for the Study of Labor (IZA).

Dustmann, C. (1994). Speaking fluency, writing fluency and earnings of migrants. *Journal of Population Economics*, 7(2):133–156.

Dustmann, C. and Van Soest, A. (2001). Language fluency and earnings: Estimation with misclassified language indicators. *The Review of Economics and Statistics*, 83(4):663–674.

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.

Gilbertson, G. A. (1995). Women's labor and enclave employment: The case of Dominican and Colombian women in New York City. *International Migration Review*, 29(3):657–670.

Goel, D. and Lang, K. (2009). The role of social ties in the job search of recent immigrants. CLSSRN working paper 5, Vancouver School of Economics.

Hahn, M. H. and Haisken-DeNew, J. P. (2013). PanelWhiz and the Australian longitudinal data infrastructure in economics. *Australian Economic Review*, 46(3):379–386.

Haisken-DeNew, J. P. and Frick, J. R. (2005). DTC - desktop companion to the german socio-economic panel (SOEP). Version 8.0.

Haisken-DeNew, J. P. and Hahn, M. H. (2010). PanelWhiz: Efficient data extraction of complex panel data sets – An example using the German SOEP. *Schmollers Jahrbuch*, 130(4):643–654.

Hou, F. P. and Picot, G. (2003). Visible minority neighbourhood enclaves and labour market outcomes of immigrants. Analytical Studies Branch Research Paper Series 204, Statistics Canada, Analytical Studies Branch.

Ihlanfeldt, K. R. and Sjoquist, D. L. (1989). The impact of job decentralization on the economic welfare of central city blacks. *Journal of Urban Economics*, 26(1):110–130.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *Quarterly Journal of Economics*, 82(2):175–197.

Kalter, F. (2006). Auf der Suche nach einer Erklärung für die spezifischen Arbeitsmarktnachteile von Jugendlichen türkischer Herkunft. *Zeitschrift für Soziologie*, 35(2):144–160.

Kanas, A., Chiswick, B. R., van der Lippe, T., and van Tubergen, F. (2011). Social contacts and the economic performance of immigrants: A panel study of immigrants in germany. IZA Discussion Paper 5775, Institute for the Study of Labor (IZA).

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

Lancee, B. (2010). The economic returns of immigrants' bonding and bridging social capital: The case of the Netherlands. *International Migration Review*, 44(1):202–226.

Lazear, E. P. (1999). Culture and language. *Journal of Political Economy*, 107(S6).

Le, A. T. (2000). The determinants of immigrant self-employment in Australia. *International Migration Review*, 34(1):183–214.

Lehmer, F. and Ludsteck, J. (2014). Wage assimilation of foreigners: Which factors close the gap? Evidence from Germany. *Review of Income and Wealth*.

Logan, J. R., Alba, R. D., and Stults, B. J. (2003). Enclaves and entrepreneurs: Assessing the payoff for immigrants and minorities. *International Migration Review*, 37(2):344–388.

McManus, W. S. (1990). Labor market effects of language enclaves: Hispanic men in the United States. *Journal of Human Resources*, 25(2):228–252.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. Columbia University Press, New York, London.

Pedace, R. and Rohn, S. (2008). A warm embrace or the cold shoulder: Wage and employment outcomes in ethnic enclaves. Working Paper 08-09, Center for Economic Studies, U.S. Census Bureau.

Portes, A. (1987). The social origins of the Cuban enclave economy of Miami. *Sociological Perspectives*, 30(4):340–372.

Portes, A. and Jensen, L. (1989). The enclave and the entrants: Patterns of ethnic enterprise in Miami before and after Mariel. *American Sociological Review*, 54(6):929.

Portes, A. and Shafer, S. (2007). Revisiting the enclave hypothesis: Miami twenty-five years later. *Research in the Sociology of Organizations*, 25:157–90.

Portes, A. and Zhou, M. (1996). Self-employment and the earnings of immigrants. *American Sociological Review*, 61(2):219–230.

Razin, E. and Langlois, A. (1996). Metropolitan characteristics and entrepreneurship among immigrants and ethnic groups in Canada. *International Migration Review*, 30(3):703–727.

Sanders, J., Nee, V., and Sernau, S. (2002). Asian immigrants' reliance on social ties in a multiethnic labor market. *Social Forces*, 81(1):281–314.

Schupp, J. (2009). 25 Jahre Sozio-oekonomisches Panel – Ein Infrastrukturprojekt der empirischen Sozial- und Wirtschaftsforschung in Deutschland. *Zeitschrift für Soziologie*, 38(5):350–357.

SOEP (2013). Socio-economic panel (SOEP): Data for years 1984-2012. DOI: 10.5684/soep.v29.1.

Sousa, L. D. (2013). Human capital traps? enclave effects using linked employer-household data. Working Paper 13-29, Center for Economic Studies, U.S. Census Bureau.

Statistisches Bundesamt (2013). Verdienststrukturerhebung 2010. Fachserie 16.

Statistisches Bundesamt (2014a). Ausländische Bevölkerung. Fachserie 1.2.

Statistisches Bundesamt (2014b). Registrierte Arbeitslose - Arbeitslosenquote nach Gebietsstand. URL: `https://www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/Arbeitsmarkt/lrarb003.html`.

Tu, J. (2010). The effect of enclave residence on the labour force activities of immigrants in canada. IZA Discussion Paper 4744, Institute for the Study of Labor (IZA).

van Ham, M., Manley, D., Bailey, N., Simpson, L., and Maclennan, D. (2011). *Neighbourhood Effects Research: New Perspectives*. Springer, Dordrecht; New York, auflage: 2012 edition.

Waldfogel, J. (2003). Preference externalities: An empirical study of who benefits whom in differentiated-product markets. *RAND Journal of Economics*, 34(3):557–568.

Warman, C. (2007). Ethnic enclaves and immigrant earnings growth. *Canadian Journal of Economics*, 40(2):401–422.

Wilson, K. L. and Portes, A. (1980). Immigrant enclaves: An analysis of the labor market experiences of Cubans in Miami. *American Journal of Sociology*, 86(2):295–319.

Xie, Y. and Gough, M. (2011). Ethnic enclaves and the earnings of immigrants. *Demography*, 48(4):1293–1315.

Zenou, Y. (2013). Spatial versus social mismatch. *Journal of Urban Economics*, 74:113–132.

Zhou, M. and Logan, J. R. (1989). Returns on human capital in ethic enclaves: New York City's Chinatown. *American Sociological Review*, 54(5):809–820.

A Appendix

Table A1: List of Countries in Sample

Ethno: Italy	
	Italy
Ethno: Turkey	
	Turkey
Ethno: Greece	
	Greece
Ethno: Spain & Portugal & Latin America	
	Argentina, Bolivia, Brazil, Chile, Columbia, Cuba, Dominican Republic, El Salvador, Mexico, Nicaragua, Paraguay, Peru, Portugal, Spain, Venezuela
Ethno: The Balkans	
	Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Ex-Yugoslavia, Hungary, Kosovo-Albania, Macedonia, Moldavia, Romania, Serbia, Slovenia
Ethno: Russia	
	Armenia, Azerbaijan, Belarus, Czech Republic, Eastern Europe, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Lithuania, Poland, Russia, Slovakia, Tajikistan, Ukraine, Uzbekistan
Ethno: Islam	
	Afghanistan, Algeria, Chad, Egypt, Eritrea, Iran, Iraq, Israel, Jordan, Lebanon, Morocco, Pakistan, Palestine, Syria, Tunisia
Ethno: Asia	
	Bangladesh, Cambodia, China, India, Japan, Korea, Philippines, Singapore, Taiwan, Thailand, Vietnam
Ethno: Western & Other	
	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Great Britain, Ireland, Luxembourg, The Netherlands, Sweden, Switzerland, USA
Ethno: Aussiedler	
	Albania, Belarus, Croatia, Czech Republic, Estonia, Ex-Yugoslavia, Great Britain, Hungary, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldavia, Poland, Romania, Russia, Slovakia, Tajikistan, Turkey, Turkmenistan, Ukraine, Uzbekistan

Table A2: Variables Description

Ethnic Share	Share of foreigners with same ethnic background in percent (source: microm)
Ethno Dummies	Set of dummies for ethnic groups defined in table A1
Female	Dummy = 1 if female
Age	Age in years
Disabled	Dummy = 1 if disabled
Married	Dummy = 1 if married
Children in HH	Dummy = 1 if children live in household
Single parent	Dummy = 1 if person is single parent
YSM	years since migration
Education:	
Low	Dummy = 1 if education level is equivalent to International Standard Classification of Education (ISCED) 1 or 2
Medium	Dummy = 1 if education level is equivalent to ISCED 3 or 4
High	Dummy = 1 if education level is equivalent to ISCED 5 or 6
Employed	Dummy = 1 if employed
Gross wage p/m	Gross monthly wage in Euros
ln(wage)	Log of gross monthly wage
Hours p/w	Hours worked per week
Tenure	Tenure with a firm in years
SME	Dummy = 1 if person is working in a small or medium enterprise
Partner's net wage p/m	Partner's net monthly wage in Euros
Job Type:	
High-skilled white collar	Dummy = 1 if International Standard Classification of Occupations (ISCO) is 1, 2 or 3
Low-skilled white collar	Dummy = 1 if ISCO is 4 or 5
High-skilled blue collar	Dummy = 1 if ISCO is 6 or 7
Low-skilled blue collar	Dummy = 1 if ISCO is 8 or 9
Unemployment LLMA	Unemployment rate in LLMA (source: INKAR)
Mean wage LLMA	Mean gross hourly wage in LLMA
West	Dummy = 1 if person is living in Western Germany
Oral German	Self-reported oral German proficiency on a 1-5 scale
Written German	Self-reported written German proficiency on a 1-5 scale

The data source for all variables is the SOEP unless otherwise indicated.

Table A3: Summary Statistics for Natives

Variable	**Mean**	**SD**	**Min**	**Max**	**N**
Personal:					
Female	0.500	0.500	0	1	65045
Age	44.320	11.734	17	64	65045
Disabled	0.099	0.299	0	1	65045
Married	0.533	0.499	0	1	65045
Children in HH	0.423	0.785	0	8	65045
Single parent	0.026	0.160	0	1	65045
YSM	0.000	0.000	0	0	65045
Education:					
Low	0.105	0.307	0	1	65045
Medium	0.598	0.490	0	1	65045
High	0.297	0.457	0	1	65045
Employment:					
Employed	0.799	0.401	0	1	65045
Gross wage p/m	2536.226	1536.223	100	9968	46341
ln(wage)	7.612	0.763	5	9	46341
Hours p/w	37.848	11.144	1	50	46341
Tenure	11.621	10.469	0	50	46341
SME	0.541	0.498	0	1	46341
Partner's net wage p/m	735.311	1218.910	0	36000	65045
Region:					
Unemployment LLMA	7.891	3.315	1	20	65045
Mean wage LLMA	17.724	2.758	10	29	46341
West	0.764	0.425	0	1	65045
German:					
Oral German	4.931	0.254	4	5	176
Written German	4.868	0.401	2	5	176

Observations have been weighted using population weights. Data source: SOEP.

Table A4: Results of the Flat Price Regression

	(1)		(2)	
	OLS		IV	
Age	-0.030***	(0.000)	-0.026***	(0.001)
Age Squared	0.407***	(0.003)	0.303***	(0.022)
Age Cubed	-0.002***	(0.000)	-0.001***	(0.000)
ln(Area)	1.122***	(0.001)	1.134***	(0.009)
State – like new as base				
Renovated	0.056***	(0.002)	0.037***	(0.014)
Modernised	-0.058***	(0.002)	-0.030***	(0.011)
Not Renovated	-0.082***	(0.002)	-0.056***	(0.013)
Fitted kitchen	0.109***	(0.001)	0.067***	(0.009)
Balcony	0.106***	(0.001)	0.089***	(0.008)
Elevator	0.058***	(0.001)	-0.007	(0.016)
Cellar	-0.041***	(0.001)	-0.036***	(0.006)
Unemployment LLMA	-0.031***	(0.000)	-0.008	(0.013)
Mean Wage LLMA	0.030***	(0.000)	0.047***	(0.006)
Share Africa	-0.207***	(0.007)	1.267	(0.778)
Share Asia	0.192***	(0.006)	1.643***	(0.471)
Share Balkans	0.078***	(0.001)	0.275***	(0.082)
Share Greece	0.081***	(0.003)	-0.188	(0.308)
Share Islam	-0.138***	(0.003)	-0.427	(0.327)
Share Italy	-0.027***	(0.001)	-0.242**	(0.103)
Share Russia	0.135***	(0.002)	0.119	(0.133)
Share Spain	-0.061***	(0.003)	-0.915***	(0.299)
Share Aussiedler	-0.560***	(0.004)	0.242	(0.379)
Share Turkey	-0.017***	(0.000)	-0.066***	(0.020)
Share Western	0.365***	(0.002)	0.673***	(0.111)
Constant	6.425***	(0.012)	5.653***	(0.221)
Year dummies	Yes		Yes	
R^2	0.808		0.620	
N	509,388		509,388	

Standard errors in parentheses and clustered on cell level for the IV estimation. IV estimation by 2SLS. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Regression of German Skills

	(1) OLS		(2) IV		(3) OLS	
Ethnic share	-0.081***	(0.023)	-0.182***	(0.033)	-0.016	(0.027)
Age	-0.081***	(0.020)	-0.073***	(0.022)	-0.053***	(0.019)
Age sqaured	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
YSM	0.049***	(0.005)	0.053***	(0.005)	0.054***	(0.005)
Female	0.222***	(0.070)	0.191***	(0.070)	0.160**	(0.062)
Disabled	-0.110	(0.083)	-0.106	(0.083)	-0.079	(0.082)
Married	-0.314***	(0.103)	-0.278***	(0.107)	-0.248***	(0.091)
Medium	0.661***	(0.084)	0.602***	(0.091)	0.470***	(0.070)
High	1.187***	(0.115)	1.098***	(0.117)	0.887***	(0.100)
Children in HH	0.030	(0.039)	0.060	(0.037)	0.028	(0.033)
Single parent	-0.099	(0.166)	-0.100	(0.167)	0.015	(0.149)
West	0.054	(0.139)	0.140	(0.137)	0.022	(0.149)
Turkish as base						
Italy × ethnic share					-0.145	(0.147)
Greece × ethnic share					0.075	(0.122)
Spain × ethnic share					-0.923***	(0.189)
Balkans × ethnic share					0.039	(0.052)
Russia × ethnic share					-0.115	(0.100)
Middle East × ethnic share					0.331	(0.238)
Asia × ethnic share					5.155	(4.562)
Western × ethnic share					0.105	(0.154)
Aussiedler × ethnic share					0.040	(0.145)
Italy					0.212	(0.205)
Greece					0.696**	(0.333)
Spain					0.699***	(0.221)
Balkans					0.457***	(0.158)
Russia					0.725***	(0.214)
Middle East					0.332	(0.256)
Asia					-0.236	(0.553)
Western					0.915***	(0.262)
Aussiedler					0.787***	(0.127)
Constant	4.683***	(0.421)	4.568***	(0.446)	3.660***	(0.440)
Year dummies	Yes		Yes		Yes	
R^2	0.339		0.308		0.407	
N	4310		4310		4310	