Regional distribution and location choices of immigrants in Germany

Kerstin Tanis

IAB – Institute of Employment Research

FAU – Friedrich-Alexander University of Erlangen-Nuremberg

- Working Paper, June 2017 -

Abstract

This paper investigates initial and subsequent location choices of European labor immigrants in Germany. Previously, it was found that immigrants settle solely in metropolitan areas whereas peripheral areas, those suffering mostly from demographic changes, are chosen very scarcely. Thus, immigration seems to accelerate spatial inequality and other regional discrepancies. While previous studies measured individual and regional effects on immigrants' location choices separately, this study brings both effects together. Thus, this study produces a more precise understanding of location choice drivers. The use of individual immigrant data from the IAB-Integrated Employment Biographies as well as regional data from the Regional Database Germany allows us to measure possible interactions effects at a low regional level (counties). In the empirical section we provide evidence of the determinants for initial and following location choices based on discrete choice models, focusing on the top European sending countries.

Introduction

Since Germany is facing a demographic change, resulting in decreasing labor supply, immigrants have become (once again) subject to national debates. Considering immigrants as potential labor source, the role of immigrants continues to grow economically and demographically. However, not all regions are affected by the same size of immigrant inflows. While a huge body on literature investigates the impact of immigration on the national economy (e.g. Borjas, 2006; Brücker & Jahn, 2011; Dustmann, Frattini, & Preston, 2013), very little research was conducted on the regional distribution on immigrants within this host country. This is surprising for several reasons (Scott, Coomes, & Izyumov, 2005). On the one hand, economic development officials at the local level benefit from the knowledge of location choice determinants when they try to stem or stimulate inflows and regional growth. On the other hand, the Federal Government learns about the regional-diverse impacts of its national immigration policies. The aim of this paper is to close this knowledge gap by investigating regional and individual determinants of immigrants during their time of residence in Germany.

Previous research findings into immigrants' location choices conducted in Northern America (Bartel, 1989; Chiswick & Miller, 2004; Kaushal, 2005; Massey & Denton, 1987; Newbold, 1999; Scott et al., 2005) and Europe (Åslund, 2005; Damm, 2009; Jayet, Rayp, Ruyssen, & Ukrayinchuk, 2016) have been consistently shown that immigrants solely settle in metropolitan areas with long migration histories. In general, previous studies explained immigrants' location patterns with two determinant groups. The main literature stream explains the unevenly distributed inflow of immigrants across regions within a host country by regional characteristics, focusing on amenities or pull factors. Especially agglomerations were found to provide certain migration-specific amenities. Therefore, the region's population size and the size of the existing immigrant stock are the most frequently mentioned regional determinants for immigrants' location choice. Concentration patterns of immigrants in the United States describe this dependency impressively: more than one half of all immigrants living in the United States concentrate on only six "gateway" cities (Camarota, 2001). Additionally, well-developed public transport or rapid, but often low-paid, job opportunities were found to attract immigrants to agglomeration centers. Apart from regional characteristics there are also some studies using socioeconomic/-demographic characteristics as explanation for location choices (Chiswick & Miller, 2004). For instance, research found that highly educated are less concentrated than lower educated within the host country (Jayet, Ukrayinchuk, & De Arcangelis, 2010). However, there are few studies investigating individual and regional location choice determinants jointly. One

exception is the study of Scott et al. (2005), who find that taking account of interactions between regional and individual characteristics has a great impact on the estimated coefficient in discrete choice models. Furthermore, they argue that from a policy perspective, analyzing regional and individual effects separately is not sufficient as one might expect that migrants select into regions where their specific characteristics are requested and migration outcomes are highest. Thus, individual and regional characteristics become jointly an explanation for location choices. Considering this important finding the present study will also focus on interaction effects.

In order to close the research gap on immigrants' location choices in Germany, this study uses the knowledge of former studies to generate an even more precise estimation. Apart from a joint investigation of individual and regional characteristics on immigrants' location studies, the underlying study considers also possible changes in the received importance of location characteristics concerning initial and subsequent location choices. Furthermore, by examining location choices on a low aggregation level, that is counties, this study is able to obtain a better and more accurate understanding of the importance of regional characteristics than by using higher and more heterogeneous aggregation levels, e.g. federal states. Finally, in a location choice framework, employment-based immigrants seem to be the most interesting group of immigrants. In contrast to other immigrant groups, they have a larger choice set and are more sensitive to regional characteristics (Bartel, 1989). Immigrants for family reasons are predetermined in their location choice by working family members and asylum seekers are regionally distributed by law in Germany. By combining data of the Federal Agency of Employment with regional data, a rich data set is created, which helps to overcome the outlined limitations of former studies.

In the next section, we will focus on the theoretical background including the standard discrete choice modeling approach and the empirical implementation. Afterwards, the required data sets are introduced in more detail. Then, preliminary results of the analysis are presented. Afterwards, the paper concludes and outlines possible limitations.

Theoretical framework

Utility maximization

According to the theory of utility maximization, a rational individual (immigrant i) will choose that location j out of a set of J locations which maximizes his/her utility. Following a "two-

stage" approach, utility is defined as a function of regional and individual characteristics as well as their interactions.

At the first stage, the utility level of a certain region is determined only by its regional characteristics. The utility function for immigrant *i* residing in region *j* is thus given by

$$U_{ij} = \sum_{j=1}^{J} \beta R_j + \varepsilon_{ij} \quad (1)$$

where U_{ij} is the degree of utility immigrant *i* generates by choosing to locate in region *j*. R_j is a vector of regional characteristics of region *j*, β a vector of estimated coefficients, and ε_{ij} is a randomly distributed error term.

At the second stage, the degree of achieved utility does not only vary between locations, but also between individuals. Therefore, individual characteristics should be considered as an additional determinant of a location's utility. In a discrete choice framework, this is done by interacting regional characteristics with characteristics of immigrant *i*

$$U_{ij} = \sum_{k=1}^{K} \beta R_j + \sum_{k=1}^{K} \sum_{l=1}^{L} \delta R_j X_i + \varepsilon_{ij} \quad (2)$$

where $R_j X_i$ defines an interaction term between the characteristics of a certain region *j* and the vector of individual characteristics *X* of immigrant *i*, δ states the corresponding vector of estimated coefficients.

For both specifications (1) und (2) immigrant *i* will choose region *j* over *c* if

$$U_{ij} > U_{ic},\ j \neq c$$

and for simplicity reasons, both utility functions (1) and (2) can be rewritten in

$$U_{ij} = Z_{ij} + \varepsilon_{ij}, with\ \varepsilon_{ij} \approx IDD. \quad (3)$$

Derivation of hypotheses

Regarding the regional determinants of utility, this analysis focuses on three variable groups: population characteristics, labor market characteristics, and housing market characteristics. Population characteristics include the variable of most interest. The immigrant share in a region is assumed to capture important network effects. It is well known from literature (e.g. Wang, De Graaff, & Nijkamp, 2016), that immigrants prefer regions where the immigrant density is high. Positive network externalities explain this preference, because choosing a region with a high share of immigrants in the population gives the newly arriving immigrants some security.

A region densely populated with immigrants does not only ensure cultural and linguistic familiarity, but also regional open-mindedness towards immigrants and the existence of migration-specific infrastructure like foreign administration agencies. Therefore, this effect is assumed to be very important for the initial location choice. However, it is assumed to lose its explanatory power for subsequent location choices within the host country. Living in immigrant enclaves for a long time hinders immigrants to integrate quickly into the host society, because of they lack important contacts to natives. The second variable regional group, which is assumed to affect the location choice of immigrants are labor market opportunities. Labor market conditions are a main determinant of deciding where to locate, because they signalize immigrants the achievable outcome regarding the decision to migrate in the most direct/monetary way. In opposite to network effects, it is hypnotized that regional labor market characteristics are an important determinant for the initial, but also for any subsequent location choices. Finally, the analysis will include housing market variables like the share of housing property in a region[1]. Compared to other countries, Germany has a very low share of people owning their own houses or flats. Therefore, immigrants are supposed to choose regions where they are able to acquire property.

The interaction effect of main interest is that between the network variables (immigrant share resp. ethnic concentration) and the educational attainment (low (0) vs. high (1) education) of the immigrant. Following the general argumentation that highly educated are less dependent on networks, because of facing less labor market and cultural barriers right after migrating, it is assumed that both network variables show a negative effect for immigrants with a higher educational attainment level compared to individuals with a lower education.

Estimation approach and choice set sampling

Empirically, the analysis is estimated using the standard framework for discrete choice frameworks. Mc Fadden's conditional logit model (McFadden, 1973) gives the probability the immigrant i chooses location j among all other locations using

$$P_{ij} = \frac{e^{Zij}}{\sum_{j=1}^{J} e^{ZiJ}}$$

The parameters (β, δ included in Z) are obtained by maximizing the log likelihood function

[1] For future analysis, regional rental charges will be included.

$$L* = \sum_{i=1}^{I} \sum_{j=1}^{J} D_{ij} log P_{ij} ,$$

where $D_{ij} = 1$ if immigrant i chooses region j and 0 otherwise (Scott et al., 2005). However, using a conditional model has two disadvantages.

Firstly, conditional logit models imply the strong assumption of the Independence of Irrelevant Alternatives (IIA). In general terms, this property states that all alternatives are assumed to be independent of each other regarding observable, but also unobservable characteristics. In other words, the decision maker chooses between unique alternatives, in this case locations. It was shown that the IIA property is less likely to be violated if characteristics of decision makers are included in the model specification (Ben-Akiva & Lerman, 1985). In a conditional logit framework, this is done by including interaction effects between individual and regional characteristics in the second-stage of the analysis.

Secondly, computational issue arise when the choice set is very large. Thus, previous studies used more aggregated levels. However, using more aggregated regions does not account for regional heterogeneity within the aggregation. The loss of estimation preciseness is a consequence. The IIA property of the conditional logit model can be exploit to use a random sample technique for the large choice set of 326 counties (Manski & Lerman, 1977). McFadden (1978) has shown that restricting the full choice set to a random subset leads to consistent parameter estimates. Therefore, equation (3) is estimated using the chosen county plus a random subset of potential not chosen counties. Following a recommendation of Nerella and Bhat (2004), who state to use at least a minimum of 1/8 of the full choice set, the final choice set consist of 40 counties (the chosen county plus 39 randomly assigned counties). In this paper, we assume that the IIA assumption holds, because location choice analyses including regional independent variables for single nationalities showed the same effect directions using the full and subset of possible choices. By the way, Scott et al. (2005) showed consistent results for an even smaller subset. They used a choice subset of n=10 for a full set of 298 MSA in the United States.

Data and Variables

The composition of immigrant inflows in Germany has dramatically changed over the last two decades due to current immigration regulations. The aim of this study is to reflect the most recent inflow, which started in 2011. Therefore, the initial location choice analysis is based on

immigrants who arrived in 2011, while the analysis on the second location choice is based on the same individuals four years later, which means 2014. Excluding forced migration, most immigrant nowadays come from the countries of the EU Enlargement since 2004. In this context, Germany is especially affected by immigration of people from: Hungary, Poland, Rumania, and Bulgaria. However, also immigration from Southern European countries recently increased due to the protracted consequences of the economic crisis in the Mediterranean area. The most important immigration countries from the South are: Greece, Italy, Spain, and Portugal. In order to gain precise insights into location choice patterns, separated analysis are conducted for those eight nationalities plus Austria. Compared with the first analysis which includes the whole immigrant group arrived in 2011, estimates using only European nationalities have the advantage that all immigrants face the same basis regarding immigration regulation. Furthermore, Austrians are an interesting immigrant group as they share the same language and culture with Germans.

As the analysis is based on both individual and regional characteristics, two data bases have to be merged. Information about individuals is detected from the IAB-Integrated Employment Biographies (IEB) provided by the German Federal Agency of Employment. Data is annually reported by the employers for employed individuals or by job centers for registered unemployed people (register data). Individuals appear in the data the first time they get employed or ask for social assistance. They leave the data when they move abroad or do not register themselves as unemployed, otherwise they remain in the data until they retire. This rich data set contains the necessary information about the nationality of an individual and the place of residence as well as other important socio-demographics like age, gender, and education. Unfortunately, it does not provide any information about the year of arrival or place of birth. Therefore, people who were born in Germany, but have a foreign nationality, are excluded from the analysis. This is done by only considering individuals aged 28[2] and older. Second-generation immigrants would have been appeared in the data before they turned 28 even if they had studied at universities. Furthermore, the data was restricted to immigrants living in the Western German counties and Berlin. Eastern counties are only very sparsely affected by immigration inflows and therefore not representative for the rest of Germany.

Table 1 reports the summary statistics of the individual data set. The data set includes 29,307 immigrants, who arrived in 2011. More than half of the newly arrived immigrants come from the new European countries. The average age is roughly 38 years, which shows the

[2] The age of 28 has been chosen, because this is a realistic age of having finished Bachelor and Master's studies.

objective fact, that immigrants are in general a young group in the most effective working age. More than 82 % chose to live first in an independent city or an urban county. Unfortunately, the characteristic "education" has a lot of missings, because it is voluntarily reported by the employers. This means, including the education variable, which is the individual variable of most interest, reduces the number of observations dramatically.

Table 1: Individual variables

Variable	Individuals	2011 (2014)
Male	29,307	0.545
Age	29,307	37.7
School education	8,025	
Low/Middle		0.519
High		0.481
Region of origin	29,307	
EU-15		0.156
EU-13		0.514
Former Yugoslavia		0.011
(Former) CIS		0.039
Turkey		0.047
Others		0.233
County type	29,307	
Independent city		0.433 (0.438)
Urban county		0.382 (0.381)
Rural county, concentration tendency		0.117 (0.114)
Rural county, densly populated		0.069 (0.064)

Source: IEB. Own estimation.

The second data base comes from the Statistic Departments of the Federation and the Federal States, which provide the required regional statistics on county level. It is assumed that newcomers in 2011 made their location choice based on regional information of the previous year (2010). Therefore, the regional statistics are time lagged, meaning regional data of 2010 is merged with individual data of 2011 and regional data of 2013 is merged with individual data of 2014 to investigate the second location choice. Again, the data was restricted to Western German counties and Berlin. This data will be divided in external regional effects (labor market relevant variables like unemployment rate or gross income per capita), which control for the economic environment, and network effects, which are linked to the regional foreign community of the same ethnicity (ethnic concentration) in the nationality-separated model or the whole migrant group (share of immigrants). The standard deviations as well as the minimum/maximum values reported in Table 2 suggest a high regional variation.

Table 2: Municipal variables, 2010

Variable	Counties	Mean	Std.Dev.	Min.	Max.
Population					
Immigrant share	326	0.084	0.042	0.027	0.258
Population size	326	211308	253370	33944	3460725
Population density	326	576	719	40	4355
Share of aged >65 years	326	0.204	0.02	0.15	0.267
Labor Market					
UR total	326	0.063	0.025	0.019	0.163
UR immigrant	326	0.143	0.054	0.035	0.298
Gross income	326	27445	3599	19980	42252
Share of production industry	326	0.233	0.096	0.069	0.534
Housing					
Resident-owned home rate	326	0.495	0.121	0.15	0.698

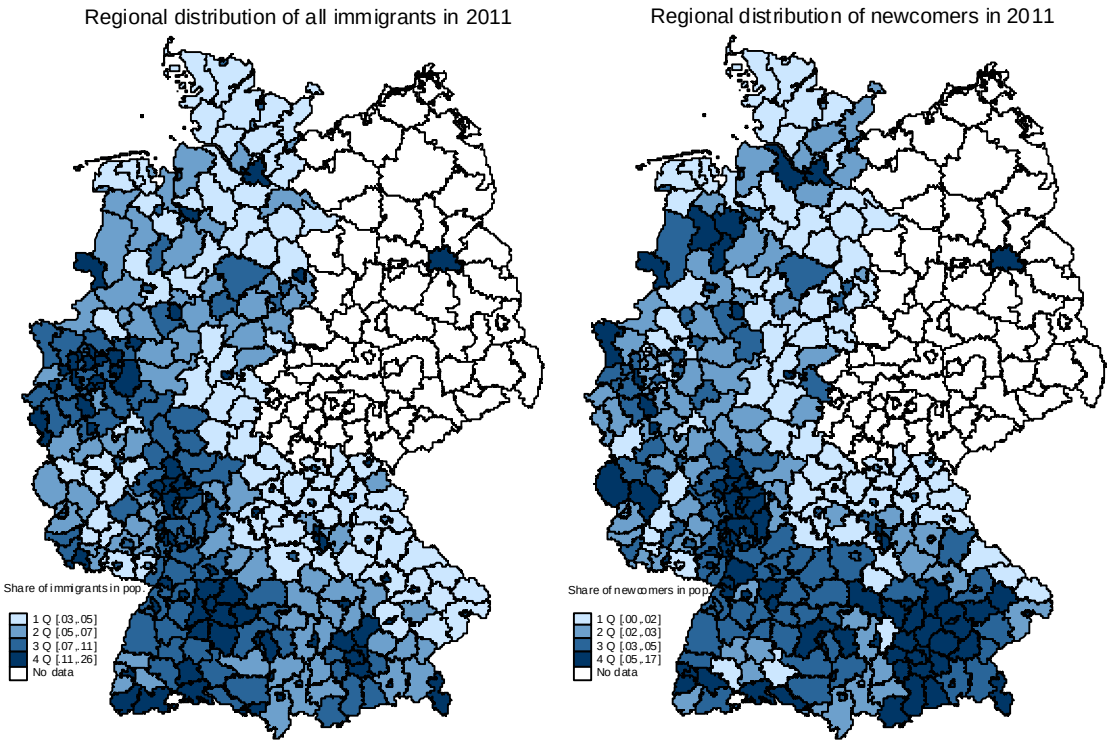
Source: Regional Database Germany. Own estimation.

Results

Regional distribution of newly arrived immigrants in 2011

The following two maps compare the general share of all immigrants with the share of newly arrived immigrants in 2011. The darker the color, the higher the share of immigrants resp. newcomers in the county. In general, there are several migration centers obvious: Munich, Stuttgart, and Frankfurt; furthermore, the city states Berlin and Hamburg are strongly affected by immigration. Counties close to the border of former Eastern Germany are less concerned with immigration inflows, the most Northern counties in Schleswig-Holstein either. Another interesting point is that immigrants tend to be more concentrated in county free cities (especially obvious Northern Bavaria). However, comparing shares of the total immigrant group to the share of newcomers shows that location preferences might have changed. Newcomers seem to be even more concentrated around Munich. North Rhine Westphalia as traditional immigration center for guest workers in the 1980s seems to lose attractiveness. This can be explained by switching economic conditions and the closeness of Munich to the countries of origin like Hungary, Bulgaria, or Rumania. In conclusion, the descriptive analysis may be a first indication for the importance of networks, but also economic conditions.

Figure 1: Regional distribution of immigrants: total immigrant group vs. newcomers



Source: IEB and Regional Database Germany. Own illustration.

Initial location choice determinants

Table 3 provides the coefficients obtained from the conditional logit model regarding the first location choice of newcomers in 2011. Columns one to three report the pure network effect. All three model specifications show a positive effect of a region's immigrant share on the probability to settle in this location. What is striking (model specification three) is that the interaction effect between high educated and the share of immigrants is positive, because one might assume that highly educated people do less rely on networks than low educated people and therefore choose regions where the immigration share is low.

Columns four and five confirm the descriptive picture from Figure 1. Immigrants prefer to settle in cities, because a higher population density increases the probability of choosing a region. A higher share of elderly has a negative impact on the location choice probability of immigrants. So does a higher unemployment rate. This is an interesting result, because immigrants tend to choose regions with a strong economy. This statement is confirmed by a positive impact of wage levels and a lower share of producing industry. Surprisingly, while the total unemployment rate of a region has the expected negative impact, immigrants' unemployment rate has a positive impact. This means, immigrants do not choose regions where the unemployment for their group is low, but where it is high. This can be explained by the fact

that a low unemployment rate leads to more competition and therefore, a higher unemployment rate for vulnerable groups like immigrants. Furthermore, it is known that immigrants from Southern Europe prefer traditionally housing property. That might explain why immigrants have a higher probability to choose a region where the housing ownership rate is higher.

Table 3: Initial location choice determinants

	(1)	(2)	(3)	(4)	(5)
Immigrant share	14.057***	12.477***	9.981***	9.477***	5.475***
Ln(Population desity)				0.269***	0.268***
Share of aged >65				-10.004***	-9.620***
UR total				-6.899***	-13.703***
UR foreign				0.345	2.427***
Ln(Gross income)				2.056***	2.308***
Share of producing industry				-2.407***	-3.057***
Housing ownership rate				3.066***	2.304***
High educ.*Im.density			6.895***		6.935***
Regional fixed effects	no	yes	yes	yes	yes
Observations	1172280	1172280	321000	1172280	321000
Prob. > Chi²	0.000	0.000	0.000	0.000	0.000
Pseudo R²	0.071	0.127	0.136	0.141	0.154

Source: IEB and Regional Database Germany. Own estimation.

In contrast to Table 3, Table 4 (see Appendix) reports the coefficients for certain nationality separately. This estimation method allows to control for an additional network variable, the share of immigrants with the same ethnicity (ethnic concentration). Including this variable into the analysis has some important impacts on the former network variable, the total immigrant share, and the interaction effects regarding education. Columns 1-4 report results for the main immigrant groups coming from the Eastern European enlargement: Bulgaria, Hungary, Poland, and Rumania. Comparing the estimated coefficients shows that there are indeed differences in location choice preferences between nations. For Hungarian and Polish immigrants, the total immigrant share remains an important determinant after including a measure of ethnic concentration – at least in the model specification without regional fixed effects. However, the effect of the immigrant share of Hungarian is positive, while it is negative for Polish. Further negative effects of the immigrant share were found for Spanish and Italian immigrants. Surprisingly, also immigrants from Austria rely on ethnic networks, even though they do not face any linguistic barriers, for example. But, there location choice is not determined by a high share of other immigrants in the region. According to these results, one might argue

that the ethnic concentration variable captures the real network effect (esp. language advantages), while the immigrant share variable captures migration-specific infrastructure effects.

Regarding local labor markets, only some nations have a higher probability to choose a location according to its unemployment rate (esp. immigrants from Greece, Italy, Portugal, Poland, and Hungary), while others (Bulgaria and Rumania) show no significant effects. Spanish even have a higher probability to choose regions with higher unemployment rates. A more homogenous result is that all nationalities have a higher probability to choose locations with higher income levels and a smaller share of employees in the producing industry.

The analysis provides interesting results in terms of the included interaction effects. The interaction effect between highly educated and the immigrant share remains positive, but the interaction effect between highly educated and the ethnic concentration in the region is negative. This can be explained by the fact that the former variable may capture some agglomeration advantages in addition to network externalities, highly educated have to rely on. The ethnic concentration measure is finally assumed to measure the real network effect, because the most useful externalities are given by language advantages. Only immigrants from the same country of origin (language) can offer this externality. Therefore, highly educated are less likely to choose regions where these network externalities are given, because they might deal better with unfamiliar situations and assimilate more quickly.

[To be continued]

Internal migration patterns of immigrants

Table 5 shows the estimated coefficients of a logit model. The dependent variable is a dummy, which becomes one when the immigrant is still observed at his/her initial place of residence. The model includes regional and individual characteristics. However, the explanatory power is very small[3]. Nevertheless, the model includes very interesting information, which is in line with the previous undertaken assumptions in the theoretical section of this paper. The share of immigrants in a region has a positive effect on the probability to relocate. To put it cautiously, this means that immigrants only rely on a high immigrant share for their initial location choice. After a certain stay in the host country, a high immigrant share in a region is a driver for relocation. In summary, the hypothesis that immigrants might realize after a certain time of

[3]Notification: The model fit will be improved by adding additional explanatory variables in future analysis.

residence that they can generate a greater outcome when they live in regions with a lower share of immigrants, is verified to some extent.

Table 4: Relocation probabilities

	Coef.	Std. Err.
Immigrant share	3.056**	1.32
Ln(Population density)	-0.089	0.08
Share of aged >65	3.633*	2.03
UR total	-6.764	4.48
UR foreign	1.492	1.93
Ln(Gross income)	0.286	0.39
Share of producing industry	-0.695	0.52
Housing ownership rate	0.991	0.66
Education	0.103	0.07
Male	0.347***	0.06
Age	-0.005	0.00
Region of origin (Ref. European country before 2004)		
European country after 2004	0.220***	0.09
Former Yugoslavia	-0.264	0.37
(Former) CIS	-0.131	0.18
Turkey	-0.931***	0.28
Others	0.017	0.09
Constant	-5.214	3.93
Observations	-5.214	
Prob. > Chi²	8025	
Pseudo R²	0.017	

Source: IEB and Regional Database Germany. Own estimation.

[To be continued: Determinants of the subsequent location choice in 2014]

Discussion

Research, which investigates immigrants' location choices, has to deal with some problematic endogeneity issues. Endogeneity arises, because the existing migrant stock is supposed to be an endogenous amenity: a concentration of migrants in a region may be the result, as well as the cause, for migration location choices. In order to estimate unbiased coefficients, the main sources of endogeneity have to be eliminated. If the behavior of the reference population, that is the existing immigrants stock in the region, is influenced by the behavior of the sample population, endogeneity appears. However, this influence can be neglected in this case as all regional variables and thus also the share of immigrants in the regions are included with a time lag. The more critical endogeneity issue is given when reference and sample population are

influenced by the same unobservable characteristics. If this influence of omitted variables is supposed to be permanent over time, it is included in the regional fixed effects and might not appear in the error term. Comparing the estimated coefficients between models with and without regional fixed effects leads to the result that also the second source of endogeneity may not dramatically bias the analysis, because they show in the same direction.

Conclusion

The aim of this paper was to investigate the determinants of immigrants' location choices during their stay in Germany. Therefore, this study provides policy remarks as well as information on location determinants, which may accelerate or hinder successful integration. Residential integration "is believed to be a crucial step in the assimilation process for immigrants" (Vang, 2010, p. 2984). Immigrants rely heavily on the existing immigrant in a region when it comes to decide where to locate for the first time. However, the importance of other immigrants seems to decrease during their residence in Germany, while good economic conditions are a time-constant driver. In sum, all included variables have a significant influence on the probability to choose a certain region. This means that even regions with a low share of immigrants may attract newcomers by advertise other important determinants like good economic conditions or focusing on internal movers. This is an important result, because it is known from research on attitudes towards ethnic minorities that a more equal distribution, that is a balanced ratio between natives and immigrants, leads to a broader acceptance of immigrants in the host population (Dustmann & Preston, 2000) and greater outcomes for both populations.

[To be continued]

References

Åslund, O. (2005). Now and forever? Initial and subsequent location choices of immigrants. *Regional Science and Urban Economics, 35*(2), 141-165.

Bartel, A. P. (1989). Where do the new U.S. immigrants live? *Journal of Labor Economics, 7*(4), S. 371-391.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9): MIT press.

Borjas, G. J. (2006). Native internal migration and labor market impact of immigration. *The Journal of Human Resources, 41*(2), S. 221-258.

Brücker, H., & Jahn, E. J. (2011). Migration and wage-setting : reassessing the labor market effects of migration. *The Scandinavian journal of economics, 113*(2), 32 S. doi: 10.1111/j.1467-9442.2010.01634.x

Camarota, S. (2001). Immigrants in the United States, 2000: A snapshot of America's foreign-born population. January 2001. Retrieved July 2, 2010.

Chiswick, B. R., & Miller, P. W. (2004). Where Immigrants Settle in the United States. *IZA Discussion paper series, 1231*.

Damm, A. P. (2009). Determinants of recent immigrants' location choices : quasi-experimental evidence. *Journal of Population Economics, 22*(1), S. 145-174. doi: 10.1007/s00148-007-0148-5

Dustmann, C., Frattini, T., & Preston, I. P. (2013). The effect of immigration along the distribution of wages. *The Review of Economic Studies, 80*(1), S. 145-173. doi: 10.1093/restud/rds019

Dustmann, C., & Preston, I. (2000). *Racial and economic factors in attitudes to immigration* (Vol. 2542): Centre for Economic Policy Research.

Jayet, H., Rayp, G., Ruyssen, I., & Ukrayinchuk, N. (2016). Immigrants' location choice in Belgium. *The Annals of Regional Science, 57*(1), 63-89. doi: 10.1007/s00168-016-0761-x

Jayet, H., Ukrayinchuk, N., & De Arcangelis, G. (2010). The Location of Immigrants in Italy: Disentangling Networks and Local Effects. *Annals of Economics and Statistics*(97/98), 329-350.

Kaushal, N. (2005). New immigrants' location choices : magnets without welfare. *Journal of Labor Economics, 23*(1), S. 59-80.

Manski, C. F., & Lerman, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica: Journal of the Econometric Society*, 1977-1988.

Massey, D. S., & Denton, N. A. (1987). Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980. *American Sociological Review, 52*(6), 802-825.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*(673).

Nerella, S., & Bhat, C. (2004). Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transportation Research Record: Journal of the Transportation Research Board*(1894), 11-19.

Newbold, K. B. (1999). Spatial Distribution and Redistribution of Immigrants in the Metropolitan United States, 1980 and 1990. *Economic Geography, 75*(3), 254-271. doi: 10.2307/144577

Scott, D. M., Coomes, P. A., & Izyumov, A. I. (2005). The location choice of employment-based immigrants among U.S. metro areas. *Journal of Regional Science, 45*(1), S. 113-145.

Vang, Z. M. (2010). Housing supply and residential segregation in Ireland. *Urban Studies*.

Wang, Z., De Graaff, T., & Nijkamp, P. (2016). Cultural diversity and cultural distance as choice determinants of migration destination. *Spatial Economic Analysis, 11*(2), 176-200.

Appendix

Table 5: Initial location choice determinants separated by nationalities

	Bulgaria		Hungary		Poland		Rumania	
Ethnic concentration	111.426	311.993***	322.776***	264.326***	117.997***	97.669***	143.738*	234.139***
Immigrant share	4.871	0.944	5.715**	-0.398	-2.764**	-0.102	4.47	2.439
Ln(Population density)	0.536**	0.168	0.232*	0.429***	0.232***	0.313***	0.869***	0.126
Share of aged >65	-13.382***	-5.794**	-4.813	-2.019	-14.141***	-8.233***	-2.248	-14.685***
UR total	-10.943	-8.573	-26.122***	-14.89	-3.101	-8.181**	-9.743	-10.109
UR foreign	6.225	-0.615	4.243	0.862	4.383***	-0.23	-3.686	3.399
Ln(Gross income)	3.351***	1.878***	1.470***	1.204**	1.427***	1.260***	2.908***	1.141**
Share of producing industry	-5.007***	-0.89	-1.684**	-1.946**	-2.715***	-2.430***	-4.157***	-0.329
Housing ownership rate	3.794**	2.789**	1.327	1.702	2.653***	3.080***	5.262**	1.764*
Highly educ.*Im.dens.	3.98	7.381***	7.609***	7.611***	6.929***	6.784***	3.237	7.694***
Highly educ.*Eth.con	-233.634*	-182.874***	-186.921**	-180.695*	-48.533***	-41.748**	-225.347*	-190.281***
Regional fixed effects	no	yes	no	yes	no	yes	yes	no
Observations	7974	21508	16042	16042	67753	67753	7974	21508
Prob. > Chi²	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R²	0.189	0.101	0.115	0.151	0.061	0.088	0.250	0.070

Table 5: Initial location choice determinants separated by nationalities (continued)

	Greece		Italy		Portugal		Spain	
Ethnic concentration	91.262***	87.255***	80.432***	54.751***	275.718***	298.815***	162.464	418.476***
Immigrant share	3.339	9.816**	-9.936***	1.461	-0.936	6.986	-12.800***	-10.512*
Ln(Population density)	0.310*	-0.032	0.484**	0.516**	1.091***	1.073***	0.358*	0.249
Share of aged >65	-15.449***	-16.142***	-24.688***	-12.007**	-24.142***	-9.827	-34.462***	-20.157***
UR total	-29.440***	-45.652***	0.177	-39.450***	-24.029*	-34.159*	14.679**	-18.582
UR foreign	16.555***	24.069***	0.731	8.041	7.565	0.874	-0.658	6.032
Ln(Gross income)	2.435***	2.239**	2.296***	2.703**	0.173	-0.139	3.904***	2.898***
Share of producing industry	-1.499	-2.036	-1.816	-2.092	0.514	0.509	-3.676***	-3.498**
Housing ownership rate	-0.934	-1.103	0.035	1.942	2.679	4.82	-2.770*	-2.259
Highly educ.*Im.dens.	7.486**	7.840**	26.522***	19.727***	10.784***	8.188**	12.105***	12.502***
Highly educ.*Eth.con	-43.553	-42.451	-163.850***	-105.167***	-421.396***	-340.785***	-232.655	-196.543
Regional fixed effects	no	yes	no	yes	no	yes	no	yes
Observations	9213	9213	10651	10651	3912	3912	10631	10631
Prob. > Chi²	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R^2	0.244	0.306	0.274	0.379	0.178	0.288	0.323	0.409

Table 5: Initial location choice determinants separated by nationalities (continued)

	Austria	
Ethnic concentration	58.367***	36.756***
Immigrant share	-6.004	-5.372
Ln(Population density)	0.249	0.292
Share of aged >65	-24.186***	-11.344*
UR total	-6.441	-17.416
UR foreign	3.759	6.957
Ln(Gross income)	1.803**	2.355***
Share of producing industry	-1.342	-2.612*
Housing ownership rate	-2.718*	-1.665
Highly educ.*Im.dens.	8.930**	8.817*
Highly educ.*Eth.con	6.582	4.486
Regional fixed effects	no	yes
Observations	7884	7884
Prob. > Chi²	0.000	0.000
Pseudo R²	0.210	0.272

Source: IEB and Regional Database Germany. Own estimation.