Meister, Moritz; Niebuhr, Annekatrin; Peters, Jan Cornelius; Stiller, Johannes

Local attributes and migration balance – evidence for different age and skill groups from a machine learning approach

Regional Science Policy & Practice

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Meister, Moritz; Niebuhr, Annekatrin; Peters, Jan Cornelius; Stiller, Johannes (2023) : Local attributes and migration balance – evidence for different age and skill groups from a machine learning approach, Regional Science Policy & Practice, ISSN 1757-7802, Wiley, Hoboken, NJ, Vol. 15, Iss. 4, pp. 794-825, https://doi.org/10.1111/rsp3.12652

Received: 9 June 2022 | Revised: 2 January 2023 | Accepted: 8 February 2023

DOI: 10.1111/rsp3.12652

ORIGINAL ARTICLE

Local attributes and migration balance – evidence for different age and skill groups from a machine learning approach

Moritz Meister[1] | **Annekatrin Niebuhr**[1,2] | **Jan Cornelius Peters**[3] | **Johannes Stiller**[3,4]

[1]Institute for Employment Research, IAB Nord, Projensdorfer Str. 82, 24106, Kiel, Germany

[2]Empirical Labour Economics and Spatial Econometrics, Department of Economics, Christian-Albrechts-Universität zu Kiel, Olshausenstr. 40, 24098, Kiel, Germany

[3]Institute of Rural Economics, Institute of Rural Studies, Johann Heinrich von Thünen Institute, Bundesallee 64, 38116, Braunschweig, Germany

[4]Commission of Experts for Research and Innovation (EFI), Pariser Platz 6, 10117, Berlin, Germany

Correspondence
Annekatrin Niebuhr, Institute for Employment Reasearch, IAB Nord, Projensdorfer Str. 82, 24106, Kiel, Germany.
Email: annekatrin.niebuhr@iab.de

Funding information
German Federal Ministry of Food and Agriculture (Federal Rural Development Scheme), Grant/Award Numbers: 2817LE054, 2817LE055

Abstract

Many European regions are currently experiencing a significant population decline and, related to this, are increasingly confronted with labour shortage. Migration is a main driver of changes in regional labour supply and the local level of human capital. A region's ability to attract residents thus becomes more and more important for its growth prospects. We use a large panel dataset for the period 2003 to 2017 to investigate the relationship between local attributes and the migration balance of regions in Germany. In particular, we examine whether the factors that determine the migration balance of regions significantly differ across age and skill groups because their contribution to regional human capital likely varies. Our econometric specification can be understood as an aggregate formulation of a two-region random utility model. The dataset includes 30 factors that might potentially influence a region's migration balance. Given this large number of explanatory variables and significant multicollinearity issues, we apply machine learning techniques [least absolute shrinkage and selection operator (LASSO), complete subset regression] to identify important local characteristics. Our results point to a robust negative relationship between the net migration rate and population density, yet locations in

close proximity to large urban centres seem to be rather attractive destination regions, and the size of the effects differs significantly across age and skill groups. Moreover, labour market conditions and some amenities are significantly correlated with the region's migration balance. However, the former and, in particular, facilities for vocational training matter primarily for young workers.

KEYWORDS
age groups, internal migration, machine learning, skill level

JEL CLASSIFICATION
C21, C63, J61, R23

1 | INTRODUCTION

Many European regions currently experience a significant population decline and, related to this, are increasingly confronted with labour shortage (see, e.g., OECD, 2022). The prospects for local economic growth deteriorate due to skill shortages, in particular in rural regions. Chen and Rosenthal (2008) note that migration is a main driver of changes in regional labour supply and the local level of human capital. Different authors emphasize, in line with this argument, that a region's ability to attract residents becomes more important for its future economic development (Malecki, 2004; Rodriguez-Pose & Ketterer, 2012). However, a differentiation by age and skill level of the workers is important in this context for several reasons. Firstly, the contribution of distinct age and skill groups to regional human capital likely differs. Moreover, age and the level of educational attainment influence the individual propensity to migrate (see Faggian et al., 2015). Finally, the preferred destinations might significantly differ across these groups, possibly adding to existing population imbalances.

The objective of this paper is to investigate the relationship between local attributes and the migration balance of regions. More specifically, we examine whether this relationship significantly differs across age and skill groups. Our regression analysis provides evidence on those characteristics that distinguish regions, which show a rather favourable migration balance for different groups of workers. This information might be helpful for the design of policy measures aiming to stabilize labour supply in structurally weak areas that are confronted with demographic change. We pay special attention to the migration behaviour of different age groups and to the variation across skill groups. Only a few studies so far investigate in detail whether the factors that determine the migration balance of regions significantly differ across age and skill groups. With increasing availability of extensive regional datasets that include detailed information on migration flows and various potential determinants, it is feasible and advisable to combine machine learning techniques with traditional approaches to investigate migration behaviour. These techniques enable us to tackle multicollinearity issues and model selection that usually arise when datasets include a large number of potentially influential factors. However, application of machine learning techniques to migration analyses is rare so far (see e.g. Micevska, 2021).

There is a vast amount of empirical literature on regional migration that provides evidence on various factors which are supposed to influence migration behaviour. While some authors stress the importance of labour market conditions (e.g., Alvarez & Royuela, 2022; DaVanzo, 1978; Scott, 2010; Shapiro, 2006), others emphasize the relevance of amenities (e.g., Glaeser & Gottlieb, 2006; Porell, 1982; Rodriguez-Pose & Ketterer, 2012). However, the set of explanatory variables included in the analyses often tends to be fairly limited. Frequently, the investigations focus on the effects of specific factors. Biagi et al. (2011) and Buch et al. (2014) are among the few studies that

consider a broader variety of factors, including economic variables and location-specific amenities as well as social and cultural characteristics. Moreover, systematic evidence on a potentially varying role of factors for the migration behaviour of different demographic groups is still scarce, and findings tend to be ambiguous.

Germany is particularly suitable for an empirical study that aims at providing evidence on population imbalances, internal migration and the determinants of labour migration. The country shows striking disparities with respect to regional migration balances, labour market conditions, local infrastructure and other amenities. In addition to a rather persistent East–West gap for different socioeconomic indicators, we observe significant differences between rural areas and large urban regions. Furthermore, demographic change, that is, a declining and aging work force, is already a challenge for the economic perspectives of many regions, in particular for rural areas in East Germany.

The paper is organized as follows. The next section provides a brief survey of the relevant literature. Section 3 describes the data and the econometric approach. We describe the results of the regression analysis in Section 4 and provide a more detailed discussion of some important findings in Section 5. Section 6 concludes the paper.

2 | LITERATURE

Migration theory discusses a variety of factors that are thought to influence the migration behaviour of individuals and households. The majority of models treat the migration decision as resulting from the evaluation of local labour market conditions and location-specific amenities. The basic idea is that the observed migration behaviour is based on the maximization of the utility of the individuals/households. The utility level of all potential places of residence, which is influenced by various local characteristics, is compared, taking into account relocation costs (see, e.g., Faggian et al., 2015). Sjaastad (1962) notes that the migration balance of a region can, thus, be understood as a function of the sum of individual utility levels (and therefore local attributes).

Numerous studies provide evidence on the significance of various factors that are supposed to influence migration behaviour and, thus, the migration balance of regions. There is robust evidence that interregional migration flows respond to (changes in) regional wage differentials and unemployment disparities (e.g., Etzo, 2011) as well as employment growth (Buch et al., 2014). Amenities that reflect local living conditions may also influence the attractiveness of regions as places of residence. Several studies show that first nature amenities such as a pleasant climate, a nice landscape and recreation areas positively correlate with the net migration rate of regions (see, e.g., Buch et al., 2014; Porell, 1982). Furthermore, second nature characteristics of a region, which include public infrastructure, cultural facilities and touristic sites, likely matter for its migration balance (Alperovich et al., 1977; Buettner & Ebertz, 2009). A detailed survey of the vast literature is beyond the scope of this paper. Our literature review focuses therefore on empirical studies that examine internal migration and consider differences in migration behaviour across skill and age groups.

Often the focus of this specific literature is on the relationship between individual attributes and the probability of migrating (see Bernard et al., 2014, and Faggian et al., 2015 for a survey) rather than examining the factors behind the migration balance of regions for distinct age and skill groups. Moreover, frequently migration decisions of specific age or skill groups are investigated, and much of this literature has focused on high-skilled individuals. In particular, there is an extensive body of literature on graduate migration which provides robust evidence on the importance of individual, study-related and regional factors for the migration decisions of young high-skilled workers after graduation (see, e.g., Faggian et al., 2006; Haapanen & Tervo, 2012; Venhorst et al., 2011). However, while these studies offer very detailed information on various factors that influence the migration of this specific group, they do not consider differences across levels of educational attainment. However, there are good reasons to expect that the importance of factors that influence migration decisions differs across demographic groups and skill levels. The impact of regional characteristics on migration behaviour likely varies between groups of workers if migration motives and preferences differ systematically between individuals.

Examining the impact of migration determinants for various age groups provides information on the extent to which residence preferences change over the course of working life (see Clark & Onaka, 1983; Kramer & Pfaffenbach, 2016). Findings by Niedomysl (2011) indicate that the living environment and housing seem to become more important as migration motives in Sweden as the age of individuals increases. The study also points to a positive correlation between the level of education and the significance of housing as a determinant of migration behaviour. Millington (2000) investigates how the importance of factors that influence migration behaviour changes over the life cycle in the United Kingdom. The findings point to a declining responsiveness to regional disparities in labour market conditions as age increases, while the opposite is true of housing and amenities – in line with results by Niedomysl (2011). Millington (2000) emphasizes importance of disaggregating by age and laments the lack of corresponding studies.

Other studies point to a significant heterogeneity of preferences and migration motives with respect to skill groups. Chen and Rosenthal (2008) note that highly educated households in the United States and younger age groups between 20 and 35 years seem to attach great importance to favourable economic conditions, while individuals aged 55 years or older tend to move to places that offer highly valued consumer amenities. Some authors argue that the utility attached to specific (dis)amenities likely differs across skill groups. Cullen and Levitt (1999) show, for instance, that migration decisions of highly educated households and those of families with children are particularly responsive to changes in crime. Urban shopping possibilities and cultural facilities are supposed to matter primarily for highly educated individuals (see Dalmazzo & de Blasio, 2011; Shapiro, 2006). Couture and Handbury (2017) show that urban amenities like restaurants and nightlife are increasingly valued by young well-educated workers in the United States, explaining at least partly their movement towards large urban centres. Amenities as well as labour market conditions seem to influence the migration decision of high-skilled workers in Germany according to results of Buch et al. (2017). Moreover, there is no evidence that their importance varies systematically across skill levels. This is in line with findings by Arntz et al. (2021). Their results indicate that preferences for urban amenities do not differ systematically by skill level. In contrast, a study by Buettner and Janeba (2016) suggests that subsidizing theatres might be effective in attracting highly educated people to a location.

Altogether, systematic evidence on a potentially varying role of factors for the migration behaviour of different demographic groups is still scarce, and existing findings are ambiguous.

3 | DATA AND ECONOMETRIC APPROACH

3.1 | Migration data and local characteristics

We use the Integrated Employment Biographies (IEB) of the Institute for Employment Research (IAB) to generate our regional migration data. The IEB includes detailed individual-level information on the workforce (15 to 65 years old) in Germany, more precisely on all registered unemployed people and all employees subject to social security contributions. Self-employed individuals, family workers and civil servants are not covered by the data. However, the migration data should be representative with respect to labour mobility because the IEB covers about 90% of the workforce in Germany.[1] Age, gender, skill level and residence of the workers is available in the dataset since 1999.

Our analysis makes use of annual migration data for the period 2000–2017. A migration event is defined as the change of residence, that is, the county region, between two reference dates (June 30 of present and previous year). The individual migration events are aggregated at the county region level (360 regions). We use the net migration

[1]See Frodermann et al. (2021) for a detailed description of the IEB.

rate nmr_{it} of region i in year t, proposed by Mitze (2019), as a dependent variable to investigate migration behaviour[2]:

$$nmr_{it} = \ln\left[\frac{(inmig_{it} - outmig_{it}) + pop_{it-1}}{pop_{it-1}}\right], \quad (1)$$

where $(inmig_{it} - outmig_{it})$ denotes net in-migration, that is, the difference between gross in-migration and gross out-migration, and pop_{it-1} is the regional workforce in $t-1$. Thus, the variable shows the relative change of the regional workforce caused by interregional migration. A value of 0.01 indicates that net in-migration gives rise to an increase of the regional workforce by 1% (10 net immigrants per 1,000 workers). We generate migration data for three qualification groups, low-skilled workers (no formal vocational qualification), medium-skilled workers (completed apprenticeship training) and high-skilled workers (university degree), and four age groups (< 25 years, 25–29 years, 30–39 years, and 40–65 years).

The annual net migration rates the range between a migration loss of 20 individuals per 1,000 workers and a migration gain of same absolute size. Furthermore, we observe that the net migration loss of East Germany declined between 2004 and 2017. Substantial out-migration was characteristic for most East German regions in the 1990s after German reunification. As regards the group of rural areas, we detect a considerable migration loss between 2007 and 2011 (see Figure A1 in the appendix). However, they perform better before and after this period in terms of their net migration rate. In particular, rural regions close to large metropolitan areas tend to experience a favourable migration balance. We also observe important regional disparities for different age and skill groups. In particular, a strong out-migration of young workers from rural regions is noteworthy, while older workers often seem to move in opposite direction.

Migration theory considers various factors that might influence a region's migration balance. To examine the impact of potential factors on labour mobility, we merge our migration data with regional information from several data sources. In addition to a set of indicators for regional labour market conditions, we include measures of local amenities as a second group of factors that might explain spatial disparities in migration balances. Information on regional unemployment, apprenticeship training positions, the demographic composition of the population, childcare facilities, average flat size, voter turnout and fiscal capacity of the region is taken from the INKAR database of the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). We use the IEB to generate indicators for the economic structure of the regions and employment shares of gastronomy and the creative economy. The regional wage level comes from the national accounts of the Federal States, population density, recreation area, overnight stays and land prices from the regional database of the Federal Statistical Office (Destatis).[3] Regional information on the crime rate is provided by the Federal Criminal Police Office (BKA), the National Meteorological Service (DWD) offers detailed meteorological information at the county level. Finally, a survey by Stiller and Ohlhoff (2021) provides evidence on return migration initiatives implemented by local governments (see Table A1 for detailed description of the variables and Table A2 for summary statistics).

Availability of time series information is potentially limiting the inclusion of further explanatory factors. We apply fixed effects regression models to account for the influence of unobserved regional characteristics. Therefore, we can only investigate the importance of migration determinants that show a sufficient variation in the time dimension. The impact of all time-constant factors is captured by region fixed effects. Moreover, local characteristics that change very slowly over time tend to suffer from weak identification. These issues are relevant for attributes

[2]We choose the net migration rate as our dependent variable because a significant correlation of local attributes with this outcome directly reflects whether regional labour shortage can potentially be reduced via policy-induced changes of the determinants. An alternative specification is the gravity model which might provide more detailed information on migration behaviour, as it differentiates between inflows and outflows. However, the approach does not refer directly to a change in the regional labour force. Moreover, estimation of spatial interaction models is subject to some methodological challenges such as the endogeneity of the spatial system, a lack of valid instruments and an absence of independence of irrelevant alternatives (see Alvarez & Royuela, 2022; Beine et al., 2016).

[3]We use the housing price index of the RWI – Leibniz Institute for Economic Research (see Klick et al., 2019) as an alternative housing market indicator.

like accessibility or first nature characteristics such as the landscape. The final dataset is an unbalanced panel because not all explanatory variables are available annually for the period 2003 to 2017. In particular, information on childcare facilities and housing prices is only available from 2008 onwards.

3.2 | Econometric approach

To investigate the importance of various potential determinants of the regional net migration rate, we apply a regression model given by:

$$nmr_{it} = \alpha + \sum_{k=1}^{K} \beta_k \ln\left(\frac{x^k_{it-1}}{x^k_{Gt-1}}\right) + \delta_i + \theta_t + \varepsilon_{it}. \quad (2)$$

We consider different local labour market conditions and other local characteristics (x^k_{it-1}; see, e.g., Buch et al., 2014; Mitze, 2019). All explanatory variables enter as logarithms of ratios, which measure the relative deviations of the local conditions from the respective German average, excluding the region under consideration (x^k_{Gt-1}). Our econometric specification can therefore be understood as an aggregate formulation of a two-region random utility model of labour migration between region *i* and the rest of Germany (see Mitze, 2019). All regressors are predetermined to account for potential endogeneity of explanatory variables. The panel specification includes regions-specific effects δ_i that capture the impact of unobserved time-constant determinants of interregional labour migration. θ_t denotes time-effects and the white noise error term is given by ε_{it}.[4]

It is important to keep in mind that a fixed effects estimation does not provide a perfect solution to the identification problems that arise from unobserved heterogeneity. The fixed effects model makes use of the within variation only. However, the cross-sectional variation often makes up the major part of the variation in regional datasets. Effects of explanatory variables that show only a minor variation in the time dimension might therefore be weakly identified (see Hausman & Taylor, 1981).

Our dataset includes 30 factors that, according to theoretical considerations outlined in Section 2, might potentially influence a region's migration balance. Moreover, we also consider spatial lags of the variables to allow for spillover effects from characteristics of neighbouring locations.[5] Given this large number of explanatory variables and significant multicollinearity issues, we apply machine learning techniques (LASSO, complete subset regression) to identify important local characteristics. Regional datasets often suffer from multicollinearity problems, as many variables exhibit a similar (cross-sectional) variation. Population density, housing prices and rents show, for instance, a strong positive correlation, and it is difficult to precisely identify the impact of a specific factor if explanatory variables co-vary.[6] Panel data tends to alleviate multicollinearity problems because co-variation is primarily related to cross-sectional variation. However, often local characteristics exhibit only small variation over time. Thus, regression analyses that consider many factors, which potentially affect the outcome variable, face a tradeoff between bias and precision. Including a comprehensive set of explanatory factors will help to avoid an omitted variable bias, but reduces the precision of the estimates as measured by the standard errors.

Model selection is challenging, in particular when there are many factors that are assumed to be important. Ahrens et al. (2020) note that iterative selection procedures, such as the general-to-specific approach, often result in pretesting biases and hypothesis tests frequently lead to false positives. They argue in favour of machine learning techniques as a model selection approach because methods like the least absolute shrinkage and selection operator

[4]The time-effects will capture to some extent changes in the average country-wide level of the factors included in the analysis. However, the denominator is equal to the national measure *excluding* the region under consideration. The variation of the denominator is small, but the time-effects will not absorb the denominator completely. Moreover, the overlap of controls should not affect the coefficient estimates.

[5]To compute spatial lags, we use a row-standardized binary contiguity weights matrix.

[6]Table A3 in the appendix shows the correlation between all explanatory variables considered in the regression analysis.

(LASSO) set some coefficients to exactly zero, thereby excluding these predictors from the model. Cochrane et al. (2022) use machine techniques to identify predictors of regional resilience because stepwise regression approaches often lead to a selection of over-fitted specifications and machine learning techniques can be used to select a robust set of explanatory variables from a large set of potential predictors.

We apply two machine learning approaches to deal with the tradeoff between bias and precision. These techniques support the selection of important explanatory variables, which should show a robust effect in many different specifications, ensure a precise estimation and avoid biased estimates. Identification of key determinants becomes increasingly difficult as the number regressors in the model grows (see Hastie et al., 2009). The least absolute shrinkage and selection operator (LASSO) extends the OLS estimator by a factor that penalizes the inclusion of additional regressors (see Tibshirani, 1996). We use the method proposed by Belloni et al. (2016) for panel models to determine the penal factor and apply LASSO to detect variables that stand out due to their relatively high explanatory power for the net migration rate.

We use LASSO to identify sparse models which can be precisely estimated. However, these models may suffer from a comparatively high risk of an omitted variable bias. The focus of machine learning techniques is on prediction rather than on producing good estimates of parameters. As a consequence, these approaches often give rise to inconsistent coefficient estimates. A main weakness of machine learning is that it produces stable parameter estimates only under strong and mostly unverifiable assumptions (see Mullainathan & Spiess, 2017).

We consider these risks by combining LASSO with a second approach, complete subset regression (CSR). Moreover, we do not completely rely on machine learning alone but in fact use LASSO and CRS to select explanatory variables that we include in our fixed effects models. CSR enables us to identify important factors in addition to the variables selected via LASSO. We use CSR also to check the robustness of the results across various specifications (see Elliott et al., 2013; Sala-i-Martin, 1997). CSR estimates the model given by Equation (2) for all different specifications (combinations of explanatory variables) for a fixed number of regressors to be included. If we choose to consider 5 explanatory variables in every specification, this implies that we estimate 142,506 different specifications $\left[\binom{30}{5} = \frac{30!}{(30-5)!5!}\right]$, given that the overall number of potentials factors is 30. The advantage of CSR is that it allows us to mitigate multicollinearity while obtaining information on the robustness of a specific coefficient across numerous specifications.

4 | RESULTS

Table 1 summarizes the regression results for different age groups. As a reference we also include the estimates for all workers in column (1). All coefficients are semi-elasticities. For instance, the coefficient of the regional wage level in column (1) indicates that an increase in the regional wage level by 10% relative to the rest of the country gives rise to a change of the net migration rate by 0.05 percentage points or, in other words, an increase of net in-migration of 5 workers per 10,000 workers who reside in the region. Furthermore, it is important to keep in mind that coefficient estimates should be interpreted as conditional correlations rather than causal effects. Thus, they provide information on how we can describe regions, which suffer from a net migration loss or benefit from an important net in-migration of workers. Some explanatory variables are likely endogenous although we reduce the risk of omitted variable bias via a large number of regional characteristics and region fixed effects included in the regression analysis. However, economic theory suggests that the relationship between migration and regional disparities is interdependent. This applies in particular to labour market/economic conditions. Regional differences in wages and unemployment might influence migration behaviour, but at the same time migration likely affects local labour market conditions (see, e.g., Granato et al., 2015).

The results show that labour market conditions as well as amenities correlate with the net migration rate. However, there are considerable differences across age groups when the impact of specific factors is concerned. The estimates for labour market indicators that are selected by our machine learning approach suggest that primarily

TABLE 1 Correlation between net migration rates and local characteristics across age groups, 2004–2017[a]

Dependent variable: net migration rate	(1) All workers	(2) < 25 years	(3) 25–29 years	(4) 30–39 years	(5) 40–65 years
Wage level[(1),(2)]	0.00522*	0.0163*	0.0194**		
	(0.00225)	(0.00754)	(0.00658)		
Vocational training positions[(1),(2)]	0.00328*	0.0117*			
	(0.00154)	(0.00549)			
Share of primary sector[(1),(2)]	−0.000980**	−0.00332**	−0.00152		−0.000583**
	(0.000327)	(0.00119)	(0.000807)		(0.000181)
Share of other low-knowledge sectors	−0.00142	−0.00615*			−0.000572
	(0.000988)	(0.00305)			(0.000546)
Share of low knowledge-intensive services	0.00356*		0.00958*	0.00402	0.00207**
	(0.00144)		(0.00400)	(0.00217)	(0.000788)
Population density[(1),(2),(3),(4),(5)]	−0.0507***	−0.0352*	−0.0940***	−0.0585***	−0.0258***
	(0.00569)	(0.0176)	(0.0134)	(0.00793)	(0.00388)
Spatial lag of population density[(1),(2),(4)]	0.00648	−0.0859***	0.0375**	0.0418***	−0.000325
	(0.00541)	(0.0163)	(0.0121)	(0.00769)	(0.00311)
Share of population < 25 years[(1),(3),(4)]	−0.000624	0.00968	−0.0229***	−0.0130**	0.00349
	(0.00301)	(0.0116)	(0.00677)	(0.00397)	(0.00216)
Share of foreign population	0.000948*			0.00158*	
	(0.000457)			(0.000774)	
Share of gastronomy	0.00149**	0.00482**	0.00350	0.00143*	0.000463
	(0.000534)	(0.00157)	(0.00179)	(0.000664)	(0.000292)
Share of creative economy[(2)]	0.00154***	0.00586***	0.00328**		
	(0.000459)	(0.00168)	(0.00126)		
Overnight stays	0.00148**	0.00263	0.00260	0.00113	0.000782**
	(0.000549)	(0.00161)	(0.00133)	(0.000800)	(0.000291)
Recreation area		0.00340*			
		(0.00159)			
Precipitation				−0.00148*	
				(0.000661)	
Public financial capacity[(1),(2)]	0.000824	0.00680*			
	(0.000835)	(0.00279)			
Voter turnout				0.00991**	0.00362*
				(0.00374)	(0.00168)
Crime rate	−0.000720			−0.00117	
	(0.000447)			(0.000614)	
Flat size	−0.0175***	−0.0624***	−0.0232**		−0.00992***
	(0.00336)	(0.0109)	(0.00849)		(0.00173)
Constant	0.00661***	0.0263***	0.00477*	−0.000679	0.00434***
	(0.000939)	(0.00302)	(0.00216)	(0.00129)	(0.000553)
N	4,911	4,911	4,911	4,911	4,911
Number of county regions	360	360	360	360	360
R^2 within	0.206	0.144	0.125	0.164	0.131

Note: With robust standard errors in parentheses. All models include region and year fixed effects. The selection of explanatory variables relies on a LASSO regression, which is combined with CSR. The superscripts (1)–(5) indicate all variables, which are selected based on LASSO, if the (sub-)sample in the respective column is considered.

[a]We restrict the regression analysis to the period 2004–2017 because many explanatory variables are only available from 2003 onwards.

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

young workers attach high importance to these attributes, confirming evidence for the United Kingdom provided by Millington (2000). We detect a significant positive correlation between the regional wage level and the migration balance only for the two youngest age groups. The older workforce seems to bring other local characteristics to the foreground when deciding on the place of residence. This is in line with previous findings for the United States by Chen and Rosenthal (2008) and Clark and Hunter (1992). Moreover, we observe for the youngest workers (< 25 years) that they seem to prefer regions that offer an extensive supply of apprenticeship training positions. The sectoral structure of the regions also matters. An increasing share of the primary sector and other low-knowledge industries tends to go along with a declining net migration rate. This applies in particular to the age group below 25 years. Being specialized in low knowledge-intensive services, in contrast, seems to increase the attractiveness of a region for specific age groups.

Furthermore, the estimates point to a robust negative correlation between changes in the net migration rate and changes in population density across age groups. For workers between 25 and 39 years, we also detect an important positive effect of density in neighbouring regions.[7] This implies that suburbanization trends distinguish the migration behaviour of these age groups. They tend to leave dense metropolitan areas and often move to the urban hinterland of large cities. This is in stark contrast to the migration behaviour of the youngest age group. Workers aged below 25 years tend leave regions, which border on large metropolitan areas. At the same time, there is also a negative impact of population density in the region itself on the net migration rate of these workers. However, the latter effect is less robust compared with other age groups.[8]

There is also a negative correlation between the net migration rate of the age groups 25–39 years and the share of young inhabitants (< 25 years). The latter result might be driven by the migration of young graduates who leave the region of study/vocational training after completion and a first phase of their career. Workers in their thirties also seem to prefer regions showing a high share of foreign inhabitants and a high voter turnout. The latter also applies to the oldest age group. These findings may be interpreted as indicating the role of political and societal participation and the benefits of cultural diversity in the region of residence.

Evidence on crime being a disamenity that affects migration behaviour is rather weak. The variable is only selected for the age group 30–39 years, and the corresponding negative effect is not precisely estimated. However, there are also estimates, which point to an impact of (urban) amenities. The importance of gastronomy, the creative economy, places of (touristic) interest and the availability of recreation area correlate positively with the net migration rate of different age groups, with a slight indication that amenities might be more important for younger workers, which is in contrast to findings by Niedomysl (2011) for Sweden. For instance, the public financial capacity of the region, which is used in our analysis to approximate the provision of public services and infrastructure, is only selected as an influential factor for the youngest age group.

As regards the housing market, only the flat size is selected as an influential factor for the period from 2004 to 2017. It is noteworthy that regions with an above-average supply of small flats seem to be rather attractive for almost all age groups, but especially for the youngest workers. This might reflect the availability of affordable housing, which is likely important, in particular for households that possess only a small budget. Restricting the analysis to the period after 2008 enables us to examine the role of housing prices and childcare facilities. Evidence on relevant effects is, however, fairly weak for these variables. We observe a negative correlation between changes in the net migration rate and changes in childcare infrastructure (significant for workers < 25 years) for all age groups apart from the workforce between 30 and 39 years, but most effects are not precisely estimated. We also detect a negative correlation between the house price index and a region's migration balance across all age groups, which is also precisely estimated at the 5% level for the workers aged 30 to 39 years.[9]

[7]When attributes of neighbouring regions are concerned, only the spatial lag of the population density turns out to significantly influence the net migration rate of different age groups. Other spatially lagged explanatory variables are not selected by the machine learning techniques.

[8]Population density does not correlate with the net migration rate of worker aged below 25 years in a regression model that only includes population density and its spatial lag.

[9]See Table A4 in the appendix.

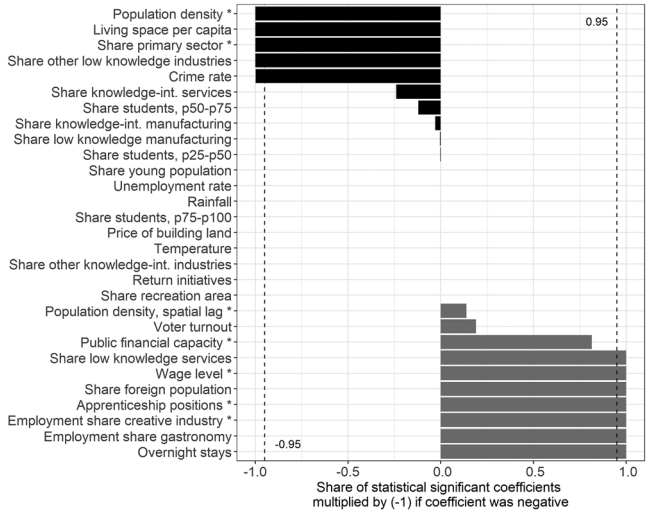

FIGURE 1 Share of significant coefficient estimates – results of CSR for the age group < 25 years. *Notes*: Share of statistically significant coefficient estimates ($p < 0.05$) in total number of 20,349 regressions. Each regression model includes all variables marked by an asterisk and five additional variables. *Source*: Own calculation using the R-package rrsim by Thomas de Graaff, IEB and the regional database described in Table A1 in the appendix.

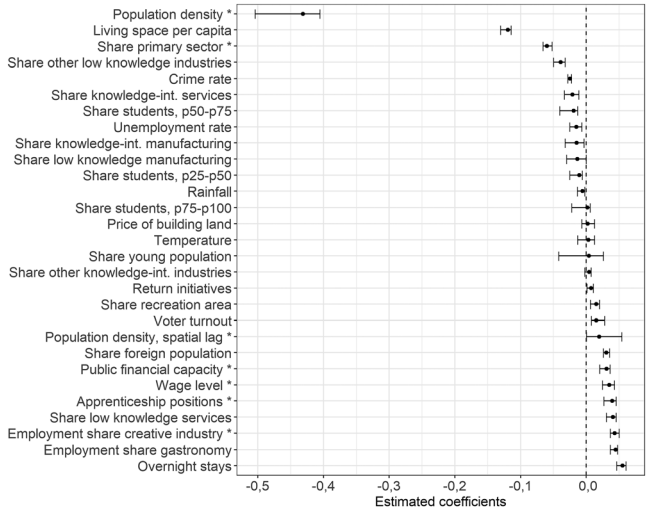

FIGURE 2 Variation of coefficient estimates – results of CSR for the age group < 25 years. *Notes*: Range of coefficient estimates from 20,349 regressions, excluding the most extreme values (5%). Each regression model includes all variables marked by an asterisk and five additional variables. The dot indicates the mean estimate. *Source*: Own calculation using the R-package rrsim by Thomas de Graaff, IEB and regional database described in Table A1 in the appendix.

To evaluate the robustness of the findings in more detail, we apply CSR, which provides information on the percentage of significant estimates and on the variation of the coefficient estimates across various specifications of our migration model. In the following, we focus on the youngest workers aged below 25 years.[10] Figure 1 and Figure 2 summarize the results of 20,349 regressions of the migration model given by Equation (2). In every regression,

TABLE 2 Correlation between net migration rates and local characteristics across skill groups, 2004–2017[a]

Dependent variable: net migration rate	(1) Low-skilled	(2) Medium-skilled	(3) High-skilled	(4) All workers
Share of primary sector		−0.000526 (0.000275)		−0.000463 (0.000294)
Share of knowledge-intensive services		−0.000995 (0.000795)		−0.00133 (0.000999)
Share of low-knowledge services	0.00449* (0.00176)	0.00356** (0.00113)		0.00381** (0.00119)
Population density[(1),(2),(3),(4)]	−0.0504*** (0.00561)	−0.0390*** (0.00377)	−0.0637*** (0.00754)	−0.0399*** (0.00368)
Spatial lag of population density[(1),(3)]	0.0186** (0.00668)		0.0402*** (0.00801)	0.00859* (0.00408)
Share of population < 25 years		0.00382 (0.00270)		
Share of foreign population [(4)]	0.00272*** (0.000746)	0.00150*** (0.000391)	0.00237* (0.000947)	0.00167*** (0.000413)
Share of gastronomy	0.00197** (0.000735)	0.000979** (0.000371)		0.00103* (0.000398)
Overnight stays		0.000811* (0.000399)		0.000664 (0.000427)
Crime rate		−0.000915* (0.000372)	−0.00206* (0.000935)	−0.000946* (0.000384)
Flat size		−0.0121*** (0.00287)	−0.0246** (0.00770)	−0.0126*** (0.00278)
Constant	0.00303** (0.00116)	0.00575*** (0.000641)	0.000572 (0.00153)	0.00400*** (0.000773)
N	4,911	4,911	4,911	4,911
Number of county regions	360	360	360	360
R^2 within	0.0844	0.133	0.0868	0.141

Note: With robust standard errors in parentheses. All models include region and year fixed effects. The selection of explanatory variables relies on a LASSO regression, which is combined with CSR. The superscripts (1)–(4) indicate all variables, which are selected based on LASSO, if the (sub-)sample in the respective column is considered.

[a]We restrict the regression analysis to the period 2004 to 2017 because many explanatory variables are only available from 2003 onwards.

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

the main explanatory factors that have beforehand been selected based on LASSO (cf., Table 1) and are marked by an astrisk, and five additional variables are included. While Figure 1 shows the incidence of significant estimates at the 5% level and the sign of the correlation, Figure 2 indicates the size and range of standardized coefficient estimates.

Figure 1 reveals that the results of those explanatory variables that we identified as important factors in Table 1 are also characterized by a high degree of robustness. Altogether there are 12 variables, which are precisely estimated with the same sign in every regression. Some additional characteristics turn out to show a significant correlation only in some specification such as the fiscal capacity of the region and the share of knowledge-intensive services. A third group of variables does not correlate with the net migration at all. It is important to take notice of the potential determinants that are not chosen at all with our approach, such as the regional unemployment rate, meteorological indicators and recreation area in case of the young workforce. Figure 2 more or less confirms the previous results. The variation of coefficient estimates is moderate for the majority of factors identified as influential.

Rare exceptions include the population density and its spatial lag. Thus, we notice that robustness also refers to the size of the 'effect' for the majority of variables.

Table 2 shows the findings of a regression analysis that differentiates between three skill groups. To provide results by skill group, we restrict our sample to employed workers aged above 25 years. A significant percentage of the age group below 25 years has not yet completed university education or apprenticeship training. These workers would be assigned to the low-skilled workforce, thus introducing a substantial measurement error. The number of relevant labour market indicators declines once we exclude workers below age 25 years. This corresponds with the above average importance of labour market conditions detected in Table 1 for the youngest workers. However, the economic structure still matters for migration in the reduced sample. The low- and medium-skilled workforce tends to prefer regions specialized in low-knowledge services, which might thus offer many job opportunities for these skill groups. We do not detect important effects of the economic structure on the net migration rate of high-skilled workers.

The coefficient estimates for population density and its spatial lag confirm the results for different age groups. We detect a significant negative correlation for all skill levels. However, there are important differences in the size of the 'effect' in absolute terms.[11] Changes in population density seem to correlate in particular with changes in the migration balance of high-skilled workers, while we observe the smallest coefficient in absolute terms for the medium-skilled employees. A robust positive correlation between the share of the foreign population and the migration balance is also visible for all skill groups. An above average importance of gastronomy seems to increase the attractiveness of locations for low- and medium skilled workers. This result might point to the influence of an amenity (Buch et al., 2017), but could also be driven by corresponding job opportunities relevant to these skill groups. The latter interpretation is in line with weak evidence on a robust relationship for the high-skilled workers. In contrast, there is some indication that the crime rate becomes more important for the net migration rate as the skill level of the workforce increases. Cullen and Levitt (1999) provide corresponding evidence for the United States. Finally, we observe a significant negative correlation between the net migration rate and the average flat size per capita for the medium- and high-skilled workforce.

5 | DISCUSSION

Our results point to a robust negative relationship between the net migration rate and population density, yet locations in close proximity to large urban centres seem to be rather attractive destination regions. Thus, agglomeration disadvantages seem to prevail when labour migration in Germany is concerned, conditional on a number of covariates. From a theoretical perspective, it is ambiguous how density influences the migration balance (see Glaeser & Gottlieb, 2006). Congestion of the local infrastructure and environmental problems may accompany high density and act as centrifugal forces, giving rise to low net in-migration (Brown & Scott, 2012). Moreover, high housing costs may contribute to net migration losses that we tend to observe for large cities. On the other hand, density may give rise to agglomeration advantages such as intense interaction and knowledge transfer from which the region's migration balance should benefit as they represent centripetal forces (see, e.g., Buch et al., 2017). Alvarez and Royuela (2022) note that new economic geography models also have implications for interregional migration and the role of centrifugal and centripetal forces in this context. For instance, Crozet (2004) describes migration decisions as a function of wages, unemployment, transport costs and the region's access to markets. He shows that access to markets significantly affects interregional migration in Europe, in line with the forward linkage in discussed in Krugman (1991). In this analysis, in contrast, agglomeration disadvantages seem to outweigh benefits when migration behaviour is concerned. However, the negative net effect of agglomeration might at least partly be caused by the inclusion of other factors in the regression model that correlate with agglomeration benefits such as the wage level and different amenities.

[11]As the results are based on different regressions, we do not test whether the differences are significant. However, comparing the differences between the coefficient estimates and the size of standard errors suggests that there are significant differences across skill groups.

In contrast, the utility that a location outside large cities offers as a place of residence seems to increase as the distance to major metropolitan areas declines. These findings suggest that, for rural regions, low density is a factor that promotes net in-migration of workers conditional on other determinants of migration behaviour. However, it is primarily those rural areas close to metropolitan regions that benefit from this constellation, and it does not apply to same extent to all age groups. Areas bordering on large cities offer good access to agglomeration advantages while workers at the same time avoid locational factors that negatively affect quality of life in metropolitan regions. The underlying suburbanization processes are primarily driven by high-skilled workers and the age group between 25 and 39 years. In contrast, for young workers aged below 25 years these regions do not seem to offer a particularly high utility level. For the youngest workforce the negative correlation between density and migration balance is less robust, suggesting that they might benefit more than other age groups from agglomeration advantages not captured by other explanatory variables. Large cities offer a broad range of training opportunities and jobs, which are of special importance for early career workers.

Moreover, labour market conditions and some amenities are significantly correlated with the region's migration balance. The role of specific factors for migration behaviour seems vary across age groups. We observe that the wage level and, in particular, facilities for vocational training matter primarily for young workers (see also Chen & Rosenthal, 2008 for corresponding US evidence). However, even for the youngest workers, these factors are not more important than amenities that reflect the recreational value of the location. It is also noteworthy that we do not detect an impact of regional unemployment on the net migration rate. Interestingly, this applies to all groups of workers considered in our analysis. This is in line with evidence provided by Buch et al. (2017) and, at least partly, by Arntz (2010).

We identify heterogenous effects for the sectoral structure of the region. While the youngest workers show disproportionate out-migration from (rural) regions with an above average share of the primary sector, low- and medium-skilled workers seem to prefer locations specialized in low-knowledge services, which probably offer many employment opportunities, especially for the former group.

Within the group of amenities, indicators for cultural facilities, places of interest and a variety of gastronomic offerings show a robust positive correlation with a region's migration balance. Locations with a relatively high share of foreign population are also often among those regions with a significant net in-migration of workers (see also Buch et al., 2014, for corresponding evidence on German cities). We interpret this variable as indicating a diverse supply of consumption goods and services (see Alesina & La Ferrara, 2005; Ottaviano & Peri, 2005). However, the relationship between cultural diversity and the net migration rate seems to be driven exclusively by the age group between 30 and 39 years. The migration balance of workers beyond their twenties also correlates positively with our measure for political and societal participation, the voter turnout. And finally, the regression results point to an adverse effect of crime on the liveability that a location offers (see also Cullen & Levitt, 1999 for the United States).

Altogether, we should not overrate the strength of all 'effects'. The size of the coefficient estimates indicates that the impact of a specific factor on the net migration rate is moderate. For instance, an increase of a region's wage level by 10% relative to all other regions improves the migration balance by 5 net immigrants per 10,000 workers living in the region. This is small in comparison to the migration losses that many areas face. On average the migration loss amounted to 20 persons per 10,000 workers among the quarter of regions, which experienced the highest net out-migration between 2014 and 2017.

6 | CONCLUSIONS

Our findings point to substantial spatial disparities in the net migration rate. However, rural areas do not show net out-migration of workers in Germany per se. There is a considerable heterogeneity between rural regions when labour migration is concerned. While rural areas close to large metropolitan regions tend to experience high net migration gains, rural sites in the periphery suffer from considerable net out-migration, as do cities on average.

This suggests that access to large urban centres might be a key factor. The results of our regression analysis also indicate that there is a robust relationship between a region's net migration rate and different labour market/economic conditions as well as (dis)amenities. However, the importance of specific factors varies significantly between age and skill groups.

Rural regions close to large metropolitan areas have great appeal, especially to high-skilled workers and those aged between 25 and 39 years, that is, parts of the workforce that are assumed to play a decisive role in the economic prospects of regions. Thus, there seems to be no immediate need for political action for this category of rural areas. In view of high population growth, caused by net migration gains, that might increase housing prices and capacity utilization of local facilities, policy might rather focus on preserving favourable conditions in these regions. In contrast, accessibility is likely an important starting point for policy action in those rural locations, which are not within commuting distance to a large metropolitan area. However, improving the transport infrastructure might not be the first best option in this context. Progressive suburbanization has severe environmental consequences (urban sprawl, land consumption). Moreover, rural areas do not necessarily benefit from improved physical accessibility via investments in transport infrastructure. Some studies suggest, in fact, it is primarily the large urban areas that benefit from improved traffic links between rural and urban regions (e.g., Faber, 2014). However, there might be alternative options to increase accessibility in rural sites if working from home becomes more common.

Another issue refers to the question of whether policy can effectively influence important determinants of migration behaviour at all. It is not feasible or extremely difficult to change some of the local attributes such as the wage level and the sectoral structure, which show a robust correlation with the net migration rate. Of course, well-designed (regional) economic policy might help to create attractive employment opportunities in rural areas. However, numerous studies provide evidence on persistent spatial wage disparities, which are partly due to agglomeration effects (Hamann et al., 2019, and Peters, 2019, provide corresponding evidence for Germany). Moreover, effectively improving some local attributes might only influence the migration behaviour of specific groups of workers. This also implies that the cost-effectiveness of some policies might be relatively high simply because they focus on local attributes that matter for age groups, which attach a relatively high utility to rural locations anyway (e.g., workers between 25 and 39 years). Corresponding starting points for policy design include the recreational value of the region, public safety and measures to strengthen political and societal participation. In contrast, we detect no important effects of childcare facilities and return initiatives of local authorities for the considered groups of workers. Hence, at least on average these factors apparently do not significantly impact internal labour migration in Germany.

When net out-migration of young workers from rural areas is concerned, labour market conditions and, in particular, facilities for vocational training might be crucial. However, in view of demographic change and declining population figures in many rural areas, the financial feasibility of providing training infrastructure close to a young worker's place of residence is increasingly put into question (OECD, 2021). Our results underline the significance of access to training positions to the migration behaviour of young workers. A policy that aims at improving the migration balance of rural areas in this age group should therefore consider how to stabilize a sustainable educational infrastructure in these regions. We refrain from a detailed presentation of different approaches (e.g., training networks of SMEs, branch offices of universities) that are discussed in this context (see Daniel et al., 2019, for the German context). Moreover, findings by Teichert et al. (2020) suggest that providing opportunities to gain knowledge about the local labour market and establishing labour market contacts before and during studies might be a possible strategy to deepen the ties of young skilled workers to the region of residence.

There is a considerable variation within the group of rural regions in Germany when it comes to the endowment with factors that positively influence the migration balance. This implies that a 'one-size-fits-all' policy for rural regions will not work. The starting point of a local strategy to improve the migration balance should therefore be a thorough analysis of locational advantages and disadvantages vis-à-vis other (rural) areas. Based on the evidence provided by such an analysis, region-specific strategies might be developed that allow for the strengths and weaknesses of the region under consideration. However, in the end we need to keep in mind that small improvements in

some fields will probably not be sufficient to achieve a fundamental change of a region's migration balance, as the size of the effects, as indicated by the coefficient estimates, tends to be rather small.

ACKNOWLEDGEMENTS

Financial support from the German Federal Ministry of Food and Agriculture for the project 'The spatial mobility of workers over the course of their working lives – Analyses for rural areas in Germany' within the framework of the Federal Rural Development Scheme is gratefully acknowledged (grants no. 2817LE054 and no. 2817LE055). We are grateful to Thomas de Graaff for providing his R-package rrsim and valuable advice. We have benefited from useful discussions with seminar participants at the SETTLE International Conference 2022 'Population imbalances in Europe' at the Universidad de Alcalá and thank editor Tomaz Ponce Dentinho as well as three anonymous referees for helpful comments and suggestions. The usual disclaimer applies. Open Access funding enabled and organized by Projekt DEAL.

ORCID

Moritz Meister https://orcid.org/0000-0003-1947-4911
Annekatrin Niebuhr https://orcid.org/0000-0001-8218-0954
Jan Cornelius Peters https://orcid.org/0000-0001-9328-3078

REFERENCES

Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2020). Lassopack: Model selection and prediction with regularized regression in Stata. *The Stata Journal*, *20*(1), 176–235. https://doi.org/10.1177/1536867X20909697

Alesina, A., & La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, *43*(3), 762–800. https://doi.org/10.1257/002205105774431243

Alperovich, G., Bergsman, J., & Ehemann, C. (1977). An econometric model of migration between US metropolitan areas. *Urban Studies*, *14*(2), 135–145. https://doi.org/10.1080/00420987720080291

Alvarez, M., & Royuela, V. (2022). The effect of labor-market differentials on interregional migration in Spain: A meta-regression analysis. *Journal of Regional Science*, *62*, 913–937. https://doi.org/10.1111/jors.12579

Arntz, M. (2010). What attracts human capital? Understanding the skill composition of interregional job matches in Germany. *Regional Studies*, *44*, 423–441. https://doi.org/10.1080/00343400802663532

Arntz, M., Brüll, E., & Lipowski, C. (2021). *Do preferences for urban amenities really differ by skill?* (ZEW Discussion Paper No. 21-045). ZEW Mannheim. https://doi.org/10.2139/ssrn.3871649

Beine, M., Bertoli, S., & Fernández-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. *The World Economy*, *39*(4), 496–512. https://doi.org/10.1111/twec.12265

Belloni, A., Chernozhukov, V., Hansen, C., & Kozbur, D. (2016). Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics*, *34*(4), 590–605. https://doi.org/10.1080/07350015.2015.1102733

Bernard, A., Bell, M., & Charles-Edwards, E. (2014). Life-course transitions and the age profile of internal migration. *Population and Development Review*, *40*(2), 213–239. https://doi.org/10.1111/j.1728-4457.2014.00671.x

Biagi, B., Faggian, A., & McCann, P. (2011). Long and short distance migration in Italy: The role of economic, social and environmental characteristics. *Spatial Economic Analysis*, *6*(1), 11–131. https://doi.org/10.1080/17421772.2010.540035

Brown, W. M., & Scott, D. M. (2012). Human capital location choice: Accounting for amenities and thick labor markets. *Journal of Regional Science*, *52*(5), 787–808. https://doi.org/10.1111/j.1467-9787.2012.00772.x

Buch, T., Hamann, S., Niebuhr, A., & Rossen, A. (2014). What makes cities attractive? The determinants of urban labour migration in Germany. *Urban Studies*, *51*(9), 1960–1978. https://doi.org/10.1177/0042098013499796

Buch, T., Hamann, S., Niebuhr, A., & Rossen, A. (2017). How to woo the smart ones? Evaluating the determinants that particularly attract highly qualified people to cities. *Journal of Urban Affairs*, *39*(6), 764–782. https://doi.org/10.1080/07352166.2017.1282765

Buettner, T., & Ebertz, A. (2009). Quality of life in the regions: Results for German counties. *Annals of Regional Science*, *43*(1), 89–112. https://doi.org/10.1007/s00168-007-0204-9

Buettner, T., & Janeba, E. (2016). City competition for the creative class. *Journal of Cultural Economics*, *40*(4), 413–451. https://doi.org/10.1007/s10824-015-9258-1

Chen, Y., & Rosenthal, S. (2008). Local amenities and life-cycle migration: Do people move for jobs or fun? *Journal of Urban Economics*, *64*(3), 519–537. https://doi.org/10.1016/j.jue.2008.05.005

Clark, D. E., & Hunter, W. J. (1992). The impact of economic opportunity, amenities and fiscal factors on age-specific migration rates. *Journal of Regional Science*, *32*(3), 349–365. https://doi.org/10.1111/j.1467-9787.1992.tb00191.x

Clark, W. A., & Onaka, J. L. (1983). Life cycle and housing adjustment as explanations of residential mobility. *Urban Studies*, *20*(1), 47–57. https://doi.org/10.1080/713703176

Cochrane, W., Poot, J., & Roskruge, M. (2022). *Urban resilience and social security uptake: New Zealand evidence from the global financial crisis and the COVID-19 pandemic*. IZA - Institute of Labor Economics. IZA discussion paper 15510.

Couture, V., & Handbury, J. (2017). *Urban revival in America, 2000 to 2010*. National Bureau of Economic Research, Inc. NBER working paper no. w24084.

Crozet, M. (2004). Do migrants follow market potentials? An estimation of a new economic geography model. *Journal of Economic Geography*, *4*(4), 439–458. https://doi.org/10.1093/jnlecg/lbh029

Cullen, J. B., & Levitt, S. D. (1999). Crime, urban flight, and the consequences for cities. *Review of Economics and Statistics*, *81*, 159–169. https://doi.org/10.1162/003465399558030

Dalmazzo, A., & de Blasio, G. (2011). Amenities and skill-biased agglomeration effects: Some results on Italian cities. *Papers in Regional Science*, *90*, 503–527. https://doi.org/10.1111/j.1435-5957.2010.00327.x

Daniel, H. D., Hannover, B., Köller, O., Lenzen, D., McElvany, N., Rossbach, H. G., Seidel, T., Tippelt, R., & Wössmann, L. (2019). Region und Bildung. In E. V. Vereinigung der Bayerischen Wirtschaft (Ed.), *Mythos Stadt-land*Report. vbw. Waxmann Verlag, Münster.

DaVanzo, J. S. (1978). Does unemployment affect migration? Evidence from micro data. *Review of Economics and Statistics*, *60*, 504–514. https://doi.org/10.2307/1924242

Elliott, G., Gargano, A., & Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, *177*(2), 357–373. https://doi.org/10.1016/j.jeconom.2013.04.017

Etzo, I. (2011). The determinants of the recent interregional migration flows in Italy: A panel data analysis. *Journal of Regional Science*, *51*, 948–966. https://doi.org/10.1111/j.1467-9787.2011.00730.x

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*, *81*(3), 1046–1070. https://doi.org/10.1093/restud/rdu010

Faggian, A., Corcoran, J., & Partridge, M. (2015). Interregional migration analysis. In C. Karlsson, M. Andersson, & T. Norman (Eds.), *Handbook of research methods and applications in economic geography* (pp. 468–490). Edward Elgar Publishing. https://doi.org/10.4337/9780857932679.00030

Faggian, A., McCann, P., & Sheppard, S. (2006). An analysis of ethnic differences in UK graduate migration behaviour. *Annals of Regional Science*, *40*(2), 461–471. https://doi.org/10.1007/s00168-006-0061-y

Frodermann, C., Schmucker, A., Seth, S., & Vom Berge, P. (2021). *Sample of integrated labour market biographies (SIAB) 1975–2019*(FDZ Data Report No. 01/2021). Institute for Employment Research.

Gehrke, B., Rammer, C., Frietsch, R., & Neuhäusler, P. (2010). *Listen wissens- und technologieintensiver Güter und Wirtschaftszweige – Zwischenbericht zu den NIW/ISI/ZEW-Listen 2010/2011*. (Studien zum deutschen Innovationssystem Nr. 19-2010). Expertenkommission Forschung und Innovation (EFI).

Glaeser, E. L., & Gottlieb, J. D. (2006). Urban resurgence and the consumer city. *Urban Studies*, *43*, 1275–1299. https://doi.org/10.1080/00420980600775683

Granato, N., Haas, A., Hamann, S., & Niebuhr, A. (2015). The impact of skill-specific migration on regional unemployment disparities in Germany. *Journal of Regional Science*, *55*(4), 513–539. https://doi.org/10.1111/jors.12178

Haapanen, M., & Tervo, H. (2012). Migration of the highly educated: Evidence from residence spells of university graduates. *Journal of Regional Science*, *52*(4), 587–605. https://doi.org/10.1111/j.1467-9787.2011.00745.x

Hamann, S., Niebuhr, A., & Peters, J. C. (2019). Does the urban wage premium differ by pre-employment status? *Regional Studies*, *53*(10), 1435–1446. https://doi.org/10.1080/00343404.2019.1577553

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-84858-7

Hausman, J., & Taylor, W. (1981). Panel data and unobservable individual effects. *Econometrica*, *49*, 1377–1398. https://doi.org/10.2307/1911406

Klick, L., Schaffner, S., & RWI. (2019). Regional real estate price index [Data set]. In *RWI-GEO-REDX (Version 1)*. RWI – Leibniz Institute for Economic Research. https://doi.org/10.7807/IMMO:REDX:V2

Kramer, C., & Pfaffenbach, C. (2016). Should I stay or should I go? Housing preferences upon retirement in Germany. *Journal of Housing and the Built Environment*, *31*(2), 239–256. https://doi.org/10.1007/s10901-015-9454-5

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, *99*(3), 483–499. https://doi.org/10.1086/261763

Küpper, P. (2016). *Abgrenzung und Typisierung ländlicher Räume*Thünen Working Paper 68. Johann Heinrich von Thünen Institute.

Malecki, E. J. (2004). Jockeying for position: What it means and why it matters to regional development policy when places compete. *Regional Studies*, *38*(9), 1101–1120. https://doi.org/10.1080/0034340042000292665

Meister, M., Stiller, J., Peters, J. C., & Birkeneder, A. (2019). *Die Binnenwanderung von Arbeitskräften in Deutschland: eine deskriptive Analyse für ländliche Räume auf Basis der Integrierten Erwerbsbiografien (IEB) des IAB*Thünen Working Paper 132. Johann Heinrich von Thünen Institute.

Micevska, M. (2021). Revisiting forced migration: A machine learning perspective. *European Journal of Political Economy*, *70*, 102044. https://doi.org/10.1016/j.ejpoleco.2021.102044

Millington, J. (2000). Migration and age: The effect of age on sensitivity to migration stimuli. *Regional Studies*, *34*(6), 521–533. https://doi.org/10.1080/00343400050085648

Mitze, T. (2019). The migration response to local labour market shocks: Evidence from EU regions during the global economic crisis. *Oxford Bulletin of Economics and Statistics*, *81*(2), 271–298. https://doi.org/10.1111/obes.12271

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*, 87–106. https://doi.org/10.1257/jep.31.2.87

Niedomysl, T. (2011). How migration motives change over migration distance: Evidence on variation across socio-economic and demographic groups. *Regional Studies*, *45*(6), 843–855. https://doi.org/10.1080/00343401003614266

OECD. (2021). *Delivering quality education and health care to all: Preparing regions for demographic change*. (OECD Rural Studies). OECD Publishing. https://doi.org/10.1787/83025c02-en

OECD. (2022). *The contribution of migration to regional development*. (OECD Regional Development Studies). OECD Publishing. https://doi.org/10.1787/57046df4-en

Ottaviano, G. I., & Peri, G. (2005). Cities and cultures. *Journal of Urban Economics*, *58*(2), 304–337. https://doi.org/10.1016/j.jue.2005.06.004

Peters, J. C. (2019). Dynamic agglomeration economies and learning by working in specialised regions. *Journal of Economic Geography*, *20*(3), 629–651. https://doi.org/10.1093/jeg/lbz022

Porell, F. W. (1982). Intermetropolitan migration and quality of life. *Journal of Regional Science*, *22*(2), 137–158. https://doi.org/10.1111/j.1467-9787.1982.tb00741.x

Rodriguez-Pose, A., & Ketterer, T. D. (2012). Do local amenities affect the appeal of regions in Europe for migrants? *Journal of Regional Science*, *52*(4), 535–561. https://doi.org/10.1111/j.1467-9787.2012.00779.x

Sala-i-Martin, X. (1997). I just ran two million regressions. *American Economic Review*, *87*(2), 178–183.

Scott, A. J. (2010). Jobs or amenities? Destination choices of migrant engineers in the USA. *Papers in Regional Science*, *89*, 43–64. https://doi.org/10.1111/j.1435-5957.2009.00263.x

Shapiro, J. M. (2006). Smart cities: Quality of life, productivity, and the growth effects of human capital. *The Review of Economics and Statistics*, *88*, 324–335. https://doi.org/10.1162/rest.88.2.324

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, *70*, 80–93. https://doi.org/10.1086/258726

Söndermann, M., Backes, C., & Arndt, O. (2009). *Gesamtwirtschaftliche Perspektiven der Kultur-und Kreativwirtschaft in Deutschland*. (BMWi-Forschungsbericht 577). Bundesministerium für Wirtschaft.

Stiller, J., & Ohlhoff, M. (2021). *Rückkehrinitiativen in Deutschland: Eine Erhebung für den Zeitraum 2001 bis 2018*. (Thünen Working Paper 177). Johann Heinrich von Thünen-Institut.

Teichert, C., Niebuhr, A., Otto, A., & Rossen, A. (2020). Work experience and graduate migration: An event history analysis of German data. *Regional Studies*, *54*(10), 1413–1424. https://doi.org/10.1080/00343404.2020.1716965

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B: Methodological*, *58*(1), 267–288.

Venhorst, V., Van Dijk, J., & Van Wissen, L. (2011). An analysis of trends in spatial mobility of Dutchgraduates. *Spatial Economic Analysis*, *6*(1), 57–82. https://doi.org/10.1080/17421772.2010.540033

How to cite this article: Meister, M., Niebuhr, A., Peters, J. C., & Stiller, J. (2023). Local attributes and migration balance – evidence for different age and skill groups from a machine learning approach. *Regional Science Policy & Practice*, *15*(4), 794–825. https://doi.org/10.1111/rsp3.12652

APPENDIX A

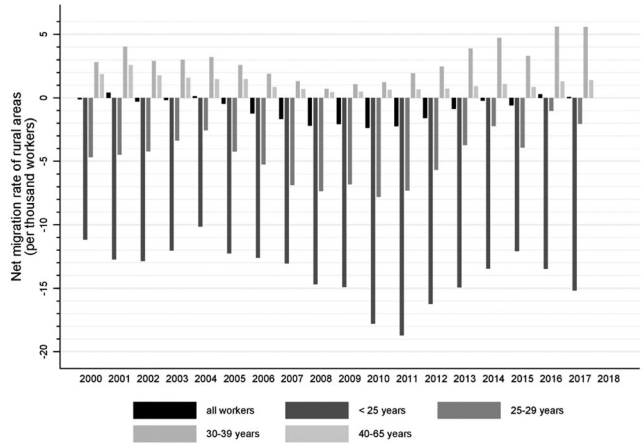

FIGURE A1 Net migration rate of rural areas in Germany by age groups. *Notes*: The net migration rate refers to the migration of all workers captured by the IEB. *Source*: Own illustration based on Meister et al. (2019: Figure 6); definition of rural areas according to Küpper (2016).

TABLE A1 Variable definition and data sources

Variable	Definition	Source	Period
Net migration rate	See equation (1)	Integrated employment biographies (IEB)	2000–2017
Wage level	Gross annual wage per capita (in 1,000 euros)	National accounts of the Federal States	2000–2017
Unemployment rate	Number of unemployed persons divided by labour force (in %)	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	1999–2017
Apprenticeship training positions	Supply of training positions per 100 school graduates	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	1999–2017
Sector shares	Share of sectors in total employment for NIW/ISI/ZEW-categories (in %); see Gehrke et al. (2010)	IEB	1999–2017
Population density	Inhabitants per km^2 (in 1,000)	*Regionaldatenbank Deutschland* of the Federal Statistical Office of Germany	1999–2016
Share of inhabitants < 25 years	Share of inhabitants below age 25 years in total population	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	1999–2017
Share of foreign population	Share of foreign inhabitants in total population	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	1999–2017

(Continues)

TABLE A1 (Continued)

Variable	Definition	Source	Period
Share of gastronomy	Share of gastronomic occupations in total employment	IEB	1999–2017
Share of creative economy	Share of cultural industry and creative economy in total employment[a]	IEB	1999–2017
Overnight stays	Number of overnight stays per inhabitant	*Regionaldatenbank Deutschland* of the Federal Statistical Office of Germany	1999–2017
Recreation area	Recreation area divided by the total area (in %)	*Regionaldatenbank Deutschland* of the Federal Statistical Office of Germany	2000, 2004, 2008–2016
Precipitation	Annual amount of precipitation (in mm/m^2)	National Meteorological Service (DWD)	1999–2017
Temperature	Mean annual temperature (in degrees Celsius)	National Meteorological Service (DWD)	1999–2017
Childcare facilities	Share of preschool children in full-time care	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	2008–2017
Crime rate	Number of cases (street crime) per 100,000 inhabitants	Crime statistics of Germany's Federal Criminal Police Office (BKA)	2003–2016
Return initiatives	Existence of active return initiative at the municipality level or the county level in a respective year	Own survey (see Stiller & Ohlhoff, 2021)	2001–2017
Flat size	Average flat size per inhabitant (in m^2)	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	1999–2017
Land price	Average land price (land ready for building; in euro per m^2)	*Regionaldatenbank Deutschland* of the Federal Statistical Office of Germany	1999–2016
Housing price index	Price index for flats (offer price)	RWI based on Immobilienscout24 (see Klick et al., 2019)	2008–2017
Public financial capacity	Fiscal capacity of municipalities per capita (in thousands of euros)	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	1999–2017
Voter turnout	Share of voters among eligible voters, bundestag elections	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	2002, 2005, 2009, 2013, 2017
Share students (3 quartiles: 25–50, 50–75, 75–100)	Number of students per 100,000 inhabitants, categorial variable based on quartile, 0–25 quartile as reference	INKAR database of the Federal Institute for Research on building, Urban Affairs and spatial development (BBSR)	1999–2017

[a]The creative economy includes literature, music, the performing arts, film, broadcasting service, design, architecture, the press, advertising and gaming (see Söndermann et al., 2009).

TABLE A2 Summary statistics

Variable	Mean	Standard deviation	Within Standard deviation	Min.	Max.
Net migration rate, all workers	−0.000582	0.00424	0.00262	−0.0205	0.0198
Net migration rate, < 25 years	−0.00625	0.0230	0.00833	−0.0740	0.140
Net migration rate, 25–29 years	−0.00352	0.0123	0.00729	−0.0608	0.0557
Net migration rate, 30–39 years	0.00101	0.00853	0.00391	−0.0382	0.0358
Net migration rate, 40–65 years	0.000515	0.00248	0.00141	−0.0104	0.0151
Net migration rate, low-skilled	0.000115	0.00557	0.00418	−0.0292	0.0318
Net migration rate, medium-skilled	0.000277	0.00424	0.00218	−0.0182	0.0265
Net migration rate, high-skilled	0.000347	0.0109	0.00619	−0.0391	0.0452
Unemployment rate	−0.165	0.460	0.0998	−1.739	0.914
Wage level	−0.0730	0.140	0.0215	−0.376	0.470
Apprenticeship training positions	0.00644	0.0467	0.0323	−0.195	0.253
Share of primary sector	−0.252	1.235	0.159	−4.046	2.334
Share of knowledge-intensive manufacturing	−0.205	0.690	0.120	−3.256	1.712
Share of knowledge-intensive sectors (other)	−0.341	0.720	0.188	−3.711	2.259
Share of knowledge-intensive services	−0.228	0.308	0.0459	−1.124	0.721
Share of low-knowledge manufacturing	0.0659	0.584	0.0673	−2.535	1.310
Share of low-knowledge sectors (other)	0.0956	0.380	0.0646	−1.389	1.249
Share of low-knowledge services	−0.0553	0.183	0.0344	−0.793	0.584
Population density	0.108	1.065	0.0234	−1.852	3.029
Spatial lag of population density	0.232	0.814	0.0186	−1.686	2.346
Share of population < 25 years	−0.00121	0.0992	0.0316	−0.349	0.270
Share of foreign population	−0.402	0.692	0.112	−2.537	1.315
Share of gastronomy	−0.0724	0.348	0.0835	−1.105	1.430
Share of creative economy	−0.403	0.459	0.0987	−1.826	1.231
Overnight stays	−0.307	0.836	0.0949	−2.739	2.377
Recreation area	0.0285	0.946	0.124	−1.678	2.694
Precipitation	−0.000209	0.190	0.0810	−0.590	0.726
Temperature	0.00891	0.0925	0.0202	−0.588	0.221
Public financial capacity	−0.119	0.316	0.0824	−1.090	1.195
Voter turnout	−0.00482	0.0557	0.0170	−0.231	0.132
Crime rate	−0.341	0.513	0.110	−1.774	1.131
Share of students (quartile 25–50 years)	0.0764	0.266	0.156	0	1
Share of students (quartile 50–75 years)	0.247	0.431	0.170	0	1
Share of students (quartile 75–100 years)	0.254	0.435	0.0929	0	1

(Continues)

TABLE A2 (Continued)

Variable	Mean	Standard deviation	Within Standard deviation	Min.	Max.
Land price	−0.291	0.856	0.192	−3.302	2.714
Flat size	0.0255	0.0881	0.0199	−0.244	0.308
Return initiatives	0.159	0.366	0.180	0	1
Childcare facilities	−0.287	0.767	0.154	−3.221	1.226
Housing price index	−0.128	0.184	0.0305	−0.868	0.864

Note: Summary statistics refer to the variables used in the regression analysis. All explanatory variables are logarithms of the ratio of the indicators in the respective regions to the value in the rest of the country. Results for categorial variables (return initiatives, share of students) refer to the value of the corresponding region.
Source: IEB and regional characteristics (Table A1), own calculations.

TABLE A3 Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(2)	−0.33	1.00														
(3)	0.15	−0.37	1.00													
(4)	−0.52	0.33	0.00	1.00												
(5)	0.50	−0.40	0.19	−0.29	1.00											
(6)	−0.03	0.35	−0.21	0.01	−0.19	1.00										
(7)	0.44	0.10	−0.07	−0.39	−0.23	0.10	1.00									
(8)	−0.33	−0.28	0.07	0.10	0.01	−0.27	−0.60	1.00								
(9)	−0.60	0.13	0.05	0.51	−0.32	0.04	−0.49	0.23	1.00							
(10)	−0.07	0.30	−0.16	0.09	−0.49	0.15	0.15	−0.51	0.01	1.00						
(11)	−0.36	0.44	−0.16	0.17	−0.60	0.20	0.27	−0.34	0.06	0.17	1.00					
(12)	0.49	0.31	−0.20	−0.39	−0.05	0.24	0.66	−0.49	−0.47	0.21	0.08	1.00				
(13)	0.35	0.03	−0.28	−0.28	0.03	0.09	0.27	−0.23	−0.21	0.21	−0.14	0.45	1.00			
(14)	0.27	−0.55	−0.08	−0.48	0.27	−0.07	0.13	0.08	−0.29	−0.15	−0.27	0.06	0.14	1.00		
(15)	0.74	−0.14	−0.02	−0.54	0.23	0.08	0.57	−0.33	−0.58	0.07	−0.23	0.69	0.49	0.41	1.00	
(16)	−0.10	−0.01	0.17	0.12	−0.31	−0.01	0.19	−0.30	0.08	0.34	0.28	−0.07	−0.18	−0.16	−0.05	1.00
(17)	0.37	0.03	−0.03	−0.38	−0.15	0.03	0.70	−0.37	−0.44	0.15	0.15	0.59	0.16	0.14	0.47	0.07
(18)	−0.15	−0.07	0.21	0.10	−0.17	−0.02	0.05	−0.08	0.11	0.14	0.10	−0.16	−0.32	−0.07	−0.13	0.80
(19)	0.36	0.43	−0.24	−0.31	−0.11	0.26	0.55	−0.50	−0.37	0.29	0.18	0.88	0.34	−0.06	0.49	−0.04
(20)	0.17	−0.38	0.19	−0.26	0.14	−0.14	0.01	0.15	−0.08	−0.14	−0.20	−0.08	−0.02	0.31	0.23	0.25
(21)	0.24	0.18	−0.20	0.02	−0.07	0.17	0.19	−0.30	−0.13	0.23	0.12	0.32	0.38	−0.01	0.27	−0.19
(22)	−0.08	0.50	0.01	0.31	−0.26	0.03	0.13	−0.20	0.06	0.16	0.31	0.17	−0.03	−0.72	−0.19	0.02
(23)	0.13	0.55	−0.41	−0.13	−0.27	0.32	0.51	−0.54	−0.31	0.43	0.33	0.65	0.30	−0.03	0.31	0.00
(24)	−0.39	0.46	0.15	0.51	−0.23	0.07	−0.21	0.00	0.32	0.08	0.24	−0.17	−0.26	−0.67	−0.46	0.06
(25)	−0.07	−0.50	0.16	0.05	0.05	−0.19	−0.31	0.35	0.21	−0.23	−0.04	−0.53	−0.31	0.10	−0.30	0.15
(26)	0.44	−0.11	0.05	−0.26	0.07	0.03	0.46	−0.30	−0.35	0.09	−0.08	0.46	0.30	0.17	0.51	0.06
(27)	0.60	−0.27	0.15	−0.36	0.14	−0.04	0.63	−0.50	−0.51	0.21	−0.12	0.47	0.37	0.30	0.66	0.17

(Continues)

TABLE A3 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(28)	0.84	−0.49	0.21	−0.43	0.43	−0.11	0.37	−0.24	−0.52	−0.01	−0.42	0.31	0.27	0.36	0.62	−0.04
(29)	0.20	−0.47	−0.15	−0.31	0.11	−0.13	0.22	−0.02	−0.28	−0.02	−0.14	0.01	0.33	0.53	0.23	−0.02
(30)	−0.02	0.02	−0.05	0.05	−0.08	−0.02	−0.02	0.01	0.05	0.12	−0.07	−0.03	0.18	−0.06	−0.03	−0.05
(31)	0.03	−0.02	−0.02	−0.01	0.13	0.02	−0.15	0.01	0.03	0.03	−0.11	−0.14	0.04	0.05	0.00	−0.04
(32)	0.26	0.25	−0.07	−0.29	−0.10	0.12	0.53	−0.40	−0.43	0.12	0.32	0.50	0.00	0.05	0.33	0.02
	(17)	**(18)**	**(19)**	**(20)**	**(21)**	**(22)**	**(23)**	**(24)**	**(25)**	**(26)**	**(27)**	**(28)**	**(29)**	**(30)**	**(31)**	**(32)**
(17)	1.00															
(18)	−0.03	1.00														
(19)	0.50	−0.15	1.00													
(20)	0.04	0.24	−0.17	1.00												
(21)	0.08	−0.28	0.34	−0.42	1.00											
(22)	0.11	−0.09	0.26	−0.33	0.13	1.00										
(23)	0.42	−0.12	0.71	−0.24	0.41	0.21	1.00									
(24)	−0.16	0.07	−0.06	−0.27	−0.05	0.66	−0.02	1.00								
(25)	−0.33	0.23	−0.47	0.16	−0.04	−0.32	−0.49	−0.26	1.00							
(26)	0.41	−0.03	0.32	0.13	0.11	−0.01	0.19	−0.21	−0.24	1.00						
(27)	0.54	−0.01	0.29	0.21	0.10	−0.06	0.23	−0.28	−0.35	0.58	1.00					
(28)	0.36	−0.07	0.17	0.22	0.15	−0.21	0.02	−0.41	0.01	0.45	0.66	1.00				

Notes: Variables: (1) wage level, (2) unemployment rate, (3) apprenticeship training positions, (4) share primary sector, (5) share of knowledge-intensive manufacturing, (6) share of knowledge-intensive sectors (other), (7) share of knowledge-intensive services, (8) share of low-knowledge manufacturing, (9) share of low-knowledge sectors (other), (10) share of low-knowledge services, (11) share of other sectors, (12) population density, (13) spatial lag of population density, (14) share of inhabitants < 25 years, (15) share of foreign population, (16) share of gastronomy, (17) share of creative economy, (18) overnight stays, (19) share of recreation area, (20) precipitation, (21) temperature, (22) childcare facilities, (23) crime rate, (24) return initiatives, (25) flat size, (26) land price, (27) housing price index, (28) public financial capacity, (29) voter turnout, (30) share of students (2nd quartile), (31) share of students (3rd quartile), (32) share of students (4th quartile).

TABLE A4 Correlation between net migration rates and local characteristics across age groups, 2009–2017

Dependent variable: net migration rate	(2) < 25 years	(3) 25–29 years	(4) 30–39 years	(5) 40–65 years
Wage level[(1),(2)]	0.00899 (0.0102)	0.0115 (0.0111)		
Vocational training positions[(1),(2)]	−0.00699 (0.00687)			
Share of primary sector[(1),(2)]	−0.00215 (0.00195)	−0.00213 (0.00160)		−0.000168 (0.000202)
Share of other low-knowledge sectors	−0.00336 (0.00334)			−0.000129 (0.000432)
Share of low knowledge-intensive services		0.00532 (0.00608)	0.00287 (0.00341)	0.00212* (0.00107)
Population density[(1),(2),(3),(4),(5)]	−0.0732** (0.0249)	−0.121*** (0.0226)	−0.0725*** (0.0122)	−0.0238*** (0.00381)
Spatial lag of population density[(1),(2),(4)]	−0.000902 (0.0213)	0.0689*** (0.0194)	0.0523*** (0.0111)	0.00323 (0.00353)
Share of population < 25 years[(1),(3),(4)]	−0.0898*** (0.0174)	−0.0696*** (0.0135)	−0.0415*** (0.00830)	−0.00657** (0.00243)
Share of foreign population			0.00150 (0.000896)	
Share of gastronomy	−0.000174 (0.00223)	0.00221 (0.00211)	0.00121 (0.000819)	0.000204 (0.000282)
Share of creative economy[(2)]	0.00368 (0.00204)	0.00354* (0.00155)		
Overnight stays		−0.000150 (0.00236)	−0.000457 (0.00129)	−0.0000788 (0.000461)
Recreation area	−0.00155 (0.00184)			
Precipitation			−0.00133 (0.000806)	
Childcare facilities	−0.00380*** (0.00109)	−0.000911 (0.00113)	0.000334 (0.000558)	−0.0000643 (0.000177)
Public financial capacity[(1),(2)]	−0.000649 (0.00355)			
Voter turnout			0.0177*** (0.00495)	0.00619*** (0.00159)
			0.0000484 (0.000857)	
Flat size	−0.0195 (0.0127)	−0.0000812 (0.0102)		−0.00337 (0.00194)
Housing price index	−0.0124* (0.00595)	−0.0158** (0.00561)	−0.00807** (0.00261)	−0.00352*** (0.00104)
Constant	0.00338 (0.00395)	−0.00661 (0.00395)	−0.00387 (0.00217)	0.00208** (0.000746)
N	3,146	3,146	3,146	3,146

(Continues)

TABLE A4 (Continued)

Dependent variable: net migration rate	(2) < 25 years	(3) 25–29 years	(4) 30–39 years	(5) 40–65 years
Number of county regions	360	360	360	360
R^2 within	0.125	0.158	0.199	0.138

Note: With robust standard errors in parentheses. All models include region and year fixed effects. The selection of explanatory variables relies on a LASSO regression, which is combined with CSR. The superscripts (1)–(5) indicate all variables, which are selected based on LASSO, if the (sub-)sample in the respective column is considered.
*p < 0.05, **p < 0.01, and ***p < 0.001.

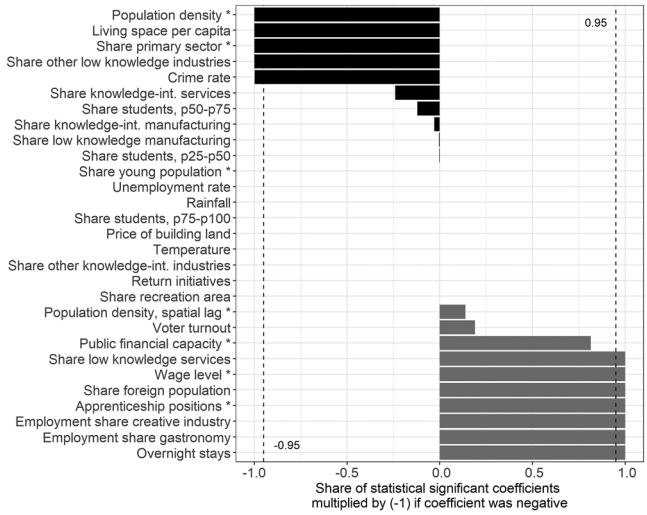

FIGURE A2 Share of significant coefficient estimates – results of CSR for all workers. *Notes*: Share of statistically significant coefficient estimates (p < 0.05) in a total number of 20,349 regressions. Each regression model includes all variables marked by an asterisk and five additional variables. *Source*: Own calculation using the R-package rrsim by Thomas de Graaff, IEB and regional database described in Table A1 in the appendix.

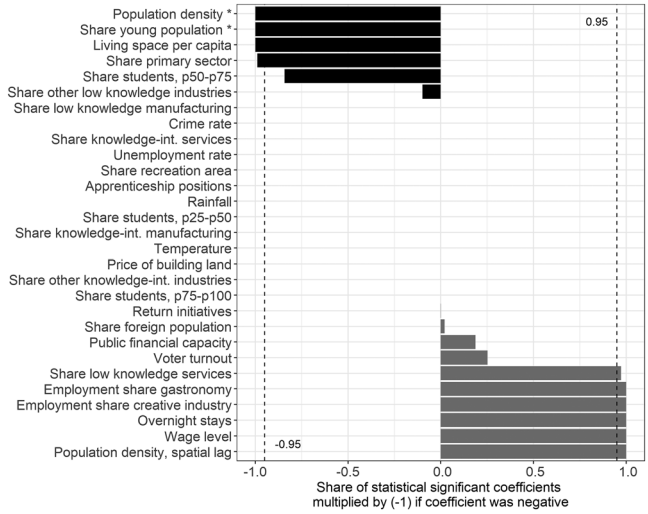

FIGURE A3 Share of significant coefficient estimates – results of CSR for the age group 25–29 years. *Notes and source*: See Figure A2.

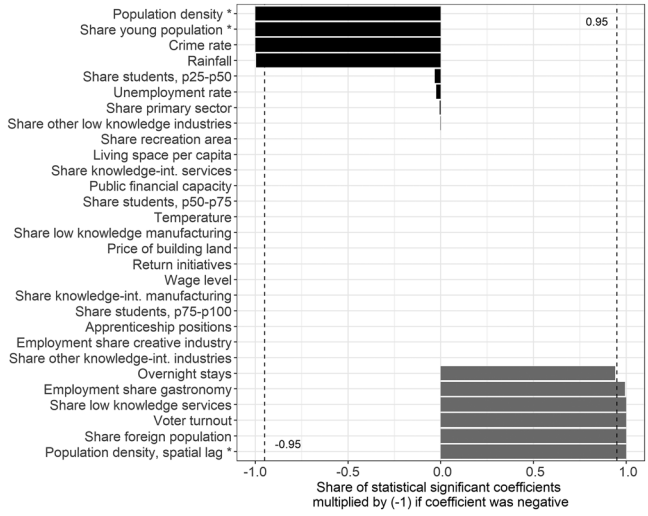

FIGURE A4 Share of significant coefficient estimates – results of CSR for the age group 30–39 years. *Notes and source*: See Figure A2.

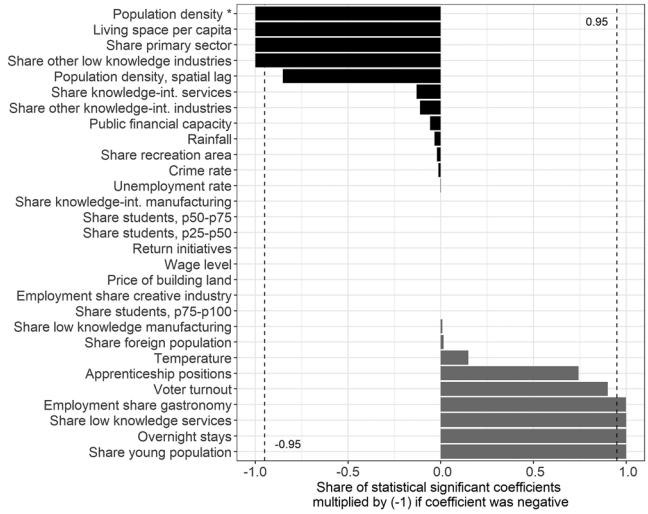

FIGURE A5 Share of significant coefficient estimates – results of CSR for the age group 40–65 years. *Notes and source*: See Figure A2.

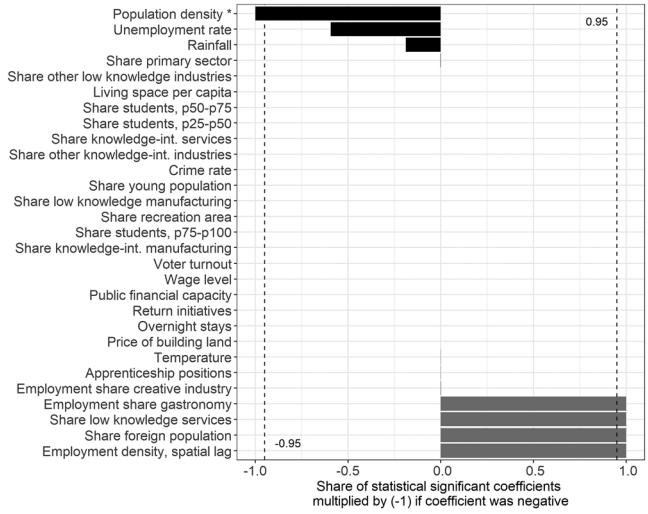

FIGURE A6 Share of significant coefficient estimates – results of CSR for low-skilled workers. *Notes and source*: See Figure A2.

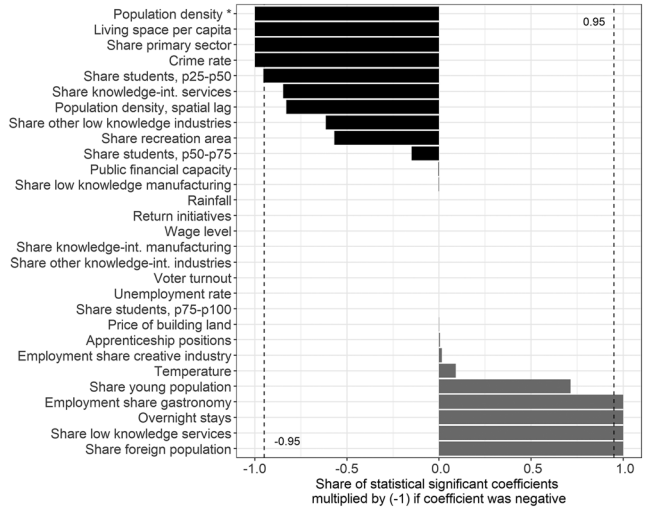

FIGURE A7 Share of significant coefficient estimates – results of CSR for medium-skilled workers. *Notes and source*: See Figure A2.

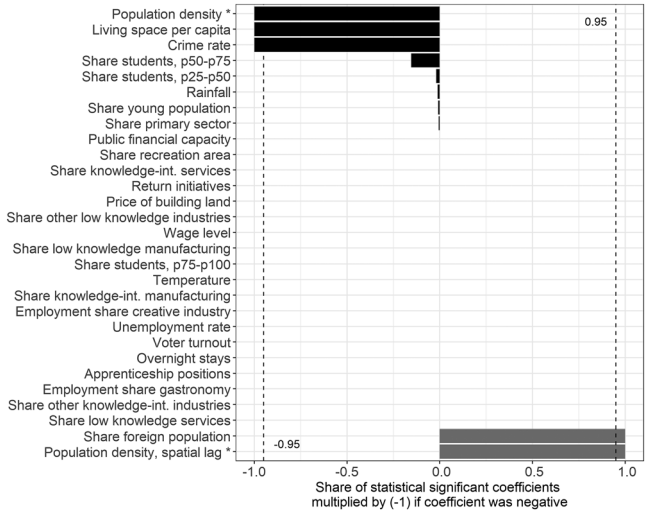

FIGURE A8 Share of significant coefficient estimates – results of CSR for high-skilled workers. *Notes and source*: See Figure A2.

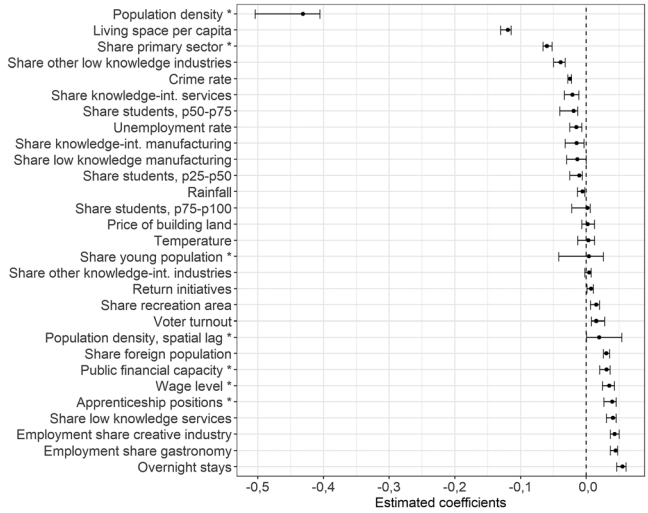

FIGURE A9 Variation of coefficient estimates – results of CSR for all workers. *Notes*: Range of coefficient estimates from 20,349 regressions, excluding the most extreme values (5%). Each regression model includes all variables marked by an asterisk and five additional variables. The dot indicates the mean estimate. *Source*: Own calculation using the R-package rrsim by Thomas de Graaff, IEB and regional database described in Table A1 in the appendix.

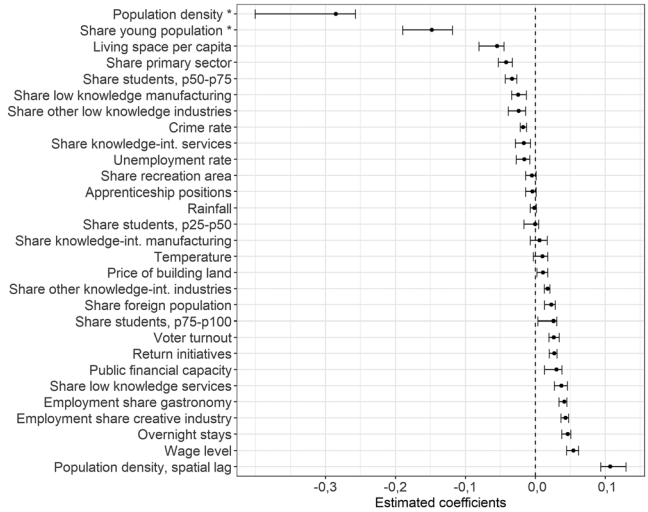

FIGURE A10 Variation of coefficient estimates – results of CSR for the age group 25–29 years. *Notes and source*: See Figure A9.

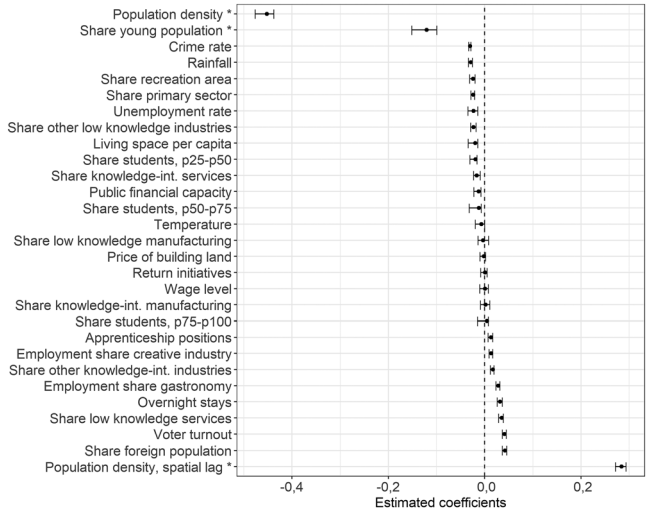

FIGURE A11 Variation of coefficient estimates – results of CSR for the age group 30–39 years. *Notes and source*: See Figure A9.

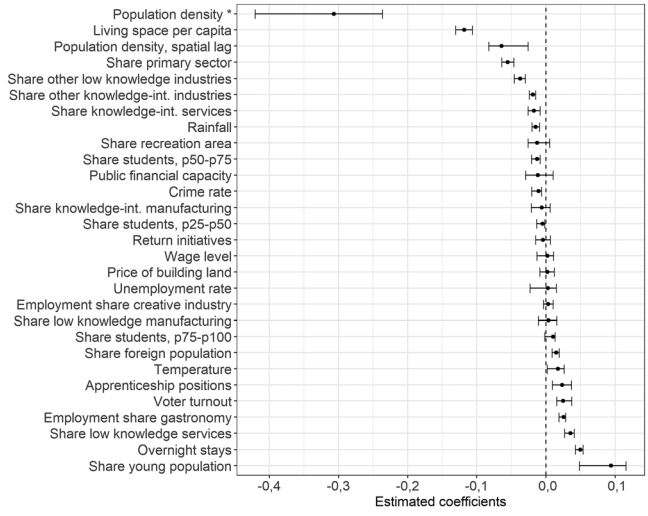

FIGURE A12 Variation of coefficient estimates – results of CSR for the age group 40–65 years. *Notes and source*: See Figure A9.

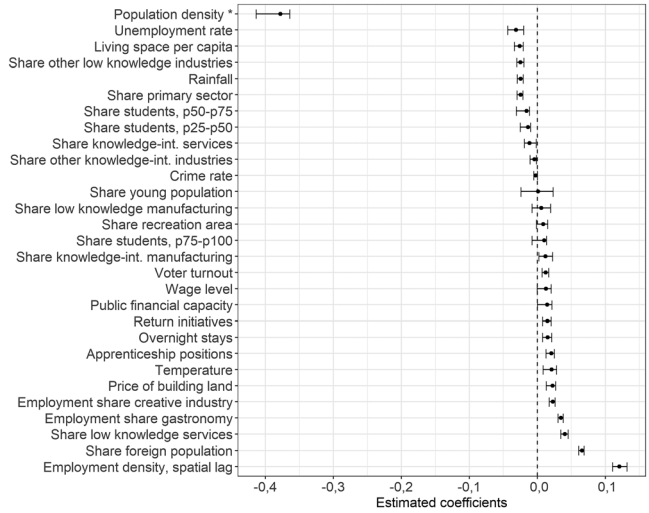

FIGURE A13 Variation of coefficient estimates – results of CSR for low-skilled workers. *Notes and source*: See Figure A9.

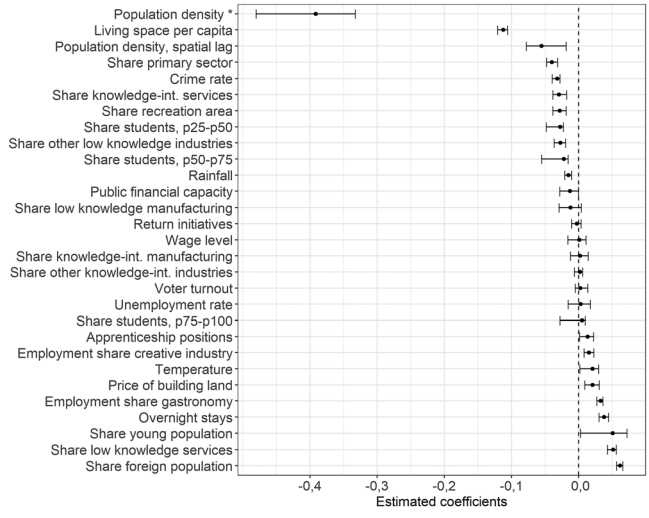

FIGURE A14 Variation of coefficient estimates – results of CSR for medium-skilled workers. *Notes and source*: See Figure A9.

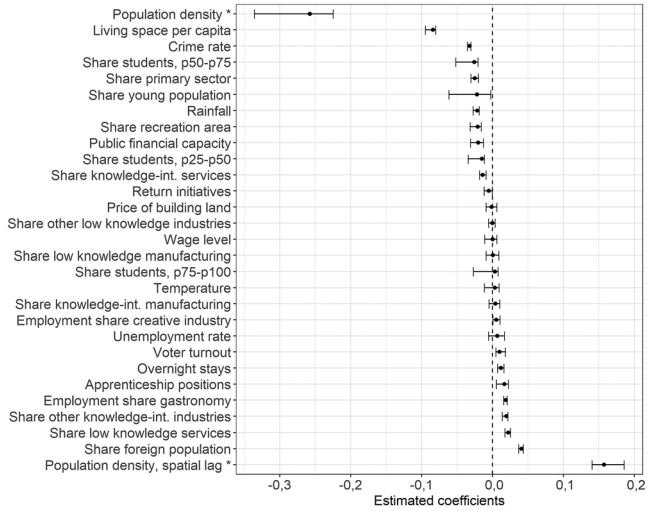

FIGURE A15 Variation of coefficient estimates – results of CSR for high-skilled workers. *Notes and source*: See Figure A9.

DOI: 10.1111/rsp3.12652

Resumen. Muchas regiones europeas experimentan actualmente un importante descenso de la población y, en relación con ello, se enfrentan cada vez más a la escasez de mano de obra. La migración es uno de los principales factores que impulsan los cambios en la oferta regional de mano de obra y en el nivel local de capital humano. Así pues, la capacidad de una región para atraer residentes es cada vez más importante para sus perspectivas de crecimiento. Con el fin de investigar la relación entre los atributos locales y el saldo migratorio de las regiones de Alemania se utilizó un extenso conjunto de datos de panel para el período 2003 a 2017. En concreto, se examinó si los factores que determinan el saldo migratorio de las regiones difieren significativamente según los grupos de edad y cualificación, ya que es probable que varíe su contribución al capital humano regional. La especificación econométrica desarrollada puede entenderse como una formulación agregada de un modelo de utilidad aleatoria de dos regiones. El conjunto de datos incluye 30 factores que podrían influir en el saldo migratorio de una región. Dado este gran número de variables explicativas y los significativos problemas de multicolinealidad, se aplicaron técnicas de aprendizaje automático [operador de reducción y selección mínima absoluta (LASSO, por sus siglas en inglés), regresión completa de subconjuntos] para identificar las características locales importantes. Los resultados apuntan a una sólida relación negativa entre la tasa neta de migración y la densidad de población, aunque las localidades próximas a los grandes centros urbanos parecen ser regiones de destino bastante atractivas, y la magnitud de los efectos difiere significativamente según los grupos de edad y cualificación. Además, las condiciones del mercado laboral y algunos servicios están correlacionadas significativamente con el saldo migratorio de la región. Sin embargo, las primeras y, en particular, las facilidades para la formación profesional, tienen importancia sobre todo para los trabajadores jóvenes.

抄録: 現在ヨーロッパの多くの地域では人口が大幅に減少しているが、これに関連して労働力不足の問題に直面する地域が増えている。移住は、地域の労働力の供給及び人的資本の地域レベルの変化を促す主な要因となっている。以上から、地域の住民誘致力の可能性は、地域の成長見通しにおいてますます重要になっている。2003~2017年までの大規模パネルデータセットを使用して、ドイツの各地域における地域属性と移住のバランスの関連性を検討した。年齢やスキルグループの人的資本への貢献度が各地域で異なる可能性が高いため、特に、地域の移住バランスを決定する要因が年齢やスキルグループによって大きく異なるかどうかを検討する。我々の計量経済学的モデルの仕様は、2領域のランダム効用モデルの集約的定式化であると解釈できる。データセットには、地域の移住バランスに影響を与える可能性のある30の要因が含まれている。この多数の説明変数と重大な多重共線性の問題を考慮して、重要な局所特性を特定するために機械学習の手法[最小絶対収縮および選択演算子(LASSO)、完全な部分集合回帰]を適用する。結果から、純移住率と人口密度には頑健な負の関連性があることが示されるが、大きな都市中心部に近い場所は、かなり魅力的な移住先地域とみられ、その影響の規模は年齢層及びスキルグループによって大きく異なる。さらに、労働市場の条件とアメニティの一部は、地域の移住バランスと有意な相関関係にある。しかし、労働市場の条件及び、特に職業訓練の施設は、主に若年労働者にとって重要である。