IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10734

Demographic Aging and Employment Dynamics in German Regions: Modeling Regional Heterogeneity

Thomas de Graaff
Daniel Arribas-Bel
Ceren Ozgen

APRIL 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10734

Demographic Aging and Employment Dynamics in German Regions: Modeling Regional Heterogeneity

Thomas de Graaff
Vrije Universiteit Amsterdam

Daniel Arribas-Bel
University of Liverpool

Ceren Ozgen
University of Birmingham,
Vrije Universiteit Amsterdam and IZA

APRIL 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Demographic Aging and Employment Dynamics in German Regions: Modeling Regional Heterogeneity*

Persistence of high youth unemployment and dismal labour market outcomes are imminent concerns for most European economies. The relationship between demographic ageing and employment outcomes is even more worrying once the relationship is scrutinized at the regional level. We focus on modelling regional heterogeneity. We argue that an average impact across regions is often not very useful, and that – conditional on the region's characteristics – impacts may differ significantly. We advocate the use of modelling varying level and slope effects, and specifically to cluster them by the use of latent class or finite mixture models (FMMs). Moreover, in order to fully exploit the output from the FMM, we adopt self-organizing maps to understand the composition of the resulting segmentation and as a way to depict the underlying regional similarities that would otherwise be missed if a standard approach was adopted. We apply our proposed method to a case-study of Germany where we show that the regional impact of young age cohorts on the labor market is indeed very heterogeneous across regions and our results are robust against potential endogeneity bias.

JEL Classification: J21, J61, J01

Keywords: demographic aging, employment, finite mixture models, self-organizing maps, youth cohorts, immigrant workers

Corresponding author:
Ceren Ozgen
University of Birmingham
Department of Economics and IRiS
JG Smith Building, Room G28
Edgbaston, B15 2TT
United Kingdom
E-mail: c.ozgen@bham.ac.uk

* Daniel Arribas-Bel and Ceren Ozgen gratefully acknowledge research funding by "Population Ageing and Regional Labour Market Development" project. An edited version of the paper will appear in Modelling Aging and Migration Effects on Spatial Labor Markets book by Springer-Verlag.

1 Introduction

Many European countries witness the persistence of high youth unemployment rates over decades despite a steady decline in the ratio of youth population to working-age population. The decrease in relative youth shares due to demographic aging did not improve the chronic unemployment of youth across Europe. Long-term unemployment rates are even more worrying, being record high after the 2008 financial crisis. The youth unemployment rate is twice as high as the overall unemployment rate in the EU where the aggregate statistics actually mask large differences between countries (e.g., Germany sees the lowest youth unemployment rate of 7 percent and Greece the highest with 50 percent).[2] These trends are contrary to the expectations that the decline in the supply of youth workers would lead to lower youth unemployment rates. Moreover, country level analysis makes mapping the relationship between demographic aging and labour market outcomes of the youth workers particularly more challenging, as there is a large variation across regions in the way they withstand demographic shocks. This is because a supply impact of youth population on employment and unemployment rates can be materialized in a number of ways. Youth population is typically unexperienced and lack specific information about labour markets. This not only leads to poor matching of workers with employers, but also increases on the job search due to skills mismatch of workers' qualifications with job-specific requirements. Therefore, an increase in the youth share would directly influence the employment opportunities available to others in the same age cohorts. Additionally, differently aged cohorts may impact each other to the extent of the substitutability between the workers, and between those in different skill groups (Biagi and Lucifora 2008). The magnitude of the impact would, however, depend on the degree of substitution among these groups. Labor market policies, economic downturns and business cycles, and rigidity of labour market institutions are other factors which intervene with this supply-demand adjustment.

Various studies already addressed the possible mechanisms through which the increase in youth share of total working age population may impact the employment opportunities of their own cohort while at the same time impact other age groups as well. Empirical evidence have repeatedly found cohort size to be an important determinant of (un-)employment. One of the very early studies by Bloom et al. (1988), by documenting the findings of 18 studies, launched a wide discussion on cohort size effects of the youth population on labour markets. They show that there is general agreement in the literature on the wage and employment impacts: entry of large cohorts of a certain age group adversely effected the wage and employment opportunities of the same cohort in relative terms. Korenman and Neumark (2000), in another influential work, extend the analysis of Bloom et al. (1988), by using panel data on 15 OECD countries for the period 1970–1994. They predict elasticities between youth unemployment and cohort size of around 0.5. A conflicting but influential result from the US case provided by Shimer (2001) presents much larger and *negative* impacts of large youth cohorts on both youth and adult unemploy-

[2] http://ec.europa.eu/social/main.jsp?catId=1036, Accessed on September 23rd, 2016

ment rates. His theoretical model rationales his findings such that, assuming labour market frictions, employee's on the job search behaviour will benefit himself, but also firms as the hiring costs are lower in younger labour markets. This will then consequently lead to job creation which would also improve employment of older workers. A significant contribution to emphasise from these conflicting findings is that in different country contexts not only the impacts can vary, but also the mechanisms through which youth cohorts alter employment prospects of their own and other cohorts.

A number of studies from a European perspective produced mixed results in favour of both Shimer (2001) and Korenman and Neumark (2000). For example, by using a long panel of population and (un-)employment data with ample information on age and gender groups, Garloff et al. (2013) show that labour market entry cohort size is an important determinant of employment and unemployment rates in Western Germany. In constrast to Shimer (2001) they find that small entry cohorts are likely to decrease the unemployment rates and small youth cohort entry increases the employment rates. Foote (2007) reports similar findings from the US labour market, while he demonstrates that the findings are sensitive to correcting for spatial correlation. The cross-sectional dependencies should be taken into account, as the regions' response to demographic shocks can be rather similar based on commuting or similarities in labour market structure of the adjacent regions. Skans (2005) points out contrasting findings from Sweden such that youth workers benefit from being in labour markets with large youth cohorts, where his findings confirms those of Shimer (2001). Biagi and Lucifora (2008) extend these analyses by introducing the role of education. In the period 1975–2002 in European countries, they disaggregate the data by education level and cohorts to analyse whether the unemployment rates are impacted differently by cohort size and education shocks simultaneously. A significant point they raise is the importance of demand in accommodating demographic and education shocks and the imperfect substitution between different skill groups. In advance economies, a demographic shock, for example a higher share of more educated workers can be accommodated better, if it coincides with a positive aggregate demand for skilled labour. They indeed show that higher educated and adult workers experience lower unemployment rates. Finally, Moffat and Roth (2013) study how the probability of being unemployed changes with the nationally and regionally defined age-cohort size. They use a more flexible (wider) definition to identify the age cohorts and utilise data from the European countries. They report that once the analysis is conducted at the regional level rather than the national level, the age-cohort size effect on the probability to become unemployed is stronger. Subsequently, the studies show a large hetefogeneity in a number of dimensions from spatial scales to methods, from characteristics of the labou rmarkets to characteristics of the youth cohorts.

A shortcoming of the literature has been its inadequacy to reconcile with variation of findings at differing level of spatial aggregation. Furthermore, though many studies are conducted at the regional level, it has still been not clear how regions' heterogeneous responses to demographic aging should be taken on board in policy making. An original contribution we make to this literature is to study regional heterogeneity, while showing regions even within the same country can be differently impacted by the same demographic shocks. In other words, the focus is on the possible variation in regions' responses to the changes in the relative share of the youth population. Theoretically, each region within a country can react uniquely to a supply shock of labour, while it is plausible to assume a certain degree of generality among sub-group of regions considering the similarity in production structure, location characteristics and demographic attributes. Our innovative methodological approach which employs a latent class analysis combined with self-organizing maps (SOM) displays a powerful segmentation and analysis of these sub-groups of homogeneous regions that show a similar pattern towards supply of youth population share in working-age workers. The next section discusses how we can model and map out regional heterogeneity, and especially how we can interpret finite mixture output using self-organizing maps. Section 3 provides an application of our proposed techniques by looking at the impact of youth shares on employment rates in Germany. The final concludes and offers suggestions for further research.

2 Modeling Regional Heterogeneity

The standard approach to model regional heterogeneity is to apply a fixed effects model where each region is modeled with its own level effect. In subsection 2.1 we discuss this approach and argue that the standard fixed effects model can be easily extended by a model with varying slope parameters— sort of like a multilevel model— if there are repeated observations over time. Varying slope parameters are appropriate when unobserved variables interact with the independent variables. For example, the impact of regional population growth on regional GDP growth might interact with the educational level of the regional population. Such a modelling approach has two large disadvantages. First, the estimation produces inefficient and usually inconsistent parameter estimates. Namely, most fixed effects (and most slope parameters for that matter) are not statistically different from each other and it is well known that when the time period is relatively short (as is usually the case) fixed effects suffer from an inconsistency problem. Secondly, if one is interested in what drives the underlying (regional) heterogeneity, then using a fixed effects approach is not appropriate as well as the fixed effects are typically discarded from the analysis.[3] In subsection 2.2 we deal with these disadvantages by employing a finite mixture model (henceforth as well denoted as FMM), a latent class cluster analysis that enables us to group region in clusters with similar parameter estimates. This is similar to the method of spatial regimes, although with a finite mixture estimation we are not restricted to assign regions *exogenously* into groups. One general drawback of cluster analysis, is that the resulting clusters are hard to interpret and to visualise. Therefore, we finally apply in subsection 2.3 a self-organising map ap-

[3] One can apply second-stage models, where the first stage estimates the fixed effects and the second stage analyses the determinants of those fixed effects. However, note again, that this is only an analysis on the levels and not on the slopes.

proach which allows us to display the varying multivariate regional characteristics in geographic space (see, e.g., Spielman and Folch 2015).

2.1 Regional heterogeneity

We start by assuming that one is interested in the effect of a regional input variable x on a regional output variable y and that she has repeated observations over a set of regions. Then a straightforward and intuitive appealing model would be the following linear regression model:

$$y_{rt} = \beta_0 + \beta_1 x_{rt} + \varepsilon_{rt}, \tag{1}$$

where r denotes the region ($r \in 1, \ldots R$), t the year ($t \in 1, \ldots T$) and ε an i.i.d. error term. The parameter β_0 denotes a level effect and β_1 is our parameter of interest and gives the marginal effect of x on y or $\frac{dy}{dx}$.

If there is another regional variable z that is correlated with both y and x, then our estimation of β_1 is *biased*, or $\mathbf{E}(\hat{\beta_1}) \neq \beta_1$ (see, e.g., Stock and Watson 2003). If z is known then including z in model (1) removes the bias. Unfortunately, z is very often not known or difficult to measure. However, by assuming that z enters the model *linearly* and does not vary over time—thus as $y_{rt} = \beta_0 + \beta_1 x_{rt} + \gamma z_r + \varepsilon_{rt}$—, one can control for z by using the following fixed model:

$$y_{rt} = \beta_{0,r} + \beta_1 x_{rt} + \varepsilon_{rt} \tag{2}$$

Here, $\beta_{0,r}$ now controls for all variables (including z) which enters the model linearly and do not vary over time.

However, the unobserved variable z might as well *interact* with the impact of x on y, so that in the most simplified version the model in fact reads as: $y_{rt} = \beta_0 + \beta_1(z_r \times x_{rt}) + \gamma z_r + \varepsilon_{rt}$. Given that there are repeated regional observations the model can now be estimated as:

$$y_{rt} = \alpha_{0,r} + \beta_{1,r} x_{rt} + \varepsilon_{rt}, \tag{3}$$

where both the level and the slope parameter is regional specific. Note that this is rather data demanding. Given that there are R regions, one need at least $2R+1$ observations. Or the number of years should be at least 3 when there is a symmetric panel. Moreover, for both *consistent* level and slope parameters, the temporal dimension should be sufficiently large (say more than 40 time periods). Usually, the latter is not the case. Therefore models such as (3) are seldom applied, although the possibility of significant regional heterogeneity in the effect parameter $\beta_{1,r}$ is widely recognized.

One way to overcome this inefficiency is a cluster analysis in the form of multivariate mixture model, which is dealt with in the next subsection.

2.2 Regional finite mixture modelling

Instead of estimating separate parameters for each region, it is far more (statistically) efficient and even consistent to estimate separate parameters for *groups* of regions. Assume that there are c more or less homogeneous groups of regions, then model (3) becomes:

$$y_{rt} = \beta_{0,c} + \beta_{1,c} x_{rt} + \varepsilon_{rt}, \tag{4}$$

where the number of parameters now amount to $c \times k$, with c being the number of groups and k the number of parameters.

To do so, we adopt a finite (or multivariate) mixture modelling approach. A statistical technique which became especially popular since the 1990s in marketing (a seminal contribution is Wedel and Kamakura 2012), but since then permeated in other economic fields, although mainly applied in the econometric realm (see, e.g., Deb et al. 1997, Arcidiacono and Jones 2003, Alfo et al. 2008) and in tackling heterogeneity in discrete choice modelling (Greene and Hensher 2003).[4] The approach works as follows.

We divide our sample of regions into an, a priori unknown, number of subsamples (clusters) of regions. So we assume that our sample consists of a mixture of C clusters, with proportions $\pi_1, \dots, \pi_C$. We can now decompose the density function of y conditional on the parameter vector β as follows:

$$f(y|\beta) = \sum_{c=1}^{C} \pi_c f(y|\beta_{\mathbf{c}})). \tag{5}$$

To estimate model (5) usually the expectation maximization (EM) procedure as introduced by Dempster et al. (1977) is applied. It starts by introducing a latent variable, u_{rc}, denoting whether region r belongs to cluster c. Thus:

$$u_{rc} = \begin{cases} 1, & \text{if region } r \text{ belongs to cluster } c \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

We then assume the following distribution for u_{rc}:

$$f(\mathbf{u}_r|\pi) = \prod_{r=1}^{R} \pi_r^{z_{rc}}, \tag{7}$$

with $\mathbf{u}_r$ a vector of (6) for each cluster, c. Let $\mathbf{U}$ denote the matrix of all $\mathbf{u}_r$ then the complete log-likelihood can now be written as:

[4] Interestingly, the underlying algorithm and implementation where only the constants α_c are allowed to vary over groups is heavily applied in labour economics by, e.g., Lancaster (1992), Munch et al. (2006), and De Graaff and Van Leuvensteijn (2013), usually in a multivariate setting where the constants α_c are then argued to remove unobserved heterogeneity.

$$\ln[\mathscr{L}(\beta|y,\mathbf{U})] = \sum_{r=1}^{R}\sum_{c=1}^{C} u_{rc} f(y|\beta_{\mathbf{c}}) + \sum_{r=1}^{R}\sum_{c=1}^{C} u_{rc}\ln(\pi_c). \quad (8)$$

The EM algorithm now proceeds as follows:

E-step : We estimate the cluster probabilities for each region, so that the probability that region r belongs to cluster c is:

$$\hat{u}_{rc} = \frac{\pi_c \prod_{c=1}^{C} f(\ln(y_{rc}|\beta_c))}{\sum_{g=1}^{C} \pi_g \prod_{c=1}^{C} f(\ln(y_{rc}|\beta_g)}. \quad (9)$$

So, in this step, all $\hat{u}_{rc}$ are estimated so that they can be used in the log likelihood given by (8).

M-step : First, we need to derive the proportions π_c by applying the equality: $\hat{\pi}_c = \frac{1}{R}\sum_{r=1}^{R}\hat{u}_{rc}$. Using now both $\hat{u}_{rc}$ and $\hat{\pi}_c$ enables us to estimate β by using (5) and conventional likelihood procedures. We repeat the E- and M-step until the log likelihood (5) stops improving.

Although the EM algorithm is computationally cumbersome it is parsimonious as well. It results in a parameter estimation of C endogeneously and robustly formed clusters. However, cluster analyses are notoriously difficult to interpret and analyse because one not only gets a different set of regression parameters for each cluster, but a full set of probabilities of each region belonging to each cluster. To help visualize, explore, and clarify the clustering outcome we therefore propose to use a self-organising maps as a novel approach for the interpretation of the output of an FMM regression, hence facilitating the understanding of complex regional characteristics in geographic space.

2.3 Interpreting Mixture Modeling Output with Self-organising maps

A self-organising map (SOM, Kohonen 2001) is a kind of computational neural network that is able to simultaneously reduce the number of dimensions (*projection*) as well as observations (*quantization*) in a multidimensional data set. Although the mathematics underlying the algorithm are more intricate, the intuition is relatively straightforward. Given that, in this context, it is only required to be able to interpret its output, not necessarily the mechanism by which the algorithm reaches it, this is what we will focus on.[5] The essence of the SOM is to translate the statistical properties of the original dataset (Ω) onto a network of interconnected neurons represented by a two-dimensional grid of hexagons (H). Each of these neurons has a vector of as many dimensions as Ω whose values, after the process of *training* the

[5] For a detailed explanation of the underlying learning mechanism and its implementation, please refer to Kohonen (2001) The analysis in this paper was carried out using the `kohonen` library in the statistical software platform R (Wehrens et al. 2007).

network, capture the statistical variation contained in Ω. The power of the SOM resides in the fact that, once the network has been trained and its neurons have *learned* the properties of the original dataset, it is possible to map the original observations onto the network. Because H preserves information topologically, statistical similarity is turned into spatial relationships. This allows to represent multi-dimensional relationships and make intuitive comparisons through a visual display that maps the original observations to the network's neurons. In a sense, this property of the SOM is akin to other projection techniques such as Principal Component Analysis (PCA) or Multi-Dimensional Scaling (MDS), with the advantage that the output space onto which the data is projected—the network H—is limited and known. Additionally, the non-parametric and learning nature of the SOM algorithm has been shown to be more robust when it comes to capturing complex, non-gaussian relationships (Yan and Thill 2009). In the context of this paper, we use the SOM to explore the distribution of the probability that each German region belongs to each of the clusters specified by the FMM model. It is important to note that this is in essence a multi-dimensional dataset: we have several probabilities associated to every region. FMM output returns probabilities for each region to belong to each cluster (though in some cases probability for a region can be zero in well segmented distributions). For a given region r, the probability of belonging to different clusters can vary in magnitude. This makes visualizing it at once difficult simply because regions may belong to multiple clusters at the same time, while with varying probabilities. The usual approach in the literature to work around this challenge is to implicitly reduce the dimensionality to a single one, the cluster for which every region displays a highest probability of belonging. In other words, this is equivalent to "rounding up" the highest probability of each region to one, and setting all the others to zero, then focusing only on the former one. Although convenient this approach implies simplifying the FMM output greatly and it imposes an artificial degree of certainty about each region's cluster membership. In cases where the set of probabilities are not very far from this case (i.e., when there is only one cluster with a high probability and all the others are negligible), this assumption is reasonable and valid. However, in cases where the situation is less clearcut, this can be a problem, and neglecting the nuances of the distribution of probabilities can lead to incorrect interpretations. In this context, our approach is to feed the set of probabilities (a matrix of R rows and C columns) to the SOM algorithm and use its output to explore how each region r relates to others when it comes to membership to each of the clusters identified by the FMM. This is articulated through the visual display the SOM offers. By plotting in a single graph the similarity between regions based on their probabilities, *as well as* the cluster each region would have received under the traditional methodology, our approach will enable the exploration of nuances and cases where cluster membership is not a clearcut decision. This approach is novel and produces results that help the interpretation of an otherwise complex and obscure output.

3 Empirical application: Ageing in Germany

3.1 Data

The empirical application of this paper focuses on the impact of regional youth share on employment in German regions. Furthermore, it explores the degree of variation of this impact across the regions. To this end, we have collected repeated observations of regional employment shares—the total number of the employed relative to the population aged between 18 and 64 —and youth shares— the number of individuals aged between 18 and 24 relative to the population in the age group 18–64—in Germany.

The data we employ has some distinctive attractive features. First of all, it presents a geographically complete picture of employment and aging dynamics in Germany by covering both Eastern and Western regions in our empirical analysis. Secondly, we use labour market areas defined on the basis of commuting distances, meaning each region represents a self-contained labour market area. Labor market areas can be formed by one or multiple districts (kreise) which are equivalent to a NUTS 3[6] level region. The demarcation of the regions is in line with the definition provided by Kosfeld and Werner (2012). The employment data is obtained from the Institute for Employment Research (IAB), and population variables are constructed by using German Statistical Office data.

Our analysis include 141 labour market areas in the period 2000–2010. Due to a number of restrictions on data availability our sample period is confined to a decade. We control for possible sectoral demand-side shocks regions face by including a measure (the Bartik index)[7] which resorts on occupation data which is broken down on the basis of the complexity of tasks required in each occupation. Using task complexity levels rather than the standard yet very broad sector division has the advantage of properly accounting for the common nature of tasks, that cuts across the sectors and the respective demand conditions. Unfortunately, the occupation data that we use to construct the Bartik index is not available after 2011.

Figure 1 displays the geographical distribution of both employment rates (5a) and youth shares (5b) across German regions. Clearly, there is much variation in regional employment levels ranging from more than 60% in Frankfurt and München to less

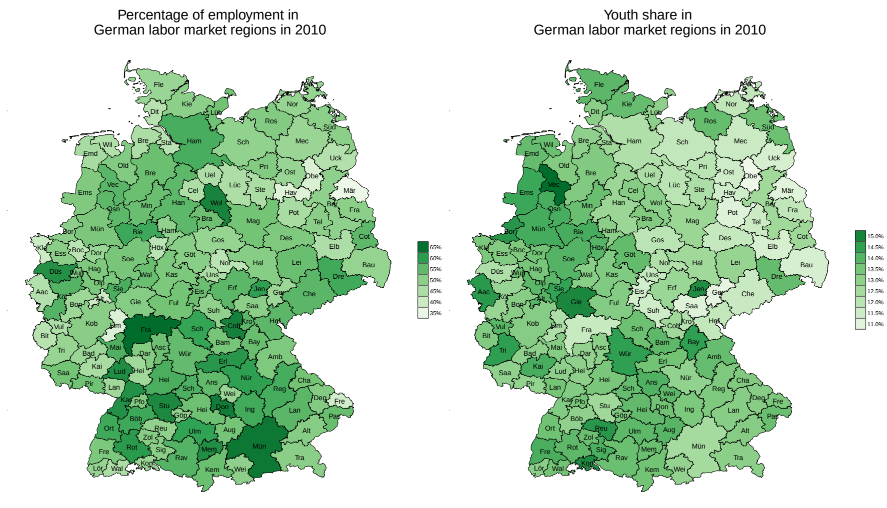
---

[6] Nomenclature of Territorial Units for Statistics

[7] The particular measure we employ reads as:

$$\hat{L}_{rt} = \sum_{k} \left[\frac{E_{k,t}}{E_{k,t-1}} E_{rk,t-1} \right], \tag{10}$$

where $\hat{L}$ is the weighted sum of employment across all sectors k in region r in and year $t-1$, with the weights being given by the rate of sector-specific employment in year t and year $t-1$ at the national level. As such, this variable represents the level of employment in region r that is predicted for the case in which employment in each sector grows at the same rate as the corresponding sector at the national level indicated by E. This variable is used as an exogenous measure for demand changes for labour.



(a) Employment rate (b) Youth share

Fig. 1 Employment rate and youth share across German regions

than 40% in regions as Freiburg and Oberhaven. In general, employment rates are higher in Southern Germany (with some exceptions) and lower in Eastern Germany. There is much less variation in youth shares which range from 11% to 15%. The lowest youth shares in Germany can be found in Eastern Germany (and Frankfurt) and the regions that host the highest share of youth is in Western Germany.

3.2 The generic impact of youth shares on regional employment rates

We start our analysis by first applying our linear model (1) in natural logarithmic form as follows:

$$\ln(E_{rt}) = \beta_0 + \beta_1 \ln(YS_{rt}) + \beta_2 B_{rt} + \nu_t + \varepsilon_{rt}, \tag{11}$$

where E_{rt} defines the employment rate in region r, measured as the number of workers between 18 and 64 years old, divided by the working population between 18 and 64 years old. YS denotes the youth share in region r, and it is defined as the number of individuals between 18 and 25 years old divided by the working population between 18 and 64 years old in the same region. B is the Bartik index which controls for regional specific demand effects and ν_t denotes a vector of year specific effects. Model 1 in Table 1 provides the estimation results of the model (11). Clearly, there is a statistically significant negative impact of youth share on employment rate, where

Table 1 Generic impact of youth share (18–64) on log(employment rate)

	log(employment)			
	OLS	First-diff	First-diff	First-diff IV
	Model 1	Model 2	Model 3	Model 4
log(Youth share)	-0.202^{***}	-0.153^{***}	-0.174^{***}	-0.255^{***}
	(0.060)	(0.018)	(0.023)	(0.004)
Bartik index	0.058^{***}	0.634^{***}	0.260^{***}	0.285^{***}
	(0.007)	(0.023)	(0.030)	(0.021)
Region first-differenced	No	Yes	Yes	Yes
Time dummies	No	No	Yes	Yes
N	1,410	1,269	1,269	1,269
R^2	0.272	0.561	0.755	0.748

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

in this model the elasticity is around -0.2. Allowing for regional fixed effects by estimating model (2) leads to Models 2 and 3 of Table 1. [8]

Youth shares still have a significant negative impact on regional unemployment rates, but the size of the elasticity increased slightly to around -0.15. Thus, allowing for regional heterogeneity that has a linear impact on employment rates has a moderate impact on the size of the estimate.

An important issue is the potential endogeneity that would bias the OLS estimation. Young people are likely to sort systematically to regions with better employment opportunities. We address this by instrumental variables (IV) estimation. Our identification relies on predicting the size of youth cohort that currently resides in labour market areas in our sample with the young cohort lagged by 15 years. So our instrument is the log of the number of individuals aged between 3 and 9 relative to the population in age groups 3–49, 15 years prior to the study period. It is likely that the cohort 15 years past cannot be attracted by economic opportunities of today while the size of the age groups of past is likely to strongly correlate with the size of the age groups of today. The first-stage statistics strongly confirm our expectation. Model 4 of Table 1 displays the IV regression. The result is in line with the impact found in OLS estimations; youth share has a negative and statistically significant impact on employment rate at 1 percent level. The magnitude of the predicted coefficients is fairly similar, which reduces our concerns for sorting of youth.

In addition, although we correct for demand side effects, one might be concerned with the possibility that youth cohorst are influenced by employment rates, mostly by interregional migration. Using instrumental variables, we reconfirm Garloff et al. (2013) and show that the potential endogeneity bias (where one of which underlying

[8] Because of strong temporal autocorrelation in both the employment rates and the youth shares, we estimate model (2) by applying first differencing. Thus, we estimate: $\ln(E_{r,t}) - \ln(E_{r,t-1}) = \beta_1(\ln(YS_{r,t}) - \ln(YS_{r,t-1})) + (\varepsilon_{r,t} - \varepsilon_{r,t-1})$. Although less efficient than the usual within estimator, first differencing requires less strong identification assumptions. For the linear model, this should only affect the standard errors and indeed, both fixed effects estimation strategies lead to similar results. It matters however for the clustering analysis.

drives can be migration of young employees to highly prosperous areas) does not overturn our findings neither in terms of magnitude nor significance of the estimated relationship.[9]

Combining clustering and instrumental variable estimations is still however cumbersome. Therefore we do not focus on IV estimations within FMM context, given the IV-panel estimation results of above. Next, we assess whether unobserved variables might actually interact with the impact of youth share.

3.3 Region slope parameters

The results of an estimation of model (3) are depicted in Figure 2. So, every region now is associated with a value of $\beta_{1,r}$. Taken at face-value, there is an enormous regional variation in $\beta_{1,r}$, ranging from -1 to 1. Unfortunately, most of these estimates are not statistically significant and even not consistent, given that the total number of time periods T per region r is 10. So, every $\beta_{1,r}$ is based on 10 observations, which is too few to comply with the usual properties of ordinary least squares.

Clearly, however, there is evidence that there is a large regional variation in the impact of regional youth shares on employment rates. To reveal this spatial pattern in a consistent manner, we therefore resort to estimating our clustering model of (4).

3.4 Finite Mixture Results

The exact model we estimate deviates slightly from (4) and boils down to:

$$\ln(E_{rt}) = \beta_{0,r} + \beta_{1,c}\ln(YS_{rt}) + \beta_{2,c}B_{rt} + \nu_{t,c} + \varepsilon_{rt}, \qquad (12)$$

where t denotes a vector of year dummies. Note that this is a very exhaustive model, where we allow for regional fixed effects with $\beta_{0,r}$ and cluster specific impact of youth share $\beta_{1,c}$, regional demand effects $\beta_{2,c}$ and year effects $\nu_{t,c}$. Before we estimate an FMM, we first difference our data and effectively remove the regional fixed effects.

The number of clusters or components in finite mixture modelling is determined by the researcher herself. However, using information criteria the optimal number of clusters from a statistical point of view can be assessed. Figure 3 provides three information criteria: the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the integrated classification likelihood (ICL) criterion. The

[9] Sander (2014) points out that the internal migration patterns in Germany have been predominantly within East Germany, while significant trends to urban cores from nearby suburban areas as well as metropolitan hinterlands during our study period. At the same time young adults with families out-migrated to urban agglomerations in many non-metropolitan cities. We expect that using labour market areas which includes daily commuting patterns and FMM for our analysis to some extent should tackle with potential bias internal migration patterns might cause.

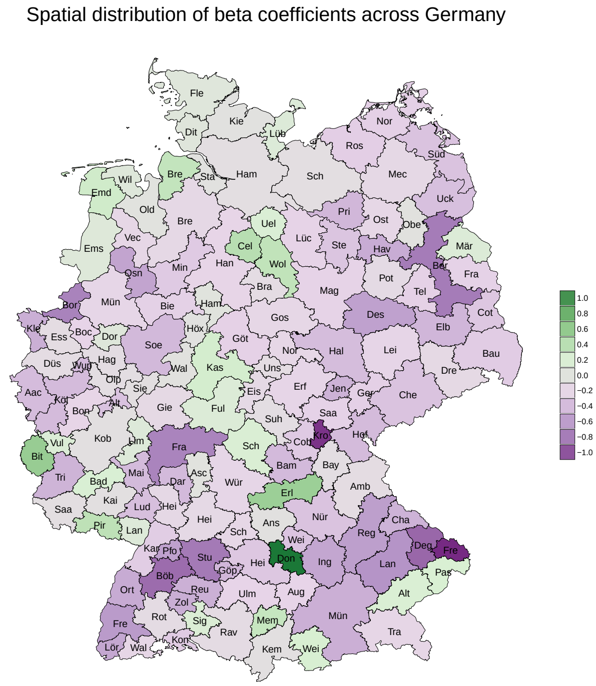


Fig. 2 Regional specific varying impacts of the youth share on the employment rate

latter two are well known to 'punish' the criterion severely for the number of parameters used. Strikingly, according to the BIC and the ICL the optimal number of clusters is 2 (in our case, these are the largest city regions versus the rest). According to the AIC the optimal number of clusters is 8. For illustration purposes, we choose to settle in the middle and opt for 4 clusters.

When allowing for four clusters we get the estimation results of model (12) as displayed in Table 2.

As we focus on the the impact of the youth share on employment rate, we see that the impact differs from -0.33 (cluster 2) to -0.06 (cluster 3). To visualise these clusters, Figure 4 displays the cluster with the largest probability for each region. Clearly, there is spatial autocorrelation except for cluster 2. This cluster contains the largest and most important cities of Germany, including Berlin, Hamburg, Munich and Frankfurt. Cluster 1 is formed predominantly by clusters in the northern and western part of the cluster (and some in the periphery). Cluster 3 displays mostly regions in Eastern Germany and some in the periphery and cluster 4 is a very distinctively southern Germany and Ruhr area phenomenon. This spatial autocorrelation is most likely caused by spatial unobserved heterogeneity where variables as local institutions, history and sector structure might play an important role. Note that the impact of youth shares is statistically similar for these clusters 1 and 4. The differ-

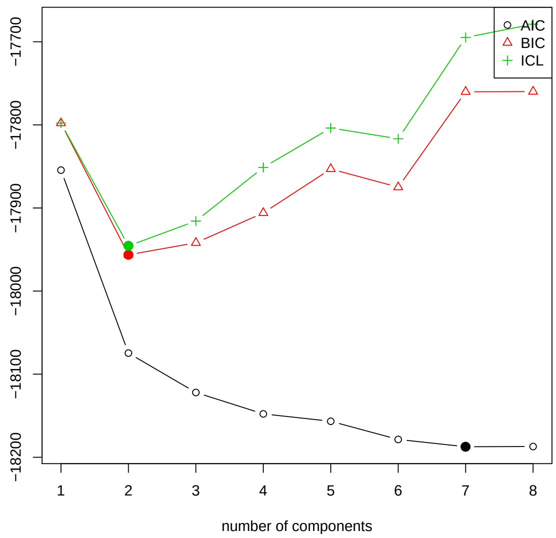


Fig. 3 Number of clusters and information criteria

Table 2 Finite mixture estimation results (dependent variable: ln(Employment rate))

	Cluster			
	1	2	3	4
ln(Youth share)	-0.206^{***}	-0.329^{**}	-0.056^{*}	-0.258^{***}
	(0.033)	(0.108)	(0.026)	(0.033)
Bartik index	0.184^{***}	0.272^{**}	0.277^{***}	0.336^{***}
	(0.060)	(0.085)	(0.042)	(0.040)
Region fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
N	1,269	1,269	1,269	1,269

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

ence is formed by the Bartik index, where cluster 4 seems to be more affected by regional demand effects.

As FMM works each region received a probability to 'belong' to a cluster. Typically, these probabilities are close to 0 or 1 (about 65% of the regions have a dominant probability larger than 0.8). However, some regions are more difficult to classify and have significant probabilities for two clusters or more (in this case, usually for cluster 1 and 4). To visualise these probabilities, the next subsection applies a SOM analysis.

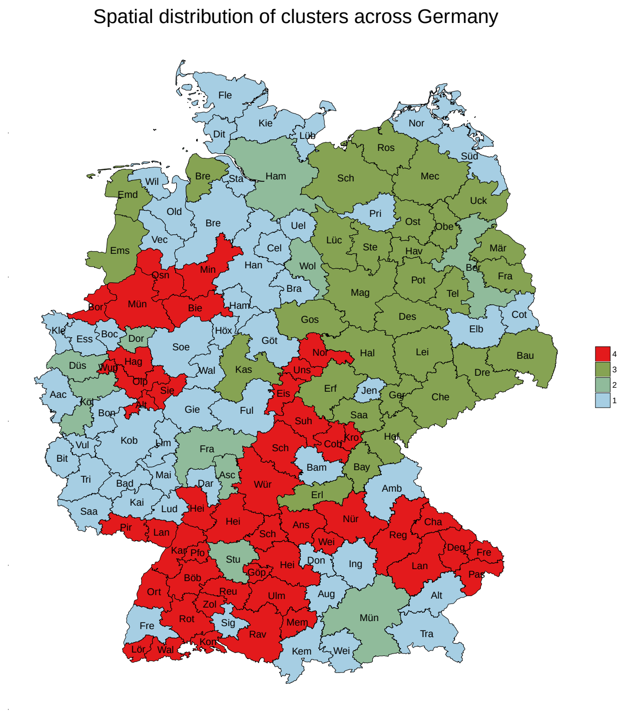
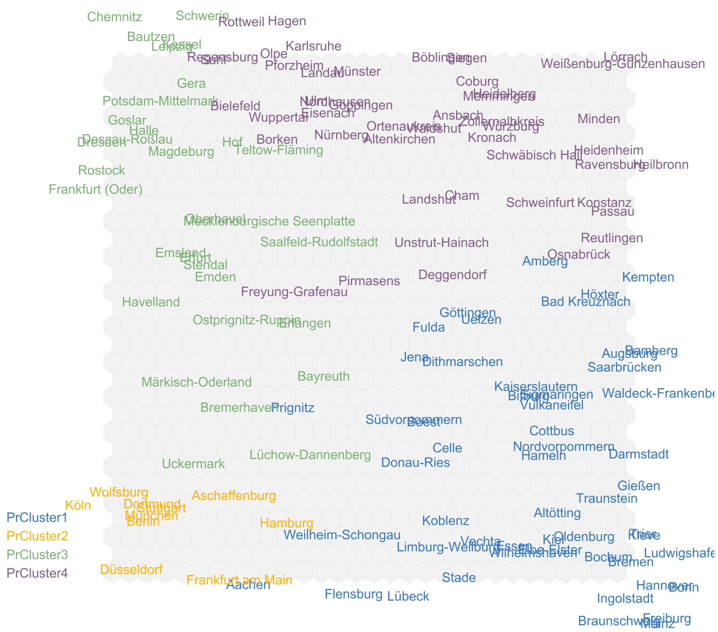


Fig. 4 Clusters in Germany of the impact on employment rates

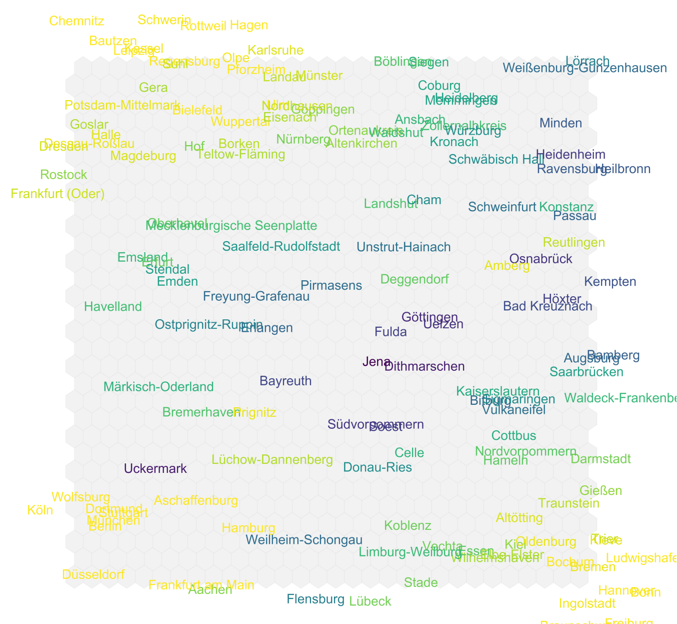
3.5 *A Self-organising Map of Ageing regions in Germany*

Figure 5 (a) displays the SOM output. We use 900 neurons laid out as a 30 by 30 grid in order to allow the 141 regions to spread freely within the output space. The figure presents the location where the region has been assigned in the network although, for visibility regions, we slightly alter their location randomly. As explained above, the properties of the SOM algorithm imply that the location of each observation in the network is representative of its statistical attributes. Thus, being similar in the attribute space translates into being located nearby in the SOM space. This means that, for instance, Chemnitz and Bautzen, both regions located in the top-left corner of the SOM can be assumed to have a very similar distribution of probabilities of belonging to each clustering. Equally, both can also be assumed to have a rather different profile than Ingolstadt, in the bottom-right corner.

The name of each region is coloured using a scheme that follows the traditional approach of assigning each observation into the cluster with largest probability (and that is displayed geographically in Figure 4). As expected, most regions that the traditional approach would group into the same cluster locate in the same part of the SOM. This allows for an additional advantage of the SOM: the clusters themselves can be explored further by considering their distribution across the network. For



(a) Cluster memberships



(b) Stability of membership

Fig. 5 SOM of probabilities

example, cluster 2 (yellow) is more similar to cluster 4 (purple) than to cluster 3 (green).

The most interesting aspect of using a SOM to interpret results from a FMM model comes when we consider cases where the probability profile is not as clear-cut as we would like it to be. For example, Prignitz (in the middle of the network) has a probability of 0.48 of belonging to cluster 1 but also one of 0.46 of belonging to cluster 3. As such, it is coloured as cluster 1 (blue), but located close to many others labeled as cluster 3 (green). This degree of nuance and detail is completely eliminated when we adopt the traditional approach, as we would only observe the region is assigned into cluster 1. However, the SOM is capable of representing it in an intuitive and useful way, allowing the researcher to explore the FMM output much more richly.

A complementary way to understand the value of the SOM in this context is provided in Figure 5 (b), which colors the name of each region on a yellow to purple gradient by the difference in probability of belonging to the two most likely clusters. In other words, a region in yellow features a large difference between the probability of belonging to its most likely cluster and the next one (i.e. the difference between the largest and second largest probabilities for that region is large). In contrast, a region in purple displays a relatively high probability of belonging to more than one cluster and can thus be considered as "on the border". On-the-border regions, which we could only identify through combining FMM with SOM, are those that are impacted by a youth shock similarly while in the FMM output they appear in completely different clusters. This approach helps highlight regions for which the traditional simplification of selecting the cluster of largest probability is a valid approach (regions in yellow) and those for which a significant amount of information may be lost with such simplification (regions in purple).

4 In Conclusion

Demographic aging is a significant concern for many developed countries. In order to correctly address the associated problems, it is crucial to properly identify the needs of the labour markets. In this paper we focus on the impact of youth on regional employment rates in 141 German labour market areas. We show that fixed effects models with varying β coefficients are not adequate to reflect and handle the heterogeneity. By employing an innovative methodology that is combining a latent class analysis with self-organising maps, this research depicts a great deal of variation of how ageing impacts employment opportunities of the working age population at the regional level. Although our OLS and IV estimations are in line with previous research on German local labour markets (e.g., with Garloff et al. (2013))—although contradicts with the work of Shimer (2001), we offer further explanations for a number of issues not yet addressed in the literature.

First, although the OLS predicts an aggregate elasticity for the youth impact on employment rate around -0.2, we find that the elasticities actually vary in the order of -0.06 to -0.33 in 4 types of regions in Germany. FMM analysis partitions Germany into four unique clusters of broader regions, namely metropolitan areas, Southern Germany, West Germany with industrial core and finally Eastern Germany (excluding Berlin). Second, we show that the labour market areas with highest employment rates are hurt the most by a demographic shock of youth share. These are large metro regions like Munich, Frankfurt, Düsseldorf, Stuttgart, Hamburg and Berlin. In contrast, eastern German labour markets which experience significantly large internal migration within, are those which are affected the least from increasing youth population. Note that however, as a result of FMM estimation, typically some regions receive probabilities that allow them to be assigned in more than one of these clusters. SOM helps visually mapping the distribution of these probabilities across all the regions in the analysis. Therefore, this extension allows us to interpret the FMM output in further detail and to identify regions for which a single cluster membership might not reflect the output of the model. Through this approach we are able to exactly pinpoint which regions better embody the characteristics of the cluster and which ones are found to be "on the border" between two clusters. Finally, our results also show that regions exhibit different levels of resilience to regional demand shocks.

Our results imply that policy challenges for demographic ageing require to look beyond a country as a whole. Given the extension in longevity does not meet increase in active labour period, an economic perspective taking regional labour market heterogeneity into account is crucial. Our results are suggestive for policy-makers to consider the possible impacts of aging on employment opportunities in varying regional economic contexts both for adults and also for the youth. As shown in our analysis local resources, regions sector structure and characteristics of the local youth workforce are important factors to influence employment rate.As shown in our analysis local resources, region's sector structure and characteristics of the local youth workforce are important factors to influence employment rate. We hope that our methodological application sheds some light on the contrasting empirical findings in the literature and opens new avenues for research to analyse further the determinants behind the differing impacts found, possibly based on the economic character of the labour markets. On the methodological side, this application can be useful for studies trying to uncover a range of issues where there is significant underlying heterogeneity in a number of dimensions of locations, workers, firms and regions.

References

Alfo, M., Trovato, G. and Waldmann, R. J.: 2008, Testing for country heterogeneity in growth models using a finite mixture approach, *Journal of applied Econometrics* **23**(4), 487–514.

Arcidiacono, P. and Jones, J. B.: 2003, Finite mixture distributions, sequential likelihood and the em algorithm, *Econometrica* **71**(3), 933–946.

Biagi, F. and Lucifora, C.: 2008, Demographic and education effects on unemployment in europe, *Labour Economics* **15**(5), 1076–1101.

Bloom, D. E., Freeman, R. B. and Korenman, S. D.: 1988, The labour-market consequences of generational crowding, *European Journal of Population/Revue Européenne de Démographie* **3**(2), 131–176.

De Graaff, T. and Van Leuvensteijn, M.: 2013, A european cross-country comparison of the impact of homeownership and transaction costs on job tenure, *Regional Studies* **47**(9), 1443–1461.

Deb, P., Trivedi, P. K. et al.: 1997, Demand for medical care by the elderly: a finite mixture approach, *Journal of applied Econometrics* **12**(3), 313–336.

Dempster, A. P., Laird, N. M. and Rubin, D. B.: 1977, Maximum likelihood from incomplete data via the em algorithm, *Journal of the royal statistical society. Series B (methodological)* pp. 1–38.

Foote, C. L.: 2007, Space and time in macroeconomic panel data: young workers and state-level unemployment revisited.

Garloff, A., Pohl, C. and Schanne, N.: 2013, Do small labor market entry cohorts reduce unemployment?, *Demographic Research* **29**, 379.

Greene, W. H. and Hensher, D. A.: 2003, A latent class model for discrete choice analysis: contrasts with mixed logit, *Transportation Research Part B: Methodological* **37**(8), 681–698.

Kohonen, T.: 2001, *Self-organizing maps, vol. 30 of Springer Series in Information Sciences*, ed: Springer Berlin.

Korenman, S. and Neumark, D.: 2000, *Cohort Crowding and Youth Labor Markets: A Cross-National Analysis*, number January, University of Chicago Pres.

Kosfeld, R. and Werner, D.-Ö. A.: 2012, Deutsche arbeitsmarktregionen–neuabgrenzung nach den kreisgebietsreformen 2007–2011, *Raumforschung und Raumordnung* **70**(1), 49–64.

Lancaster, T.: 1992, *The econometric analysis of transition data*, number 17, Cambridge university press.

Moffat, J. and Roth, D.: 2013, The cohort size-wage relationship in europe, *Technical report*, Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics (Volkswirtschaftliche Abteilung).

Munch, J. R., Rosholm, M. and Svarer, M.: 2006, Are homeowners really more unemployed?, *The Economic Journal* **116**(514), 991–1013.

Sander, N.: 2014, Internal migration in germany, 1995-2010: New insights into east-west migration and re-urbanisation, *Comparative Population Studies* **39**(2).

Shimer, R.: 2001, The impact of young workers on the aggregate labor market, *The Quarterly Journal of Economics* (August), 969–1007.

Skans, O. N.: 2005, Age effects in Swedish local labor markets, *Economics Letters* **86**, 419–426.

Spielman, S. and Folch, D. C.: 2015, Social area analysis and self-organizing maps, *in* C. Brunsdon and A. Singleton (eds), *Geocomputation: A Practical Primer*, Sage, Los Angeles, USA, chapter 9, pp. 152–168.

Stock, J. H. and Watson, M. W.: 2003, *Introduction to econometrics*, Vol. 104, Addison Wesley Boston.

Wedel, M. and Kamakura, W. A.: 2012, *Market segmentation: Conceptual and methodological foundations*, Vol. 8, Springer Science & Business Media.

Wehrens, R., Buydens, L. M. et al.: 2007, Self-and super-organizing maps in r: the kohonen package, *J Stat Softw* **21**(5), 1–19.

Yan, J. and Thill, J.-C.: 2009, Visual data mining in spatial interaction analysis with self-organizing maps, *Environment and Planning B: Planning and Design* **36**(3), 466–486.