

Classifying vocational training markets

Corinna Kleinert^{1,2,3}  · Alexander Vosseler⁴ ·
Uwe Blien^{2,3}

Received: 29 August 2016 / Accepted: 26 October 2017 / Published online: 15 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract The German educational system is characterized by a large sector of dual vocational training, which facilitates integration into the labour market. This system creates a specific training market for school leavers, which is characterized by strong regional disparities. These differences as well as their consequences have not been systematically analysed in previous research. In a theory-guided analysis this paper examines empirically which structural ‘handicaps’ affect regional transition rates from school to training and how regional training markets may be classified according to these structural factors. To this end, a new method is applied which combines regression and cluster analysis to avoid arbitrariness in the selection of classification variables. It generates a well-interpretable classification of vocational education markets, which is of broad use in research and labour market policy. The method may be applied to solve a broad variety of similar research problems in regional science.

JEL Classification I21 · J24 · R23

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00168-017-0856-z>) contains supplementary material, which is available to authorized users.

✉ Corinna Kleinert
corinna.kleinert@lifbi.de

¹ Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany

² Institute for Employment Research, Regensburger Str. 104, 90478 Nuremberg, Germany

³ University of Bamberg, Bamberg, Germany

⁴ Siemens Bank GmbH, Otto-Hahn-Ring 6, 81739 Munich, Germany

1 Introduction

Since the financial crisis in 2009 many European countries are plagued by high rates of youth unemployment. In contrast, in Austria, Denmark, Germany and Switzerland youth unemployment rates have been relatively low (Eurostat 2017). All four countries share a distinctive feature: a substantial part of post-school education is organized via a market-mediated vocational training system, also called ‘dual training’ because learning takes place in firms and in schools. One advantage of the dual system is its institutionalized link to the labour market. Due to the fact that educational curricula are directly related to the production process of goods and services, many employers hire their apprentices after training. Thus transitions from training to work are smooth (Gangl 2003; Pollmann-Schult and Mayer 2004). Because of this feature the German vocational training system receives considerable international attention (see e.g. Jacoby 2014; Williams 2017).

Since in dual training market imbalances arise at an earlier phase in young people’s lives than in other educational systems (Kleinert and Jacob 2013), previous research on vocational training concentrated on transition problems from school to training and focussed on either demand- or supply-side explanations on the microlevel. Most studies overlooked that there is also systematic spatial variation in transition outcomes. This is particularly surprising as descriptive data show that vocational training markets in Germany are characterized by strong regional disparities (Mohr et al. 2014). So far, empirical evidence on their structure, patterns and consequences is rare in the training literature as well as in regional science, whereas regional disparities in labour markets have been widely studied (see e.g. Dauth 2013; Blien et al. 2010).

Regional rates of placement into apprenticeships depend on numerous structural conditions. Thus, a typology of training market regions is required to map the diverse combinations of characteristics into some manageable types. In order to identify such a pattern of regional training market disparities, the relevant structural conditions have to be determined and condensed by an empirical strategy. Such classification analyses have a long tradition in regional science (Aumayr 2007; Baum 2007; Kronthaler 2005; Romano et al. 2015; Stimson et al. 2003) and usually rely on exploratory methods such as cluster analysis. Here, the researcher is not provided with criteria that help to decide which variables should be included in the classification process and how to weight them.

Against this background, this article has three central objectives: first, it examines which structural characteristics of regional training markets contribute to differences in regional transition rates. To this end, we describe the scattered empirical evidence on this issue and combine it into a coherent framework, which is then tested empirically. We thus contribute to the literature on vocational training markets by adding a genuine regional perspective and to regional science by analysing the field of dual training which has not got much attention so far. Second, this article adopts a newly developed method to the classification problem at hand, which combines regression and cluster analysis and provides exact criteria for the selection of variables and their weights which are theory-guided (Blien et al. 2010). We show that methods of spatial econometrics can be included in this approach in case of regional dependencies. We thus contribute to regional science by steering the tradition of regional classification

analyses in a new direction. Third, we present new insights on the regional pattern of vocational training markets over Germany. To our knowledge this is the first spatial analysis of these markets in regional science.

The article proceeds as follows. In order to understand the classification problem in this particular institutional setting, the next section portrays the German dual-training system. Subsequently, we present theoretical considerations and previous research on structural factors determining school-to-training transitions in order to justify our selection of regional characteristics. In the third section, we describe the data and the regression-based cluster approach. Afterwards, the results on the two steps of empirical examination, regression and cluster analysis, are shown. The article concludes with a summary, discussion and outlook.

2 Institutional and theoretical framework

2.1 The German system of vocational training

Germany has a three-tier post-school education system, which consists of dual vocational training (or apprenticeship training), full-time vocational schools and academic education (Franz and Soskice 1995). The 'dual system' of vocational training comprises a large part, whereas university entrance rates are low compared to other countries. The dual system is quite attractive for school leavers because it is the only post-school track open for leavers from all school tracks and a vocational training certificate is regarded as minimum prerequisite in the German labour market (Shavit and Mueller 1998; Solga and Konietzka 1999).

Dual training is market-mediated, i.e. employers may freely decide whether they offer training, how many positions and which occupations they provide, and which applicants they hire. Apprentices participate in financing by accepting wage cuts, and the government provides accompanying education in vocational schools. Employers bear the largest part of training costs, which are relatively high compared to other countries (Dionisius et al. 2009). Nonetheless, investments in dual training may be attractive for employers in the long run: first, firm-based contents are directly related to the production process of goods and services. Second, employers are able to recoup their investments by keeping their apprentices as skilled workers because worker mobility is reduced by labour market regulations (Acemoglu and Pischke 1999). Through the chambers, employers also participate in designing and adapting the vocational school curricula. Firms thus often use dual vocational training to provide for their long-term firm-based stock of human capital.¹ In this sense, the training market can be understood as submarket of the labour market (Schweri and Mueller 2007). Nevertheless, there are differences: first, vocational training is not used in all economic sectors and occupational fields. Second, it is highly regulated in terms of contents, duration and certificates (Wolter and Ryan 2011). For the nearly 330 different occupations currently

¹ An additional motivation for firms' investments in training is apprentices' low labour costs and fixed-term contracts (Lindley 1975). For analytical purposes it would thus be helpful to compare labour costs with the local price level, but this is not available (Blien et al. 2009).

offered in the dual system, there are detailed nation-wide curricula and their length is fixed.

Dual training ends with a final practical and theoretical examination which is certified by chambers and vocational schools. Successful graduates acquire a highly standardized diploma that is widely acknowledged among employers. The majority of employers who provide training hire their apprentices subsequently as regular employees (Seibert and Kleinert 2009). The biggest advantage of firm-based training thus is the smooth labour market transitions it produces, which are reflected in low rates of youth unemployment. In dual-training systems market imbalances show up earlier, in transitions from school to training. Since most school leavers searching for training positions are still required to attend education and not eligible for unemployment benefits, the amount of transition problems is not reflected in unemployment rates. Dual-training systems are only efficient motors of school-to-work transitions if they succeed in a balanced matching of school leavers and training firms in quantitative and qualitative terms (Kleinert and Jacob 2013).

This is the reason why vocational training in Germany also involves the Federal Employment Agency. Its main duty is to support the matching process in the vocational training market by helping employers and applicants with placement.² The practical purpose of our typology is to support this duty by clustering regions with different structural ‘handicaps’ regarding the matching of training positions and applicants. Thus, it is intended to represent both the magnitude and the nature of training market problems labour market policy has to deal with.

2.2 Regional determinants of demand and supply in training markets

To date, there is no comprehensive theory on vocational training markets (Wolter and Ryan 2011). Existent approaches have either focused on the question why firms invest in training or why some school leavers do not succeed in entering training. Both approaches only analyse one side of vocational training markets, usually from a microperspective. Accordingly, there are only a few empirical studies on the effects of regional characteristics on training markets, which we discuss in the following (Hillmert 2001; Muehlemann and Wolter 2007, 2011; Schweri and Mueller 2007).

Since the dual system of vocational training in Germany is market-oriented, it is more vulnerable to fluctuations in supply and demand than school-based education (Wolter and Ryan 2011). The supply of apprentices is closely tied to *demographic developments*. The more students leave school in a certain year, the fiercer they compete for training positions. While Hillmert (2001) merely finds small negative effects of school leavers’ cohort size on transition rates in a longitudinal analysis, Kleinert and Jacob (2013) show that youth cohort shares in the regional population have a negative effect on transition chances, particularly in periods with large or growing cohorts. The fact that employers’ training decisions depend on their business expectations (Trotsch

² In practical terms, a large part of school leavers and training firms are using the local employment agencies’ services for finding open positions and suited applicants. Additionally, employment agencies offer vocational guidance for young people, which is also widely used.

and Walden 2010) means that spatial and temporal fluctuations in *economic cycle* affect the demand side of training markets. Studies from various countries focused on business cycle effects on the provision of training positions, in sum with ‘a significant, but modest impact’ (Wolter and Ryan 2011).

Apart from cyclical changes, there are structural differences in regional training markets which change over a longer time span. On the supply side, this accounts for the school leavers’ *educational composition*. The higher the share of school leavers with university entrance certificate (*Abitur*), the more of them will enrol in university instead of vocational training (Schweri and Mueller 2007). The same is true if there are many full-time vocational schools, colleges or universities in a region (Muehlemann and Wolter 2007). Sociological research has shown that *social characteristics* may work as powerful cues that signal expected problems during training and thus prevent employers to hire respective candidates. In particular young people from socially disadvantaged families and men with migration background have difficulties in entering training (Aybek 2011; Solga 2002). On an aggregate level this means that employers may hire apprentices from other regions or stop offering training if the regional supply of school leavers is over-represented with these groups.

On the demand side, the literature on the question why firms invest in training gives some hints on relevant regional differences in firm characteristics. In the view of the ‘new training literature’ (Acemoglu and Pischke 1998, 1999; Leuven 2005) Germany is characterized by frictional labour markets with information asymmetries, compressed wage structures and industry and firm monopsonies. These factors explain why investments in vocational training, with its large shares of general and occupational human capital, may be profitable for firms. First, *unionized firms* are more likely to train than non-unionized firms because unions impose wage floors that lead to wage compression (Dustmann and Schoenberg 2008). Thus, the lower degree of firm unionization in East Germany might explain why less training positions are provided there. Second, *large and older firms* can profit more from training than small or recently founded firms (Dustmann and Schoenberg 2008). Since vocational training is heavily regulated, it is easier and cheaper for them to fulfil requirements. They are more likely to have enough suitable work for apprentices and vacancies for skilled workers (Schweri and Mueller 2007), and they make better use of information on their apprentices’ skills (Dustmann and Schoenberg 2008). Empirically, establishment size has a substantive positive effect on the propensity to offer training, while its effect on training intensity, i.e. the amount of training positions relative to its workforce, is negative (Neubaeumer and Bellmann 1999). In general, employers only invest in training if they expect to need skilled workers (and if training is cheaper than external hiring). This may be one reason why empirical research observes pronounced *sectoral differences* in training (Neubaeumer and Bellmann 1999). While traditionally the production sector had been the core of vocational training in Germany (Hillmert 2008), training positions in service occupations have grown in recent years and positions in production have declined due to enduring structural problems and increasing cost pressure from international competition (Thelen and Busemeyer 2008). Besides, several studies show that high net training costs hinder employers to offer training (Schoenfeld et al. 2010). Cost–benefit analyses illustrate large differences between occupations and sectors, with particularly low costs in agriculture, personal services,

medical assistant occupations, hotel and catering, and sales in Germany (Schoenfeld et al. 2010).

Employers' motives to train may also differ in *rural and urban regions* (Harhoff and Kane 1997): in rural areas reputation has a bigger impact on training decisions because to train apprentices signals a high-quality workplace as well as social commitment (Sadowski 1980) and thus ensures 'the smooth running of the business' (Franz and Soskice 1995: 232). Finally, school leavers may influence regional training markets by their search behaviour. Large firms are more attractive for applicants than small firms due to higher employment security and better career chances (Neubaeumer 1999). Similarly, applicants prefer trade, technical and clerical occupations to 'dirty' blue-collar occupations and personal services (Franz and Soskice 1995). Accordingly, school leavers in regions with a high share of unattractive training positions may extend their search to other regions. Despite apprentices' young age commuting is common in vocational training in Germany (Bogai et al. 2008). Thus, the composition of training firms with regard to size and sectors in a region itself as well as in *neighbouring regions* with high commuting flows may affect a region's aggregate matching outcome.

In sum, theories and empirical studies on training markets suggest that several factors may contribute to differing regional transition rates to training. On the supply side of school leavers, factors such as cohort size, educational and social composition as well as school-based alternatives may play a role. On the demand side of firms, the economic situation, the share of old, large and unionized firms, and the sectoral mix might be important. Besides, regional conditions such as urbanization and characteristics of neighbouring regions have to be considered. In the following, it is tested empirically whether these factors have a measurable effect on regional transition rates to vocational training.

3 Methods

3.1 Data and variables

Since the local employment offices support employers and school leavers in finding suitable applicants and training positions, the 156 regional employment office districts in Germany form the spatial units used in this analysis.³ The data used for our typology stem from 2009/2010. Where monthly or daily information was available, we aggregated data for the so-called training year (*Ausbildungsjahr*), which started in October 2009 and ended in September 2010. This time frame follows the firms' yearly apprentice hiring process. In sum, the data set used here contains aggregated data for 154 regional units in one single training year.⁴ Information stems from various official sources, such as the Federal Institute for Vocational Training (BIBB), the Federal Statistical Office, and the Statistical Service of the Federal Employment Agency.

³ These districts unite several of the 402 German counties (*Landkreise*) and cities (*kreisfreie Städte*). In our analysis, we distinguish 154 regions because for the three Berlin districts only aggregate information is available.

⁴ An earlier version of the typology was calculated for the training year 2007/2008 (Heineck et al. 2011).

In order to estimate the effects of structural conditions on vocational training markets, we generated a *target variable* that maps the outflow of school leavers who search for training positions to vocational training. Since the total amount of persons searching for training is unknown,⁵ the transition rate to training is approximated by dividing the number of non-subsidized training contracts through the number of school leavers plus applicants from previous school-leaving years. In the numerator, subsidized training contracts are excluded in order to generate an unbiased picture of (exogenous) market conditions. In the denominator, also applicants who left school in earlier years are considered to account for the fact that a varying number of applicants do not find a training position directly after leaving school and many register as applicants at the employment agencies again in later years. In 2009/2010, there were pronounced regional differences in the transition rates to training (Fig. 1 in the online appendix). Low transition rates were found in Saxony, North Rhine-Westphalia and Hesse, in contrast to high rates in Schleswig–Holstein, Mecklenburg–West Pomerania and Bavaria. Particularly high rates showed up in metropolitan areas such as Frankfurt, Cologne, Hamburg, Stuttgart or Munich, but also in urban regions in Eastern Germany like Dresden, Leipzig, Halle or Chemnitz.

Besides the target variable, we selected indicators for its *determinants*, which represent regional influences of demand and supply discussed in the previous section. The variables include demographic pressure and business cycle, the school leavers' educational and social composition, the structure of training establishments⁶ in terms of size and sectors and population density. We use the share of non-Germans in the population as proxy for school leavers with migration background. For other factors spatially inclusive data are not available. This regards the welfare dependency of school leavers, alternatives to dual training, as well as age structure and unionization of training establishments. For an overview of the dimensions included in the models, indicators and quantities see Table 1.

3.2 A regression-based clustering approach

The approach applied here is based on a method developed by [Blien et al. \(2010\)](#), who propose a regression-based clustering approach, which consists of two steps: variable selection and classification. Since this combined method is of a general nature, it may be used for different classification problems in regional science ([Blien et al. 2010](#)).

In the first step, a pre-defined target variable, in our context the transition rate to firm-based vocational training, serves as response variable in a Gaussian linear regression model in order to select a subset of statistically significant predictor variables. By using a stepwise selection algorithm it is reduced to a final model, which only includes empirically significant variables. The initial set of variables which enters the model is theory-guided (see Sect. 2.2). Spatially or time-lagged endogenous variables are not

⁵ Federal Employment Agency data on reported applicants and their placement are not useful, because reporting behaviour of training firms and applicants is highly selective.

⁶ In the empirical analyses, we use establishments instead of firms because they are productive organizations localized in defined places, which typically take decisions on personnel.

Table 1 Indicators for structural determinants of demand and supply in regional training markets

Determinant	Indicator	Quantities used
Demographic pressure	Relative cohort size	N school leavers/ N population in working age (15–64)
Business cycle	Unemployment rate (in dep. workforce)	N unemployed/(N unemployed + N dependent employed)
School leavers' educational composition	Share of school leavers with <i>Abitur</i>	N school leavers eligible for higher education/ N school leavers
School leavers' social composition	Share of non-German population	N non-German population/ N population
Training market situation	Density of training positions	N employees in training establishments/ N employees in all establishments
Establishment size structure	Share of large training establishments	N training establishments with 500 employees or more/ N training establishments
Sectoral structure	Share of training establ. in industry/construction	N training establishments in services/ N training establishments
Urban/rural areas	Population density	(ln) inhabitants/km ²

allowed to be included as predictors, because the possibility of conducting a classification on the response variable should be ruled out. Consequently, in our case the final model only includes regressors that are theoretically and statistically meaningful in explaining regional variation in the transition rate to vocational training.

Two measures are taken to capture potential spatial dependencies: first, diagnostic tests for the presence of a spatial lag structure and spatially correlated regression errors are applied (Anselin et al. 1996). For this purpose, the following structural model, imposing either $\psi = 0$ ('lag') or $\phi = 0$ ('errors') below, is estimated by feasible generalized least square:

$$(1) \mathbf{y} = \phi \mathbf{W} \mathbf{y} + \mathbf{X} \beta + u$$

$$(2) u = \psi \mathbf{W} u + \epsilon, \quad \epsilon \stackrel{\text{i.i.d.}}{\sim} N_n(0_n, \sigma^2 \mathbf{I}_n)$$

Here $\mathbf{y} = (y_1, \dots, y_n)'$ denotes an n -vector of observations on the response variable, i.e. the transition rate to vocational training, \mathbf{X} is an $n \times k$ matrix of exogenous variables, β is the k -vector of regression coefficients, ϕ and ψ are the scalar autoregressive coefficients of the spatially lagged endogenous variable and the lagged error term, respectively. \mathbf{W} denotes an $n \times n$ spatial weight matrix with positive elements, which represents the 'degree of potential interaction' between neighbouring locations and is scaled such that each row sums to one (Anselin et al. 1996). In our case, \mathbf{W} is a commuting matrix of apprentices between all 154 regions. Second, characteristics of neighbouring regions are included as variables in the regression model. These variables are derived by pre-multiplying all the exogenous factors \mathbf{X} with matrix \mathbf{W} , which is accordingly used as weighting matrix. To control for spatial dependencies the model is estimated again, this time including the additional matrix-weighted regressors in Eq. (1) and setting $\phi = \psi = 0$ in (1) and (2).

Given a final specification indicated by a set of predictors $X^* = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ with corresponding estimates β , each variable in X^* is standardized and multiplied by the absolute value of the realized t-statistic $|t_{\beta_j}|$. It is easy to show that the usual t-values from a linear regression model convey the same relative information as the standardized regression coefficients (Bring 1994).⁷ To emphasize this notion, note that the t-value of a regressor z is related to the increment in R^2 obtained by adding z to a model that already contains $k - 1$ variables, summarized by the matrix \mathbf{X} , i.e.

$$|t_{\beta_j}| = \sqrt{\frac{R_{X_z}^2 - R_X^2}{(1 - R_{X_z}^2) / n - k}} \tag{1}$$

where $R_{X_z}^2$ denotes the new R^2 after variable z is added (Greene 2003).

In the second step of the analysis, a cluster analysis is performed with the set of standardized, t-multiplied predictors, which were selected in the first step, to classify regional entities. Two methods are successively combined: first, a hierarchical-agglomerative cluster analysis according to Ward is applied. Since this method does not necessarily produce a final partition \mathcal{C} of objects that minimizes the within-cluster

⁷ A proof of this proposition is available from the authors upon request.

variance, K -means clustering is utilized subsequently to optimize the final cluster solution \mathcal{C}^W . The centroids of the clusters obtained in the Ward step are used as initial partitions for K -means clustering (Everitt et al. 2011; Mirkin 2005). The final cluster solution due to K -means \mathcal{C}^{KM} can be evaluated by regressing \mathbf{y} on a set of \mathbf{P} indicator variables, where the p variable equals 1 if observation i falls in this cluster. By doing so, the usefulness of the solution can be assessed in terms of its variance ‘explanation’ with respect to the response variable of the regression step, which was used to determine the relevant structural factors.

From a statistical viewpoint, this approach can be distinguished from model-based clustering approaches using mixture models (for an overview, see Fraley and Raftery 2002) as well as from clustering approaches with variable selection (see for example Witten and Tibshirani 2010; Celeux 2014 for an overview). Although variable selection, i.e. determination of the cluster space, in our approach is model-based, clustering itself is not, since both Ward and K -means are deterministic clustering methods. In contrast, mixture models assume an explicit probabilistic model with respect to the unconditional distribution of (unlabelled) data \mathbf{X} , whereas our approach assumes a probabilistic model within a Gaussian linear regression framework for the conditional distribution of \mathbf{y} given \mathbf{X} .

4 Results

4.1 Selecting regional determinants

The regression analysis started with including all the exogenous variables described in Table 1. Then statistically insignificant and collinear variables were dropped, one at a time, to find the sparsest model with the highest ‘explanatory’ power (in a statistical sense). In Table 2, Model 1, the final estimation results are shown. This model consists of five exogenous variables with highly significant coefficients and theoretically expected signs.

Since we did not use functional regions (Karlsson and Olsson 2006) which are characterized by internal interaction, it is important to control for interregional spillovers. Besides the usual diagnostic tests, two robust Lagrange multiplier (LM) tests for the presence of a spatial lag and a spatial autoregressive error of order one [AR(1)] were conducted (see last two rows of Table 2 for the LM test statistics). For model 1, both test results clearly lead to a rejection of the null hypothesis. To control for spatial dependencies the model was estimated again, this time including characteristics of neighbouring regions in form of additional matrix-weighted regressors. The extended regressions were again reduced stepwise by omitting insignificant and multicollinear covariates. It turned out that the inclusion of a single additional variable, the share of large training establishments in surrounding regions, is sufficient to account for spatial dependencies (Table 2, Model 2). The LM tests show that both null hypotheses cannot be rejected now.

Moreover, Model 2 has a significantly higher explanatory power: nearly 70% of the variation in the regional transition rate to training can be explained by the six variables in the model, which again all show the theoretically expected signs. The additional

Table 2 Final regression model with and without a spatially lagged exogenous variable

Exogenous variables	Model 1		Model 2		Weight ^a
	<i>B</i>	<i>t</i>	<i>B</i>	<i>t</i>	
Constant	0.504***	81.29	0.504***	103.89	–
Relative cohort size	– 0.052***	– 6.06	– 0.059***	– 8.76	21.4
High educated school leavers	– 0.028***	– 5.20	– 0.028***	– 4.25	10.4
Unemployment rate	– 0.033***	– 3.61	– 0.040***	– 5.53	13.5
Secondary sector training establishments	– 0.058***	– 7.05	– 0.051***	– 7.87	19.2
Large training establishments	0.029***	4.07	0.029***	4.75	11.6
Large train. est. in surrounding regions	–	–	– 0.049***	– 9.73	23.8
Adjusted R^2	0.499***		0.693***		
Spatial error (LM test statistic) ^b	34.722***		0.873***		
Spatial lag (LM test statistic)	37.387***		0.034***		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

^a Relative importance in explaining the transition rate, measured by the absolute t value (in %).

^b Robust Lagrange multiplier test

variable has the greatest relative importance overall, measured by its t -value. The more large training establishments are in surrounding regions, the fewer applicants start training in their own region. The second most important explanatory variable is the relative cohort size of school leavers. The larger is the share of school leavers relative to the resident working age population, the fewer of them manage to find training positions. The share of training establishments in the secondary sector (manufacturing and construction) also has a negative effect on the regional transition rate. Compared to these three factors the unemployment rate in a region is less important. In regions with high levels of unemployment the transition rate tends to be lower. The share of large training establishments and the share of high educated school leavers have the smallest explanatory power. Since large establishments offer not only job opportunities, but also potential training positions, their share has a positive impact. The more school leavers in a region are highly educated, the more of them enter academic education, and the lower is the transition rate to training.⁸

4.2 Clustering regional training markets

In the second analysis step, the determinants selected in Table 2, Model 2, were z -transformed, weighted by their t -values and included in a Ward and in a K -means cluster analysis. We decided for a final solution of twelve clusters, which jointly describe 79% of the six classification variables' variance. This solution was regarded as satisfactory concerning the coherence of the variables' combinations and the range

⁸ Due to high multicollinearity the density of training positions and population density have not been included in the final specification. The share of the non-German population did not enter because it showed no significant effect.

Table 3 Values of classification variables in the training market types 2010^a

	N	Cohort size	High educated	Unemployment	Sec. sector est.	Large establ.	Surrounding
I: Eastern German districts with very few school leavers and high unemployment							
Ia: Rural districts with large secondary sector	17	-	(-)	+	++	(-)	(+)
Ib: Rural districts with average training market conditions	10	-	++	+	0	0	0
Ic: Differing districts with favourable training market conditions	9	-	+	++	-	±	-
II: Dynamic metropolitan areas in the West							
IIa: Metropolitan districts with favourable training market conditions and low competition	11	(-)	+	(+)	-	++	-
IIb: Urban districts with strong large-establishment neighbourhood	12	(+)	(+)	-	0	-	++
III: Western districts with large-establ. neighbourhoods							
IIIa: Urban districts with average conditions	17	+	+	0	0	0	+
IIIb: Rather urban districts with very low unemployment and high competition	16	+	-	-	+	(-)	+
IIIc: Metropolitan districts with high unemployment	5	(+)	+	++	-	+	+
IV: Western districts with no large-establ. neighbourhood and low unemployment							
IVa: Rather urban districts favourable training market conditions and medium competition	23	(+)	0	(-)	(-)	(-)	-
IVb: Rural districts with large secondary sector and high competition	25	++	-	-	+	-	-
IVc: Rural districts with very weak large-establ. neighbourhood and high competition	9	+	-	-	0	-	-

^a Legend: - Strongly below average, - below average, (-) slightly below average, 0 on average, (+) slightly above average, ++ above average, +++ strongly above average, ± heterogeneous

of variables' values in the single clusters, whereas graphical tools and stopping rules, such as the Calinski and Harabasz pseudo-F index or the Duda–Hart $Je(2)/Je(1)$ -index, showed no clear preference for a particular cluster solution. Since one of the twelve clusters contained only two regions, it was aggregated with the closest neighbouring cluster, resulting in a final classification of eleven training market types. The typology's effectiveness of discrimination was tested by an analysis of variance with regard to the regional transition rate to training.⁹ It shows a highly significant value of the F -statistic and an adjusted R^2 of about 48%. This implies that about half of the regional variation of the transition rate is taken over by the classification.

Table 3 depicts the levels of the classification variables in the eleven training market types, which were combined to four higher-ranking groups. Training market type I is restricted to East German regions characterized by high unemployment and few school leavers. It consists of three subtypes that differ from each other regarding the size of the secondary sector and the urban/rural divide. Type II is primarily represented by large metropolitan areas in West Germany, such as Hamburg, Cologne, Frankfurt/Main and Munich, and their surrounding commuter belts. Type IIa contains the urban centres, whereas Type IIb regions are found in the urban 'hinterland' of some of the large cities in Type IIa. They are characterized by an extraordinary high number of large training establishments in neighbouring regions (the urban centres), which attract many school leavers living in these commuting areas. Type III consists of urban regions in Western Germany with an above-average share of large training establishments in their neighbouring districts. Three subtypes are found here, which mainly differ from each other by their unemployment rate. Type IIIc is the smallest cluster with only five regions in the densely populated Ruhr area (*Ruhrgebiet*), which are characterized by very high unemployment rates. Type IV mostly consists of rural regions in West Germany with low unemployment rates and a small number of large training establishments in their neighbourhoods. This group consists of three subclusters, which mainly differ by the size of secondary sector as well as by spatial location. Type IVc is much smaller than Type IVa and IVb and mainly contains economically isolated regions at the country borders.

The spatial distribution of the eleven training market types is presented in Fig. 1, which shows some interesting patterns. Though it was to be expected that the training markets in East and West Germany are different, it is a surprise that there is a complete separation. All the regional units in Eastern Germany belong to types Ia, Ib and Ic, which are exclusively located in the East. Even 20 years after unification the social and economic reality of Eastern Germany is still different from the West. Within Eastern Germany there is a North/South divide, while this is not the case in Western Germany. This again is surprising, since the labour market performs better in the Southwest than in the Northwest (Blien et al. 2010). Apart from these features there is no large-scale spatial division within the country, i.e. the map shows no large connected areas belonging to the same training market type (apart from type IIIc), and some types are distributed over the whole area of Western Germany. Finally, there is a clear distinction

⁹ This is equivalent to a regression with the regional transition rate to training as response variable; only that now the eleven clusters form the regressors.

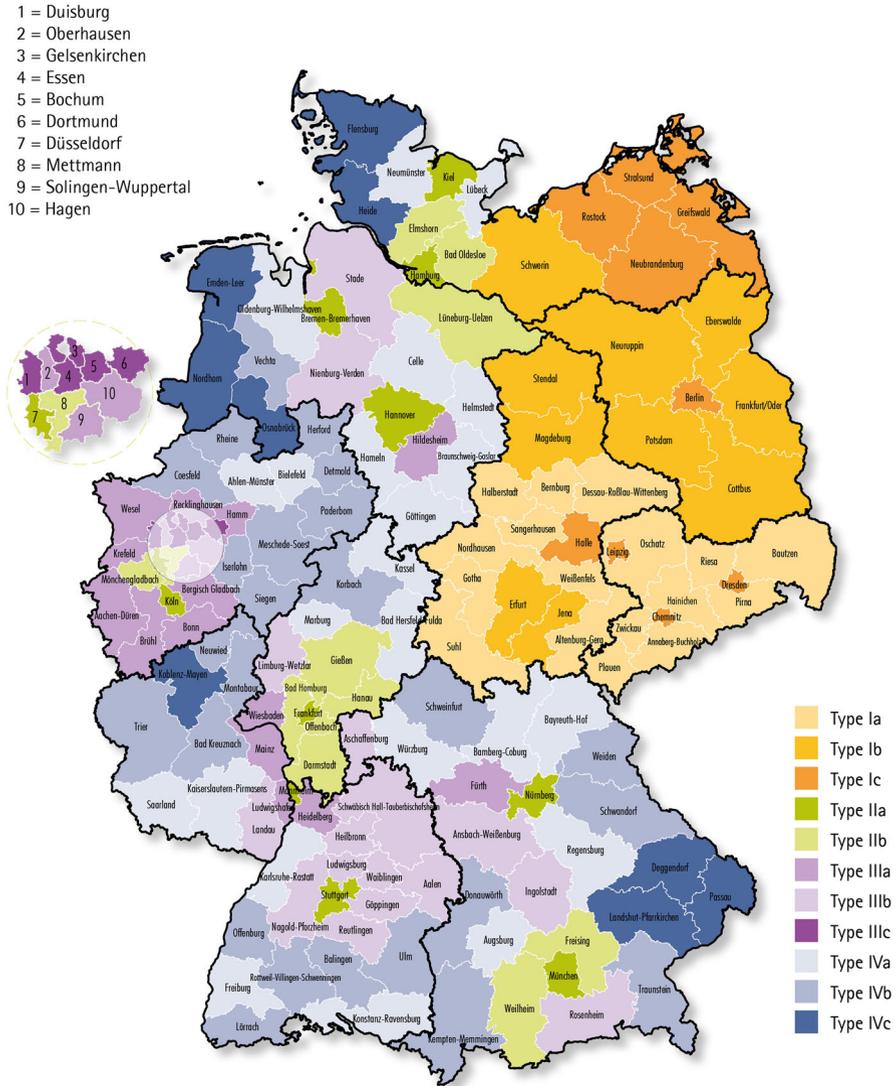


Fig. 1 Training market types 2010 by employment agency districts

between metropolitan, urban and rural training markets all over Germany, despite the fact that population density was not considered in clustering.

5 Summary, discussion and outlook

Germany is characterized by a specific form of post-school education, dual vocational training, which facilitates smooth labour market integration and creates a specific

market for school leavers characterized by strong regional disparities. Hence, this article aimed at characterizing regional training markets with respect to the structural ‘handicaps’ they represent for placing young people into training positions. To this end, we combined the scattered evidence on regional determinants for entry into training into a coherent framework and applied a new method for clustering heterogeneous training regions, which overcomes arbitrary selection of classification variables by combining regression and cluster analysis. Therefore, it is a form of regression-based clustering, linking a theory-guided analysis of determinants of regional disparities with standard classification approaches. This method helped to identify six highly significant demand- and supply-side factors, which affect regional transition rates to training. It generates a well-interpretable classification of vocational training markets. Finally, this article showed that methods of spatial econometrics can be included in this approach in case of regional dependencies.

Nevertheless, our study has some limitations. First of all, the two-step approach can only be applied if an external criterion is available which enables to select cluster factors by regression. If no such criterion can be found, other methods of clustering are to be recommended. A second point appears as a limitation, but is lying in the nature of the problem at hand. Since we do not make any assumptions about the ‘reality’ of the identified clusters, the classification represents an optimal division of a multidimensional cloud of cases. Small changes in design (e.g. in regression weights) can thus result in a substantially different cluster solution. However, if there are ‘real clusters’, in the sense that there are ‘gaps’ in the cloud of cases or that some variables are highly correlated within a specific cluster, the probability is high that these clusters are empirically identified. An example is the group of Ruhr cities (Type IIIc), which is stable over time and across different specifications. Finally, usual limitations of statistical analysis have to be mentioned, e.g. the fact that not all theoretically relevant factors could be measured due to lacking data availability. However, by providing the R^2 of the regression and the cluster partition we were able to assess the quality of the included information.

Future avenues of research could take up these deficits. First, it seems promising to collect data on and test effects of structural factors neglected so far, such as the regional supply of educational alternatives to dual training. Second, it would be fruitful to address the modifiable area unit problem (MAUP) by using differently sized administrative units. Such a comparison could reveal interesting results regarding regional interrelations and spillovers and contribute to the knowledge on the spatial nature of vocational training markets. Third, the statistical method applied here, regression-based clustering, might be combined with novel statistical developments, e.g. with Bayesian clustering.

The described classification does not only expand empirical knowledge, but also serve practical purposes and may inform future research. In practical terms, it is used by the Federal Employment Agency to manage local agencies by generating customized goal indicators and to manage their budgets, to exchange experiences in similar regions about best practices, to adopt target-oriented training measures and to assess how effective they are—functions that contribute to maintain the advantages of vocational

training.¹⁰ Beyond practical applications, the clusters may be used in future research on young people's individual transitions chances, where they offer a parsimonious instrument to examine effects of regional opportunity structures and their interplay with individual supply-side factors.

These applications suggest that it might be worth to transfer the approach demonstrated here to other fields of regional disparities such as social benefits or traffic control as well as to school-to-work transitions in other countries. For example, regional labour market conditions and education structures may determine early employment integration in countries which strongly rely on general education and provide more unstructured school-to-work transitions than in Germany. In countries with a stronger regionally segregated pattern of schools and universities and higher student fees, regional disparities in population as well as in educational institutions may explain levels of educational attainment. For these examples, the proposed method of regression-based clustering may pose as well a useful instrument to practically decide how to address region-specific constellations of hurdles.

Acknowledgements The authors would like to thank the participants of the workshop 'Chances and risks of demographic change for vocational training in the regions' (Bonn, September 2013) for useful comments on a previous version of the paper. Special thanks are due to Guido Heineck and Thomas Kruppe for their co-work on different versions of the classification, to Laura Hahn and Gunther Müller for their help with data preparation and to Wolfgang Dauth, Holger Seibert and Daniel Werner for critical commentaries and insightful suggestions. The usual disclaimer applies.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Acemoglu D, Pischke J-S (1998) Why do firms train? *Q J Econ* 113:79–119
- Acemoglu D, Pischke J-S (1999) Beyond Becker: training in imperfect labour markets. *Econ J* 109:112–142
- Alecke B, Mitze T, Untiedt G (2013) Growth effects of regional policy in Germany: results from a spatially augmented multiplicative interaction model. *Ann Reg Sci* 50:535–554
- Anselin L, Bera AK, Florax R, Mann JY (1996) Simple diagnostic tests for spatial dependence. *Reg Sci Urban Econ* 26:77–104
- Aumayr C (2007) European region types in EU-25. *Eur J Comp Econ* 4:109–147
- Aybek C (2011) Varying hurdles for low-skilled youth on the way to the labour market. In: Wings M, Windzio M, de Valk H, Aybek C (eds) *A life-course perspective on migration and integration*. Springer, Heidelberg
- Baum S et al (2007) Considering regional socio-economic outcomes in non-metropolitan Australia: a typology building approach. *Pap Reg Sci* 86:261–286
- Blien U, Gartner H, Stueber H, Wolf K (2009) Regional price levels and the agglomeration wage differential in western. *Ger Ann Reg Sci* 43:71–88
- Blien U, Hirschenauer F, Phan Thi Hong V (2010) Classification of regional labour markets for purposes of labour market policy. *Pap Reg Sci* 89:859–881

¹⁰ Of course, the classification might be also helpful for other public programs, such as regional policy (see e.g. Alecke et al. 2013).

- Bogai D, Seibert H, Wiethoelter D (2008) Duale Ausbildung in Deutschland: Die Suche nach Lehrstellen macht junge Menschen mobil. IAB-Kurzbericht 09/2008. <http://doku.iab.de/kurzber/2008/kb0908.pdf>. Accessed 26 August 2016
- Bring J (1994) How to standardize regression coefficients. *Am Stat* 48:209–213
- Celeux G et al (2014) Comparing model selection and regularization approaches to variable selection in model-based clustering. *J Soc Fr Stat* 155:57–71
- Dauth W (2013) Agglomeration and regional employment dynamics. *Pap Reg Sci* 92:419–435
- Dionisius R, Muehleemann S, Pfeifer H, Walden G, Wenzelmann F, Wolter SC (2009) Costs and benefits of apprenticeship training. A comparison of Germany and Switzerland. *Appl Econ Q* 55:7–37
- Dustmann C, Schoenberg U (2008) Why does the German apprenticeship system work? In: Mayer KU, Solga H (eds) *Skill formation, interdisciplinary and cross-national perspectives*. Cambridge University Press, Cambridge, pp 85–108
- Eurostat (2017) Unemployment statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Unemployment_statistics. Accessed 2 August 2017
- Everitt B, Landau S, Leese M, Stahl D (2011) *Cluster analysis*, 5th edn. Wiley, Chichester
- Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 97:611–631
- Franz W, Soskice D (1995) The German apprenticeship system. In: Buttler F, Franz W, Schettkat R, Soskice D (eds) *Institutional frameworks and labor market performance*. Routledge, London
- Gangl M (2003) Returns to education in context: individual education and transition outcomes in European labor markets. In: Müller W, Gangl M (eds) *Transitions from education to work in Europe*. Oxford University Press, Oxford, pp 156–185
- Greene WH (2003) *Econometric analysis*, 5th edn. Pearson Education, New Jersey
- Harhoff D, Kane T (1997) Is the German apprenticeship system a panacea for the U.S. labor market? *J Popul Econ* 10:171–196
- Heineck G, Kleinert C, Vosseler A (2011) Regionale Typisierung: was Ausbildungsmärkte vergleichbar macht. IAB-Kurzbericht 13/2011. <http://doku.iab.de/kurzber/2011/kb1311.pdf>. Accessed 26 August 2016
- Hillmert S (2001) Cohorts and competition. Transitions from school to work in the context of economic and demographic change. Max Planck Institute for Human Development. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.8802&rep=rep1&type=pdf>. Accessed 26 August 2016
- Hillmert S (2008) When traditions change and virtues become obstacles. In: Mayer KU, Solga H (eds) *Skill formation, interdisciplinary and cross-national perspectives*. Cambridge University Press, Cambridge, pp 50–81
- Jacoby T (2014) Why Germany is so much better at training its workers. *The Atlantic* 10 2014. <https://www.theatlantic.com/business/archive/2014/10/why-germany-is-so-much-better-at-training-its-workers/381550/>. Accessed 16 July 2017
- Karlsson C, Olsson M (2006) The identification of functional regions: theory, methods, and applications. *Ann Reg Sci* 40:1–18
- Kleinert C, Jacob M (2013) Demographic changes, labor markets and their consequences on post-school-transitions in West Germany 1975–2005. *Res Soc Strat Mobil* 32:65–83
- Kronthaler F (2005) Economic capability of East German regions: results of a cluster analysis. *Reg Stud* 39:739–750
- Leuven E (2005) The economics of private sector training: a survey of the literature. *J Econ Surv* 19:91–111
- Lindley RM (1975) The demand for apprentice recruits by the engineering industry: 1951–1971. *Scot. J Polit Econ* 22:1–24
- Mirkin B (2005) *Clustering for data mining. A data recovery approach*. Chapman & Hall, London
- Mohr S, Troltsch K, Gerhards C (2014) Regional matching problems and establishments with declining training places. *BWP* 2/2014. <https://www.bibb.de/en/33335.php>. Accessed 18 August 2017
- Muehleemann S, Wolter S (2007) Regional effects on employer-provided training: evidence from apprenticeship training in Switzerland. *Z Arb* 40:135–147
- Muehleemann S, Wolter C (2011) Firm sponsored training and poaching externalities in regional labor markets. *Reg Sci Urban Econ* 41:560–570
- Neubaeumer R (1999) *Der Ausbildungsstellenmarkt der Bundesrepublik Deutschland*. Duncker & Humblot, Berlin

- Neubaeumer R, Bellmann L (1999) Ausbildungsintensität und Ausbildungsbeteiligung von Betrieben. In: Beer D, Frick B, Neubaeumer R, Sesselmeier W (eds) Die wirtschaftlichen Folgen von Aus- und Weiterbildung. Hampp, München and Mering, pp 1–43
- Pollmann-Schult M, Mayer K-U (2004) Return to skills: vocational training in Germany 1935–2000. *Yale J Sociol* 4:73–99
- Romano E, Mateu J, Giraldo R (2015) On the performance of two clustering methods for spatial functional data. *ASTA Adv Stat Anal* 99:467–492
- Sadowski D (1980) Berufliche Bildung und betriebliches Bildungsbudget. Poeschel, Stuttgart
- Schoenfeld G, Wenzelmann F, Dionisius R, Pfeifer H, Walden G (2010) Kosten und Nutzen der dualen Ausbildung aus Sicht der Betriebe. Ergebnisse der vierten BIBB-Kosten-Nutzen-Erhebung. W, Bertelsmann, Bielefeld
- Schweri J, Mueller B (2007) Why has the share of training firms declined in Switzerland? *Z Arb* 40:149–167
- Seibert H, Kleinert C (2009) Duale Berufsausbildung. Ungelöste Probleme trotz Entspannung. IAB-Kurzbericht 10/2009. <http://doku.iab.de/kurzber/2009/kb1009.pdf>. Accessed 26 August 2016
- Shavit Y, Mueller W (eds) (1998) From school to work. A comparative study of educational qualifications and occupational destinations. Clarendon Press, Oxford
- Solga H (2002) Stigmatization by negative selection: explaining less-educated persons' decreasing employment opportunities. *Eur Sociol Rev* 18:159–178
- Solga H, Konietzka D (1999) Occupational matching and social stratification. *Eur Sociol Rev* 15:25–47
- Stimson R, Baum S, O'Connor K (2003) The social and economic performance of Australia's large regional cities and towns. *Aust Geogr Stud* 41:131–147
- Thelen K, Busemeyer MR (2008) From collectivism towards segmentalism. Institutional changes in German vocational training. MPIfG Discussion Paper 08/13. http://www.mpifg.de/pu/mpifg_dp/dp08-13.pdf. Accessed 26 August 2016
- Troltsch K, Walden G (2010) Beschäftigungsentwicklung und Dynamik des betrieblichen Ausbildungsangebotes. *Z Arb* 43:107–124
- Williams A. (2017) Why Other Countries want to import Germany's dual-education system. *Handelsblatt Global* 2017-04-25. <https://global.handelsblatt.com/companies-markets/why-other-countries-want-to-import-germanys-dual-education-system-754411>. Accessed 16 July 2017
- Witten DM, Tibshirani R (2010) A framework for feature selection in clustering. *J Am Stat Assoc* 105:713–726
- Wolter SC, Ryan P (2011) Apprenticeship. In: Hanushek EA, Machin SM, Woessmann L (eds) *Handbook of Economics of Education*, vol 3. North Holland, Amsterdam, pp 521–576