Lett Spat Resour Sci (2018) 11:17–26
https://doi.org/10.1007/s12076-017-0194-x

ORIGINAL PAPER

A new classification of regional labour markets in Germany

Uwe Blien[1] · Franziska Hirschenauer[1]

Received: 23 January 2017 / Accepted: 16 October 2017 / Published online: 26 October 2017

Abstract In Germany there are considerable disparities between regional labour markets. As local labour market conditions influence the success of labour market policy measures, they must be taken into account when the performance of employment agencies is compared. In its system of management by objectives, the Federal Employment Agency therefore uses classifications of employment agency districts based on a specially designed classification approach. This method is a quite powerful combination of regression and cluster analysis, because the types identified are relatively homogeneous not only with respect to the classification variables but also with respect to the response variable in the regression step. The method can be used to address a wide range of research questions.

Keywords Regional labour markets · Labour market policy · Regional typology · Cluster analysis methods

JEL Classification R23 · R11 · R50

1 Introduction and background

Germany's labour market has seen positive development in recent years compared to other European states. Despite this favourable trend, regional economic disparities in Germany remain considerable. Even today—more than 25 years after reunification—the eastern German regions have not yet fully caught up with their western German

✉ Franziska Hirschenauer
franziska.hirschenauer@iab.de

[1] Institute for Employment Research of the Federal Employment Agency, Regensburger Str. 100, 90478 Nuremberg, Germany

counterparts in economic terms. This is shown by various indicators, such as the unemployment rate (7.4% East, 5.3% West in August 2017) or GDP per inhabitant. The western German regional values vary more considerably than those in eastern Germany, indicating relatively pronounced regional disparities within western Germany that evolved long before reunification in 1990.

As a consequence of these striking regional labour market disparities, there are also differences in the integration results that can be achieved by means of labour market policy measures in Germany's individual regions. It is easier to get unemployed people back into the labour market permanently if the labour market situation is favourable than is the case if it is poor. The headquarters of the Federal Employment Agency (Bundesagentur für Arbeit) has to take the different labour market conditions into account when comparing the performance of the individual employment agencies. In its system of management by objectives it therefore uses regional classifications based on a specially designed classification approach (Blien et al. 2010). This is also applied in this paper. The key property of the approach is that the employment agencies belonging to one and the same group are similar with respect to the "handicap" that the labour market situation poses for the integration of the unemployed. The members of *one* group—also termed "comparison type" in this paper—resemble each other with regard to the factors that influence the success of labour market policy.

In contrast to traditional classification approaches, a two-step method is used. In the first step a regression analysis is carried out with the integration rate as the response variable. This variable relates the number of cases of integration into employment to the total population of unemployed people. Only the variables that are found to be significant influencing factors of this response variable are used as classification variables in the second step, when the population of agencies is split into groups. Unlike conventional classification procedures, the classification variables are not selected arbitrarily or on the basis of expert opinions but are based on empirical findings. This ensures that the groups are differentiated with respect to the response variable in the regression step. As will be shown later, the types identified discriminate nearly as good with respect to the response variable as do all the exogenous variables together.

This response variable, the integration rate, is not included among the classification variables, because it is among those variables that are later used to assess the performance of an agency. In the system of management by objectives used by the Federal Employment Agency, several target variables are compared when benchmarking individual agencies: if an agency has relatively high values compared to the average of its respective type, its performance is regarded as good.

The same two-step regression-based classification method can be used to analyse a wide range of research problems. It is also possible to assess other forms of regional policy or to classify the performance of schools, hospitals or firms (in a specific market) in a similar way. The method used could be applied to many different classification problems in regional science, economics, the social sciences and even the natural sciences. The information on the types can be included as additional controlling variables in evaluation studies using individual data.

In the following, the details of the classification method are explained, taking the latest classification of employment agencies as an application, and then the classification results are presented.

2 The two-step approach

Our classification approach is new in its combination of regression and cluster analysis. It can be applied if an external criterion is available for the selection of classification variables. Variables are included in the second step, the cluster step, if they are significant as exogenous variables in the first, the regression step. In our case the variables are examined to see whether they are related to the integration rate, which is the response variable of the regression. This procedure differs from traditional (Mirkin 2005; Everitt et al. 2011) and more recent cluster analysis approaches (Fraley and Raftery 2002), which do not use an external criterion to decide whether a classification variable should be included. In our case, we are interested in whether the classification variables are relevant for the "handicap" of the local employment agency posed by its labour market situation.

The greater the influence of an exogenous variable on the response variable in the regression analysis, the higher is the weight given to this variable in the cluster analysis. This is guaranteed by including significant variables only. In addition, the individual variables are weighted by the t-values obtained in the regression analysis. This is related to a basic theorem of the regression method (Bring 1994). The t-value of a variable shows the increase in the R^2 that is related to this variable when it is the last one to be included in the regression equation. This is the reason why the t-value represents the "contribution" of a single variable to the R^2 of the regression analysis. It therefore shows the relative importance of the specific variable. This is visible from the following equation (slightly modified from Bring 1994):

$$|t_k| = \left| \sqrt{\frac{R^2_{1,2,3,\ldots k} - R^2_{1,2,3,\ldots(k-1)}}{(1 - R^2_{1,2,3,\ldots k})/n - k - 1}} \right|$$

The integer n gives the number of cases. $R^2_{1,2,3,\ldots k}$ is the R^2 in a regression with k exogenous variables, $R^2_{1,2,3,\ldots(k-1)}$ is that calculated with a set of regressors reduced by one variable. Since the denominator of the term below the square root is constant, the contribution of all regressors can be compared. Therefore, comparing t-values is equivalent to considering the reductions in the R^2 obtained by excluding one variable.

The second, or classification, step is again a combination of two procedures: Ward's method and k-means clustering. With Ward's hierarchical method, in each step of the clustering process those agencies that lead to the smallest possible increase in the variance criterion F are grouped together. For group (or cluster) p it is:

$$\mathrm{F_p} = \sum_{\mathrm{i=1}}^{\mathrm{n_p}} \sum_{\mathrm{j=1}}^{\mathrm{J}} \left(\mathrm{x_{ij}} - \bar{\mathrm{x}}_{\mathrm{ij}}\right)^2$$

Here, $\bar{x}_{ij}$ is the mean of the j variable in cluster i, in other words $\bar{x}_{ij} = \frac{1}{n_p} \sum_{i=1}^{n_p} x_{ij}$, where $\mathrm{n_p}$ represents the number of cases in cluster p. Ward's method has the advantage over other clustering procedures that it tends to result in clusters of similar size and that singletons (clusters containing only one spatial unit) are less likely than is the case with other methods.

The second procedure—the k-means method—is applied because Ward's method, like all agglomerative hierarchical procedures, may generate suboptimal clusters in which agencies exhibit a greater distance from the centroid of their own cluster than to the centroid of a different one. k-means is used to correct such classifications so that all agencies belong to the comparison type to whose centroid they are closest (Mirkin 2005; Everitt et al. 2011).

3 The application of the chosen method

The spatial units of investigation in this analysis are the 156 employment agency districts (Arbeitsagenturbezirke) in Germany that have been valid since 2013. They constitute an administrative—not a functional or spatial—territorial structure. Berlin consists of three employment agency districts, but they are combined to form one spatial unit here owing to the high level of commuting within the city.

The numerator of the integration rate comprises the annual total number of cases in which employment-agency clients (registered unemployed persons and participants in employment and training measures) were integrated into employment covered by social security or self-employment.

The denominator of the integration rate contains the so-called client potential for the same year, in other words all the individuals who were clients for all or part of the year under observation. In the annual period of 2012, the client potential of the employment agencies comprised 3424 million persons. 1546 million of them took up employment in the course of the year. The integration rate was thus 45.1%. The regional values, measured at the level of employment agency districts, ranged between 36.6 and 56.5%.

What labour market conditions are associated with the regional variation in the integration rate? Of a large number of conceivable influencing factors, the regression analysis revealed seven that had a significant impact on the regional integration rate: the unemployment rate, the seasonal span, the share of the labour force with no vocational qualifications, the degree of tertiarisation, the share of workers in establishments with fewer than 100 employees, the job density and the spatial variable of the seasonal span. Like the response variable, these seven variables are included in the regression as logarithms. Together, they explain (in a statistical sense) 86.1% of the regional variation in the integration rate in 2012. This large explained share of the variance makes it clear how strongly the agencies' integration results are affected by regional circumstances.

As expected, the *unemployment rate* plays a key role. It has a negative impact on the integration rate (Table 1). *The seasonal span* reflects the fact that some labour markets are subject to strong seasonal fluctuations. The stronger these dynamics are, the more cases of integration can be counted. The *share of the labour force with no vocational qualifications* exerts a significant impact on the integration rate. The larger the regional share of low-skilled persons is, the lower the regional integration rate falls. This finding reflects the fact that getting people with no vocational qualifications (back)

Table 1 Classification variables and weighting of the agency classification 2014. *Source*: Statistics of the Federal Employment Agency, own calculations

Classification variables	Direction of influence	Weighting	
		abs. (=|t-value|)	rel. (%)
Results obtained from the regression of the integration rate			
Annual average unemployment rate for 2012 (%)			
Unemployed persons in relation to the entire civilian labour force	Negative	8.6	25.1
Seasonal span 7/11–6/12 (percentage points)[a]			
Difference between the maximum and minimum seasonal factor of a 12-month period. The seasonal factor is the relation between the unemployment figure of a particular month and the unemployment figure in the moving annual average	Positive	7.7	22.4
Share of labour force with no vocational qualifications in 2012 (%)			
Unemployed persons (annual average 2012) and employees covered by social security (as of June 2012) aged 25–64 with no vocational qualifications in relation to all unemployed and employed persons of this age	Negative	5.9	17.2
Degree of tertiarisation as of 30.6.2012 (%)			
Employees covered by social security in respective sections of economic activities in relation to all employed persons	Positive	3.9	11.4
Share of workers in establishments with fewer than 100 employees as of 30.6.2012 (%)			
Employees covered by social security in establishments with fewer than 100 employees in relation to all employed persons	Positive	2.6	7.6
Job density as of 30.6.2012 (%)			
Employees covered by social security in the agency district in relation to the population aged 15–64	Positive	2.2	6.4
Spatial variable of the seasonal span 7/11–6/12 (percentage points)			
Arithmetic mean of the seasonal span of the neighbouring employment agency districts weighted by the shares of individuals commuting out of their own district	Positive	3.4	9.9

[a] The seasonal span was included in the cluster analysis as a logarithm to avoid outlier problem

into employment is a difficult task, which can be seen not least in the high specific unemployment rates of this group. The *degree of tertiarisation*, the employment share of the services sector, provides clear indications of a region's industry structure. A large degree of tertiarisation has a positive impact on the integration rate due to the pronounced employment dynamics of the services sector.

The *share of workers in establishments with fewer than 100 employees* has a significant effect on the integration rate. A regional firm-size structure with a predominance of small establishments has a positive impact on the integration into employment. This can be explained by the stronger labour turnover in small and medium-sized establishments and the disproportionately large number of hires in these establishments as a result. *The job density* reflects the demand side of the labour market and measures a region's supply of jobs. It is positively correlated with the integration rate.

In addition to these influencing factors, it must be taken into consideration that employment agency districts are open regions that are interconnected via commuter movements. If there is a large degree of commuter interconnections, then the labour market conditions of the interconnected regions are also of importance. In regional analyses they may then no longer be examined separately, as this would lead to biased results. Statistical tests have shown that the spatial dependencies between the agency districts can be taken into account by using the *spatial variable of the seasonal span*. It has a positive impact on the integration rate.

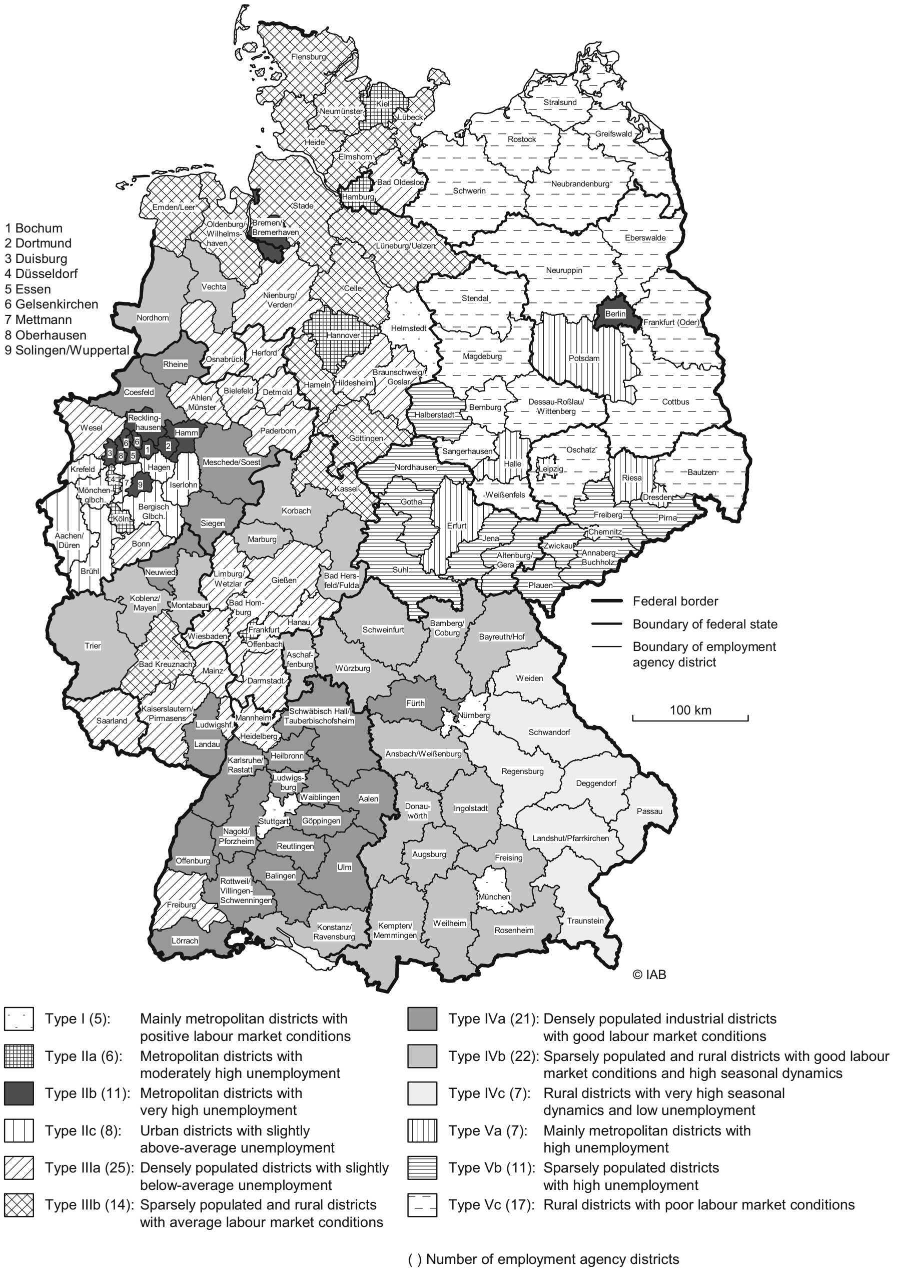
These seven variables are subjected to z-transformation and weighting and are then included as classification variables in a two-step cluster analysis that is used to group the agencies into comparison types. The first step of the cluster analysis is an agglomerative hierarchical clustering procedure following Ward's method. The second step is a non-hierarchical iterative clustering process following the k-means procedure.

4 Results

The outcome of the classification comprises 12 clusters, called "comparison types", nine in western and three in eastern Germany (see Fig. 1). Their size ranges from 5 to 25 employment agency districts. The R^2 of the corresponding variance analysis or dummy regression (with a reference category and 11 dummies representing the 12 types) stands at 79.9%, i.e. a large share of the entire variation in the regional integration rate can be attributed to the affiliation to different comparison types. This shows the advantage of combining regression and cluster analysis: the R^2 of the cluster solution is only slightly lower than the R^2 of the regression (86.1%), although the continuous variables of the latter transport more information.

The twelve comparison types can be combined into five superordinate "strategy types", which exhibit pronounced differences in terms of unemployment, settlement structure and spatial distribution. The types are combined according to content and not by means of a statistical procedure.

Strategy types I and II both cover agency districts that are metropolitan or urban in nature but differ with regard to their labour market situation: the regional values of the unemployment rate in comparison type I are lower than the national average and therefore lower than in comparison types IIa to IIc, where the rates are generally above average. The situation in comparison type IIb, which includes numerous old industrial cities in the Ruhr area, among others, is especially unfavourable. The employment agency districts of IIb have the highest unemployment rates in western Germany and fall into a similar range of values as those of eastern German towns and cities



Source: Statistics of the Federal Employment Agency, own calculations

Fig. 1 Comparison types of the employment agencies in 2014. *Source*: Statistics of the Federal Employment Agency, own calculations

(Type Va), sometimes even higher. In addition to the high unemployment rates, two other indicators are characterised by relatively poor values, namely high values for the share of the labour force with no vocational qualifications and low values for the job density.

Strategy type III comprises densely populated and rural districts in western Germany with average unemployment. The two comparison types that are combined to form this strategy type occur almost only outside of the southern German *Länder* (Federal States) of Bavaria and Baden-Württemberg. With regard to the unemployment rate, comparison type IIIa, the largest of all twelve comparison types, exhibits somewhat better values than IIIb. As regards the seasonal span and the share of workers in establishments with fewer than 100 employees, the regional values in IIIb are above average and higher than in IIIa.

Strategy type IV consists of densely populated and rural districts in western Germany that are characterised by a good labour market situation. Comparison type IVa can be found mainly in Baden-Württemberg. All members of this type exhibit unemployment rates that are not only below the national average but also below the lower western German average. What is also characteristic is the pronounced industrial orientation, which is reflected in low regional values for the degree of tertiarisation. Type IVb is found mainly in Bavaria. This comparison type, too, is characterised by low unemployment rates. The strong seasonal dynamics, which are surpassed only by type IVc, are also noteworthy. In the agency districts of type IVc, all of which are located in eastern Bavaria, industries with a strong seasonal component, such as construction, hotels and restaurants, and agriculture and forestry play a considerable role, which is reflected in maximum values for the seasonal span.

With the exception of Berlin, which belongs to IIb, all the eastern German employment agency districts in which unemployment is still above the national average belong to strategy type V. The members of comparison type Va are mainly agency districts with an urban character whose unemployment rates are more or less higher than the national average, in other words, some above, some below the higher eastern German average. Unlike the urban comparison types IIb and IIc, type Va is characterised by low regional values for the share of the labour force with no vocational qualifications. Type Vb is found almost only in Thuringia and Saxony. It is the only eastern German comparison type whose members all have unemployment rates below the eastern German average. This type is also characterised by above-average values for the seasonal span and the share of workers in establishments with fewer than 100 employees as well as below-average values for the share of the labour force with no vocational qualifications and the degree of tertiarisation. The latter indicates a relatively pronounced importance of the manufacturing industry. Type Vc, the largest of the three eastern German comparison types, can be found above all in Mecklenburg-West Pomerania, Brandenburg and Saxony-Anhalt. With regard to the unemployment rate, the majority of the members of this type with a rural structure have rates above the eastern German average, a not inconsiderable number of them even well above it. Characteristics of this type are also above-average values for the seasonal span and the proportion of workers in establishments with fewer than 100 employees and below-average shares of the labour force with no vocational qualifications. Above-average values for the degree of tertiarisation have to be seen in some cases against the background of the lower supply of jobs. They do not reflect a successful sectoral structural change but indicate a lack of regional job opportunities in the secondary sector.

5 Conclusion

Germany's regional labour markets exhibit considerable heterogeneity. Differences still exist not only between the regions of western and eastern Germany but also within these two parts of the country. In order to assess the performance of the employment agencies fairly, it is important to take these labour market disparities into account, as meeting labour market policy objectives depends not only on the actions of the agencies but also on the regional labour market conditions. To be able to take this into consideration in the management system of the Federal Employment Agency, employment agency districts with similar labour market conditions are grouped into a typology of comparison types.

The classification of agency districts is based on a two-step classification procedure. First, the context conditions are identified, i.e. the variables underlying the regional differences in the integration of unemployed and other employment-agency clients into employment. Of the potential influencing factors considered, it is possible to identify seven actual determinants of the integration rate, which are then used to form the comparison types: the unemployment rate, the seasonal span, the share of the labour force with no vocational qualifications, the degree of tertiarisation, the share of workers in establishments with fewer than 100 employees, the job density and the spatial variable of the seasonal span.

The current classification comprises twelve comparison types whose basic structure is characterised by differences between eastern and western Germany, between north and south within western Germany and between urban and rural areas across the entire country.

The method will be useful again for future updates of the classification. The combination of regression and cluster analysis transcends the bounds of standard cluster procedures. In the terminology of modern approaches to Artificial Intelligence, clustering techniques are regarded as methods of "unsupervised learning" (see Russell and Norvig 2016: 695ff.), because there is no external criterion for assessing the classification. In our case the situation is different, because calculating an R^2 for the classification shows its quality. The presented classification method is therefore enriched by a regression step and possesses some features of "supervised learning".

Because the R^2 of the regression analysis in the first step is only slightly larger than that of a regression with only dummies representing the types, the method applied can be characterised as quite "powerful". The information included in the exogenous variables of the regression step is almost entirely transferred to the cluster step. It is possible to use the proposed two-step method for analysing other forms of regional policy and for classifying many objects. These could be regions or firms or even objects in the natural sciences. The requirement is the availability of an external criterion that can be used as the response variable in the regression step.

References

Blien, U., Hirschenauer, F., Van Thi Hong, P.: Classification of regional labour markets for purposes of labour market policy. Pap. Reg. Sci. **89**(4), 850–880 (2010)

Bring, J.: How to standardize regression coefficients. Am. Stat. **48**(3), 209–213 (1994)

Everitt, B., Landau, S., Leese, M., Stahl, D.: Cluster Analysis. Wiley, Chichester (2011)

Fraley, C., Raftery, A.E.: Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. **97**(458), 611–631 (2002)

Mirkin, B.: Clustering for Data Mining. A Data Recovery Approach. Chapman & Hall/CRC, Boca Raton (2005)

Russell, S., Norvig, P.: Artificial Intelligence. A Modern Approach, 3rd edn. Pearson, Boston (2016)