

Iwona Bąk

West Pomeranian University of Technology in Szczecin

e-mail: iwona.bak@zut.edu.pl

ORCID: 0000-0001-8959-7269

Katarzyna Wawrzyniak

West Pomeranian University of Technology in Szczecin

e-mail: katarzyna.wawrzyniak@zut.edu.pl

ORCID: 0000-0003-4161-3877

THE USE OF REGRESSION TREES IN THE STUDY OF DEPENDENCIES BETWEEN THE EMPLOYMENT EFFICIENCY INDEX AND THE STRUCTURE OF THE UNEMPLOYED IN COUNTY EMPLOYMENT OFFICES

DOI: 10.15611/pn.2020.4.01

JEL Classification: C38, J49

© 2020 Iwona Bąk, Katarzyna Wawrzyniak

This is an open access article distributed under the Creative Commons Attribution-Non-Commercial-NoDerivs license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>)

Quote as: Bąk I., and Wawrzyniak, K. (2020). The use of regression trees in the study of dependencies between the employment efficiency index and the structure of the unemployed in county employment offices. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 64(4).

Abstract: The aim of the article is to analyse the relationship between the employment efficiency index and the structure of unemployment in county ('powiat') employment offices in Poland. Five research hypotheses connected with the research goal were formulated. The most important features characterizing the unemployed were: education, age, the duration of unemployment, gender, the place of residence. Regression trees were used as the research tool, which enabled to group county employment offices with differing values of the employment efficiency index with a simultaneous indication of independent variable variants (diagnostic features) affecting its level. After conducting the research, the authors managed to verify positively and unquestionably the hypothesis that the higher the percentage of the unemployed with higher education, the easier it is to achieve higher employment efficiency. In a less explicit way, only under certain conditions, hypotheses regarding the age of the unemployed and the place of residence were verified. Nevertheless, it was not possible to verify positively the hypotheses regarding the impact of gender and the duration of unemployment on the employment efficiency index.

Keywords: employment efficiency index, the structure of unemployment, county ('powiat') employment offices, regression trees.

1. Introduction

Pursuant to the Act of 20 April 2004 on employment promotion and labour market institutions (Dz. U. z 2019 r. poz. 1482 z późn. zm., art. 6, ust. 1, 2, p. 26) in the forefront among the labour market institutions are public employment services, which constitute employment authorities together with county (Polish ‘powiat’) and regional (voivodship) employment offices, an office supporting the minister responsible for labour matters and regional offices that carry out tasks specified in the Act. From the perspective of direct contact with the unemployed, an important place in the structure of public employment services is occupied by ‘powiat’ employment offices, which, using the appropriate forms of professional activation understood as services and instruments of the labour market financed by the Labour Fund, try to restore the unemployed to the open labour market. Since the efficiency of these offices in this respect is assessed annually, each of them tries to make the indicators being assessed as best as possible. Pursuant to the Act on employment promotion and labour market institutions (Dz. U. z 2019 r., poz. 1482 z późn. zm., art. 4, ust. 1¹, pkt. 2b, 2c, p. 19) these indicators include the employment efficiency index and the cost-effectiveness indicator.

If one assumes that the efficiency (effectiveness) of the functioning of county employment offices is perceived primarily through the number of the unemployed who commenced employment after completing their participation in the selected form (or forms) of professional activation, then the employment efficiency index should be considered as the basic indicator. This indicator is a stimulus, so the higher its value, the higher the effectiveness of activation of the unemployed in a given office¹. Therefore, such offices try to achieve the highest values of this indicator so that their effectiveness is assessed as highly as possible. Unfortunately, this is not always possible and there are offices for which achieving employment efficiency above 50% is a big challenge. What are the reasons for this? In order to obtain an answer to this question, and some others, a survey was conducted in 2018 among the staff of public employment services (Siewiera, 2019, pp. 34-46). Most of the respondents indicated that the level of the employment efficiency index is primarily determined by the characteristics of the unemployed registered in the area of the functioning of a given county employment office. The most important features were: education, age, the duration of unemployment, gender, and place of residence. According to the respondents, the most difficult situation is in those offices where predominant clients were the unemployed residing in the countryside, long-term unemployed, people aged 55 and older, and those without professional qualifications.

¹ It is worth quoting the results of the research by Piwowarski (2017, pp. 187-202), which showed that for 16 voivodships in 2011-2014 there was no significant relationship between the employment efficiency index and the rate of outflow from unemployment due to employment after activation. So despite the increase in the employment efficiency index in the years under review, no increase in employment was observed as a consequence of participation in forms of professional activation.

The article presents the results of research aimed at verifying the following working hypotheses²:

1) the employment efficiency index is higher in those offices where the percentage of unemployed women is lower than the percentage of unemployed men,

2) the higher the education of the unemployed, the higher the employment efficiency index,

3) the higher the percentage of long-term unemployed, the lower the employment efficiency,

4) the higher the percentage of the unemployed aged 55+ and 24 years and less, the lower the employment efficiency index,

5) the employment efficiency index is higher in those offices which operate in areas with a low percentage of unemployed residing in the countryside.

In order to verify the above-mentioned hypotheses³, it was decided to apply regression trees which enabled to group county employment offices with differing values of the employment efficiency index with a simultaneous indication of independent variable variants (diagnostic features) affecting its level.

The survey was conducted based on 340 'powiat' employment offices operating in Poland in 2018.

2. Review of research on employment efficiency

Problems regarding the active labour market policy (ALMP) and how to measure it are of interest to countries all around the world. As noted in (Immervoll and Scarpetta, 2012) over the past decade, many OECD countries have introduced or strengthened strategies for 'activating' the unemployed and other people receiving benefits. These strategies were called 'activation strategies'. According to the authors of this publication, "an effective policy strategy in this area is based on a perfectly tailored balance between financial support, incentives for work, as well as mutual commitments and active labour market programs". According to Frederiksson and Johansson (2003), properly designed ALMPs should be a useful tool in the fight against unemployment.

In Western countries a significant portion of public money is spent on active labour market policies. In 2013, average public expenditure on ALMP represented

² Working hypotheses were formulated based on the suggestions of representatives of public employment services (Siewiera, 2019, pp. 34-46).

³ For verification of individual hypotheses, e.g. correlation coefficients (Pearson or Spearman) calculated between the employment efficiency index and selected demographic characteristics characterizing the structure of the unemployed can be used separately. Research on the strength and direction of the relationship between the employment efficiency index and the features characterizing the socio-economic situation in the area of functioning of 'powiat' employment offices by voivodship and offices can be found, among others in (Bąk, Wawrzyniak, and Sobolewski, 2018, 2019; Bąk, Wawrzyniak, and Oesterreich, 2019; Siewiera, 2019).

0.5% of gross domestic product (GDP) in OECD countries, which corresponds to 1.0% of total public expenditure (OECD, 2017). National governments have various reasons to invest in ALMP. In addition to the individual negative aspects of unemployment – for example, loss of income and the depreciation of human capital – unemployment benefits also have a significant impact on the state budget (Vooren, Haelermans, Groot, and Brink 2019). Expenditure on active labour market programmes varies considerably across OECD countries. Given their share in GDP, they range from 0.15% to 0.20% in the US and Japan, respectively, to around 1% in France, Germany and 1.5% or more in the Scandinavian countries. According to Immervoll (2012), in high-income and developing countries there is a growing demand for well-designed ‘active’ or ‘activating’ social and labour market policies that can have a positive impact on employment. In the paper (Kluve et al., 2019), the authors reviewed 113 impact assessments of programmes activating young people on the labour market around the world and found that these are more successful in middle and low income countries. According to (Card, Kluve, and Weber, 2010), subsidized employment programs in the public sector are generally less effective than other types of ALMP.

Active labour market policies, including measures such as job search assistance, training, wage subsidies for the private sector and direct job creation in the public sector, are an important element in the efforts of European countries to combat unemployment. For the Member States of the European Union, ALMPs form a central part of the European employment strategy, which is one of the key objectives of the common economic policy. There is also a growing awareness of the need to develop scientifically sound measures of the effectiveness of different ALMPs (Kluve, 2006). According to Guzman (2014), active European labour market policies, especially the information on employment and training agencies, are important in reducing unemployment. In his opinion, Central Europe shows more efficiency in applying active labour market policies compared to other geographical areas.

The study of the effectiveness of an active labour market policy is also of great interest in Polish scientific publications (Elastyczność..., 2006; Błądowski, 2008; Kisiel and Nowińska, 2010; Rękas, 2011; Wiśniewski and Zawadzki, 2011; Drela, 2013). The discussions on the effectiveness of instruments of an active labour market policy emphasize, irrespectively of the attitude towards the methods of its measurement, the importance of their use for increasing the work opportunities of the unemployed (Błądowski, 2008). It is important to monitor the effectiveness of active support instruments on the labour market and to analyse it, so that knowledge in this field also contributes to increasing the effectiveness of these activities.

3. Material and research method

The use of regression trees for the classification (grouping) of objects characterized by many variables requires indicating among them a dependent variable measured

on strong (metric) scales, i.e. on a range or quotient scale (Gatnar and Walesiak, 2004, pp. 56-57).

In connection with this condition and forms of working hypotheses which were formulated in the study, the index of total employment efficiency in % was adopted as the dependent variable (Y). This indicator – in accordance with the Act on employment promotion and labour market institutions (Dz. U. z 2019 r., poz. 1482 z późn. zm., art. 4, ust. 1¹, pkt 2b, p. 19) – should be understood as the percentage share of the number of persons who, during or after their participation in the basic forms of professional activation, have been shown to be employed, in relation to the number of persons who have completed their participation in the basic forms of professional activation⁴.

However, the set of independent (explanatory) variables includes the following variables characterizing the unemployed in the area of functioning of county employment offices⁵:

X_1 – the share of unemployed women in the total number of unemployed in %,

X_2 – the share of the unemployed with higher education in the total number of unemployed in %,

X_3 – the share of the unemployed with basic vocational education in the total number of unemployed in %,

X_4 – the share of the unemployed with junior secondary higher education and lower in the total number of unemployed in %,

X_5 – the share of the unemployed aged 24 and less in the total number of unemployed in %,

X_6 – the share of the unemployed aged 55-64 in the total number of unemployed in %,

⁴ The catalogue of basic forms of professional activation, understood as services and instruments of the labor market financed from the Labour Fund, is published annually by the minister responsible for labour on the website of the office supporting the minister (Ustawa o promocji ..., Dz. U. z 2019 r. poz. 1482 z późn. zm., art. 4, ust. 11, pkt. 1, p. 19). This catalogue has been functioning unchanged since 2014 and includes the following six support instruments for the unemployed, to which most funds of the Labor Fund are allocated (*Efektywność podstawowych form...*, 2019, p. 8): training, internships, intervention works, public works, the reimbursement of the costs of furnishing or equipping the workplace for the referred unemployed, one-off funds for co-financing of undertaking business activity.

⁵ Data on the total employment efficiency indicators in 'powiat' employment offices come from publications (*Efektywność podstawowych form...*, 2019), while data on selected characteristics of the unemployed registered in the area of their competence were taken from the Central Data Bank of the Central Statistical Office. The use of data from two different sources resulted from the fact that data on the structure of the unemployed who took advantage of the forms of professional activation in a given local employment office in a given year are not generally available. Therefore, in the study it was decided to take into account the structure of all the unemployed registered in the area of these employment offices. This made it possible to approximate the characteristics of the unemployed population, also the local unemployed who were taken into account when calculating the employment effectiveness index.

X_7 – the share of the unemployed residing in rural areas in the total number of the unemployed in %,

X_8 – the share of the unemployed remaining unemployed for a period from 12 to 24 months in the total number of unemployed in %,

X_9 – the share of the unemployed remaining without work for more than 24 months in the total number of unemployed in %,

X_{10} – the share of the unemployed aged 18-24 remaining unemployed for more than 6 months in the total number of unemployed in %,

X_{11} – the share of the unemployed aged 55-64 remaining unemployed for more than 1 year in the total number of unemployed in %,

X_{12} – the share of the unemployed aged 55-64 remaining unemployed for more than 1 year in the number of unemployed aged 55-64 in %,

X_{13} – the share of the unemployed without qualifications (without seniority) in the total number of unemployed in %,

X_{14} – the share of the disabled unemployed in the total number of the unemployed in %.

Additionally, the variable X_{15} was introduced, which differentiates county employment offices according to the area in which they operate. Three variants were adopted for this variable: 1 – offices operating in urban counties ('powiat'), i.e. towns with 'powiat' rights, 2 – offices operating in rural and urban counties, which in addition to rural communes and smaller cities also include towns with 'powiat' rights, 3 – offices operating in rural counties, generally comprising several communes, consisting of rural areas and smaller towns. The first fourteen independent variables are so-called quantitative predictors (measurement on a metric scale), while the variable X_{15} – is a qualitative predictor (measurement on a non-metric scale).

Having gathered statistical data on the variables selected for the study, the regression tree was determined, which is a graphic representation of the model in the form of (Gatnar, 2008, pp. 37-39):

$$Y = f(\mathbf{x}_i) = \sum_{k=1}^K \alpha_k \mathbf{I}(\mathbf{x}_i) \in R_k, \quad (1)$$

where: Y – dependent variables, R_k ($k = 1, \dots, K$, K – the number of segments) are subspaces (segments) of exogenous variable spaces \mathbf{X}^L (X_1, X_2, \dots, X_L , L – the number of exogenous variables), $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iL}]$ – observations from recognizable set, α_k – model parameters, \mathbf{I} – indicator functions.

When the variables X_1, \dots, X_L are metric, each of the segments R_k is defined by its boundaries in space \mathbf{X}^L as follows:

$$\mathbf{I}(\mathbf{x}_i \in R_k) = \prod_{l=1}^L \mathbf{I}\left(v_{kl}^{(d)} \leq x_{il} \leq v_{kl}^{(g)}\right), \quad (2)$$

where the values $v_{kl}^{(d)}$ and $v_{kl}^{(g)}$ mean respectively the upper and lower limit of the segment in l -th dimension of space.

However, if the variables X_1, \dots, X_L are non-metric, the subspace R_k can be defined as

$$I(\mathbf{x}_i \in R_k) = \prod_{l=1}^L I(x_{il} \in B_{kl}), \quad (3)$$

where B_{kl} is a subset of the variable category set X_l , i.e. $B_{kl} \subseteq V_l$.

The parameters of the regression model (1) are calculated according to the formula:

$$\alpha_k = \frac{1}{N(k)} \sum_{\mathbf{x}_i \in R_k} y_i, \quad (4)$$

where: $N(k)$ – the number of observations (objects) located in the segment R_k , y_i – values adopted by the dependent variable in the segment R_k .

To assess the quality of the space division of the explanatory variables \mathbf{X}^L obtained on the basis of the regression model (1), the variance of the dependent variable is used.

The regression model discussed above is the result of a recursive division of the space of explanatory variables \mathbf{X}^L . This division consists in the fact that in each step the output fragment of space \mathbf{X}^L is optimally divided into two or more parts by one of the variables. The choice of the variable and the place of space division is made so that the obtained segments R_k are the most homogeneous due to the dependent variable (Y), i.e. in the case of a metric dependent variable they should have the smallest variance of this variable. The course of the space division process \mathbf{X}^L shows a tree, which due to the form of the model, was called regression⁶.

It is worth mentioning that the model in the form of a regression tree generates classification rules, which are in the form of conjunctions of the values of explanatory variables that decide about the division of space \mathbf{X}^L . In other words, based on the regression tree, one obtains a set of logical partitioning conditions, type *if ... then ...*, leading to a clear classification of the examined objects (Gatnar and Walesiak, 2004, pp. 108-109; Internetowy Podręcznik Statystyki). Owing to this, by interpreting the obtained results, one can unambiguously characterize the separated groups of objects due to both the values of the dependent variable and the values of the explanatory variables that were used in the model.

An important problem associated with the use of regression trees in practice is the selection of the optimal model (regression tree) so that its prediction error is as low as possible (Walesiak and Gatnar, 2009, p. 242). One of the most commonly used selection methods is the edge trimming method. This reduces the size of the tree by removing some of its fragments – this changes the structure and properties of the tree, because those divisions that were not of great importance in terms of correct

⁶ An example of creating a tree in two-dimensional space and discussing the procedure of recursive division of the space of explanatory variables can be found, among others in (Gatnar and Walesiak, 2004, pp. 104, 108-114; Gatnar, 2001, 2008).

classification are removed. The best known method of pruning a tree is to find such a tree that at the same time has a small classification error and low complexity. This is done in two stages (Gatnar and Walesiak, 2004, pp. 115-117):

- firstly, as a result of gradual pruning of the maximum tree D_{max} , i.e. the tree with the largest number of end nodes, a sequence of trees of smaller and smaller size is created (ever smaller complexity, fewer and fewer nodes), for which the assessment of resubstitution (replacement) of classification error is increasing,
- then the authors choose a tree of the right size from the tree sequence according to the error assessment based on the T test set or cross-validation procedure.

Breiman et al. (cited in Walesiak and Gatnar, 2009) suggest that choosing the smallest (the least complex) tree for the right size, for which the costs of the cross-validation do not differ significantly from the minimum costs of this verification. As a basis for selection, they propose the principle of one standard error, i.e. as a tree of the right size we choose a tree of the smallest size, whose cross-validation costs do not exceed the minimum cross-validation costs plus 1 standard error of cross-validation costs for a tree with minimum cross-validation costs. This can be written by the formula:

$$D^{opt} = \underset{K_t}{\operatorname{arg\,min}}(Q(D_t) + \sigma), \quad (5)$$

where: D^{opt} – optimal tree, K_t – number of terminal nodes (leaves) in the tree t , $Q(D_t)$ – cross-validation costs for a tree with minimum cross-validation costs, σ – the standard error of cross-validation costs for a tree with minimum cross-validation costs.

4. Research results

In this article the *CART* procedure (*C&RT*) was used to determine the regression trees, and the calculations were made in the *Statistica 13.0* program, having adopted the following assumptions: stop criterion – cut according to variance, stopping parameters: minimum number in the node – 34, i.e. 10% of the total number, maximum number of nodes – 1000, validation – V – fold cross check⁷ ($v = 10$).

As a result of the procedure used – under the adopted assumptions – a sequence of 20 regression trees was obtained characterized by the fact that for each tree size in the sequence there is no other tree of the same size that would have lower costs.

⁷ The *V*-cross test available in procedure *CART* is used when there is no test sample and the learning sample is too small to separate from it a separate test sample. The value v provided by the user for this test specifies the number of sub-samples, if possible of equal numbers, which are randomly created from the training sample. The tree of the given size is calculated v times, each time taken for the calculation excluding one sub-sample, which serves as a test sample. Thus, each of the sub-samples is used $(v - 1)$ times in the training sample, and only once as a test sample. The cost of the cross-valuation (CV) is calculated as the average cost of v test samples and this average is an estimate of the cost of the cross-valuation(CV) (Internetowy Podręcznik Statystyki).

It follows that each tree in the sequence is optimal, and this is ensured by the method of pruning the tree according to the minimum complexity measured at the cost of the cross-validation (Internetowy Podręcznik Statystyki).

Table 1 shows the obtained tree sequence along with information on the complexity (size) of the tree, the costs of the cross-validation and the costs of re-substitution.

Table 1. Sequence of regression trees for the model of employment efficiency index set against variables characterizing the structure of the unemployed in 'powiat' employment offices

Tree number	Number of end nodes	Cross validation costs	Standard error of the cross-validation costs	Re-substitution cost	Node complexity
1	26	56.149	5.291	17.163	0.000
2	24	55.873	5.305	17.618	0.227
3	23	55.426	5.312	17.916	0.298
4	21	54.268	5.163	18.582	0.333
5	20	53.652	5.135	18.937	0.355
6	19	52.320	5.115	19.323	0.386
7	18	50.428	4.978	19.741	0.418
8	17	48.788	4.882	20.246	0.505
9	16	47.745	4.959	20.820	0.573
10	15	46.821	4.953	21.459	0.639
11	13	45.705	4.876	22.756	0.649
12	12	45.404	4.783	23.480	0.725
13	11	45.022	4.766	24.240	0.759
14	10	44.317	4.525	25.052	0.813
15	7	42.692	4.225	27.913	0.953
16	6	41.871	4.216	28.903	0.990
17	5	41.127	4.181	30.018	1.115
18	3	39.037	4.013	32.837	1.410
19	2	37.040	3.785	34.527	1.690
20	1	37.547	3.944	37.310	2.783

Source: own calculations.

Choosing the right size tree from the optimal tree sequence started by applying the criterion written with formula (5). Table 1 shows that the minimum costs for the cross-validation are for tree number 20 and amount to 37.547. For this tree, the standard error of the cross-validation costs is 3.944. Therefore, the tree of the right size will be a tree for which the cost of the cross-validation does not exceed 41.481 ($37.545 + 3.944$). This condition is met by trees with numbers 19, 18 and 17.

From these trees, one should choose tree number 17, i.e. the tree with the largest number of terminal nodes, because then two conditions will be met simultaneously: the maximum complexity of the tree and the relatively lowest costs of the cross-validation. As tree number 17 has only 5 end nodes (leaves) formed using four explanatory variables (predictors: X_2, X_4, X_5, X_7), and the number of county employment offices in individual terminal nodes is: 1, 1, 97, 204, 37, it was also decided to check trees number 16 and 15, for which the costs of cross-validation are slightly higher than 41.481. In these trees, the complexity increases, and thus the number in terminal nodes decreases – this applies to those nodes that in tree number 17 were from 37 to 204. The number of predictors used in the construction of these trees is also increasing, which is important from the point of view of the research goal. The decision to check these trees is also confirmed by the cross-validation and re-substitution costs graph, which clearly shows that the difference between these costs begins to increase more clearly only from tree number 14 and reaches the highest value for tree number 1 (see Figure 1).

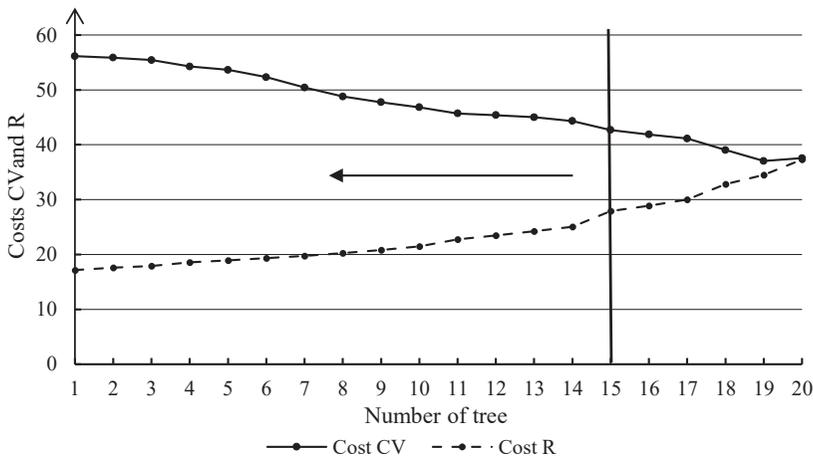


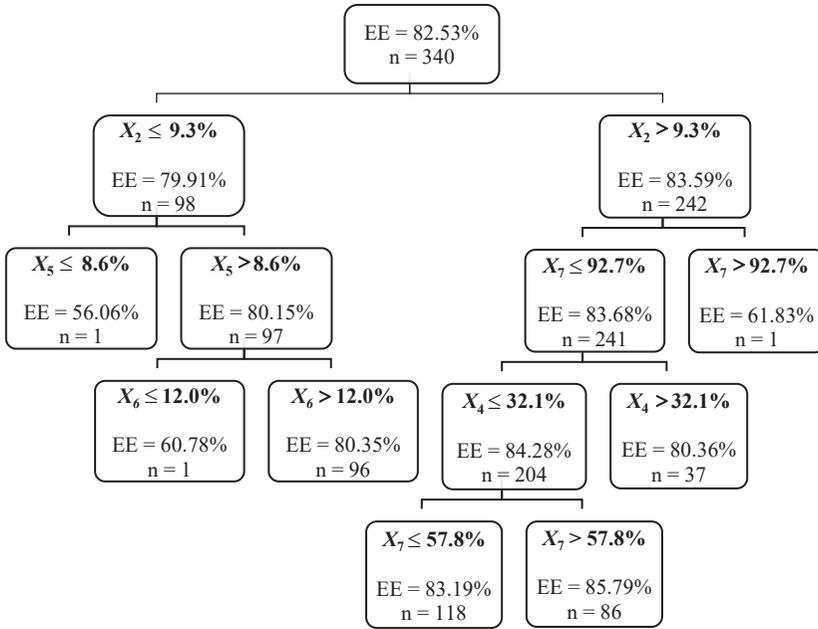
Fig. 1. Costs of cross-validation (CV) and re-substitution (R) for a sequence of 20 trees

Source: own elaboration pursuant to Table 1.

After checking the structure of trees 16 and 15, and the spreads of these predicted values in relation to the observed employment efficiency index, tree number 15 was chosen, for which the spread chart shows how the model clearly separates groups of ‘powiat’ employment offices. Figure 2 presents regression tree number 15.

By interpreting the end nodes (leaves) of tree number 15 using logical partitioning conditions, one can conclude that in 2018:

1) in 96 county employment offices the average employment efficiency index was 80.35% – these offices conducted professional activation of the unemployed,



Where: EE – the average employment efficiency index in %, X_2 – the share of the unemployed with higher education in the total number of the unemployed in %, X_4 – the share of the unemployed with lower secondary education and below in the total number of unemployed in %, X_5 – the share of the unemployed aged 24 years and younger in the total number of the unemployed in %, X_6 – the share of the unemployed aged 55-64 in the total number of unemployed in %, X_7 – the share of the unemployed residing in rural areas in the total number of unemployed in %.

Fig. 2. The regression tree number 15

Source: own elaboration.

for whom the percentage of persons with higher education did not exceed 9.3%, the percentage of persons aged 24 and less was above 8.6%, and the percentage of persons aged 55-64 was above 12%,

2) in 37 county employment offices the average employment efficiency index was 80.36% – these offices conducted professional activation of the unemployed, for whom the percentage of persons with higher education was above 9.3%, the percentage of persons living in rural areas did not exceed 92.7%, and the percentage of persons with at most lower secondary education was above 32.1%,

3) in 118 county employment offices the average employment efficiency index was 83.19% – these offices conducted professional activation of the unemployed, for whom the percentage of persons with higher education was above 9.3%, the percentage of persons with at most lower secondary education was not higher than 32.1%, and the percentage of persons living in rural areas did not exceed 57.8%,

4) in 86 county employment offices the average employment efficiency index was 85.79% – these offices conducted professional activation of the unemployed, for whom the percentage of persons with higher education was above 9.3%, the percentage of persons with at most lower secondary education was not higher than 32.1%, and the percentage of persons living in rural areas was between 57.8% and 92.7%,

5) in three county employment offices, the employment efficiency index was clearly below the average value of this index in other groups, but the reasons for this were different in these offices: in the first the low-efficiency was connected with the percentage of the unemployed with higher education not higher than 9.3%, and the percentage of young unemployed not exceeding 8.6%, while in the second group the reasons of low efficiency should be sought in the low percentage of the unemployed with higher education, the percentage of young unemployed above 8.6% and the percentage of the unemployed aged 55-64 at a level not exceeding 12%, while the third office had low efficiency, because despite the percentage of the unemployed with higher education above 9.3%, the percentage of unemployed living in rural areas exceeded 92.7%.

5. Conclusion

The tool used in the study allowed the separation of groups of county ('powiat') employment offices differing in the level of the employment efficiency index with the simultaneous indication of those demographic features that affected the index in this particular study, i.e. for selected variables in 2018. It turned out that out of 15 explanatory variables (predictors), only the following five can be considered important (according to importance): the share of the unemployed with lower secondary education and below in the total number of unemployed in % (X_4), the share of the unemployed residing in rural areas in the total number of unemployed in% (X_7), the share of the unemployed aged 24 years and younger in the total number of the unemployed in% (X_5), the share of the unemployed with higher education in the total number of the unemployed in% (X_2), the share of the unemployed aged 55-64 in the total number of unemployed in% (X_6). The last of the variables, even though it is only 9th place in the ranking of importance, appeared to be significant in dividing node 5 and enabled the separation of one county employment office, which clearly differed in the level of the employment efficiency index from the others included in this node.

Combining the interpretation of the tree with the purpose of the study, based on the data from 2018 it should be stated that the hypothesis that the higher the percentage of the unemployed with higher education, the easier it is to achieve higher employment efficiency was positively and unambiguously verified (compare the right and left branches of the tree). On the other hand, the hypotheses regarding the age of the unemployed and the place of residence were verified less categorically.

The age of the unemployed has an impact on the level of efficiency in those offices, where the percentage of the unemployed with higher education does not exceed 9.3% and then an increase in efficiency can be observed when the percentage of young unemployed and the percentage of the unemployed aged 55-64 increases. The place of residence has an impact on the employment efficiency in those offices where the percentage of unemployed with higher education exceeds 9.3%. It can then be observed that a very high level of the percentage of the unemployed living in rural areas reduces employment efficiency, while when the level of this percentage is not exceeding 92.7%, an additional feature of the unemployed that affects the efficiency index is the percentage of the unemployed with junior secondary education or a lower one – the level of this percentage above 32.1% causes a decrease in the employment efficiency ratio.

Unfortunately, this study was not able to positively verify the hypotheses regarding the impact of gender and the duration of unemployment for the unemployed on the employment efficiency index. It seems that in this particular case these features did not have a significant impact on the level of employment efficiency.

References

- Bąk, I., Wawrzyniak, K., and Oesterreich, M. (2019). The application of statistical methods to identify factors determining employment effectiveness in district labour offices in Poland. *Acta Universitatis Lodzianis Folia Oeconomica*, 343(4), 235-255.
- Bąk, I., Wawrzyniak, K., and Sobolewski, A. (2018). Przestrzenne zróżnicowanie sytuacji społeczno-gospodarczej a efektywność zatrudnieniowa w powiatowych urzędach pracy w Polsce. *Folia Pomeranae Universitatis Technologiae Stetinensis Oeconomica*, 347(93)4, 17-28.
- Bąk, I., Wawrzyniak, K., and Sobolewski, A. (2019). Modelowanie ekonometryczne efektywności zatrudnieniowej i kosztowej w powiatowych urzędach pracy. In *Mikroekonometria. Teoria i praktyka* (pp. 77-91). Warszawa: Difin.
- Błądowski, P. (ed.). (2008). *Efektywność usług i instrumentów rynku pracy służących podnoszeniu kwalifikacji bezrobotnych w Polsce i wybranych krajach Unii Europejskiej*. Warszawa: Ministerstwo Pracy i Polityki Społecznej. Departament Rynku Pracy.
- Card, D., Kluve, J., and Weber, A. (2010). Active labour market policy evaluations: a metaanalysis. *The Economic Journal*, (120).
- Drela, K. (2013). Efektywność programów wsparcia osób bezrobotnych w Szczecinie. *Ekonomiczne Problemy Usług*, (103). Szczecin: Zeszyty Naukowe Uniwersytetu Szczecińskiego.
- Efektywność podstawowych form aktywizacji zawodowej, realizowanych w ramach programów na rzecz promocji zatrudnienia, łagodzenia skutków bezrobocia i aktywizacji zawodowej w 2018 roku.* (2019). Warszawa: Ministerstwo Rodziny, Pracy i Polityki Społecznej, Departament Funduszy. Retrieved from <https://www.gov.pl/web/rodzina/efektywnosc-form-promocji-zatrudnienia>
- Elastyczność i sprawność rynku pracy. (2006). *Zeszyty BRE Bank – CASE* (87). Warszawa.
- Frederiksson, P., and Johansson, P. (2003). *Employment, mobility, and active labor market programs* (IFAU Working Paper 2003:3). Uppsala. Retrieved from <http://ftp.iza.org/dp2018.pdf>
- Gatnar, E. (2001). *Nieparametryczna metoda dyskryminacji i regresji*. Warszawa: Wydawnictwo Naukowe PWN.

- Gatnar, E. (2008). *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*. Warszawa: Wydawnictwo Naukowe PWN.
- Gatnar, E., and Walesiak, M. (eds.). (2004). *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej.
- Guzman, G. (2014). How effective are active employment policies to reduce unemployment in EU countries? *Atlantic Review of Economics*, (2). Retrieved from <https://www.google.com/url?sa=t&rc=t&ct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKewipmuSG2u3hAhUIZ1AKHwMQBTcQFjAAegQIAhAC&url=https%3A%2F%2Fdialog.unirioja.es%2Fdescarga%2Farticulo%2F5318291.pdf&usq=AOvVaw210SUE4DeeyOw62Xe4Vli>
- Immervoll, H. (2012). *Activation policies in OECD countries: an overview of current approaches*. Labour Market 14. The World Bank. Retrieved from https://www.researchgate.net/publication/254456336_Activation_policies_in_OECD_countries_an_overview_of_current_approaches
- Immervoll, H., and Scarpetta, S. (2012). Activation and employment support policies in OECD countries. An overview of current approaches. *Immervoll and Scarpetta IZA Journal of Labor Policy*, 1(9), 1-20. Retrieved from <https://izajolp.springeropen.com/track/pdf/10.1186/2193-9004-1-9%20>
- Internetowy Podręcznik Statystyki. Drzewa klasyfikacyjne i regresyjne. Copyright Stat-Soft, Inc., 1984-2011. Retrieved from www.statsoft.pl/stathome.html
- Kisiel, R., and Nowińska, S. (2010). Efektywność aktywnej polityki rynku pracy na tle doświadczeń międzynarodowych. *Polityka Społeczna*, (5-6), 7-12.
- Kluve, J. (2006). The effectiveness of European active labor market policy. *IZA Discussion Paper Series*. Retrieved from <http://ftp.iza.org/dp2018.pdf>
- Kluve, J., Puerto, S., Robalino, D., Romero, J. M., Rother, F., Stoter, J., Weidenkaff, F., and Witte, M. (2019). Do youth employment programs improve labor market outcomes? A quantitative review. *World Development*, 114, 237-253. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0305750X18303905>
- Piowowski, R. (2017). Czy wskaźnik efektywności zatrudnieniowej ulega paradoksowi efektywności. *Acta Universitatis Lodzianae. Folia Oeconomica*, 1(327), 187-202.
- Rękas, M. (2011). Efektywność aktywnej polityki rynku pracy (ALMP) w Polsce w latach 2005-2009. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu (Ekonomia)*, 168(2), 163-175.
- Siewiera, K. (ed.). (2019). *Model badania długoterminowej skuteczności i efektywności działań powiatowych urzędów pracy*. Szczecin: Stowarzyszenie Czas Przestrzeń Tożsamość.
- Ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy (Dz. U. z 2019 r., poz. 1482 z późn. zm.)
- Vooren, M., Haelermans, C., Groot, W., and Brink H. M. (2019). The effectiveness of active labor market policies: a metaanalysis. *Journal of Economic Surveys*, 33(1), 125-149. Retrieved from <https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12269>
- Walesiak, M., Gatnar, E. (eds.). (2009). *Statystyczna analiza danych z wykorzystaniem programu R*. Warszawa: Wydawnictwo Naukowe PWN.
- Wiśniewski, Z., and Zawadzki, K. (eds.). (2011). *Efektywność polityki rynku pracy w Polsce*. Toruń: Wojewódzki Urząd Pracy, Uniwersytet Mikołaja Kopernika.

WYKORZYSTANIE DRZEW REGRESYJNYCH W BADANIU ZALEŻNOŚCI POMIĘDZY WSKAŹNIKIEM EFEKTYWNOŚCI ZATRUDNIENIOWEJ A STRUKTURĄ BEZROBOTNYCH W POWIATOWYCH URZĘDACH PRACY

Streszczenie: Celem artykułu jest analiza zależności pomiędzy wskaźnikiem efektywności zatrudnieniowej a strukturą bezrobotnych w powiatowych urzędach pracy w Polsce. Sformułowano pięć hipotez badawczych związanych z realizacją celu badania. Za najważniejsze cechy charakteryzujące bezrobotnych uznano: wykształcenie, wiek, czas pozostawania bez pracy, płeć, miejsce zamieszkania. Jako narzędzie badawcze zastosowano drzewa regresyjne, dzięki którym możliwe jest pogrupowanie powiatowych urzędów pracy różniących się między sobą wartościami wskaźnika efektywności zatrudnieniowej z jednoczesnym wskazaniem wariantów zmiennych niezależnych (cech diagnostycznych) wpływających na jego poziom. Po przeprowadzeniu badań udało się pozytywnie i jednoznacznie zweryfikować hipotezę, że im wyższy odsetek bezrobotnych z wykształceniem wyższym, tym łatwiej osiągnąć wyższą efektywność zatrudnieniową. W sposób mniej jednoznaczny, tylko przy określonych warunkach, zweryfikowano hipotezy dotyczące wieku bezrobotnych oraz miejsca zamieszkania. Natomiast nie udało się pozytywnie zweryfikować hipotez dotyczących wpływu płci oraz czasu pozostawania bez pracy bezrobotnych na wskaźnik efektywności zatrudnieniowej.

Słowa kluczowe: wskaźnik efektywności zatrudnieniowej, struktura bezrobotnych, powiatowe urzędy pracy, drzewa regresyjne.