ANNALS OF THE POLISH ASSOCIATION
OF AGRICULTURAL AND AGRIBUSINESS ECONOMISTS

received: 07.11.2019
acceptance: 20.11.2019
published: 15.12.2019

Annals PAAAE • 2019 • Vol. XXI • No. (4)

JEL codes: Q10

DOI: 10.5604/01.3001.0013.5836

JAKUB STANISZEWSKI

Poznań University of Economics and Business, Poland

CAN STRUCTURAL GENOTYPES OF AGRICULTURE BE DISTINGUISHED IN INDIVIDUAL REGIONS OF THE EUROPEAN UNION?[1]

Key words: structural genotype, latent profile analysis, structural changes

ABSTRACT. The aim of this study is to verify the hypothesis that structural genotypes exist in EU agriculture. To verify the hypothesis, Eurostat data from the 2010 and 2016 Farm Structure Survey for 258 NUTS 2 regions were used. 21 structural variables were calculated on the basis of the data and then they were used to carry out a latent profile analysis. At the last stage of the research, the influence of structural genotypes on the productivity of production factors was verified using the MANOVA method. The obtained results confirm the existence of 5 structural genotypes: (1) small farms and a polarized agrarian structure, (2) very small, non-specialized, labour-intensive farms and a polarized agrarian structure, (3) large farms and non-polarized agrarian structure, (4) average farms, (5) very large farms – outliers. What differed the most among farms was the concentration of production. Additionally, the relative stability of the adherence of the regions to a given genotype was confirmed, as well as the influence of this relation on the productivity, in particular on labour productivity. Any transitions between clusters concerned, in the vast majority, those regions where the scale of production was higher.

INTRODUCTION

This paper is devoted to an attempt to determine the lasting impact of structural conditions on the processes of agricultural development in the European Union. These considerations were inspired by the concept of functional genotypes of cities presented by Krzysztof Gwosdz [2013]. They were originally defined as a set of functions having a decisive impact on the formation of a city or the location system of cities and its / their development [Krzysztofik 2012, cited in: ibidem, p. 22]. In the context of agricultural sector research, a set of structural features of the agricultural sector determining its development can be considered its structural genotype [Staniszewski, Czyżewski 2018].

Apart from the academic value, knowledge on the existence and stability of structural genotypes also has a number of practical applications. Firstly, it helps to identify regions with similar development factors but better economic results, which may constitute a point of reference and a model of "good practices" in the process of strategic planning. Secondly, it can provide information on the effectiveness of agricultural and structural

---

[1] The project was funded by the National Science Centre (Poland) (Dec. 2018/29/N/HS4/01799).

policies in eliminating structural constraints on development. Thirdly, it is possible to use the identified similarities between regions as a basis for the implementation of common political postulates on the European Union forum.

Existing research in the area of agricultural typology of EU regions has not been comprehensive. The regions were clustered on the basis of characteristics regarding only one of the production factors: labour [Giannakis, Bruggeman 2018] or land [Loughrey et al. 2016, Estel et al. 2016], or only one of the directions of agricultural production [Guth 2015]. Studies were conducted to, among others, identify peripheral regions [Camaioni et al. 2013], similarities in the implementation models of the common agricultural policy [Czyżewski, Smędzik-Ambroży 2017], agricultural models [Kryszak 2018], or factors affecting the implementation of sustainable intensification of agricultural production [Scherer et al. 2018]. Three strategies can be identified in terms of the research approach and data used. The first one was based on data not taking into account the administrative division of the continent, available with an accuracy of 1 km2. Such research was conducted using the Kohonen network method [Estel et al. 2016, Van der Zanden et al. 2016]. The second strategy is based on generally available data at the level of FADN and NUTS 2 or 3 regions, subjected to statistical clustering procedures using hierarchical methods [Camaioni et al. 2013], the *k*-means method [Giannakis, Bruggeman 2018] or the Ward method [Czyżewski, Smędzik-Ambroży 2017, Guth 2015, Kryszak 2018, Scherer et al. 2018, Staniszewski, Czyżewski 2018]. In the third approach, Moran's autocorrelation coefficient is used for clustering [Loughrey et al. 2016].

The innovativeness of the presented studies as compared to the previous ones results from three aspects. Firstly, the structures are approached in a comprehensive way. They will be analysed in terms of three basic dimensions – concentration, specialization and focus of production - identified in earlier review studies [Staniszewski, Czyżewski 2019]. Secondly, the innovative latent profile analysis (LPA), with an algorithm for selecting clustering variables, will be applied [Marbac, Sedeki 2017]. Compared to the methods used so far, its advantages include the possibility of a parametric determination of the optimal number of clusters and optimal set of clustering variables, as well as the assessment of the accuracy of the fit. Thirdly, the research is based on the latest farm structure survey (FSS) data for 2016, published in early 2019.

MATERIAL AND METHODS

To verify the hypothesis on the occurrence of structural genotypes of agriculture in EU regions, it is necessary to verify three sub-hypotheses:

- H1: It is possible to specify internally consistent clusters of regions which are, at the same time, significantly different in terms of their structural characteristics of agriculture.
- H2: The resulting clustering pattern is stable over time.
- H3: The clustering determines the development of the agricultural sector.

In order to confirm the main hypothesis, all three of the above-mentioned sub-hypotheses should also be confirmed.

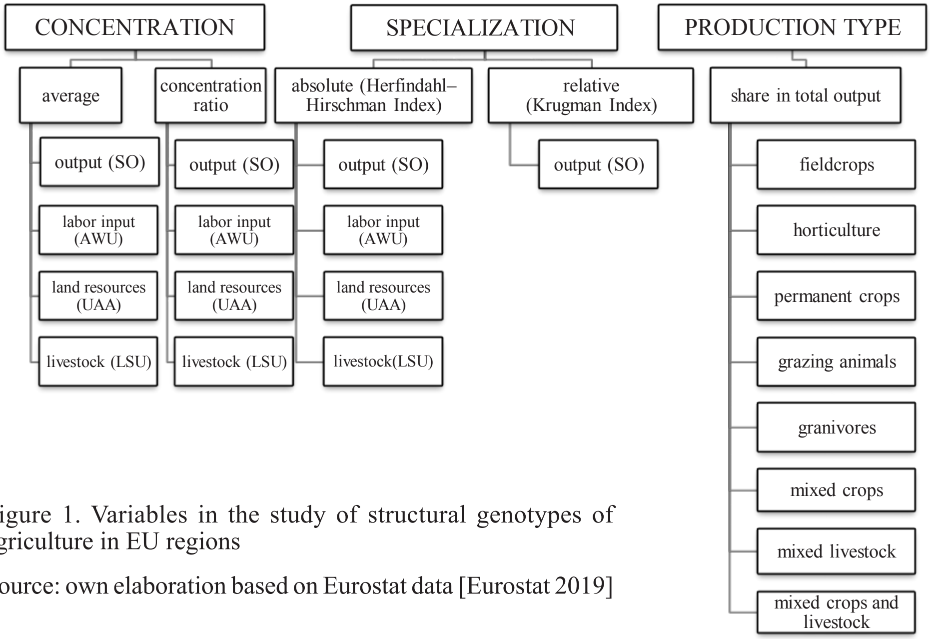


Figure 1. Variables in the study of structural genotypes of agriculture in EU regions

Source: own elaboration based on Eurostat data [Eurostat 2019]

In the first stage of the research, a database was created containing 21 variables describing the structure of agriculture and agricultural production in 258 NUTS 2[2] regions in 2010 and 2016. 2010 was selected as the baseline year due to methodological changes introduced, then in relation to previous editions. These mainly involved the change in the minimal size of a farm eligible for the survey, which makes the data from the survey editions before and after 2010 not fully comparable. The structural variables applied are presented in Figure 1.

The indicators of concentration, Hirschman-Herfindahl and Krugman, used at this stage of work are standard indicators used for structure analysis. The rationale for applying them and relevant formulas can be found in the work of Jakub Staniszewski and Andrzej Czyżewski [2019].

On the basis of the data processed this way, clusters were formed using the method of latent profile analysis proposed by Matthieu Marbac and Mohammed Sedki. This method focuses on the selection of an optimal set of clustering variables, i.e. one that only contains essential variables and excludes those that are independent from them. In the applied approach, subsequent combinations of variables are evaluated using MICL (Maximum Integrated Complete-data Likelihood) criterion. The advantage of this approach over the use of classical measures such as BIC or ICL is that it does not require estimating all combinations of clustering variables using the highest reliability method, which significantly accelerates calculations. The algorithm for selecting variables is described

---

[2] The “urban” regions were excluded from the survey: Brussels, Prague, Paris, Berlin, Bremen, Hamburg, Vienna, Inner-London and Outer-London, as well as the overseas territories of Guyana, Réunion and Martinique.

in detail by Matthieu Marbac and Mohammed Sedki [2017], and can be applied as part of the R VarSelLCM package available in the CRAN repository.

In the next stage of the study, the stability of structural genotypes over time was determined. This was done using the adjusted Rand index (ARI) [Hubert, Arabie 1985]. It takes values from the range of 0-1 and grows with increasing similarity of clustering results.

In the last stage of the research, the links between identified structural genotypes and the development of the agricultural sector were determined. The level of development was evaluated on the basis of productivity of land, labour and livestock, understood as the value of standard output per unit of input. The multivariate analysis of variance (MANOVA) was used for contrasts.

RESULTS OF THE STUDIES

Latent profile analysis indicated that only 4 out of 21 proposed structural variables are relevant for clustering. These are the variables describing the concentration of production in light of the average value of standard output and available labour, land and livestock. The influence of particular variables on the results of clustering is presented in Figure 2.

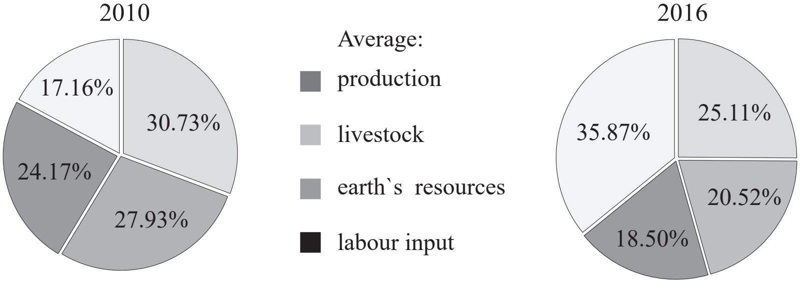


Figure 2. Discriminatory power of clustering variables

Source: own elaboration based on Eurostat data and the VarSelLCM package

At this stage of the research, it is also necessary to determine the number of analysed clusters. For this purpose, the results of clustering were simulated for models with 2 to 6 groups. The model fit indicators for the estimated models are presented in Table 1.

The obtained results suggest a division into 5 clusters. In the case of 2010 data, choosing the 6 cluster model causes a decrease in the MICL criterion value, which is undesirable. In the case of models for 2016, although parametric analysis justifies using 6 clusters, such a strategy would lead to the formation of a cluster with only 2 regions, which impedes interpretation and is inaccurate. The final results of the clustering are presented in Figure 3.

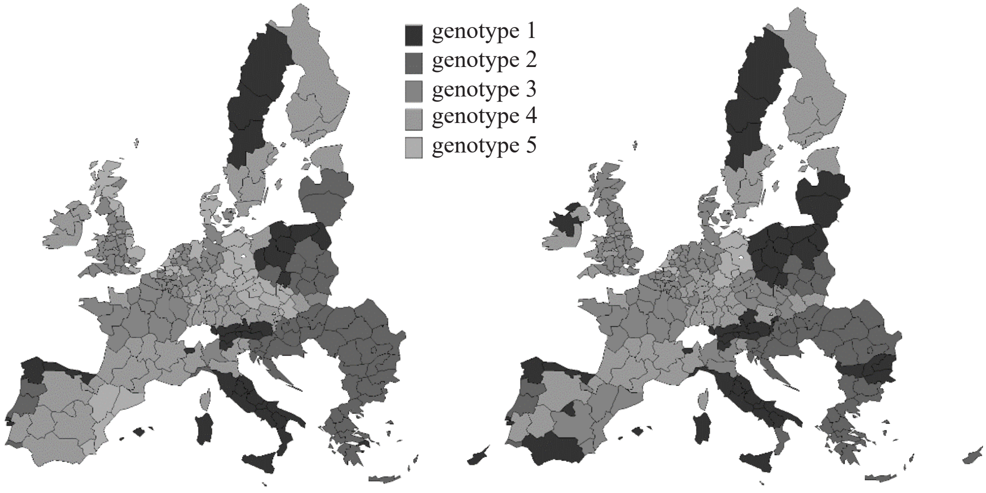
The resulting structural genotypes can be summarized on the basis of features detailed in Figure 4. Genotype 1 covers regions with a small average farm size and a polarized structure of resources, which is reflected in high concentration ratios linked to low average values. This means that there are many small farms which lower the average and few large farms which concentrate most of the resources (apart from labour). These regions are also characterised by specialisation in permanent crops and grazing animal production.

Table 1. Model fit indicators with a different number of clusters

Clasters	AIC	BIC	MICL	Average probability of misclassification	LRT
2010					
2 clusters	-2,338.12	-2,428.72	-3,331.11	0.008272771	
3 clusters	-2,120.35	-2,226.94	-3,152.93	0.02870452	$p < 0.001$
4 clusters	-2,038.31	-2,160.89	-3,103.67	0.03180165	$p < 0.001$
5 clusters	-1,970.48	-2,109.04	-3,075.44	0.04606123	$p < 0.001$
6 clusters	-1,932.45	-2,087.01	-3,077.81	0.04469715	$p < 0.001$
2016					
2 clusters	-2,656.10	-2,746.70	-3,735.23	0.01232289	
3 clusters	-2,334.99	-2,441.58	-3,452.80	0.0233902	$p < 0.001$
4 clusters	-2,239.18	-2,361.76	-3,372.63	0.02965636	$p < 0.001$
5 clusters	-2,152.49	-2,291.06	-3,314.48	0.03375963	$p < 0.001$
6 clusters	-2,103.43	-2,257.98	-3,298.68	0.04274379	$p < 0.001$

Signs: AIC – Akaike Information Criterion, BIC – Schwarz Information Criterion, MICL – Maximum Integrated Complete-data Likelihood, LRT – Likelihood Ratio Test

Source: own elaboration based on Eurostat data and the VarSelLCM package

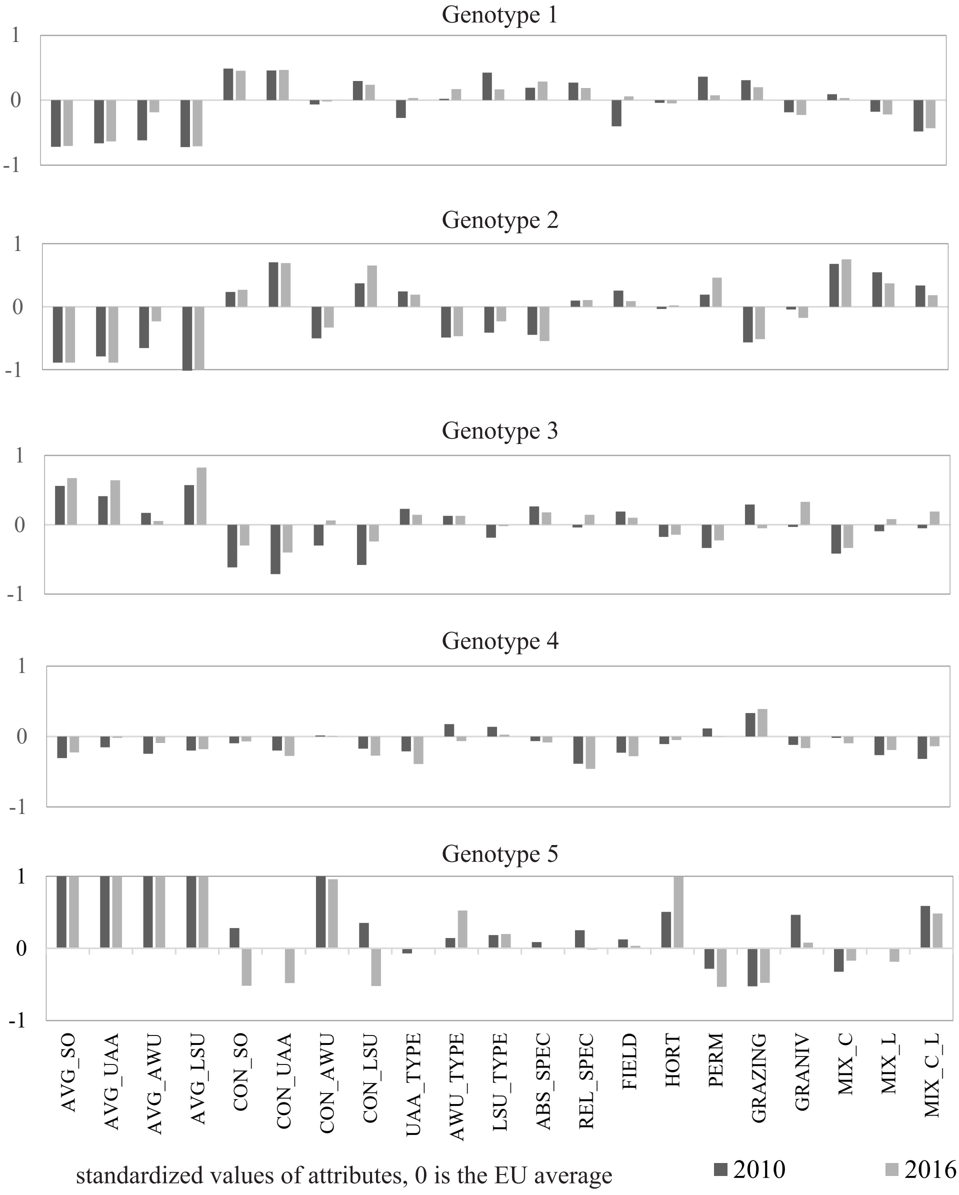


2010

2016

Figure 3. Results of clustering EU regions by structural genotypes of agriculture in 2010 and 2016

Source: own elaboration based on Eurostat data and the VarSelLCM package



AVG_ – standard output (SO), utilised agricultural area (UAA), labour input (AWU), livestock (LSU); CON_ – concentration of output (SO), utilised agricultural area (UAA), labour input (AWU), livestock (LSU); _TYPE – utilised agricultural area (UAA), specialisation in labour input utilisation (AWU), livestock (LSU), ABS_SPEC – absolute specialisation of production, REL_SPEC – relative specialisation of production; FIELD – share of field crop production in SO, HORT – horticultural production, PERM – permanent crops, GRAZING – grazing animals, GRANIV – granivores, MIX_C – mixed crop production, MIX_L – mixed livestock production, MIX_C_L – mixed crop and livestock production

Figure 4. Characteristics and variability of structural genotypes of agriculture in EU regions

Source: own elaboration based on Eurostat data and the VarSelLCM package

Genotype 2 includes regions where farms are even smaller and more polarised and production is labour-intensive. This can be concluded by comparing the degree of concentration of labour input with the concentration of other factors. It is much smaller, which means that the largest farms absorb most of the resources and the labour factor remains “immobile” in small farms. Furthermore, since most farms there are characterised by mixed production types, they are not specialised.

Genotype 3 is representative of regions where agricultural production takes place on large farms, among which the resources are evenly distributed. They are also characterised by a relatively low share of permanent crops. Genotype 4 can be defined as regions where agricultural production is average. The values of all structural indicators are close to the average and the relative specialisation indicator is low, which shows the deviation of the production structure in the region from the EU average. Genotype 5 comprises the outlying regions, mainly in terms of above-average large scale production in farms. This group is also characterised by a high level of concentration of the labour factor and a high share of horticultural crops and granivores.

The next stage of the research included the identification of grouping stability within structural genotypes. The value of the corrected Rand index for clustering in the two analysed periods was 0.5806. Out of 258 regions, 59 changed their categorization. Detailed information on the shifts is presented in Table 2. The largest changes took place within genotype 5, from which most regions moved to genotype 3. In the end, only eastern German states remained within the “outlying” group, characterized by a significant concentration in agriculture, linked to the surviving agrarian structure from socialist times, and some Dutch regions, where due to scarcity of land, agriculture is highly intensive. It can, therefore, be concluded that the regions of Czechia, the United Kingdom, Spain, Belgium and Denmark experienced a process of relative deconcentration. Apart from this, no mass “migration” between genotypes was observed.

The last stage of the research involved determining the relationship between structural genotypes and the level of agricultural development represented by input productivity indicators. The relationships were studied with the use of the MANOVA method. It requires fulfilling a set of assumptions:

Table 2. Shifts in the clustering of regions by structural genotypes of agriculture between 2010 and 2016

Genotype		2010	Increase					–	2016
			1	2	3	4	5		
Decrease	1	40		3				3	52
	2	53	9					9	47
	3	61				7		7	88
	4	63	6		4			10	60
	5	41			30			30	11
		+	15	3	34	7	0	59	

Source: own elaboration based on Eurostat data and the VarSelLCM package

- a normal distribution of variables in groups – the method is resistant to failure to meet this assumption if the number of observations in groups is greater than 30, as in the case of this study;
- homogeneity of variance in groups – meeting this assumption was tested using the Brown-Forsythe and Levene's test; the variance proved to be non-homogenous, which means that a more rigorous level of significance should be used for interpretation (p = 0.01);
- no outliers – the Mahalanobis distances showed 3 outlier observations in 2010 and 5 in 2016, which is acceptable at N = 258;
- lack of collinearity of variables – correlation coefficients of the dependent variables did not exceed 0.4 in 2010, and 0.43 in 2016, thus this assumption can be considered fulfilled.

Therefore, MANOVA proved to be feasible to conduct on the analysed dataset. The results of the obtained estimates are presented in Table 3.

Table 3. MANOVA results for differences in the productivity of production factors in EU regions with different structural genotypes

		2010			2016		
	Df	Pillai	approx F	Pr(>F)	Pillai	approx F	Pr(>F)
Total	4	0.851	25.028	< 2.2E-16	0.877	26.128	< 2,2E-16
	Df	F		Pr(>F)	F		Pr(>F)
Land	4	6.2699		7.96E-05	17.778		7.087E-13
Labour	4	116.87		< 2.2E-16	107.69		< 2.2E-16
Livestock	4	10.84		3.98E-08	7.82		5.91E-06

Source: own elaboration based on Eurostat data

The estimation indicates the existence of statistically significant differences between genotypes in the area of productivity of all analysed production factors. In order to obtain more detailed information on which pairs of genotypes these differences occur between, it was necessary to perform post-hoc tests, the results of which are presented in Table 4.

Tukey's HSD test indicates that half of the variation (16 out of 32 significant differences) comes from comparisons of other genotypes with genotype 5, representing outlier observations. This dictates a cautious approach to the obtained results. On the other hand, significant differences for each of the remaining pairs, in both periods can be indicated. This provides the basis for a positive verification of the third hypotheses. The analysis of the frequency of differences between the productivity of particular production factors shows that they occurred most frequently within the productivity of the labour factor (18 out of 32), less frequently in the case of livestock (8 out of 32), and least frequently in the case of the land factor (6 out of 32). The differences in the productivity of the labour factor were also the most "persistent", and did not disappear in any of the cases. The differences regarding livestock disappeared twice, and appeared in the case of one pair. Differences in the productivity of the land factor were maintained in two cases and revealed in two others.

Table 4. Results of Tukey's HSD test of differences in productivity of production factors in EU regions with different structural genotypes

Genotype	1	2	3	4
2	AWU10, LSU10, AWU16, LSU16			
3	AWU10, LSU10, AWU16	AWU10, AWU16		
4	AWU10, AWU16	AWU10, LSU10, AWU16	AWU10, AWU16	
5	AWU10,UAA16, AWU16	UAA10, AWU10 LSU10, UAA16, AWU16, LSU16	UAA16, LSU16	UAA10, AWU10 UAA16, AWU16, LSU16

Signs: UAA – utilised agricultural area, AWU – labour, LSU – livestock, 10, 16 – year of analysis, only differences for which $\alpha < 0.01$ were included, the darker the cell, the more differences are significant

Source: own elaboration based on Eurostat data

CONCLUSIONS

In the course of the research, 5 relatively stable clusters of regions were identified on the basis of variables characterising the structural conditions of agricultural production in EU regions. At the same time, adherence to the clusters significantly determined the productivity of production factors. These results allow for the positive verification of all three sub-hypotheses, and thus also the main hypothesis concerning the occurrence of structural genotypes in EU agriculture. The identified genotypes are: (1) small farms and polarised agrarian structure, (2) very small, non-specialised, labour-intensive farms and polarised agrarian structure, (3) large farms and non-polarised agrarian structure, (4) average farms, (5) very large farms – outlier observations.

At the same time, an in-depth analysis allowed to formulate a number of additional conclusions. Firstly, the feature that differentiates farms the most in EU regions is the concentration of resources and production. Ultimately, only the variables representing this feature had an impact on clustering. Secondly, structural changes taking place in agriculture consisted mainly of the transition from the group of regions with very large farms (5) to the group with a large farms (3). One of the reasons for this may be the evolution of direct payments of the common agricultural policy, which are subject to modulation mechanisms, rewarding medium-sized holdings. In the genotypes describing regions with small-scale farming (1 and 2), the movements only took place between them or consisted of moving from the group of average farms to small ones (from 4 to 2). This may be due to a lack of farms with sufficient economic strength to initiate the concentration process. Thus, it can be concluded that it is easier to achieve structural changes towards reducing the scale of production rather than increasing it. However, it is important to be aware that the conclusions on the causes of this situation are only a premise for further research and require additional empirical verification. Thirdly, particular attention should be paid to the problem of differentiation in labour productivity in agriculture, which is the most persistent and, at the same time, translates directly into variations in agricultural income.

Looking to formulate recommendations for agricultural policy, it can be pointed out that there is a need to support structural changes towards increasing the scale of production in regions within genotypes (1) and (2), as well as improving labour productivity, which constitutes the main difference. Regions in central and eastern European countries, which in the vast majority of cases belong to these genotypes, may use Italian, Austrian and Swedish regions as a model, where economic performance is clearly better under similar structural conditions.

BIBLIOGRAPHY

Camaioni Beatrice, Roberto Esposti, Antonello Lobianco, Francesco Pagliacci, Franco Sotte. 2013. How rural is the EU RDP? An analysis through spatial fund allocation. *Bio-Based and Applied Economics* 2 (3): 277-300.

Czyżewski Bazyli Smędzik-Ambroży Katarzyna. 2017. The regional structure of the CAP subsidies and the factor productivity in agriculture in the EU 28. *Agricultural Economics (Zemědělská Ekonomika)* 63 (4): 149-163. DOI: 10.17221/302/2015-AGRICECON.

Estel Stephan, Tobias Kuemmerle, Christian Levers, Matthias Baumann, Patrick Hostert. 2016. Mapping cropland-use intensity across Europe uUsing MODIS NDVI time series. *Environmental Research Letters* 11 (2): 24015. DOI: 10.1088/1748-9326/11/2/024015.

Eurostat. 2019. *Farm indicators by agricultural area, type of farm, standard output, legal form and NUTS 2 regions,* http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ef_m_farmleg&lang=en, accesss: 8.08.2019.

Giannakis Elias, Adriana Bruggeman. 2018. Exploring the labour productivity of agricultural systems across european regions: A multilevel approach. *Land Use Policy* 77 (March): 94-106. DOI: 10.1016/j.landusepol.2018.05.037.

Guth Marta. 2015. Diversity of FADN milk farms in the regions of the European Union in 2011. *Roczniki Naukowe SERiA* XVII (3): 119-124.

Gwosdz Krzysztof. *Pomiędzy starą a nową ścieżką rozwojową* (Between the old and new development paths). Kraków: IGiGP UJ, 2014.

Hubert Lawrence, Phipps Arabie. 1985. Comparing partitions. *Journal of Classification* 2 (1): 193-218. DOI: 10.1007/BF01908075.

Kryszak Łukasz. 2018. Agricultural models in EU FADN regions and changes in farm productivity and incomes. *Journal of Agribusiness and Rural Development* 4 (50): 403-413. DOI: 10.17306/J.JARD.2018.01105 Journal.

Krzysztofik Robert. 2012. *Umiastowienie obszaru Polski od XIII do XXI wieku. Interpretacja geograficzno-historyczna* (Formation of the territory of Poland from the 13th to the 21st century. Geographical and historical interpretation). Katowice: WN Śląsk.

Loughrey Jason, Trevor Donnellan, John Lennon. 2016. *The inequality of farmland size in Western Europe*. [In] Discussion paper prepared for presentation at the 90th Annual Conference of the Agricultural Economics Society. University of Warwick, England 4-6 April 2016, 8: 1149-1151. DOI: 10.1093/cid/ciy757/5087840.

Marbac Matthieu, Mohammed Sedki. 2017. Variable selection for model-based clustering using the Integrated Complete-Data Likelihood. *Statistics and Computing* 27 (4): 1049-1063. DOI: 10.1007/s11222-016-9670-1.

Scherer Laura A., Peter H. Verburg, Catharina J.E. Schulp. 2018. Opportunities for sustainable intensification in European agriculture. *Global Environmental Change* 48 (April): 43-55. DOI: 10.1016/j.gloenvcha.2017.11.009.

Staniszewski Jakub, Andrzej Czy*żewski*. 2018. Economic factors underpinning the structural genotypes of agriculture development in the European Union after 2004. *Journal of Agribusiness and Rural Development* 50 (4): 445-454. DOI: 10.17306/j.jard.2018.00436.

Staniszewski Jakub, Andrzej Czyżewski. 2019. *Rolnictwo Unii Europejskiej w procesie zrównoważonej intensyfikacji* (European Union agriculture in the process of sustainable intensification). Warszawa: PWN.

Zanden Emma H., Van der Christian Levers, Peter H. Verburg, Tobias Kuemmerle. 2016. Representing composition, spatial structure and management intensity of European agricultural landscapes: a new typology. *Landscape and Urban Planning* 150: 36-49. DOI: 10.1016/j.landurbplan.2016.02.005.

***

CZY MOŻNA MÓWIĆ O GENOTYPACH STRUKTURALNYCH ROLNICTWA W REGIONACH UNII EUROPEJSKIEJ?

Słowa kluczowe: genotyp strukturalny, latent profile analysis, zmiany strukturalne

ABSTRAKT

Celem opracowania jest weryfikacja hipotezy o występowaniu w rolnictwie UE genotypów strukturalnych. Do jej weryfikacji wykorzystano dane Eurostat: „Badania Struktury Gospodarstw Rolnych" dla 258 regionów NUTS 2 w latach 2010 i 2016. Na podstawie danych obliczono 21 zmiennych strukturalnych, które zostały następnie wykorzystane do przeprowadzenia analizy profili ukrytych. Na ostatnim etapie badań zweryfikowano oddziaływanie genotypów strukturalnych na produktywność czynników wytwórczych z wykorzystaniem metody MANOVA. Uzyskane wyniki potwierdzają istnienie 5 genotypów strukturalnych: (1) małe gospodarstwa i spolaryzowana struktura agrarna, (2) bardzo małe, niewyspecjalizowane, pracochłonne gospodarstwa i spolaryzowana struktura agrarna, (3) duże gospodarstwa i niespolaryzowana struktura agrarna, (4) gospodarstwa przeciętne, (5) bardzo duże gospodarstwa – obserwacje odstające. Cechą, która najsilniej różnicowała gospodarstwa była koncentracja produkcji. Dodatkowo potwierdziła się względna stałość przynależności regionu do danego genotypu, a także wpływ tej przynależności na wydajność czynników wytwórczych, w szczególności na wydajność pracy. Ewentualne przejścia pomiędzy grupami dotyczyły w zdecydowanej większości regionów o większej skali produkcji.

AUTHOR

JAKUB STANISZEWSKI, PHD
ORCID: 0000-0001-8074-0911
Poznań University of Economics and Business
Department of Macroeconomics and Agricultural Economics
10 Niepodległości Av., 61-875 Poznań, Poland