

Selection of the set of areal units for economic regional research on the land use: a proposal for Aggregation Problem solution

Michał Bernard PIETRZAK^{1*} and Adam P. PALCERZAK²

Authors' affiliations and addresses:

¹ Nicolaus Copernicus University in Toruń,
Poland: Faculty of Economic Sciences and
Management, Department of Econometrics and
Statistics, ul. Gagarina 13A, 87-100 Toruń,
Poland
e-mail: pietrzak.science@gmail.com

² University of Warmia and Mazury in Olsztyn,
Poland
University of Entrepreneurship and Law, Czech
Republic
Faculty Economic Sciences, Department Market
and Consumption, pl. Cieszyński 1/327, 10-720
Olsztyn, Poland
e-mail: a.balcerzak@uwm.edu.pl

*Correspondence:

Michał Bernard Pietrzak, Department of
Econometrics and Statistics, ul. Gagarina 13A,
87-100 Toruń, Poland
e-mail: pietrzak.science@gmail.com

Funding information:

The National Science Centre, Poland
Grant Number, 2015/17/B/HS4/01004

Acknowledgement:

The research is funded by the National Science
Centre, Poland under the research project no.
2015/17/B/HS4/01004.

How to cite this article:

Pietrzak, M. B. and Balcerzak, A. B. (2021).
Selection of the set of areal units for economic
regional research on the land use: a proposal for
Aggregation Problem solution. *Acta Montanistica
Slovaca*. Volume 26 (2) 222-234

DOI:

<https://doi.org/10.46544/AMS.v26i2.04>

Abstract

The knowledge of the spatial development of phenomena is crucial in the case of research in economics, geological survey, mining, earth resources and geography. In the literature one can diagnose an important methodological and implementation gap concerning the selection of the set of areal units within the Aggregation Problem. The issue relates to determining boundaries of areal units (regions), whose properties are described by spatial data. The boundaries of areas should be established in such a way that a given analyzed phenomenon is influenced by the same main causes. Only in this case, the analyzed spatial data will properly reflect the impact of main causes, the properties of phenomena and dependencies between them. This means that determining the proper boundaries of areas is a necessary condition for receiving correct conclusions (e.g. delimiting metropolitan areas, assessing mineral resource potential and deposits, or assessing the dynamics of surface processes).

From this perspective, the main objective of the article is presenting the proposal for solving the Aggregation Problem, where as the case study the economic analysis of agrarian resources and structure is used. The solution to the problem will lead to establishing the system of macroregions, where the obtained proposal of a system of four sets of areal units is important from the point of view of spatial research. The main added value of the research and its specific contribution to the literature is based on the fact that the proposed solution to the Aggregation Problem can be considered as universal, which is not limited to selected scientific disciplines. The methodology presented in the article can be effectively applied to other spatial research in the field of geology and mining, where the most appropriate research field is the issue of locating areas with appropriate properties or areas which are affected by given analysed phenomena.

Keywords

economic regional research, spatial data, Aggregation Problem solution, Modifiable Areal Unit Problem (MAUP), agrarian structure, land use



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

The subject of the article concerns the issue of ensuring the correctness of the performed spatial research. Spatial research is the basis for solving important research problems in such scientific disciplines as economics, geography, geological survey, mining and earth resources (e.g. delimitation of metropolitan areas, assessment of mineral resource potential and deposits, or assessment of the dynamics of surface processes). The issue of the correctness of spatial research will be considered in the context the selection of the set of areal units within the Aggregation Problem. The article discusses such a problem due to the fact, that the basis of each regional research is the selection of the set of areal units and the spatial data assigned to it. Only in the next steps the appropriate spatial analysis can be performed. From the methodological perspective, the issue of choosing the set of areal units is discussed in the literature as part of the Modifiable Areal Unit Problem (MAUP) (Anselin 1988; Arbia, 1989; Tobler, 1989; Paelinck, 2000; Flowerdew, 2011; Pietrzak, 2018a, 2019). The essence of the MAUP concerns the possibility of obtaining various research outcome as a result of changing the set of areal units or performing aggregation of spatial data (Anselin, 1989; Arbia, 1989). The literature considers two aspects of MAUP: Aggregation Problem and Scale Problem (Openshaw & Taylor, 1979). The current article focuses on the Aggregation Problem, which is related to the possibility of obtaining different results depending on the choice of the set of areal units at the same aggregation level (Openshaw, 1984). The solution of Aggregation Problem enables to determine the appropriate set of areal units at the selected level of aggregation, the use of which will ensure correct results for regional research. The boundaries of areal units should be established in such a way that the analyzed phenomenon is influenced by the same main causes. Only in this case, the analyzed spatial data will properly reflect the impact of main causes, the most important properties of phenomena and dependencies between them.

Managing an industrial enterprise, implementing complex geographical infrastructure projects or conducting regional policy requires from decision-makers an interdisciplinary approach. Often, knowledge from various scientific disciplines is used. To improve the effectiveness of decision making process the solutions for the encountered problems must be based on the high quality spatial research. In that case the high quality assessment of spatial differentiation of the socio-economic situation, physical conditions and the wealth of raw materials for selected areal units (regions) can be considered as the common issue. Therefore, it should be emphasized that the characteristic feature of regional research is the inclusion of the spatial dimension (Sánchez-López et al. 2019; Formánek, 2019; Semenکو et al., 2019), thus, gaining specific knowledge about selected areal units (Meyer & Meyer, 2019; Shkolnyk et al., 2019). In the case of industrial enterprises, this knowledge is one of the factors that allow to achieve a competitive advantage, introduce new innovations or make successful investments (Kijek & Matras-Bolibok, 2019). In the case of central administration or local government, it is needed for effective implementation of long term sustainable development policies or practically speaking implementing complicated infrastructural projects (Bednář, Halásková, 2018; Szopik-Depczyńska et al., 2018).

From the application perspective, the subject of the article will concern an attempt to solve the Aggregation Problem related to spatial economic research on the land resources and agrarian structure in Poland. A system of appropriate sets of areal units will be determined at subsequent levels of aggregation. A common feature of all adopted sets of areal units should be that their individual regions may differ significantly in the values of diagnostic variables, but within their own borders they are spatially homogeneous. It should be emphasized that the use of a system of appropriate sets of areal units in regional studies should ensure a correct analysis of economic phenomena, which can provide valuable information for further decision making or significant policy implications. This is due to the fact that spatial data assigned to such a set of areal units are characterized by the property of causal homogeneity. In such a situation, the analyzed spatial data correctly reflect the influence of the main causes. However, the main value added of the research and a specific contribution to the literature is not only restricted to spatial economic research concerning the discussed case study. The proposed solution to the Aggregation Problem is universal in the sense that it is not limited to selected scientific disciplines or research problems. The specific contribution of the article can be applied to such scientific disciplines as geology, geography, earth resources and mining, where significant research problems are related to the identification of areas with appropriate properties or areas which are mostly affected by given phenomena.

In order to define the boundaries of the systems of territorial units taxonomic methods are most often used (Szopik-Depczyńska et al., 2017; Kurowska-Pysz et al., 2018; Balcerzak, 2020; Marks-Bielska et al., 2020; Kuc-Czarnecka et al., 2020). In the article, an analysis of the spatial differentiation of the agrarian structure in Poland is used as the case study. Limiting the taxonomic analysis to one dimension results from the fact that a characteristic feature of agriculture in Poland in terms of the agrarian structure, is the fragmentation of farms and its significant spatial differentiation (Michna, 2007, pp. 5-21), which results, from economic, social and historical factors (Walczak, Pietrzak, 2016, pp. 468-470). In the past, these factors conditioned the shape of the agrarian structure, preserving its present form to such a high degree that it remained unchanged despite many intensive policies of the state after 1990. In Poland the agrarian structure is one of the most important driver of development of agriculture, and it significantly affects the possibilities of land using for industrial purposes,

therefore, in the case of determining the boundaries of the systems of territorial units, the taxonomic analysis may be reduced to a spatial analysis of its diversity.

In the next section, the literature review presenting the theoretical background of current paper will be given. Then, the research objective within the methodological perspective and implementation gap concerning the selection of the system of territorial units within the Aggregation Problem is presented. The next sections provide the results, discussion and conclusions.

Theoretical background

Since the end of the 1980s, the problem of using statistical methods and spatial econometrics in regional research has been intensively discussed (Anselin, 1988; Arbia, 1988, 2006; Tobler, 1989; Paelinck, 2000; Haining, 2005; Bivand et al., 2008; LeSage & Pace, 2009; Pietrzak, 2018). An additional impulse for this research is undoubtedly the development of computer software, which opens the possibility of implementing spatial quantitative methods (Bivand et al., 2008; LeSage & Pace, 2009) not only at the academic level but also in practical decision making. Therefore, in the literature on spatial economic research, the growing number of works take into account the problem of space in the analyzes performed. The spatial dimension is taken into account directly by using spatial statistics methods and spatial econometric models (Haining, 2005; LeSage & Pace, 2009; Pietrzak & Ziemkiewicz, 2017; Kiselitsa et al., 2018; Kuc, 2017b; Shuyan & Fabus, 2019) or is additionally taken into account after prior application of taxonomy methods (Kuc, 2017a; Hlavacek & Sivicek, 2017; Kljucnikov et al., 2018; Rogalska, 2018; Horska et al., 2019; Bal-Domańska et al., 2020).

The regional research literature emphasizes the need to consider the Modifiable Areal Unit Problem (MAUP). The authors most often refer to two items here: the MAUP definition and research methodology (Openshaw & Taylor, 1979; Openshaw, 1984; Anselin, 1988; Arbia, 1989; Tobler 1989; Fotheringham & Wong, 1991; Reynolds, 1998; Paelinck, 2000; Dark, Bram, 2007; Flowerdew, 2011). Openshaw and Taylor (1979) state that the MAUP problem was identified already by Gehlke & Biehl (1934, p. 170) and Yule & Kendall (1950, pp. 320-334).

Gehlke and Biehl (1934) performed a correlation analysis on the basis of two spatial data sets related to 252 territorial units of Cleveland. For the spatial data from the first set, a correlation relationship between the number of juvenile offenders in the region and the median apartment rent was determined. On the other hand, for the data from the second set, a relationship between the share of juvenile offenders among adolescents and the median rent of flats was established. In the case of both data sets, spatial data related to 252 territorial units were aggregated and assigned successively to systems consisting of 200, 175, 150, 125, 100 50 and 25 territorial units. Based on the obtained results, the authors indicated differences in the calculated values of the Pearson's linear correlation coefficient depending on the adopted aggregation level.

In turn, Yule and Kendall (1950) presented the results of spatial economic research for 48 agricultural territorial units in England. The research concerned the analysis of the correlation between the yield of wheat and the yield of potatoes. The obtained values of the Pearson's linear correlation coefficient systematically increased as a result of a change in the aggregation level, which was then an argument for Openshaw and Taylor (1979) that proved the importance of the Modifiable Areal Unit Problem.

Based on the two mentioned pioneering research articles, Openshaw and Taylor (1979) distinguished two aspects of the Modifiable Areal Unit Problem. The problem of changing the obtained results during the transition to another level of aggregation originally discussed by Gehlke and Biehl (1934) and Yule and Kendall (1950) will be referred to as the Scale Problem. In this case, changes in the results are a consequence of adopting a new set of areal units, most often at a higher level of aggregation. On the other hand, the Aggregation Problem was defined as the problem of obtaining different research results depending on the choice of the set of areal units, however, within the same aggregation level (Openshaw, Taylor, 1979, p. 128; Openshaw, 1984, p. 8). Therefore, the essence of the Aggregation Problem comes to finding the right set of areal units at the aggregation level adopted in a given study.

In the case of determining boundaries of the set of areal units, Openshaw and Taylor (1979) introduced methodological proposals in the form of building two systems. The presented systems made it possible to generate randomly a potential set of areal units (Openshaw & Taylor 1979, pp. 127, 131-132; Openshaw, 1984, pp. 8-12). The first of the systems, the zoning system, allowed for the random creation of sets, however, assuming that the generated territorial units had a continuous border. The second system proposed by Openshaw and Taylor was described as the grouping system. The use of this system also allowed for the random generation of sets of areal units, but the territorial units created within the grouping system did not have to meet the condition of border continuity (Openshaw, 1984, pp. 8-9). This means that the created territorial unit could consist of several territorial sub-units (islands), not adjacent to each other.

Openshaw and Taylor (1979); Openshaw (1984); Anselin (1988); Tobler (1989); Haining (2005) state that the key issue in the case of Modifiable Areal Unit Problem is the selection of the right set of areal units at the selected level of aggregation. In the regional research, such a set should be always adopted, where the spatial

data related to it correctly reflect the impact of causes for the studied economic phenomena (Tobler, 1989, pp. 115-116; Haining, 2005, pp. 150-151). Since the determination of the appropriate set of areal units is closely related to a given research problem undertaken, in each case only the knowledge and scientific experience of the researcher and the results of previous studies may allow for the correct determination of the boundaries of the system of territorial units. In the work by Pietrzak (2018b, pp. 75-107) it was justified that Openshaw and Taylor's proposal to use the zoning system and grouping system as part of the spatial economic research conducted is incorrect. As a result of using the zoning system and grouping system, any set of areal units can be generated randomly. Therefore, the analyzed spatial data assigned to such a set will not correctly reflect the causal relationships for the analyzed economic phenomena. The use of the zoning system and grouping system creates a potential risk of receiving incorrect conclusions based on the obtained test results. This means that if the researcher does not consider the problem within the proper set of areal units, the performed regional analysis will be incorrect, therefore, from the practical perspective, it can bring inappropriate policy implications.

Research objective and methodology

The aim of the article is to develop an original solution for the Aggregation Problem based on the case of land use and agricultural research in Poland. The solution to the problem will lead to establishing the system of sets of areal units at subsequent levels of aggregation. The result of the implementation of that objective will be a proposal of a system of four sets of areal units, which is pointed here as a contribution fulfilling methodological and implementation gap. The obtained results are important from the point of view of regional research on the land use and agriculture, as the application of the proposed systems is the condition of obtaining correct data for potential policy decision making in the case of one of the most important sector of the Polish economy and it concerns the basic strategic production factor, which is also important from the perspective of industrial development of the country. Additionally, even concentrating directly only on the agriculture, it should be remembered that this sector is currently subjected to many structural changes, which are related to common agricultural EU policy and its EU financing. However, what is most important here, the scientific and practical value added of the current research should not be only restricted to the Polish economy, but the proposed case study can be considered as a universal example, which can be generalized to regional research in other countries and sectors (e.g. in mining).

As it was already stressed, the undertaken research goal is related to the key issue in regional research. The results obtained from spatial economic analyzes are based on spatial data, which as a two-dimensional random fields, are the realizations of spatial economic processes (Pietrzak, 2010a, 2010b). The spatial data used in the research are assigned to regions in accordance with the established set of areal units. The set of areal units together with the spatial data related to them are referred to as spatial data system (Pietrzak 2018b, pp. 37-48).

There should be a hierarchical relationship between spatial data systems that allows for aggregation of spatial data. Data aggregation is defined in the literature as the process of combining numerical data on set of lower-order units, resulting in obtaining numerical information on higher-order units (Pawłowski, 1969, p. 24). Aggregation of spatial data should be assigned to the type of subjective aggregation. In this case, instead of the hierarchical criterion of economic objects, a geographical criterion is used (Pawłowski, 1969, p. 237; Pietrzak 2018b, p. 31-33). The hierarchy system used in the aggregation process is based on regional boundaries. Lower-order objects (regions) are spatially contained within the boundaries of higher-order objects (macroregions). Most often, the spatial hierarchy of regions is adopted on the basis of the boundaries of the set of areal units. As part of the conducted regional research, sets of areal units at different levels of aggregation can be adopted. As an example of potential set of areal units used in the process of aggregation of spatial data, the sets of NUTS classification developed by Eurostat (Nomenclature of Units for Territorial Statistics, Eurostat, 2015) can be given. It should be emphasized that most of the spatial data used in regional research is obtained from public statistics. In the case of the European Union, spatial data are made available under the mentioned administrative NUTS systems. The purpose of introducing the NUTS classification was to ensure the collection, compilation and sharing of comparable data across the EU for the Member States. The set of areal units NUTS 0 defines the countries of the EU. On the other hand, subsequent NUTS sets define regions of smaller and smaller area within the borders of the member states. From the methodological and policy perspective, subsequent NUTS levels are not random, and the analysis of most economic phenomena and the relationships between them in the NUTS classification should lead to correct research results.

An example of a hierarchy system based on the border criterion will be presented for two NUTS sets in Poland: the set of subregions (NUTS 3 system) and the set of voivodeships (NUTS 2 system). For the territory of Poland, the system of NUTS 3 is a set of areal units at the lower level of aggregation, and the system of NUTS 2 is a set at a higher aggregation level. The NUTS 3 and NUTS 2 sets are shown in Figure 1, which shows the spatial hierarchy according to which the lower-order regions fall into the higher-order regions.

The NUTS 3 and NUTS 2 sets of areal units described above allow to analyze phenomena at a lower and higher level of aggregation. Aggregation level results in adopting a set of areal units with appropriate

characteristics to which spatial data will be assigned. The selected set has a fixed number of areal units of a specific shape and size. Taking into account the area with established borders (e.g. the territory of Poland), it is possible to establish many potential sets of areal units with a different number of n sub-areas. The lowest aggregation level will relate to the set of areal units with the largest number of n sub-areas. On the other hand, the highest level of aggregation will relate to the set consisting of the smallest number of n sub-areas.

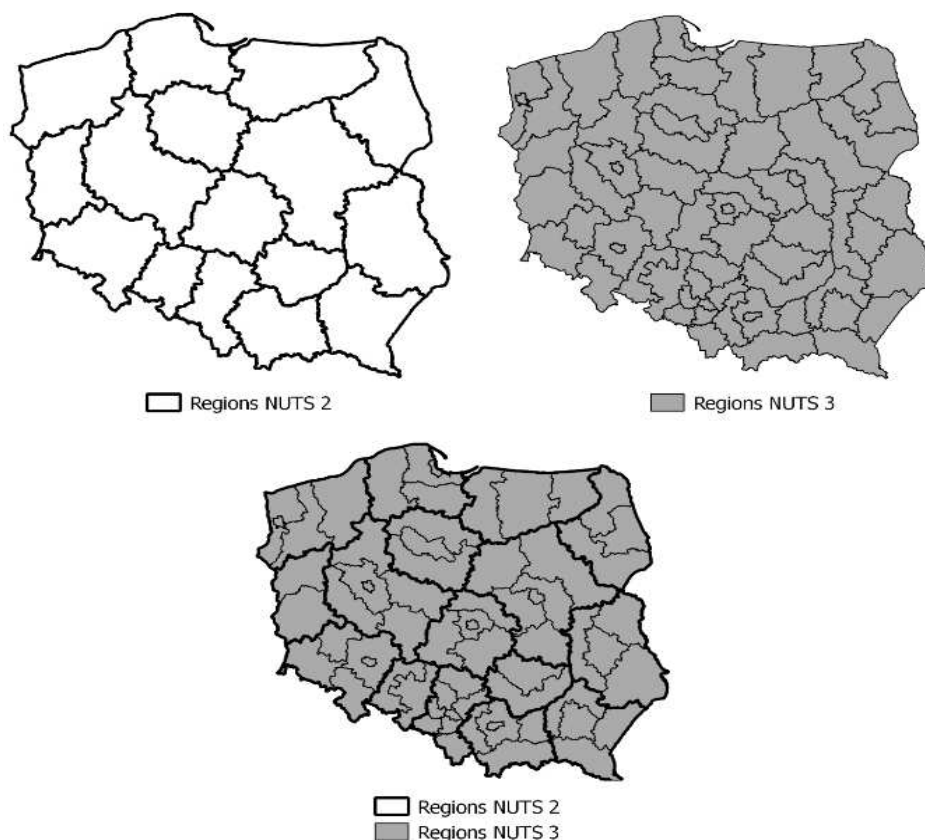


Fig. 1. Spatial hierarchy based on NUTS 3 and NUTS 2 sets

The spatial data system adopted in the regional empirical research should contain the set of areal units and the spatial data related to it, which are characterized by the property of causal homogeneity. Such a system of set and data will be defined as a causally homogeneous system of spatial data (Pietrzak, 2018b, pp. 37-48). Spatial data has the property of causal homogeneity if, for each of the regions that make up the set of areal units, the data are the result of the same set of main causes (Pietrzak, 2018b, pp. 42-48). Only cause-homogeneous spatial data can adequately reflect the effects of causes within an established spatial data pattern. This means that the use of a causally homogeneous system of spatial data is the condition for correct assessment of the studied economic phenomena (Tobler, 1989, pp. 115-116; Anselin, 1988, p. 27; Haining, 2005, pp. 150-151). If spatial data do not have the property of causal homogeneity, then in each area of the set of areal units different principal causes may interact, or different combinations of principal causes may occur. The research results obtained on the basis of such a spatial data system will be affected by a cognitive error, the weight of which will depend on the degree of interference in the interaction of the main causes.

To bring valuable information, regional studies most often concern different levels of aggregation. This necessitates the adoption of many set of areal units and the examination of the causal homogeneity of spatial data at each level of aggregation. Considering this issue allows us to introduce the concept of a homogeneous system of sets of areal units, which was defined by Pietrzak (2018b, pp. 42-48) as a system of sets of areal units at various levels of aggregation, where spatial data related to these sets have the property of causal homogeneity. In the procedure of determining a homogeneous system of sets of areal units, the researcher determines for each of the adopted levels of aggregation only one set of areal units with the causally homogeneous spatial data related to it. This is due to the fact that, within the analysed research problem, at the selected level of aggregation there is only one, causally homogeneous, spatial data system. No other set of areal units at the same level of aggregation can be used due to the lack of causal homogeneity of spatial data related to it. Therefore, determining a homogeneous system of sets of areal units is crucial for regional research, as it allows to draw correct conclusions that are used to solve a given research problem. Thus, the homogeneous system of sets of

areal units, defined within the spatial economic analysis, is a necessary condition for obtaining appropriate research results. However, one should be aware that the spatial data systems adopted at the selected levels of aggregation will never ideally reflect the interaction of real causes.

In the case of a homogeneous system of sets of areal units, it can be concluded that the selection of set of areal units is limited from the bottom and from the top in terms of the causal homogeneity property of spatial data (Pietrzak, 2018b, pp. 48-57). This limitation is the consequence of the fact that spatial data systems are most often not causally homogeneous at a very low or very high level of aggregation. Therefore, a homogeneous system of sets of areal units should be limited only to causally homogeneous spatial data systems at appropriate levels of aggregation. The bottom-up limitation results from the fact that in the case of socio-economic phenomena, the identification of the impact of the main causes is possible only within the whole region, which constitutes a complex economic system. The selection of the lowest level of aggregation in the form of spatial point data (selected consumers, enterprises) will not allow for the assessment of socio-economic phenomena, the nature of which is revealed only in the functioning of the whole region. On the other hand, the bottom-up limitation also results from the fact that for too large areas there is a spatial interaction of several main causes. Each of the main causes is the result of the functioning of relatively independent regions, located within a larger area. This means that data which has the causal homogeneity property at a selected level of aggregation may lose this property at a higher aggregation level.

The two following definitions: spatial data system and a homogeneous system of sets of areal units are part of the Aggregation Problem, which was redefined by Pietrzak (2018b, pp. 75-107). According to the proposed redefinition of Aggregation Problem, it was identified as the problem of creating a single set of areal units at aggregation level in such a way that, within the research problem undertaken, it belongs to a homogeneous system of sets of areal units (Pietrzak, 2018b, pp. 102-104). Therefore, the solution to the problem will consist in adopting an appropriate set of areal units, which can be assigned to a homogeneous system of sets of areal units.

The data used come from the 2002 General Agricultural Census (it was obtained from Statistics Poland: <https://stat.gov.pl/en/>). To determine the value of the Gini index, the following ranges of the agricultural land area were selected: (1-5 ha), (5-10 ha), (10-20 ha), (20-50 ha) and (50 ha and more). It should be emphasized that the data enabling the determination of the agrarian structure at the level of the NUTS 4 territorial units were made available by the Central Statistical Office only for the year 2002 after the publication of the results of the General Agricultural Census. For other years, the spatial data made available by the Central Statistical Office (GUS) allows for the determination of the agrarian structure only at the level of the NUTS 2. All the calculations were made in Excel and R.

Results

In line with the adopted aim of the article, the solution for the Aggregation Problem was developed, within which a set of territorial unit systems was established at subsequent levels of aggregation. According to the presented methodology, spatial data assigned to such set of areal units should be characterized by the property of causal homogeneity. Only under such conditions will the adopted spatial data correctly reflect the impact of the causes influencing changes in the land use. The already mentioned NUTS territorial units were used in the study, followed by NUTS 5, NUTS 4, NUTS 3, NUTS 2 and NUTS 1 (only when one of the set of areal units is unacceptable, the boundaries of the new set of areal units will be considered.). This choice is justified by the fact that official statistics services should provide reliable spatial data on the land use for the set of areal units, compliant with the NUTS classification.

The analysis began with considering the system of NUTS 5. At the level of NUTS 5 aggregation, most spatial data on the land use and agrarian structure in Poland is not available. This means that it is not possible to use spatial data systems at this aggregation level and the system of NUTS 5 cannot not be assigned to a homogeneous set of systems.

In the case of the NUTS 4 aggregation level, official statistics provide spatial data that can be used in the research. According to the assumptions of the Central Statistical Office in Poland, complete data on the NUTS 4 system are collected, or the collected data are generalized to individual territorial units, assuming their homogeneity. This means that the spatial data on the land use in Poland, referred to the system of NUTS 4 territorial units, should be characterized by the property of causal homogeneity. Therefore, the system of NUTS 4 territorial units will be assigned to a homogeneous system of sets of areal units.

After assigning the NUTS 4 to a homogeneous system of sets of areal units, further sets at higher aggregation levels, which may also be assigned to this system, should be considered. The following sets will be considered consecutively: the NUTS 3, NUTS 2 and NUTS 1. In the case of set of areal units at higher levels of aggregation than the NUTS 4, it should be checked whether the regions included in the mentioned sets are homogeneous in terms of agrarian structure, which will ensure the presence of the causal homogeneity property. Therefore, to evaluate the selected NUTS set in terms of the possibility of assigning them to a homogeneous system of sets of areal units, a study of the diversity of the agrarian structure was carried out. The spatial variability of the agrarian

structure in a selected area is the most commonly applied tool of assessing the long term possibilities of land use. The high concentration of farm area is most often considered as the main factor supporting the competitiveness of agriculture (Michna, 2007, pp. 5-13). In Poland, the size of a farm is one of the most important variables determining the level of its competitiveness, as farms with a small area tend to be characterized with high unit production costs due to scale effects, are not able to generate an adequate level of income to function efficiently and benefit EU financial support for the process of modernization (Michna, 2007, pp. 5-11). Therefore, it is justified to use agrarian structure analysis to discuss the construction of a homogeneous system of sets of areal units from the perspective of regional analysis on the possibilities of the land use.

In order to determine the spatial variability of the land use, an analysis of the concentration of agricultural land was used. The concentration value was measured with application of the Gini index.

On the basis of the obtained values of the Gini index, the regions from the NUTS 4 set of areal units were divided in terms of the concentration of land into four classes. Regions were assigned to classes on the basis of the natural breaks method (Jenks, 1967). Then, Figures 2, 3 and 4 show the spatial differentiation of the concentration of land for the NUTS 4 system and the boundaries of the three sets of areal units at successive, higher levels of aggregation NUTS 3, NUTS 2, NUTS 1. Visual assessment of changes in the Gini index values in the Figures 2-4 confirms significant spatial diversification of the agrarian structure in Poland. This proves significant differences in terms of the area of farms, depending on the selection of areas with a high or low concentration of agrarian structure. Therefore, the analysis of the spatial differentiation of the agrarian structure should allow to determine which of the considered NUTS sets of areal units can be assigned to a homogeneous system of sets of areal units for the purposes of regional research on the land use.

As a result of the analysis of the spatial differentiation of the agrarian structure for the NUTS 3, the following conclusions were obtained. It was found that within each of the individual regions of the NUTS 3 system (higher aggregation level), the regions from the NUTS 4 system (lower aggregation level) assigned to one of the four classes dominate (Figure 2). This means that each of the regions from the NUTS 3 system is characterized by a relatively constant agrarian structure within its borders. The observed domination of the regions of the NUTS 4 system from the same classes allows for the conclusion that spatial data related to the system of NUTS 3 territorial units should be characterized by the property of causal homogeneity. Therefore, the NUTS 3 set of areal units, like the NUTS 4 set, will be assigned to a homogeneous system of sets of areal units.

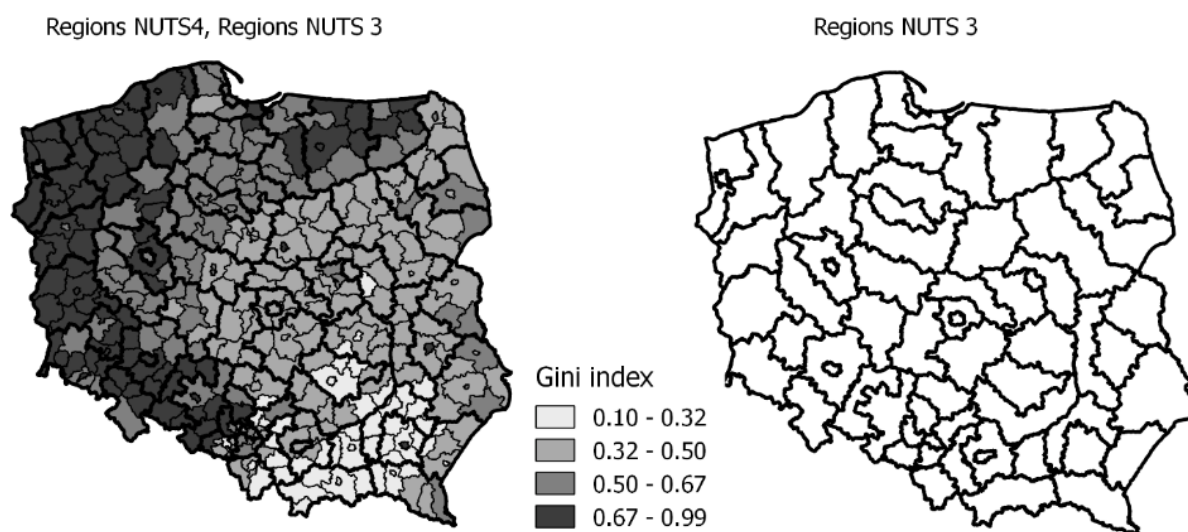


Fig. 2. Spatial differentiation of the agrarian structure in the NUTS 4 set of areal units and the boundaries of the NUTS 3 system

A similar situation occurs in the case of the NUTS 2 set or areal units. Also within the boundaries of most of the NUTS 2 regions, NUTS 4 regions dominate and can be assigned to one of the classes (Figure 3). It should be emphasized that for some regions from the NUTS 2 system there are deviations in the nature of the agrarian structure, but overall it can be considered that these regions are homogeneous due to the spatial differentiation of the agrarian structure. Therefore, the system of NUTS 2 territorial units will be assigned to a homogeneous system of sets of areal units, along with the NUTS 4 system and the NUTS 3 system.

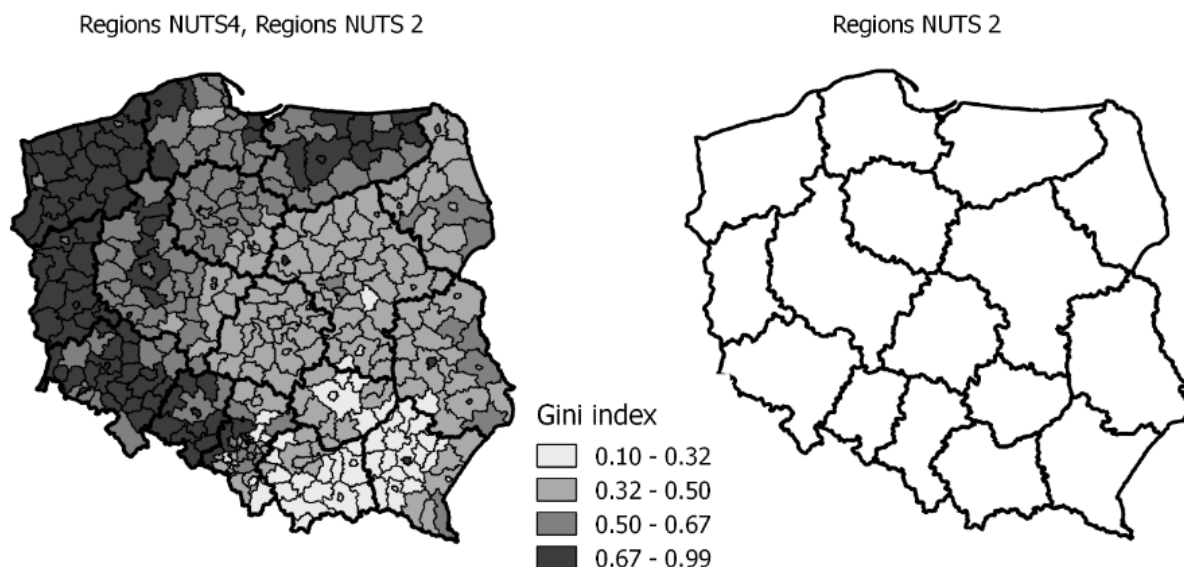


Fig. 3. Spatial differentiation of the agrarian structure in the NUTS 4 set of areal units and the boundaries of the NUTS 2 system

On the other hand, in the case of the NUTS 1 set of areal units, significant changes in the nature of the land structure within individual regions are visible (Figure 4). In the eastern region, the poviats belonging to the Podkarpackie and Świętokrzyskie voivodeships differ in terms of their agrarian structure from poviats in other voivodeships due to the lower level of agricultural land concentration. On the other hand, in the north-western region there is a lower level of agricultural land concentration in the poviats of the Wielkopolskie voivodship in comparison to poviats from other voivodships. Then, in the northern region, poviats from the Warmińsko-Mazurskie voivodship are characterized by a higher level of agricultural land concentration. This means that a NUTS 1 set of areal units cannot be assigned to a homogeneous system of sets of areal units, because the spatial data related to this system will not have the property of causal homogeneity. The regional research carried out on its basis will lead to incorrect conclusions and, consequently, inappropriate potential research policy implications.

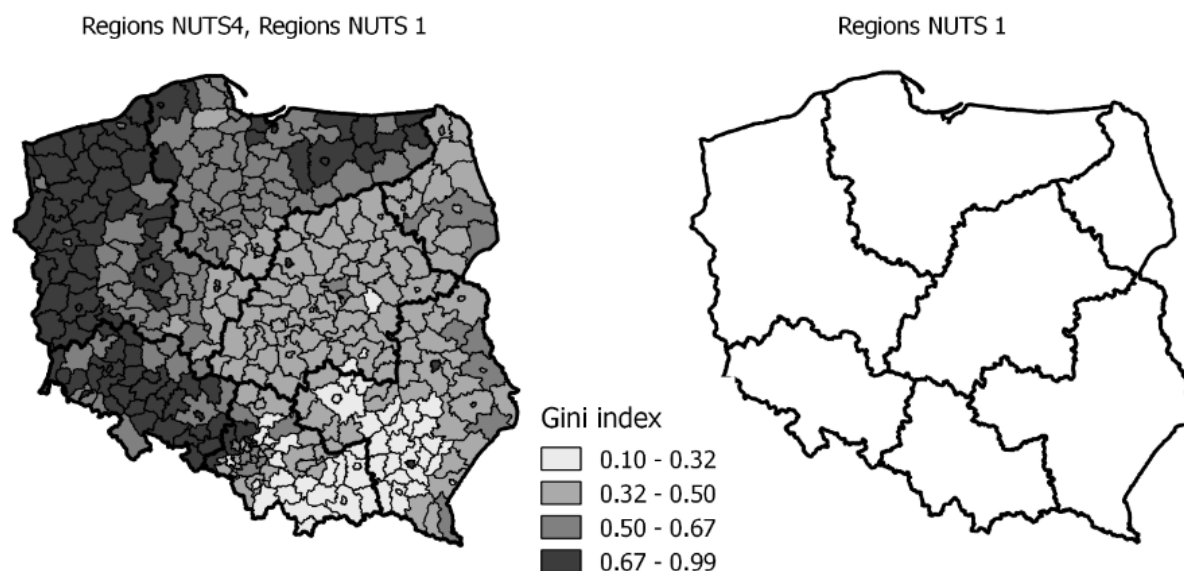


Fig. 4. Spatial differentiation of the agrarian structure in the NUTS 4 set of areal units and the boundaries of the NUTS 1 system.

Since the NUTS 1 set cannot be assigned to a homogeneous system of sets of areal units, a different set should be adopted at this level of aggregation, which can be assigned to a homogeneous system. The set of agricultural macro-regions of the SGM was taken into account. This set is used for the needs of farm statistics, which is kept within the Farm Accountancy Data Network (FADN). The obligation to use the FADN to assess the activity of agriculture was introduced in the European Economic Community as a result of the EEC Council

Regulation No. 79/65 / EEC of 1965. Then, in 1993, the FADN standards in the European Union were adopted. Therefore, during the accession process in 2000-2004, Poland was obliged to define the system of agricultural macro-regions of the SGM (Skarżyńska et al., 2005, pp. 7-16).

The SGM set was created after application of spatial aggregation on the basis of regions from the NUTS 2 set of areal units. In the first step, regions from the NUTS 2 set were divided into classes, taking into account the degree of their similarity in terms of significant agricultural features. For the division of regions from the NUTS 2 set, cluster analysis was used, where nine diagnostic variables describing the level of agricultural development in Poland were adopted (Skarżyńska et al., 2005, pp. 10-19). Then, the spatial aggregation process was performed, consisting in combining regions from the NUTS 2 set belonging to the same class, with the additional condition of mutual neighbourhood. This made it possible to obtain a set of areal units at a higher level of aggregation, the regions of which are homogeneous in terms of development of agriculture, therefore, its competitive potential. The set of agricultural macro-regions applicable in Poland is shown in Figure 5.

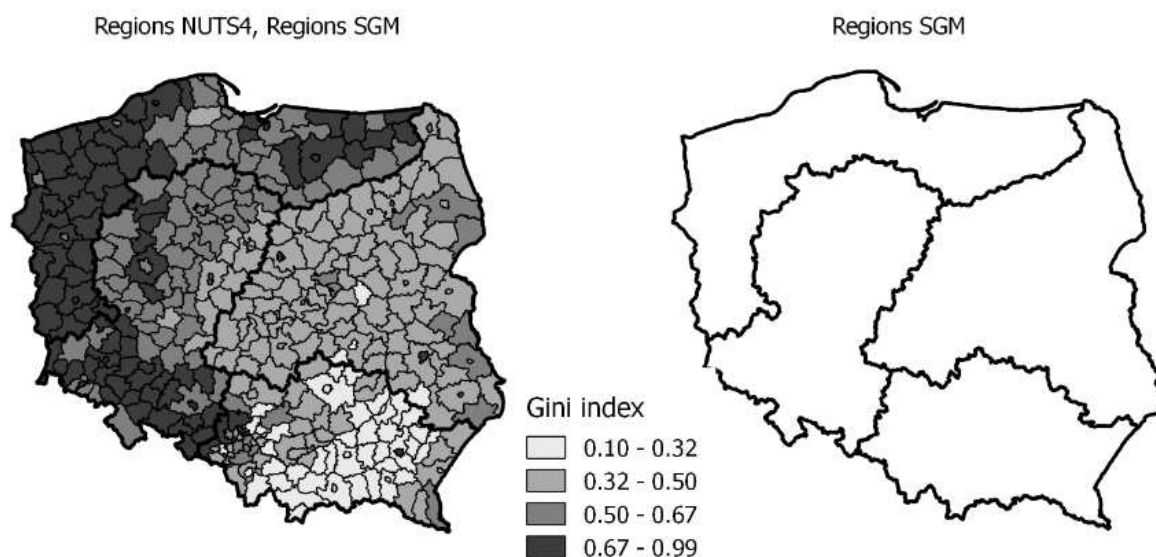


Fig. 5. Spatial differentiation of the agrarian structure in the NUTS 4 territorial unit system and the boundaries of the SGM agricultural macro-region system.

All regions of the SGM agricultural macro-region set are relatively homogeneous in terms of land use and agrarian structure. Therefore, the SGM set will also be assigned to a homogeneous system of sets of areal units within the taken research problem. The creation of a new set of SGM macro-regions in connection with Poland's accession to the European Union is an example of solving the Aggregation Problem by creating a new set of territorial units. Both the performed spatial economic analyzes and the statistics of regions based on SGM macro-regions should be effectively applied in regional research. Therefore, it can be expected that their application should lead to correct results, and applicable policy implications.

As a result of the implementation of the objective of the current research, a homogeneous system of sets of areal units was determined, consisting of the following sets of areal units: the NUTS 4 set, the NUTS 3 set, the NUTS 2 set and the SGM agricultural macro-region set. Additionally, it should be emphasized that for each of the adopted sets of areal units, the regularities between the processes relating to the land use and agrarian structure should be of relatively constant nature within the borders of individual regions.

Discussion

In line with the stated goal of current research, we propose the Aggregation Problem solution, which enables to determine a homogeneous system of sets of areal units for regional research on the land use as the basic earth resource and an important production factor, not only for agriculture, but also the resource determining the prospects for future industrial investments and its development. Thus, a system of appropriate set of areal units was established, the use of which will ensure the correct assessment of the variability of the phenomena at different levels of aggregation. A homogeneous system of sets of areal units for regional research in agrarian structure was created from the sets of areal units: NUTS 4, NUTS 3 and NUTS 2 and the system of agricultural macro-regions of the SGM. Such a defined homogeneous system of sets of areal units should be used in future research related to the issues of changes in the agrarian structure and policies affecting long term land use.

From the overall methodological perspective, the article shows how difficult it is to determine the causal homogeneity property for spatial data and the related determination of a homogeneous system of sets of areal units. An additional difficulty in determining a homogeneous system of sets of areal units and assessing its topicality in subsequent studies is also caused by trends in the spatial development of socio-economic phenomena. In the case of the analysis of the agrarian structure and the land use, a high degree of persistence of the spatial development of this phenomenon was confirmed (Walczak & Pietrzak, 2016). This means that a correctly adopted homogeneous system of sets of areal units can be applied in subsequent periods of the analysis.

On the other hand, there are also many phenomena whose level of spatial development changes significantly, even in short time. For example, the spatial structure of the residents' propensity to purchase via the Internet or use electronic banking may undergo such changes (Jibril et al., 2019), and in this case the established homogeneous system of sets of areal units may not be valid in subsequent periods. This could be especially seen during last case of Covid pandemic (Zinecker et al., 2021; Vasilyeva et al., 2021) This is an important observation, because it shows that within the framework of the given research problem, the boundaries of set of areal units that have been assigned to a homogeneous system of sets of areal units may change over time or even change suddenly due to some unexpected phenomenon. Therefore, the correction of the boundaries of the set of areal units is often necessary, if there is a change in the spatial differentiation of the considered phenomena and the relationships between them. This issue is the objective of future research based on current contribution. On the other hand, this factor confirms the universal methodological value added of the discussion presented in current paper and its universal applicability to regional studies in other national and sectorial context. This means the possibility of using the methodology proposed in the article to solve research problems in scientific disciplines, where the spatial aspect is important, in particular economics, geological survey, mining, earth resources and geography.

Conclusions

The article considers the Aggregation Problem for regional research on the example of land use in Poland. This problem concerns the possibility of obtaining different results of regional research, which is a consequence of adopting different sets of areal units within the same level of aggregation. The solution of the Aggregation Problem is based on selecting a set of areal units at a given level of aggregation in such a way that it belongs to a homogeneous system of sets of areal units from the perspective of the research problem. The spatial data assigned to such a system have the property of causal homogeneity and properly reflect the way in which the causes influencing the analyzed economic phenomena interact. The selection of an appropriate set of areal units is important, as the necessary condition for drawing correct conclusions is to perform an analysis based on causally homogeneous spatial data. This means that in the case of regional studies, only the identification of a homogeneous system of sets of areal units can allow for a correct assessment of economic relationships (Tobler, 1989, pp. 115-116).

It should be emphasized that solving the Aggregation Problem is of fundamental importance for spatial research, because each analysis based on spatial data requires their reference to a specific set of areal units. Therefore, the selection of the appropriate system within the given research problem determines the success of further research. The researcher's decision to choose the set, where the spatial data related to it will not have the property of causal homogeneity, will result in the inability to solve the research problem or will lead to incorrect conclusions and policy implications.

Despite the fact that in each case the decision on the choice of the set of areal units must be made by the researcher, which is often done in an arbitrary manner and is affected by objective factors such as data availability, it should largely take into account the research problem undertaken and be based on the available knowledge, the results of previous research and the researcher's scientific experience. Unfortunately, in practice, in many cases the sets of areal units are arbitrarily determined by researchers, without reference to the research problem undertaken and the nature of the analyzed economic phenomena (Fotheringham & Wong, 1991; Reynolds, 1998; Dark & Bram, 2007; Flowerdew, 2011), which must be considered as a fundamental methodological weakness.

Also in the case of works presenting the results of simulation analyzes for the needs of regional studies, the sets of areal units are usually determined at random using a computer algorithms/program available (Reynolds, 1998). This fact indicates the necessity to determine the appropriate set of areal units in regional studies with a specific proposal of methodological approach to this issue, which is done in the current article.

The discussed methodological approach is not free of some objective limitations, which can be of the highest importance in the case of solving decision making problems in industrial applications, where the availability of high quality low aggregation special data creates many significant technical problems and high economic costs. This issues are commonly seen in the case of mining industry and geology, where the decision-making mistakes on the projects resulting from wrong aggregation of spatial data can result not only in failures of given projects, but even bankruptcy of the previously stable enterprises.

References

- Anselin, L. (1988). *Spatial econometrics: method and models*. Boston: Kluwer Academic Press.
- Arbia, G. (1988). *Spatial data configuration in statistical analysis of regional economics and related problems*. Dordrecht: Kluwer Academic Press.
- Arbia, G. (2006). *Spatial econometrics, statistical foundations and applications to regional convergence*. Heidelberg: Springer.
- Balcerzak, A.P. (2020). Quality of institutions in the European Union countries. Application of TOPSIS based on entropy measure for objective weighting. *Acta Polytechnica Hungarica*, 17(1), 101-122. doi: <https://doi.org/10.12700/APH.17.1.2020.1.6>.
- Bal-Domańska, B., Sobczak, E., & Stańczyk, E. (2020). A multivariate approach to the identification of initial smart specialisations of Polish voivodships. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 15(4), 785–810. <https://doi.org/10.24136/eq.2020.034>
- Bednář, P., & Halásková, M. (2018). Innovation performance and R&D expenditures in western European regions: Divergence or convergence? *Journal of International Studies*, 11(1), 210-224. <https://doi.org/10.14254/2071-8330.2018/11-1/16>.
- Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York: Springer.
- Dark, S.J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, 31 (5), 471–479. doi: <https://doi.org/10.1177/0309133307083294>.
- Eurostat. (2015). *Regions in the European Union Nomenclature of territorial units for statistics NUTS 2013/EU-28*. Luxembourg: Publications Office of the European Union.
- Flowerdew, R. (2011). How serious is the Modifiable Areal Unit Problem for analysis of English census data? *Population Trends*, 145, 1-13. doi: <https://doi.org/10.1057/pt.2011.20>
- Formánek, T. (2019). GDP per capita in selected EU countries: Economic growth factors and spatio-temporal interactions examined at the NUTS2 level', *Journal of International Studies*, 12(1), 119-133. doi: <https://doi.org/10.14254/2071-8330.2019/12-1/8>.
- Fotheringham, A.S., & Wong, D.W.S. (1991). The Modifiable Areal Unit Problem in multivariate analysis. *Environment and Planning A*, 23, 1025-1044.
- Gehlke, C. E., & Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, 29, 169-170.
- Haining, R. P. (2005). *Spatial data analysis. Theory and practice*. Cambridge: Cambridge University Press.
- Hlavacek, P., & Sivicek, T. (2017). Spatial differences in innovation potential of central European regions during post-transformation period. *Journal of International Studies*, 10(2). <https://doi.org/10.14254/2071-8330.2017/10-2/4>.
- Horska, E., Moroz, S., Polakova, Z., Nagyova, Eudmila, & Paska, I. (2019). Evaluation of interaction between chosen indicators of development of regions in Ukraine. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(2). <https://doi.org/10.24136/eq.2019.016>.
- Jenks, G.F. (1967). The Data Model Concept in Statistical Mapping, *International Yearbook of Cartography*, 7, 186–190.
- Jibril, A. B., Kwarteng, M. A., Appiah-Nimo, C., & Pilik, M. (2019). Association rule mining approach: evaluating pre-purchase risk intentions in the online second-hand goods market. *Oeconomia Copernicana*, 10(4), 669-688. <https://doi.org/10.24136/oc.2019.032>.
- Kijek, T., & Matras-Bolibok, A. (2019). The relationship between TFP and innovation performance: evidence from EU regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(4), 695-709. <https://doi.org/10.24136/eq.2019.032>.
- Kiselitsa, E. P., Shilova, N. N., Liman, I. A., & Naumenko, E. E. (2018). Impact of spatial development on sustainable entrepreneurship. *Entrepreneurship and Sustainability Issues*, 6(2). [https://doi.org/10.9770/jesi.2018.6.2\(28\)](https://doi.org/10.9770/jesi.2018.6.2(28)).
- Kljucnikov, A., Sobekova-Majkova, M., Vincurova, Z., Gailiuniene, M., & Kiausiene, I. (2018). The insolvency of SMEs within the perspective of the Central European region. *Transformations in Business & Economics*, 17(2(44)).
- Kuc, M. (2017a). Is the regional divergence a price for the international convergence? The case of Visegrad group. *Journal of Competitiveness*, 9(4). <https://doi.org/10.7441/joc.2017.04.04>.
- Kuc, M. (2017b). Social convergence in Nordic countries at regional level. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(1). <https://doi.org/10.24136/eq.v12i1.2>.
- Kuc-Czarnecka, M., Lo Piano, S. and Saltelli, A. (2020). Quantitative Storytelling in the Making of Composite Indicator'. *Social Indicators Research*. Advance online publication. <https://doi.org/10.1007/s11205-020-02276-0>.

- Kurowska-Pysz, J., Szczepańska-Woszczyzna, K., Štverková, H. and Kašík, J. (2018). The catalysts of cross-border cooperation development in euroregions', *Polish Journal of Management Studies*, 18 (1), 180-193. <https://doi.org/10.17512/pjms.2018.18.1.14>.
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: Wiley.
- Marks-Bielska, R., Lizińska, W., Wojarska, M., & Babuchowska, K. (2020). Institutional efficiency versus stability of local governments in basic areas of activity: the case of Poland. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 15(3), 463–487. <https://doi.org/10.24136/eq.2020.021>.
- Meyer, N., & Meyer, D.F. (2019). Examining the impact of entrepreneurial activity on employment and economic growth: The case of the Visegrad countries. *Polish Journal of Management Studies*, 20(1), 277-292. <https://doi.org/10.17512/pjms.2019.20.1.25>.
- Michna, W. (2007). *Controlled and spontaneous changes in agrarian structure in various regions of the country*. Warszawa: IERiGŻ-PIB.
- Openshaw, S. (1984). The modifiable areal unit problem. Concepts and Techniques in *Modern Geography*, 38.
- Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.). *Statistical methods in the spatial sciences*. London: Pion.
- Paelinck, J. H. P. (2000). On aggregation in spatial econometric modelling. *Journal of Geographical Systems*, 2(2). doi: <https://doi.org/10.1007/PL00011452>.
- Pietrzak, M. B. (2018). The composition of agricultural macroregions in Eastern Poland - an empirical example of the Aggregation Problem. In M. Papież & S. Śmiech (Eds.). *The 12th Professor Aleksander Zelias international conference on modelling and forecasting of socio-economic phenomena. Conference proceedings*. Cracow: Foundation of the Cracow University of Economics.
- Pietrzak, M. B. (2018b). *Problem zmiennej jednostki odniesienia w przestrzennych badaniach*. Toruń: Polskie Towarzystwo Ekonomiczne oddział w Toruniu.
- Pietrzak, M. B. (2019). Modifiable Areal Unit Problem: the issue of determining the relationship between microparameters and a macroparameter. *Oeconomia Copernicana*, 10(3), 393-417. <https://doi.org/10.24136/oc.2019.019>.
- Pietrzak, M. B., & Ziemkiewicz, B. (2017). The use of random fields in the modifiable areal unit problem. In M. Papież & S. Śmiech (Eds.). *The 11th professor Aleksander Zelias international conference on modelling and forecasting of socio-economic phenomena. Conference proceedings*. Cracow: Foundation of the Cracow University of Economics.
- Reynolds, H.D. (1998). *The Modifiable Area Unit Problem: Empirical Analysis by Statistical Simulation*. Doctoral thesis, graduate Department of Geography, University of Toronto.
- Rogalska, E. (2018). Multiple-criteria analysis of regional entrepreneurship conditions in Poland. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 13(4). doi: <https://doi.org/10.24136/eq.2018.034>.
- Sánchez-López, C., Aceytuno, M. T., & De Paz-Bañez, M. A. (2019). Inequality and globalisation: Analysis of European countries. *Economics and Sociology*, 12(4), 84-100. <https://doi.org/10.14254/2071-789X.2019/12-4/5>.
- Semenenko, I., Halhash, R., & Sieriebriak, K. (2019). Sustainable development of regions in Ukraine: before and after the beginning of the conflict. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(2). doi: <https://doi.org/10.24136/eq.2019.015>.
- Shkolnyk, I., Kozmenko, S., Kozmenko, O., & Mershchii, B. (2019). The impact of economy financialization on the level of economic development of the associate EU member states. *Economics and Sociology*, 12(4), 43-58. <https://doi.org/10.14254/2071-789X.2019/12-4/2>.
- Shuyan, L., & Fabus, M. (2019). Study on the spatial distribution of China's out-ward foreign direct investment in EU and its influencing factors. *Entrepreneurship and Sustainability Issues*, 6(3). doi: [https://doi.org/10.9770/jesi.2019.6.3\(16\)](https://doi.org/10.9770/jesi.2019.6.3(16)).
- Skarżyńska, A., Goraj, L., & Ziętek, I. (2005). *Methodology SGM "2002" for the typology of farms in Poland*. Warszawa: IERiGŻ-PIB.
- Szopik-Depczyńska, K., Cheba, K., Bąk, I., Kiba-Janiak, M., Saniuk, S., Dembińska, I. and Ioppolo, G. (2017). The application of relative taxonomy to the study of disproportions in the area of sustainable development of the European Union. *Land Use Policy*, 68, 481-491. <http://dx.doi.org/10.1016/j.landusepol.2017.08.013>.
- Szopik-Depczyńska, K., Cheba, K., Bąk, I., Stajniak, M., Simboli, A. and Ioppolo G. (2018). The study of relationship in a hierarchical structure of EU sustainable development indicators. *Ecological Indicators*, 90, 120-131. <https://doi.org/10.1016/j.ecolind.2018.03.002>.
- Tobler, W. R. (1989). Frame independent spatial analysis. In M. F. Goodchild & S. Gopal (Eds.). *The accuracy of spatial databases*. London: Taylor & Francis.

- Vasilyeva, T., Kuzmenko, O., Kuryłowicz, M., & Letunovska, N. (2021). Neural network modeling of the economic and social development trajectory transformation due to quarantine restrictions during COVID-19. *Economics and Sociology*, 14(2), 313-330. doi:10.14254/2071-789X.2021/14-2/17
- Walczak, D., & Pietrzak, M. B. (2016). Analysis of agrarian structure in Poland in 1921 and 2002 based on the example of selected districts. In M. H. Bilgin, H. Danis, E. Demir & U. Can (Eds.). *Business challenges in the changing economic landscape. Proceedings of the 14th Eurasia Business and Economics Society conference*. Cham: Springer.
- Yule, G.U., & Kendall, M.G. (1950). *An introduction to the theory of statistics*. London: Griffin.
- Zinecker, M., Doubravský, K., Balcerzak, A.P., Pietrzak, M. B., Dohnal, M., (2021). The Covid-19 Disease and Policy Response to Mitigate the Economic Impact in the EU: An Exploratory Study Based on Qualitative Trend Analysis. *Technological and Economic Development of Economy*, 27(3), 742–762. doi: <https://doi.org/10.3846/tede.2021.14585>.