EQUILIBRIUM
Quarterly Journal of Economics and Economic Policy
2014 VOLUME 9 ISSUE 3, September
p-ISSN 1689-765X, e-ISSN 2353-3293
www.economic-policy.pl

Pietrzak M.B (2014), *Redefining the Modifiable Areal Unit Problem Within Spatial Econometrics, the Case of the Aggregation Problem*, "Equilibrium. Quarterly Journal of Economics and Economic Policy", Volume 9, Issue 3, pp. 131-151, DOI: http://dx.doi.org/10.12775/EQUIL.2014.021

Michał Bernard Pietrzak*
Nicolaus Copernicus University, Poland

Redefining the Modifiable Areal Unit Problem Within Spatial Econometrics, the Case of the Aggregation Problem**

JEL classification: *C01, C15, C21*

Keywords: *spatial econometrics, modifiable areal unit problem, scale problem, aggregation problem*

Abstract: *The paper focuses on the issue of the aggregation problem, which is frequently discussed within spatial econometrics. The aggregation problem is one of the two aspects of the modifiable areal unit problem (MAUP). The aggregation problem is connected with the volatility of the obtained results occurred when various compositions of territorial units for the same aggregation scale were applied. The objective of the present paper is to consider the redefinition of aggregation problem and showing positive solution of the aggregation problem based on the empirical example of determining agricultural macroregions.*

In the article the aggregation problem was defined as a problem of establishing a particular composition of territorial units at a selected aggregation scale in a such a way that is remains in the quasi composition of regions within the undertaken research problem. The paper also presented the procedure for determining agricultural macroregions where the analysis of the spatial volatility of the agrarian structure and the current knowledge on the agriculture in Poland were applied.

Date of submission: September 15, 2013; date of acceptance: March 26, 2014

* Contact: pietrzak@umk.pl, Nicolaus Copernicus University, ul. Gagarina 13A, 87-100 Toruń, Poland

** The project was co-financed by Nicolaus Copernicus University in Toruń within a UMK research grant no. 1481-E.

In addition, the paper considered the final areal interpretation problem connected with the incorrect determination of the area in relation to which final conclusions are drawn. The problem was presented basing on the example of the establishment of the average concentration of the area of agricultural land in Poland with the use of the Gini index calculated for districts. The paper emphasised that ignoring the final areal interpretation problem in spatial analyses may lead to an apparent identification of the modifiable areal unit problem.

Introduction

The *modifiable areal unit problem* (MAUP) is a crucial issue that is given much consideration within spatial econometrics. Spatial analyses conducted have frequently resulted in differentiated results, while modifying the aggregation scale. As observed, the volatility of the obtained results also occurred when various compositions of territorial units for the same aggregation scale were applied. The perceived problems have led to the formulation of the *modifiable areal unit problem* within which two aspects are considered – the *scale problem* and the *aggregation problem* (see Arbia, 1998).

The objective of the present paper is to consider the *aggregation problem* and to redefine it. Also, a positive solution of the *aggregation problem* will be shown based on the empirical example related to the need for designating the borderlines of SGM macroregions in Poland. This will be performed for a EU system of collecting accountancy data from farms – *the Farm Accountancy Data Network* (FADN). The paper will also emphasize the impossibility of a positive solution of the *aggregation problem* in its separation from the undertaken research problem. In addition, it will be proved that a positive solution consists in designating precisely one composition of territorial units at the aggregation scale accepted for the research. That composition, being the only one, will allow for a correct analysis of the properties of phenomena and the dependence between them within the research problem to be conducted.

The issue of the *modifiable areal unit problem* was already considered in the works of the following: Gehlke & Biehl (1934), Yule & Kendall (1950), Robinson (1950), Blalock (1964), Openshaw & Taylor (1979), Openshaw (1984a, 1984b), Reynolds (1988), Fotheringharn & Wong (1991), Holt *et al.* (1996), Tranmer & Steel (2001), Arbia (2006), Manley *et al.* (2006), Suchecki (2010), Flowerdew (2011) and Pietrzak (2014a, 2014b).

Redefining the aggregation problem

The issue of the *modifiable areal unit problem* is related to the character of irregular areas which are modifiable. That means that the arbitrarily determined shapes and space of areas may be any[1]. However, this arbitrary character of areas is only apparent. In reality the phenomena and dependence between them occur in social and economic systems that are related to the already specifically shaped areas. These systems, in turn, make up a larger system related to the phenomena and dependence occurring in areas with a higher aggregation scale.[2] Therefore, the researcher's task is to identify correctly the spatial volatility of the phenomena within the undertaken research problem and then, based on it, to determine the borderlines for the areas. The arbitrarily designated composition of territorial units is determined by the properties and dependence identified for the researched phenomena.

The issue of the *modifiable areal unit problem* appears within the conducted spatial analyses. There are four necessary conditions for ensuring the correctness of obtained results and these conditions were worked out by Pietrzak (2014a). The first condition assumes that a starting point in every analysis should be the formulation of a research problem. All the assumptions made for the research should be determined within the undertaken research problem. The basis for spatial analyses is formed of the data which are the realization of spatial processes[3]. The second condition consists in establishing an adequate aggregation scale for spatial data that would be appropriate for drawing correct conclusions on the phenomena and the relationships holding between them. The third conclusion requires the accepted data to be reliable. The reliability of spatial data is to be ensured by obtaining them from institutions specialising in public statistics. Conclusions are related most frequently to the areas with a higher aggregation scale than the aggregation scale of the data possessed. The fulfilment of the fourth and the last condition consists in determining the size (boundaries) of the region in relation to which the formulated conclusions of the spatial analysis performed will be applied.

[1] This opinion on the nature of irregular areas was expressed by Taylor and Openshaw (1979).

[2] The problem of the identification of social and economic processes and the phenomena being shaped within them requires further theoretical consideration and providing more details which means going beyond the framework of the present paper.

[3] The spatial process is understood here as a two-dimensional random field (see Arbia, 1998; Pietrzak, 2010a; Pietrzak, 2010b).

In his work, Pietrzak (2014a) introduced the concept of *the quasi composition of regions* (QCR). *The quasi composition of regions* was defined as a set of compositions of territorial units, with lower and upper limits[4], consisting of particular compositions of territorial units for further aggregation scales. Particular composition should be selected in a way that allows an appropriate analysis to be performed within the undertaken research problem. It means that within the undertaken research problem there is only one *quasi composition of regions* which allows the identification and description of the dependence holding for the analysed phenomena.

In the subject literature shown above the consideration of the *aggregation problem* consists in researching the volatility of results, which is dependent on the arbitrarily accepted composition of territorial units with the same aggregation scale. In the author's opinion, the approach presented in the literature is inappropriate, since it allows analyses made on multiple arbitrary compositions of territorial units which do not belong to *the quasi composition of regions*. It must be emphasised that, in fact, only one composition ensures correct conclusions, while for the remaining compositions the obtained results do not have a cognitive value. The implication is that the results obtained for various compositions of territorial units within one aggregation scale are incomparable. Moreover, the volatility of results does not need to be verified since it has to occur anyway. Within incorrect compositions different properties and dependence get mixed between phenomena, therefore, almost any result is possible to be obtained. From the perspective of the undertaken research problem, the described volatility of results seems obvious for the researcher.

Pietrzak's work (2014a) redefines *the aggregation problem*. Its redefinition was based on the four aforementioned conditions ensuring the correctness of analyses conducted and on the concept of *the quasi composition of regions*. *The aggregation problem* was defined as a problem of determining a particular composition of territorial units at a selected aggregation scale in a such a way that is remains in *the quasi composition of regions* within the undertaken research problem. A positive solution of the *aggregation problem* consists in designating a composition of territorial units which would ensure a correct analysis of the properties of the phenomena and of the dependence holding between them within the undertaken research problem.

[4] Setting the lower and upper limits on compositions of areas results from the fact that the correctness of conclusions does not need occur for all aggregation scales. Therefore, *a quasi composition of regions* should be limited only to appropriate aggregation scales.

The aggregation problem based on the example of SGM macroregions in Poland

The definition of the *aggregation problem* presented in the previous subchapter indicates that it needs to be related to the empirical research problem. That means that only the researcher's knowledge and scientific experience allow the composition of territorial units and a positive solution of the *aggregation problem* to be determined correctly. In the article, the *aggregation problem* will be presented in the light of the research problem which was the need for defining a composition of agricultural SGM macroregions in Poland[5].

Defining a new composition of SGM macroregions was connected with Poland's joining the European Union. The accession makes it obligatory for Poland to collect accountancy data from farms (*the Farm Accountancy Data Network* – FADN). The economic size (their profitability) of Polish farms is measured within the FADN system. The economic size of farms was expressed in ESU (*the European Size Unit*). The value of ESU for particular farms was calculated based on SGM (*Standard Gross Margin*) coefficients. SGM coefficients were next referred to SGM agricultural macroregions and, depending on the macroregion a given farm belonged to, they took different values. SGM macroregions should be differentiated by the size of agricultural production and by the factors which have a major impact on the production effects achieved by farms. In addition, SGM regions should be internally homogeneous, if considered by their agricultural development and culture. Since 2010 SO (*standard output*)[6] has been the basis for determining the economic size of farms within FADN. The standard output measure replaced SGM. The introduction of SO did not bring about any changes in the agricultural macroregions and the composition of agricultural macroregions modified due to Poland's joining the European Union is still in effect[7].

[5] It is assumed that within the undertaken research problem the selected agricultural macroregions should differ in the scope of the underlying values of the agricultural variable. SGM regions should also be internally homogeneous in their levels of agricultural development and cultures.

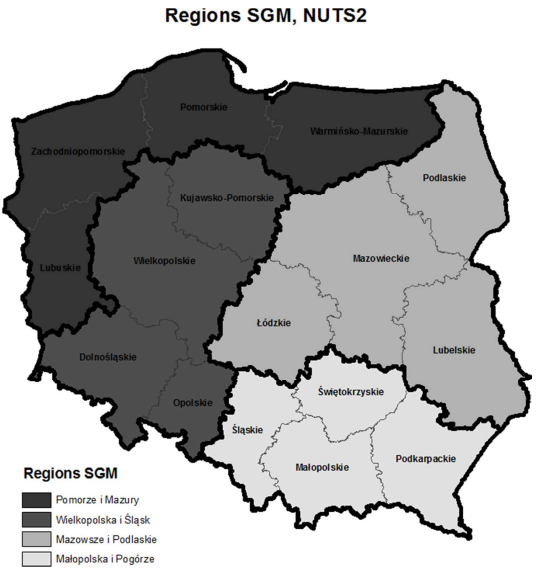
[6] The first set of the SO '2007' coefficients was provided to Eurostat by the end of 2010 (see Goraj *et al*., 2010).

[7] The problems related to agricultural macroregions, standard gross margins (SGM), standard output (SO), European Size Unit (ESU), Farm Accountancy Data Network (FADN) were also considered in the works by Goraj *et al*. (2010), Skarżyńska *et al.* (2005).

Prior to 2000 Poland had had four SGM macroregions. On 29 November, 2000 due to the changes undertaken by Poland within the preparation for EU membership, a new act on collecting and using farm accountancy data was passed (Journal of laws from 2001, No. 3, item 20). Therefore, it was necessary to reconsider the borderlines of the SGM macroregions based on which Standard Gross Margins were calculated within the FADN system. A newly determined division into agricultural macroregions should allow the statistical results of the Polish agriculture to be presented appropriately. In consequence, a new composition of the following SGM macroregions was determined in Poland (see Skarżyńska, Goraj, Ziętek 2005):

– *the Pomorze and Mazury Region*. It comprises the lubuskie, pomorskie, warmińsko-mazurskie and zachodniopomorskie provinces.
– *the Wielkopolska and Śląsk Region*. It comprises the dolnośląskie, kujawsko-pomorskie, opolskie and wielkopolskie provinces.
– *the Mazowsze and Podlasie Region*. It comprises the lubelskie, łódzkie, mazowieckie and podlaskie provinces.
– *the Małopolska and Pogórze Region*. It comprises the małopolskie, podkarpackie, śląskie and świętokrzyskie provinces.

Figure 1. Poland's division into SGM regions



Source: elaborated by the author.

The existing composition of the four SGM macroregions was a starting point for the determination of the borderlines of the macroregions. In order to ensure the homogeneity of the macroregions in their levels of agricultural development and culture, the provinces were classified taking into account their degrees of similarity[8] as regards the underlying agricultural properties[9]. Eventually, as a result of the conducted taxonomic analysis the borderlines of the SGM macroregions regions were altered. The lubuskie province was moved into the Pomorze and Mazury macroregion, and the świętokrzyskie province to the Małopolska and Pogórze macroregion. The borderlines of the Wielkopolska and Śląsk as well as of the Mazowsze and Pogórze macroregions remained unchanged. The borderlines of the macroregions and of the provinces comprised by them, which were determined in accordance with the NUTS 2 classification, are shown in Figure 1. This composition was included in the annex to the Treaty of Accession of the Republic of Poland to the European Union.

The creation of the new composition of the SGM regions due to Poland's joining the European Union is an example of a positive solution of the aggregation problem[10]. This composition is included in *the quasi composition of regions* within the undertaken research problem. The prepared statistics based on the SGM macroregions and the performed economic analyses should lead to the obtainment of correct results.

Despite the fact that the problem of determining the borderlines of SGM regions found a positive solution, the paper attempted to identify again the composition of SGM macroregions within the research problem undertaken in the same way[11]. Instead of the actually conducted taxonomic analysis, an analysis of spatial volatility of the agrarian structure in Poland was performed in the paper[12]. The size of agricultural farms and their spatial vola-

[8] For the purpose of isolating macroregions a cluster analysis was performed based on the division into provinces and nine diagnostic agriculturally important variables were assumed (see Goraj *et al.*, 2005).

[9] The areas of provinces (NUTS2) were used for the purpose of determining the borderlines of agricultural macroregions.

[10] The composition of the macroregions was adjusted to the current situation of the Polish agricultural sector which in the time period 1989-2000 underwent numerous changes. This is a significant observation since it implies that within the identified research problem the composition of territorial units may change over time. This composition will be modified if there is a change in the spatial differentiation of the considered phenomena and in the dependence between them. However, at a selected time period only one composition of territorial units can be correct.

[11] The author is of the opinion that due to that *the aggregation problem* will be presented better.

[12] The concept of the agrarian structure was defined very narrowly as the deployment of farms by their area (Bukryba-Rylska, 2008; Woś, 1998).

tility constitute an important element of rational management in agriculture. The size of farms is important since small-sized farms are not capable of generating sufficient income that would enable them to function efficiently. Too large dispersion of farms means no possibilities for making progress, as well as increasing unit production costs. This justifies the use of the analysis of the agrarian structure for the purpose of the determination of macroregions that are homogeneous in the levels of development and culture (see Michna, 2005)[13].

The spatial volatility of the agrarian structure in Poland (the concentration of the area of agricultural land) will be described by means of the Gini coefficient[14]. The agrarian structure will be considered at a low aggregation scale NUTS 4 (districts) for the year 2002.[15] The obtained values of the concentration of the area of agricultural land were initially divided into two groups by the median[16] (Figure 2). Analysis of spatial volatility of the value of the Gini index allows the isolation of two regions, differentiated by the agrarian structure. The eastern region, with the exclusion of the warmińsko-mazurski region, is characterized by a low concentration of agricultural land and the western region (including the province) is characterized by a high concentration of agricultural land. The indicated division shows the existence of a significant difference in the agrarian structure of the two regions[17].

[13] The results obtained based on the analysis of the agrarian structure should be coherent with the results obtained based on the taxonomic analysis conducted for the nine diagnostic features.

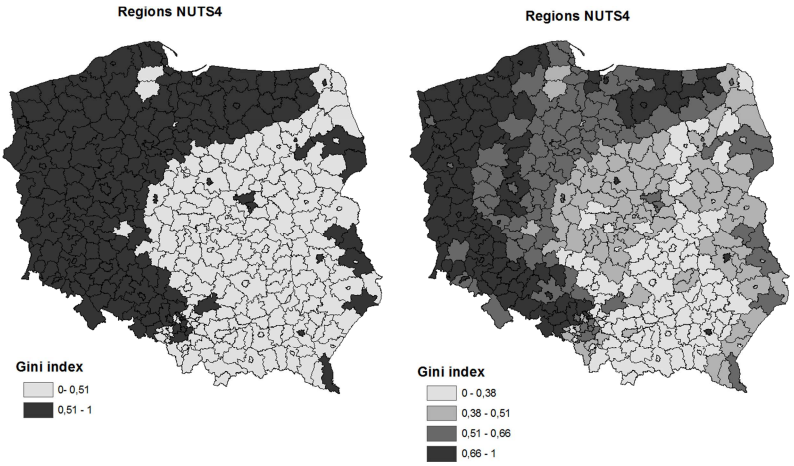
[14] The problem of the determination of the Gini coefficient was given a comprehensive analysis in the work by Ceriani & Verme (2011). The area of the agricultural land was calculated as the total area of farms minus woodland and wasteland. The following area groups were taken for the calculation of the Gini coefficient (1-5 ha), (5-10 ha), (10-20 ha), (20-50 ha) and (50 ha and more).

[15] This is the only step that allows the Gini coefficient to be determined on the level of districts. The data used were taken from the Agricultural Census made in 2002. All data are derived from the following website: www.stat.gov.pl.

[16] The first group was formed of districts with the concentration value below the designated median at the level of 0.57, and the other group of those with a higher value of the Gini coefficient.

[17] That is an important argument supporting the need for undertaking the research problem in the form of the determination of agricultural macroregions.

Figure 2. Spatial volatility of the concentration of agricultural land



Source: elaborated by the author.

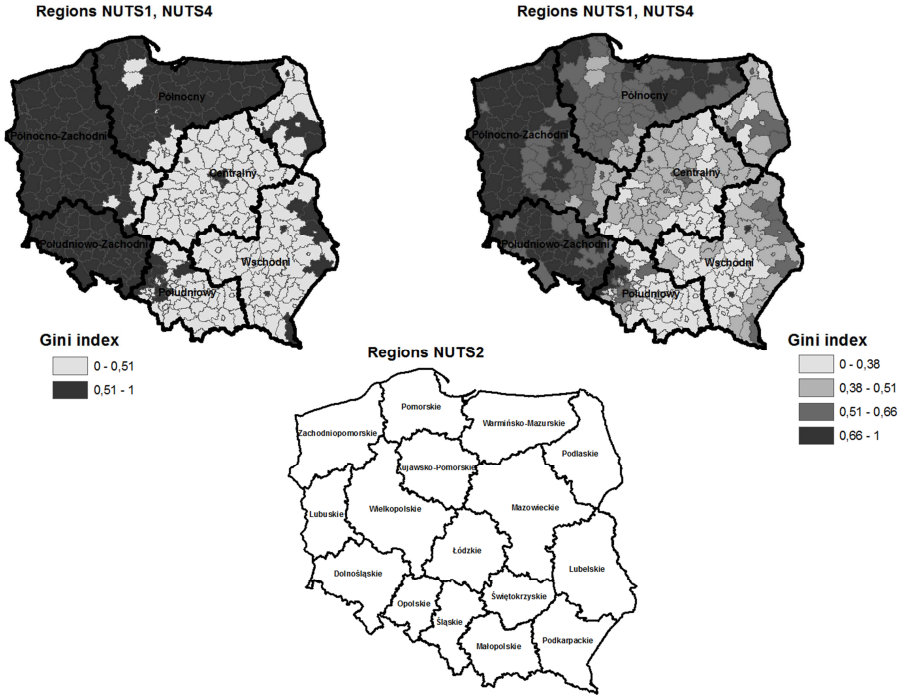
Figure 2 also presents the division of the concentration of agricultural land into four classes in accordance with the quartile values. The analysis of the four-class division allows the confirmation of the existence of even larger differentiation of the agrarian structure in Poland. That confirms again the need for creating macroregions differentiated by significant features concerning agriculture and at the same time being internally homogeneous in the level of the development of agriculture.

The first step made towards the achievement of *the aggregation problem* should be an attempt to use the already existing composition of territorial units which may turn out to be a correct composition[18]. Since the number of SGM macroregions should not be composed of too many units, the NUTS 1 (regions) will be taken as the initial composition of SGM macroregions. The composition of NUTS 1 regions was presented together with the spatial volatility of the agrarian structure shown in Figure 3. It can be observed that the NUTS 1 regions quite fit into the spatial volatility of the agrarian structure. Only in the case of regions 3 and 4 can we recognise them as heterogeneous relative to the agrarian structure.

[18] The paper assumes that the only available compositions are the compositions of territorial units within the NUTS classification. The composition of macroregions which had been used prior to 2000 will not be taken into consideration which is intended to provide a better presentation of *the aggregation problem*.

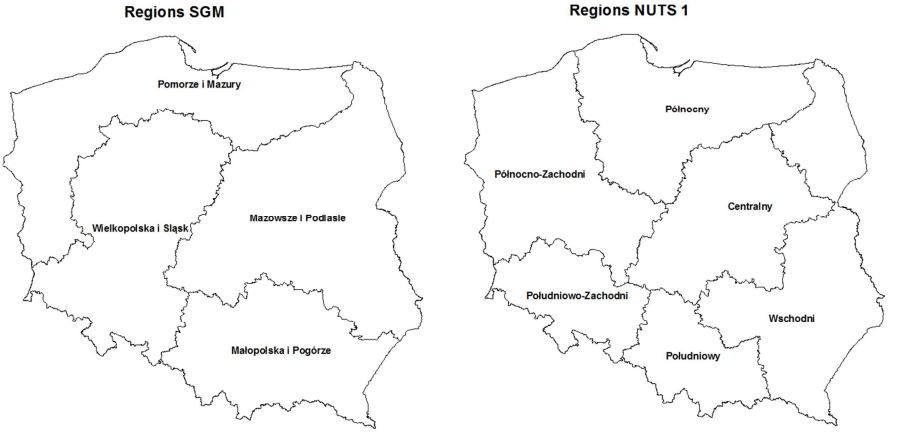
However, the additional consideration of all of the regions by their levels of agricultural development and cultures allows the formulation that within the eastern region the lubelskie province stands out form other regions of the region, in the north-western region it is the wielkopolskie province that stands out and in the northern region – the kujawsko-pomorskie province. The evaluation of further provinces leads to the isolation of four groups which share a similar degree of the development of agriculture. The lubuskie, pomorskie, warmińsko-mazurskie, and zachodniopomorskie provinces form the first group with the lowest intensity of agricultural production in Poland. The provinces have the lowest participation of agricultural land in the total possessed by private farms. That contributes to a high concentration of the area of agricultural land. What can be observed is the high share of the structure of the sown area of cereals and low livestock. Another group is composed of the dolnośląskie, kujawsko-pomorskie, opolskie, and wielkopolskie, provinces where the concentration of the sown area of cereals is recorded at a high level. Production is conducted very intensively, which is indicated by the highest level of the use of mineral fertilizers in Poland and the amount of concentrates purchased. In addition, the region is characterized by the highest level of qualifications possessed by farmers. The third group comprises lubelskie, łódzkie mazowieckie, and podlaskie provinces. Those provinces have agricultural farms with the average-sized area, a high participation of sustainable grassland and, which is related to that, high cattle population. The region is characterized by poor quality soil and a low level of fertilization. The participation of crops is high, however, the yield is low. All that contributes to the region's low intensity of agricultural production. The last group is composed of the małopolskie, podkarpackie, śląskie and świętokrzyskie provinces. This region is characterized by the agrarian dispersion and it has the lowest average farm area. The livestock population is the highest in the country, however, it is conducted on a small scale. The low intensity of production also results from the low intensity of the use of mineral fertilizers and from purchasing concentrates. Therefore, it must be stated that using a NUTS 1 composition for analysing agricultural issues will lead to incorrect results. Figure 4 shows visible differences between a NUTS 1 composition and the actual composition of SGM macroregions.

Figure 3. The composition of NUTS 1 regions and agricultural macroregions as compared with the spatial volatility of the agrarian structure



Source: elaborated by the author.

Figure 4. Agricultural SGM macroregions and the NUTS 1 classified regions



Source: elaborated by the author.

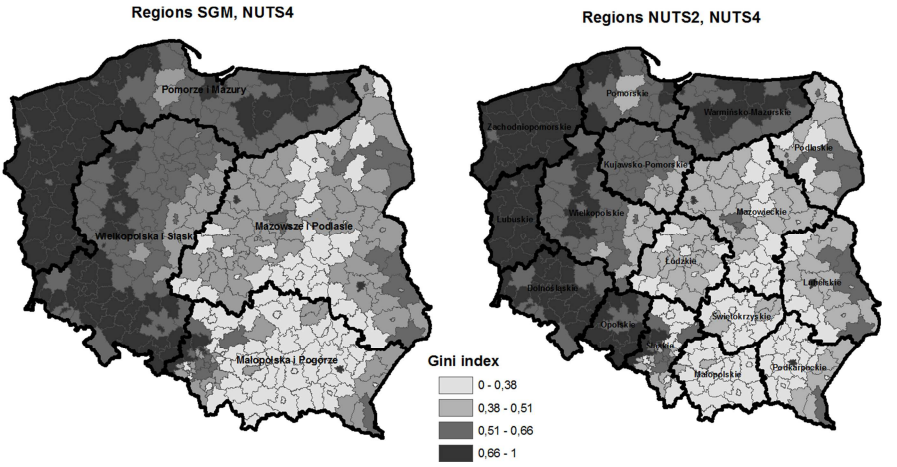
The rejection of a NUTS 1 composition makes it impossible for the existing composition of territorial units to be used. The only way to solve the situation is creating a single composition that will ensure the correctness of the results obtained within the undertaken research problem. It helps to use areas with a lower aggregation scale[19], while determining a single composition of territorial units individually. Combining such areas will give the territorial units of the composition being created. In the case of the determination of agricultural macroregions, the analysis of the spatial volatility of the agrarian structure and the analysis of the provinces in the scope of the level of the development of agriculture should be combined. The consideration of the borderlines of the aforementioned four groups of provinces, similar in the level of agricultural development and culture, indicates the homogeneity of those groups in their spatial volatility of the agrarian structure (Figure 5). That means that agricultural macroregions may be created out of the provinces belonging to the four distinguished groups. It turns out that the macroregions determined in that way (based on the analysis of the concentration of agricultural land and on the researcher's knowledge and experience) overlap with the official composition of the SGM macroregions. This consistency results primarily from the connection between a region's agrarian structure and its situation regarding agriculture. The faulty agrarian structure acts as a powerful destimulant for the appropriate changes in and development of agriculture. Unfortunately, the analysis

[19] In the case of determining agricultural macroregions it is best to use provinces (NUTS 2).

conducted shows significant disproportions in the agrarian structure in Poland (Figure 5), which is reflected in the substantial disproportions in the level of the key variables of the agricultural activity. Therefore, the agricultural policy conducted by the state should differ depending on the macroregion.

The presented procedure of the determination of agricultural macroregions led to a satisfactory solution of *the aggregation problem*. The deepened consideration of the empirical research problem formed the basis for the problem solution. The implication is that the considerations were not limited merely to the determination of the macroregions but also concerned the situation in the Polish agriculture.

Figure 5. The composition of the NUTS2 regions and of the agricultural macroregions compared against the spatial volatility of the agrarian structure



Source: elaborated by the author.

Final areal interpretation problem

The problem which may additionally occur within the analysis conducted in the paper is *the final areal interpretation problem* (see: Pietrzak, 2014a). The problem concerns the situation in which the characteristics of the phenomena or the strength of the dependency between them have been determined for excessively extensive areas. That is related to the incorrect for-

mulation of condition 4[20], where an inappropriate area was selected based on which conclusions are drawn from the conducted analysis. Spatial process within the wrongly identified area lose the preferred properties (homogeneity and systematic heterogeneity) in relation to the applied measures and models[21]. A mistake made while analysing consists in drawing conclusions which will not be correct for the selected area since the accepted measures or models will not have a cognitive value. Avoiding the final areal interpretation problem means reducing the area of analysis with a view to obtaining appropriate properties of spatial processes or accepting other adequate research tools.

In the case of the analysis performed in the paper, the final areal interpretation problem may occur while attempting to designate the average concentration of Poland's agricultural area (the country's area, NUTS0) based on the Gini coefficient for districts (NUTS4)[22]. A correct way of proceeding undertaken in order to avoid the final areal interpretation problem consists in a precise analysis of the spatial volatility of the agrarian structure in Poland[23]. The analysis of Figure 1 immediately indicates that the obtained measure (the average) does not possess a cognitive value due to a significant differentiation of the spatial concentration of agricultural land at the level of districts[24]. In the case of the phenomenon of the concen-

[20] One of the four conditions ensuring the formulation of correct conclusions based on the spatial analyses conducted. The conditions were presented in the earlier part of the paper.

[21] Spatial data describing the phenomena are regarded as the realisations of spatial processes (two-dimensional random fields, see Arbia, 2006; Pietrzak, 2010a, 2010b). Within the considered structure of the internal structure of spatial processes we can distinguish such constituents of the structure as, for instance, the property of the homogeneity (spatial dependence, see Anselin, 1988; Pietrzak *et al.*, 2014), of the systematic heterogeneity and of the unsystematic heterogeneity (see Pietrzak, 2014a). Applying selected measures or models calls for fulfilling properties specific for spatial processes (see Pietrzak, 2010a, 2010b, 2013).

[22] Most frequently the Gini coefficient based on the data aggregated for the whole territory of Poland is designated for the needs of measuring the concentration of the area of agricultural land in Poland. In such a situation it is impossible to consider the spatial volatility of the agrarian structure in Poland. Comparing countries with the application of that measure becomes problematic since two countries with various spatial volatility may obtain the same values of the measure. The measure obtained in such a way is usually used in a preliminary evaluation of the countries under examination.

[23] The need for conducting research on spatial volatility concerning the phenomena occurring within the analyses performed was discussed in the works by Wegenast (2010) and Roberts (1982).

[24] In that case it does not really matter on the basis of what composition of territorial units and at what scale of aggregation the average concentration of agricultural land was calculated. All compositions lead to the obtainment of incorrect results which is the essence of *the final areal interpretation problem.*

tration of the area of agricultural land the following conclusion may be drawn: the analysed spatial process for the area of Poland has the properties of systematic heterogeneity. The applied measure (the average) requires fulfilling the property of homogeneity and, therefore, the values obtained for Poland will not have any cognitive value (see Pietrzak, 2014a). Further analysis of the spatial volatility of the concentration of agricultural land (the agrarian structure) at the level of districts will allow the properties of the homogeneity of the provinces (NUTS2) and of the larger SGM macroregions to be identified. If, based on the Gini coefficient calculated for districts, we designate the average for a particular province or a particular SGM region, the result obtained should reflect correctly the size of the average concentration of the agricultural land for that particular area.

Due to the spatial volatility of the agrarian structure in Poland, the analysis should consist of two steps. In the first step, one should isolate that agricultural land that is homogeneous relative to the agrarian structure. These can be the four SGM areas. The second step consists in describing each area separately, as well as the ties between the areas. This is the only way in which Poland's agrarian structure can be described[25]. Table 1 contains the average values of the Gini index calculated separately for the four agricultural macroregions. The volatility in the average indicates the differentiation of the agrarian structure and the related situation in agriculture depending on the accepted macroregion.

Table 1. The values of the average concentration of the area of agricultural land and the related situation in agriculture depending on the accepted macroregion

Agrarian macroregion	Average concentration of the area of farming grounds
Region Pomorze i Mazury	0,698
Region Wielkopolska i Śląsk	0,626
Region Mazowsze i Podlasie	0,457
Region Małopolska i Pogórze	0,362

Source: elaborated by the author.

[25] For the purposes of statistics the average values of the basic economic ratios are taken into account. Then, these characteristics are used in initial comparison of countries relative to the scale of the analysed phenomena.

Due to the analysis of the agrarian structure in Poland conducted within the present paper, it is possible to formulate another research problem[26] – which composition of territorial units will allow the obtainment of correct values of the average concentration of the area of agricultural land in Poland (NUTS0)[27]? The composition searched for will be designated basing on the NUTS4[28] areas (districts), however, it is necessary to determine the number of territorial units within the composition. In the case of the analysis of the agrarian structure and in the light of the presented problem of *the final areal interpretation problem*, the research problem formulated in such a way seems to be incorrect. The incorrectness of the research problem results from the fact that none of the compositions of territorial units at any aggregation scale ensures the obtainment of correct results for the area of Poland (NUTS0). However, an attempt to solve the inadequate research problem[29] leads to the obtainment of a wide range of results for various compositions of territorial units, which will result in the identification of the MAUP issue.

Within the research problem formulated for the purpose of the determination of the searched composition of territorial units, Openshaw and Taylor (1979) propose the possibility of generating compositions of territorial units in a random way within the zoning or grouping systems. Following that way, it is possible to obtain a set of potential compositions of territorial units. Another step is choosing the best composition due to the assumed function of the purpose which is to be ensured by the use of an automatic zoning algorithm (see Openshaw, 1977a, 1977b, 1977c).

However, the random choice of compositions will result in a wide range of the obtained values of the average concentration of agricultural land. In addition, the randomly generated regions should be objected by the researcher due to their borderlines and shapes. To illustrate that problem, let us assume that the generated compositions are composed of 16 areas and the basic unit used in the creation of areas are districts (NUTS4). Figure 6 shows three exemplary compositions of territorial units. The first of them

[26] In spatial econometrics the problem was referred to by Openshaw & Taylor (1979) in the following form 'The question is simply what objects and at what scales we wish to investigate'.

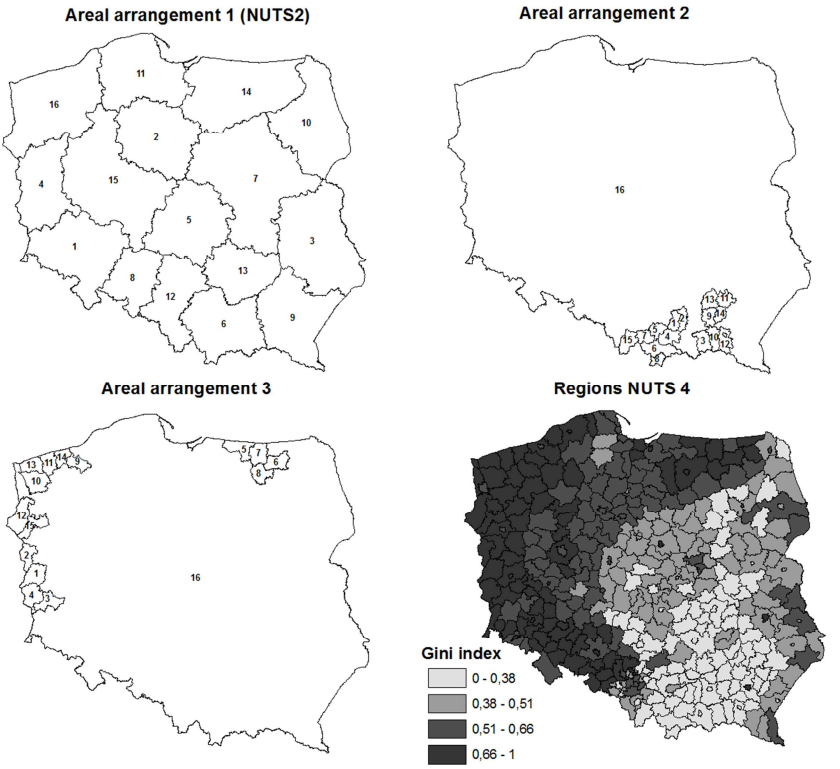
[27] The Gini index will be calculated for all selected compositions of territorial units, and then, based on the values obtained, the average for the whole territory of Poland will be designated.

[28] All areas of the new composition of territorial units will be composed of smaller NUTS4 units.

[29] The research problem determined in that way can be undertaken and solved.

corresponds to the NUTS2 classification and the further two were created arbitrarily by the author[30].

Figure 6. The compositions generated within the Zoning System



Source: elaborated by the author.

In addition, Figure 6 shows the spatial volatility of the concentration of the area of agricultural land. The arbitrarily created compositions depart from the NUTS2 composition significantly and show clearly what kinds of compositions can be obtained within the zoning system. Taking into consideration the spatial volatility of the concentration of the area of agricultural land (presented in Figure 6), the obtainment of various average values

[30] Such compositions can be obtained through a random generation of compositions of territorial units within the zoning system.

for each composition is obvious[31]. The average for the first composition is 0.51, for the second it is 023 and 0.75 in the case of third composition. The obtained wide range of average values for a set of potential compositions of territorial units leads to the conclusions on the identification of *modifiable areal unit problem*. Due to *the final areal interpretation problem*, the obtained set of potential average values will not have a cognitive value. The implication is that the identification of the MAUP is only apparent.

Conclusions

The paper presents *the aggregation problem*, which constitutes one of the aspects of the issue that is frequently discussed in spatial economics – the *modifiable areal unit problem* (MAUP), one of the issues debated within spatial econometrics. The issue of MAUP is connected with the possibility of the obtainment of various results relative to changes in the aggregation scale or in the composition of territorial units for the same aggregation scale. The objective of the paper was to consider redefining of *the aggregation problem* and to show a possibility of a positive solution of *the aggregation problem* based on an empirical example.

In the paper *the aggregation problem* was considered as a problem of the determination of a particular composition of territorial units at a selected aggregation scale in such a way that it could be encompassed by *a quasi composition of regions* within the undertaken research problem. Therefore, a positive solution of *the aggregation problem* consists in determining a composition of territorial units which will ensure a correct analysis of phenomena.

The positive solution of *the aggregation problem* was presented based on the example of the determination of SGM macroregions in Poland for the needs of a EU system of *Farm Accountancy Data Network* – FADN. The SGM macroregions should be distinctively differentiated by the values of the variables relating to agriculture as well as should be internally homogeneous as concerns the development of agriculture. The paper also presented the procedure for determining agricultural macroregions where the analysis of the spatial volatility of the agrarian structure and the current knowledge on the agriculture in Poland were applied. The determination of the appropriate composition of agricultural macroregions allowed *the aggregation problem* to be solved positively.

[31] Each of the 16 areas is composed of a different number of basic units (NUTS4 districts). The average values are determined based on the Gini indices applied for of the 16 areas of the accepted composition of territorial units.

Moreover, the paper considered *the final areal interpretation problem* connected with the incorrect determination of the area in relation to which final conclusions from the previous analysis were drawn. The problem was presented based on the example of the determination of the average concentration of the area of agricultural land in Poland (NUTS0, the country's territory) with the use of the Gini index calculated for districts (NUTS4). The paper also emphasised that ignoring *the final areal interpretation problem* in spatial analyses may lead to an apparent identification of *the modifiable areal unit problem*.

References

Anselin L. (1988), *Spatial Econometrics: Method and Models*, Kluwer Academic Publishers, Netherlands.

Arbia G. (1989), *Spatial Data Configuration in Statistical Analysis of Regional Economics and Related Problems*, Kluwer Academic Publisher, Dordrecht.

Arbia G. (2006), *Spatial Econometrics, Statistical Foundations and Applications to Regional Convergence*, Springer-Verlag, Berlin Heidelberg.

Blalock H. (1964), *Causal inferences in nonexperimental research*, University of North Carolina Press, Chapel Hill.

Bukraba-Rylska I. (2008), *Socjologia wsi polskiej*, PWN, Warszawa.

Ceriani L., Verme P. (2011), *The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini*, „Journal of Economic Inequality", Vol. 10, http://dx.doi.org/10.1007/s10888-011-9188-x.

Flowerdew R. (2011), *How serious is the Modifiable Areal Unit Problem for analysis of English census data?*, „Population Trends", No 145, http://dx.doi.org/10.1057/pt.2011.20.

Goraj L., Cholewa I., Osuch D., Płonka R. (2010), *Analiza skutków zmian we Wspólnotowej Typologii Gospodarstw Rolnych*, Instytut Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej - Państwowy Instytut Badawczy, Warszawa.

Holt D., Steel D.G., Tranmer M. (1996), *Area homogeneity and the modifiable areal unit problem*, „Geographical Systems", Vol. 3.

Fotheringharn, A.S., Wong D.W.S. (1991), *The modifiable area unit problem in multivariate analysis*, „Environment und Planning A", Vol. 23.

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", Vol. 29, http://dx.doi.org/10.2307/2277827.

Manley D., Flowerdew R., Steel D. (2006), *Scales, levels and processes: Studying spatial patterns of British census variables Computers*, „Environment and Urban Systems", Vol. 30.

Marble D.F. (2000), *Some thoughts on the integration of spatial analysis and Geographic Information Systems*, „Journal of Geographical Systems", Vol. 2, http://dx.doi.org/10.1007/s101090050026.

Michna W. (2005), *Zróżnicowanie funkcji gospodarstw rolnych w ujęciu przestrzennym*, Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy, Warszawa.

Openshaw S. (1977a), *A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling*, „Transactions of the Institute of British Geographers", New Series, Vol. 2, http://dx.doi.org/10.2307/622300.

Openshaw S. (1977b), *Algorithm 3: a procedure to generate pseudo-random aggregationsof N zones into M zones, where M is less than N'*, „Environment and Planning A", Vol. 9.

Openshaw S. (1977c), *Optimal zoning systems for spatial interaction models*, „Environment and Planning A", Vol. 9, http://dx.doi.org/10.1068/a090169.

Openshaw S., Taylor P.J. (1979), *A million or so correlation coefficients: three experiments on the modifiable areal unit problem*, [in:] Wrigley N. (ed.), *Statistical methods in the spatial sciences*, London: Pion.

Openshaw S. (1984a), *The Modifiable Areal Unit Problem*, GeoBooks, CATMOG 38, Norwich.

Openshaw S. (1984b), Ecological *fallacies and the analysis of areal census data*, „Environment and Planning A", Vol. 16, http://dx.doi.org/10.1068/a160017.

Paelinck J.H.P. (2000), *On aggregation in spatial econometric modelling*, „Journal of Geographical Systems",Vol. 2, http://dx.doi.org/10.1007/PL00011452.

Pietrzak M. B. (2010a), *Analiza danych przestrzennych a jakość informacji*, [in:] Trzaskalik T. (ed.), *Modelowanie preferencji a ryzyko '09*, Wydawnictwo Uniwersytetu Ekonomicznego, Katowice.

Pietrzak M. B. (2010b), *Problem identyfikacji struktury danych przestrzennych*, „Acta Universitatis Nicolai Copernici Ekonomia", XLI, z. 397, s. 83-98.

Pietrzak M. B. (2013), *Interpretation of Structural Parameters for Models with Spatial Autoregression*, „Equilibrium" Vol. 8 I. 2,, s. 129-155, http://dx.doi.org/10.12775/EQUIL.2013.010.

Pietrzak M. B., (2014a), *Redefining the modifiable areal unit problem within spatial econometrics, the case of the scale problem*, „Equilibrium" Volume 9, Issue 2 (forthcoming).

Pietrzak M. B. (2014b), *The modifiable areal unit problem - analysis of correlation and regression*, „Equilibrium", Volume 9, Issue 4 (forthcoming).

Pietrzak M.B, Wilk J,. Kossowski T., Bivand R. (2014), *The identification of spatial dependence in the analysis of regional economic development - join-count test application*, [in:] Papież M. & Śmiech S. (ed.), *Proceedings of 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Cracow: Foundation of the Cracow University of Economics (forthcoming).

Reynolds H. D. (1998), *The Modifiable Area Unit Problem: Empirical Analysis by Statistical Simulation*, doctoral thesis, graduate Deparment of Geography, University of Toronto.

Roberts K.D. (1982), *Agrarian structure and labor mobility in rural Mexico*, „Population and Development Review", Vol. 8, http://dx.doi.org/10.2307/1972988.

Robinson W.S. (1950), *Ecological Correlations and the Behavior of Individuals*, „American Sociological Review", Vol. 15, No. 3, http://dx.doi.org/10.2307/2087176.

Skarżyńska A., Goraj L., Ziętek I. (2005), *Metodologia SGM „2002" dla typologii gospodarstw rolnych w Polsce*, Instytut Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej - Państwowy Instytut Badawczy, Warszawa.

Suchecki B. (ed.) (2010), *Ekonometria Przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa.

Szulc E. (2007), *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*, Wydawnictwo UMK, Toruń.

Tate N., Atkinson P.M. (ed.) (2001), *Modelling scale in geographical information science*, John Wiley & Sons, Chichester.

Tranmer M., Steel D. (2001), *Using Local Census Data to Investigate Scale Effects* [in:] Tate N., Atkinson P. (ed.), *Modelling scale in geographical information science*, Chichester: John Wiley & Sons.

Woś A. (ed.) (1998), *Encyklopedia agrobiznesu*, Fundacja Innowacja, Warszawa.

Yule G.U., Kendall M.G. (1950), *An introduction to the theory of statistics*, Griffin, London.