ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Pietrzak, Michal Bernard

Working Paper

Redefining The Modifiable Areal Unit Problem Within Spatial Econometrics, The Case Of The Scale Problem

Institute of Economic Research Working Papers, No. 35/2013

Provided in Cooperation with:
Institute of Economic Research (IER), Toruń (Poland)

Suggested Citation: Pietrzak, Michal Bernard (2013) : Redefining The Modifiable Areal Unit Problem Within Spatial Econometrics, The Case Of The Scale Problem, Institute of Economic Research Working Papers, No. 35/2013, Institute of Economic Research (IER), Toruń

This Version is available at:
https://hdl.handle.net/10419/219561

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by/3.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Institute of Economic Research Working Papers

No. 35/2013

Redefining The Modifiable Areal Unit Problem Within Spatial Econometrics, The Case Of The Scale Problem

Michał Bernard Pietrzak

Toruń, Poland 2013

© Copyright: Creative Commons Attribution 3.0 License

Michał Bernard Pietrzak
pietrzak@umk.pl
Nicolaus Copernicus University in Toruń, Department of Econometrics and Statistics, ul. Gagarina 13a, 87-100 Toruń

Redefining The Modifiable Areal Unit Problem Within Spatial Econometrics, The Case Of The Scale 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 modifiable areal unit problem (MAUP), which is frequently discussed within spatial econometrics. This issue concerns the changeability of the characteristics of the analysed phenomena under the impact of the change in the composition of territorial units (see Openshaw, Taylor 1979; Arbia 1989). The article indicates four conditions which need to be fulfilled if the correctness of spatial analyses is to be maintained. Also, the paper introduces the concept of the quasi composition of regions (QCR). It was defined as a set of particular compositions of territorial units for subsequent aggregation scales. Particular compositions of territorial units are selected in a way that allows a correct analysis within the undertaken research problem to be conducted. The chief asset of the paper is the proposal to redefine the concept of the modifiable areal unit problem. Both the scale problem and the aggregation problem were linked to the accepted quasi composition of regions. The redefinition of the concept is vital for the research conducted since analysing phenomena based on compositions of territorial units which are excluded from the quasi composition of regions leads to the formulation of incorrect conclusions. Within the undertaken research problem there exists only one particular composition of territorial units which allows the identification and description of the dependence for analysed phenomena. Within the considered modifiable areal unit problem two potential problems were defined and they can occur while making spatial analyses. The first is the final areal interpretation problem (FAIP) that occurs when the characteristics of phenomena or the dependence are designated for too large region. The other issue is the aggregation scale interpretation problem (ASIP). It occurs when a quasi composition of regions is enlarged by an aggregation scale where the correctness of the results of the undertaken research problem is not preserved. In both cases it

is possible to reach a situation where the obtained characteristics will be deprived of the cognitive value.

Introduction

The paper focuses on *the modifiable areal unit problem* (MAUP) in spatial analyses conducted. The issue of *the modifiable areal unit problem* is defined within spatial econometrics as the changeability of the properties of data under the impact of a change in the composition of territorial units (areal arrangement) at the accepted aggregation scale or under the impact of a change in the aggregation scale. The research thesis formulated in the paper has the following wording: 'how to obtain correct results within analyses made for spatial data?' The answer to the question will enable us to provide a comprehensive study of the issue of *the modifiable areal unit problem* that has already been mentioned in numerous works Gehlke and Biehl (1934), Yule and Kendall (1950), Robinson (1950), Blalock (1964), Openshaw and Taylor (1979), Openshaw (1984a, 1984b), Reynolds (1988), Fotheringharn and Wong (1991), Holt, Steel, and Tranmer (1996), Tranmer and Steel (2001), Arbia (2006), Manley, Flowerdew, and Steel (2006), Suchecki(ed.) (2010), Flowerdew (2011) and Pietrzak (2014a, 2014b).

The research objective of the paper is to indicate the underlying conditions that are indispensable for the appropriateness of analyses based on spatial data. Then, based on analyses performed, *the modifiable areal unit problem* will be redefined.

Spatial economic processes create the base for analyses performed within spatial econometrics. The realizations of those processes in the form of spatial data are most frequently referred to as irregular regions (polygons), which results from the nomenclature adapted for determining boundaries of those regions. Both in the European Union and in Poland the measurement of major socio-economic characteristics of regions is made in accordance with the NUTS classification (Nomenclature of Units for Territorial Statistics). The purposes of the implementation of this nomenclature was to provide EU member states with comparable methods of data collection and interpretation as well as of making them easily available within the EU area. The NUTS 0 level defines European Union member states. In the case of Poland, lower levels of the classification of data aggregation denote the following: NUTS 1 – regions, NUTS 2 – provinces, NUTS 3 - subregions, NUTS 4 – districts, and NUTS 5 – municipalities. The order of the NUTS levels is not incidental and analyses of the majority of economic phenomena as well as dependence held

between them following the NUTS classification lead to the obtainment of correct results.

Spatial analyses of Poland, or of the European Union, carried out by various researchers, are usually based on irregular regions corresponding to the NUTS classification which results from data availability[1]. Due to this fact, the considerations made in the present paper will be limited to irregular regions. The next assumption made is to analyse data expressed in relative quantities referring to certain values characterising irregular regions (area, population). This is to be ensured by the comparability of data which also ensures the correctness of obtained results. The additional assumption of analysing data expressed in relative quantities data excludes the possibility of the occurrence of the *economic fallacy problem* and as such this problem will be omitted in the paper. The two aforementioned assumptions will definitely limit the field under research, however, they will allow many valuable conclusions to be drawn, which, otherwise, would have been diluted.

It must be emphasised that all data published under the NUTS classification are spatial data[2]. Spatial data are characterized by two properties, i.e., by spatial heterogeneity and the existence of spatial dependence (see Anselin 1988; Pietrzak et al. 2014). Any economic analysis that does not consider the above-mentioned properties of spatial data leads to cognitive errors, which undermines the reliability of its results. The issue raised indicates the need for developing and applying the tools of spatial econometrics in economic research (see Pietrzak 2010a, 2010b, Pietrzak 2013).

The conditions necessary for conducting reasonable analyses of spatial data

In this subchapter, an attempt will be made to answer the question on when an economic analysis based on spatial data referring to irregular regions gives correct results. The consideration of this issue leads to the

[1] Analyses conducted may concern both regular and irregular regions. However, due to the cost of and the time necessary for obtaining data, readily-made statistical information prepared by reliable public statistical institutions is used, which, in turn, will conduct research in the EU in accordance with the NUTS classification.

[2] Spatial data are regarded in economic analyses as cross-sectional data, which may lead to the obtainment of incorrect results of research.

identification of four underlying conditions which need to be met, if the correctness of conducted economic analyses is to be ensured.
Condition 1. A starting point in every analysis is the formulation of a research problem and taking into account all aspects relevant to the problem.
Condition 2. Establishing the aggregation scale for spatial data that would be appropriate for drawing correct conclusions. The determination must be realized within the undertaken research problem.
Condition 3. Spatial data based on which conclusions are drawn need to reliable.
Condition 4. Determining the size (boundaries) of a region in relation to which the formulated conclusions will be correct. The determination must be realized within the undertaken research problem.

In the case of an economic analysis, *condition 1* necessitates setting a starting point, which is the formulation of a research problem. Only within the formulated research problem do we make a decision as to which phenomena should be examined and we set research hypotheses related to the properties of these phenomena, or to the dependence held between them. The next assumptions made concern the time period of the analysis, the space scope and the aggregation scale of data, etc. All decisions are taken within 'the formulated research problem', where the researcher applies suitably his knowledge and scientific experience. If the researcher is going to obtain correct results, then he needs to conduct research in the way required by the research problem undertaken. It must be stressed that various research problems may require different aspects of the knowledge and experience possessed by a specific researcher. The realised research objectives and formulated hypotheses stem from the research problem formulated. It is unacceptable for the researcher to determine a research objective irrespective of the formulated research problem.

As regards *condition 2*, the choice of the aggregation scale for spatial data is made and based on it conclusions will be drawn from the analysis conducted. The aggregation scale is so determined that the researcher may state that the data[3] applied to each irregular region originate form the impact of a homogenous system within this region. Besides, a similarity of systems[4] should occur in all of the regions which shape the phenomena considered within the undertaken research problem.

[3] We assume that spatial data describe a phenomenon shaped within a specific economic system.

[4] The problem of the impact of a homogenous system within a region and a system resemblance occurring in a set of regions need to be analysed theoretically and elaborated, which significantly exceeds the framework of the present paper.

As a result of the fulfilment of *condition 2*, researchers are provided with data that set a background for formulating conclusions. In a further part of the paper *condition 2* will be extended by the concept of '*the quasi composition of regions'* (QCR)'.

The reliability of spatial data is to be ensured by providing them by specialized units, usually by public statistical units. In the majority of spatial economic studies researchers use data derived from public statistical units and analyses are conducted in accordance with the NUTS nomenclature. A problem that appears here is the lack of data for selected phenomena or providing a data aggregation scale that is too large for the defined research problem. In the case of analysis of economic phenomena, spatial data can be treated as the realisation of the $X(u)$[5] two-dimensional random field, later on referred to as 'a spatial process' (see Arbia 1989; Arbia 2006; Szulc 2007; Pietrzak 2010a, 2010b). Economic phenomena are analysed based on spatial data related to a selected aggregation scale (e.g., a province, NUTS 2). Conclusions drawn on a given phenomenon are then referred to a higher aggregation scale (e.g., a country, NUTS 0). Since the spatial data referred to a selected aggregation scale are treated as the realisations of spatial processes[6], then an appropriate identification of their *internal structure* becomes significantly important. The identification of their *internal structure* of spatial processes means a correct description of their properties[7]. In the case of spatial processes, the following elements of *the internal structure* can be distinguished: an element related to *unsystematic heterogeneity*, an element related to *systematic heterogeneity*, an element of the structure with a homogenous spatial process (*homogeneity*). The identification of *the internal structure* of a spatial process is made through the establishment of the process properties within further elements of this structure[8].

The description of *the internal structure* will be commenced with the element related to *the homogeneity* of spatial process[9]. *The homogeneity*

---

[5] u=(u1,u2) are the coordinates of the space.

[6] Since economic processes are unexperimental, we infer the properties of the process based on their individual realisations.

[7] The identification of *the internal structure* of spatial processes should constitute a base for developing spatial econometric models within spatial econometrics.

8 Not all of the elements need to be included in *the internal structure* of a spatial process.

9 The element of *the homogeneity* of *the internal structure* allows the property of spatial autocorrelation to occur. Also, the problem of spatial dependence is

of spatial process is understood in the paper as a weaker concept of stationarity[10], (stationarity is understood here in a broad sense) in the case of which the following assumptions are realised:

$$E(X(u)) = \mu$$

$$K(h) = \frac{1}{|N(h)|} \sum_{N(h)} \left([X_i(u) - \mu][X_j(u) - \mu] \right) = K(u_i, u_j') \quad , \qquad (1)$$

$$h = u_i - u_j, u_i = (u_{1i}, u_{2i}), u_j = (u_{1j}, u_{2j})$$

where E(X(u)) and K(h) are the function of the expected value and the function of the covariance of spatial process X(u) subsequently, h is the distance between site i and site j, $N(h)$ is a set of location pairs (see Szulc 2007).

The identification of *systematic heterogeneity* consists in finding properties related to systematic changes in the expected value, variance or in covariance. This element may be modelled, for instance, by means of a spatial trend, random coefficient model, spatially switching model, etc.

The last element of the internal structure of data is *unsystematic heterogeneity*, which means that a researcher is unable to determine systematic changes in the expected value, variance or in the covariance function.[11]

Condition 4 determines the boundaries (size) of the region in relation to which conclusions will be drawn within the conducted analysis. Such a region is composed of regional units with the aggregation scale defined in *condition 2*. Conclusions may be drawn only for the region whose data are characterized by *systematic heterogeneity* or *homogeneity*.

---

related to the problem of the identification of *the internal structure* of spatial process.

[10] The property of *homogeneity* denotes the unchangeability of the expected value, spatial variance, and the independence of the covariance function on spatial movement.

11 It may happen that the identification of properties related to *unsystematic heterogeneity* will be insufficient. Then, after deducting the part related to the element of *systematic heterogeneity* from the data, the identification of the element of *homogeneity* is impossible. It happens so because *the homogeneity* element is overlapped by the element related to *unsystematic heterogeneity* or to *systematic heterogeneity*. However, obtaining *the homogeneity* of a spatial process, due to the deduction from the data the part related to the element related to *systematic heterogeneity*, excludes the occurrence of the element of *unsystematic heterogeneity* in the structure.

The choice of the measures or of the model for describing the phenomena of the formulated research problem is significant[12].

Redefining the modifiable areal unit problem

Spatial economic phenomena are gathered and published predominantly by public statistics institutions within the NUTS classification. Data collected by such institutions are reliable due to the application of an appropriate methodology. Also, they are representative for examined regions due to taking suitable samples. Data are presented in an aggregated form and refer to specific irregular regions. The aggregation of data results from the mandatory requirement to keep the confidentiality of statistical data, where the surveyed entity needs to stay anonymous. In addition, research conducted by public statistics is repetitive, which gives it additional advantage. Data presented in accordance with the NUTS classification are not incidental and in the majority of studies they reflect adequately the problem under research. That means that the researcher based on his/her knowledge and scientific experience would also relate the analysed phenomena to the regions corresponding to the NUTS classification. It needs to be emphasised that obtaining data is so costly that hardly anyone can afford to commission conduction research on an arbitrarily selected composition of units with a specified aggregation scale[13]. These are the actualities of doing spatial research, where the foundation of the data reliability is created by their being published by public statistical institutions. This reality is quite distinct from the views presented in a work by Openshaw and Taylor (1979), where it is assumed that compositions of territorial units are *arbitrary in nature*. This arbitrary character consists in creating by researchers, firstly, one particular set of units and, then, based on it, conducting an analysis of specific phenomena. It must be noted that irregular regions are modifiable, which means that their boundaries and shapes may be created freely. This freedom is significantly limited by the undertaken research problem. The decision on the boundaries and shape is made arbitrarily by the researcher[14]. However,

[12] For instance, it is possible that a measure or a model requires process *homogeneity* , then the data within the region accepted need to be homogeneous.

[13] For instance, the cost of obtaining data on the unemployment rate registered for European Union member states at NUTS 4 level is estimated in millions of euros.

[14] Researchers may use the already existing composition of territorial units.

the accepted composition is related to the undertaken research problem and the researcher's scientific experience. That means that two independent researchers should take similar compositions of units within the same research problem. In order to describe such a situation the author proposes to refer to compositions of territorial units as ones *scientifically arbitrary in nature*. A research question formulated by Openshaw and Taylor (1979) had the following form: *'The question is simply what objects at what scales do we wish to investigate?'*. Attempts to answer their question unfortunately result in the arbitrariness of decisions about compositions of territorial units. Within an arbitrary composition of spatial units, spatial units may be grouped in any way. Connecting them with various shapes and various sizes of territorial units leads to a big number of potential compositions at any aggregation scale. This is a starting point for defining the *modifiable areal unit problem*, where the source of problem is the irregularity of shapes and the arbitrariness of their composition. However, a composition of territorial units at a selected aggregation scale is not random and should result from the undertaken research problem.

The above quoted work proposes two systems of an arbitrary creation of compositions of territorial units. The first system is 'a zoning system' which is a form of a contiguous territorial unit composition. The other system is 'a grouping system' which, in turn, is a form of a non-contiguous territorial unit composition. It is assumed that within the systems compositions of territorial units are multiple and the researcher is free to choose the best composition taking into account a given objective criterion. In addition, Openshaw (1977a, 1977b, 1977c) proposes an automatic zoning algorithm, within which, as a result of the purpose ascribed to the function, a composition of territorial units is obtained which is optimizing its value. However, there is only one particular set of units for a specific piece of research[15] which should be defined by the researcher within the formulated research problem. If the researcher does not consider the problem within the appropriate composition of territorial units, the performed analysis will be incorrect. Moreover, the objective criterion will not lead to choosing an inappropriate composition of territorial units, since it is not related to a specific research problem. Measuring properties and dependence between phenomena is justified only within a correct composition of territorial units. Any other composition will disturb the existing dependence. When accepting the arbitrariness of compositions[16],

[15] Phenomena cannot occur at the same time in two or more various compositions.

[16] The very idea of creating arbitrary compositions within the zoning system appears to be scientifically attractive. However, creating single compositions

we may obtain incorrect values of characteristics from a relatively wide range (see Openshaw, Taylor 1979; Reynolds 1998).

After the discussion of the arbitrary nature of the composition of territorial units and zoning grouping systems, *the modifiable areal unit problem* should be taken into account. *The modifiable areal unit problem* is considered in the subject literature in two dimensions (see. Openshaw, Taylor 1979). *The scale problem* is the first dimension. This is a matter of changing spatial data properties and the dependence linking them under the impact of a change in the aggregation scale. The problem is that while moving to higher aggregation scales it is possible to obtain different results for the properties under research as well as for the direction and strength of dependence.

The other dimension of *the modifiable areal unit problem* is *the aggregation problem*. This is the problem of changing spatial data properties and dependence held between them and linking them under the impact of accepting another *n* composition of territorial units within the accepted aggregation scale. Such a presentation of the issues of *the scale problem* and of *the aggregation problem* is inappropriate, since it allows the possibility of the arbitrariness of compositions of territorial units within the zoning and grouping systems.

Both *the scale problem* and *the aggregation problem* should be considered in accordance with the four conditions presented in the previous subchapter which allow an appropriate analysis of spatial data to be performed. That will indicate a need to redefine the concept of *the modifiable areal unit problem* already described in literature. The redefinition of the concept will be commenced with the introduction of the term of *'the quasi composition of regions'* (QCR) within *condition 2*. A *quasi composition of regions* is a set of compositions of territorial units, with lower and upper limits, consisting of particular compositions of territorial units for further aggregation scales, where all compositions allow

---

within the zoning system shows the drawback of the idea. If we consider a reasonable administrative division of a region, for example, into ten units within some research problem, then this division will concern about 10% of each region. There may occur some disturbance because of one or two regions but none of them will exceed 20%. Creating an arbitrary composition within the zoning system, however, may immediately lead to a situation where one region will have 99.1% of the country's territory and the remaining nine regions 0.1% of the territory (let us assume that a region is composed of 1,000 territorial units with the 0.1% of the territory). What kind of empirical analysis will provide sensible results? Therefore, creating territorial compositions arbitrarily within the zoning system is worth further consideration as regards the dangers they bring into scientific research.

an appropriate analysis to be performed within the undertaken research problem. Setting lower and upper limits for compositions of territorial units means that an analysis based on data form a freely selected aggregation scale does not guarantee the correctness of results obtained within the undertaken research problem. When we use the NUTS classification, then the most frequently occurring limit is the upper limit. That means that for the majority of economic phenomena the NUTS level 2 is too high for data at that aggregation scale to meet *condition 2* and to allow a correct analysis to be conducted. After determining lower and upper limits, for every aggregation scale exactly one composition of regions should be designated within the undertaken research problem. The set of those compositions of territorial units forms the *quasi compositions of regions*, which means a set of particular compositions for subsequent aggregation scales. Let us assume that we are considering territorial units at four different aggregation scales (corresponding in size, e.g., to the following classification units: NUTS 5, NUTS 4, NUTS 3, NUTS 2). Within the undertaken research problem there is only one appropriate composition of territorial units for each of the four aggregation scales. For instance, in Poland an analysis of the majority of economic phenomena based on data published for the lowest aggregation level - NUTS 5, will give correct results. Therefore, the NUTS 5 composition may be assumed to be the particular composition of units at this aggregation level. This is a composition of 2,479 municipalities. In Poland higher aggregation scales may be created by the following particular compositions: a composition of 379 district (NUTS 4), a composition of 66 subregions (NUTS 3), a composition of 16 provinces (NUTS 2), and a composition of 6 regions (NUTS 1). As was already mentioned, for the majority of economic phenomena, compositions of regional units following the NUTS classification will lead to the obtainment of correct results.

However, we face here the afore-mentioned lower and upper limits. The implication is that the correctness of results does not need to occur for all of the aggregation scales. Therefore, depending on the research problem undertaken, *a quasi composition of regions* may be composed of only a NUTS 4 composition and NUTS 3 composition or of any other combination of aggregation scales. This correct combination of compositions will never be an ideal reflection of the actual compositions for which dependence related to the undertaken research problem occur. Hence the name is *a quasi composition of regions* (QCR).

Within the undertaken research problem there exists only one *quasi composition of regions*, which allows the identification and description of dependence holding for the analysed phenomena. That means that

every single composition of territorial units, not included in *a quasi composition of regions*, will result in the obtainment of incorrect conclusions. Therefore, *the modifiable areal unit problem* is formulated in subject literature inappropriately, since in the case of an analysis based on empirical data it allows compositions existing outside *the quasi composition of regions*. For that reason the *modifiable areal unit problem* should concern a change in the properties of analysed phenomena which accompanies a change in the aggregation scale but only within a *quasi composition of regions*.

While redefining *the modifiable areal unit problem*, the *scale problem* will be determined as a problem related to a change in the properties of spatial data and casual relations for compositions of territorial units of differentiated aggregation scales that create a quasi composition of regions. However, a quasi composition of regions is designated within the undertaken research problem.

The aggregation problem, in turn, consists in creating a single composition of territorial units at any aggregation scale in such a way that it is included in a quasi composition of regions within the undertaken research problem.

The scale problem is of a significant importance for empirical analyses because usually the data that are made available to public do not concern all of the aggregation scales. Also, it happens quite frequently that the data are published for higher aggregation scales and do not represent the aggregation scales for which they were actually collected. If the properties of phenomena may have been changed due to the aggregation process, then we should bear in mind a possibility of the impact of that fact on the results of the research being conducted. Also, in a situation when researchers have access to data representing various aggregation scales, then it is worth checking the directions of changes in the properties of the phenomena under examination.

The scale problem may be solved by means of a simulation that makes it possible to identify the properties while changing the aggregation scale of data. The redefinition of *the modifiable areal unit problem* modifies the approach adapted for simulations within *the scale problem*. This is not the problem of changing properties while switching into another aggregation scale of arbitrary compositions of territorial units. The problem is about changing properties while switching into another aggregation scale of the accepted *quasi composition of regions*. In the case of the traditional definition of the *scale problem*, the simulation consists in generating the realization of processes for a specific number of various compositions of territorial units within each aggregation scale. Arbitrary compositions of territorial units were generated in accordance with the zoning system or the

grouping system (see Openshaw, Taylor 1979; Reynolds 1998)[17]. The obtained results characterised a set of arbitrary compositions of territorial units for each aggregation scale. Next, the selected characteristics were compared with each aggregation scale. Redefining *the scale problem* compels performing the simulation in a different way. One composition of territorial units for each aggregation scale needs to be selected in accordance with the determined *quasi composition of regions* and a simulation of realisations of the process should be made only for this compositions. The properties computed based on the simulation represent single composition of territorial units for a selected aggregation scale. The comparison of the obtained results will allow changes in the process characteristics within the accepted *quasi composition of regions* to be checked.

The simulation concerning *the scale problem* entails an empirical aspect in the sense that it is related to the analysis being conducted. This follows from the fact that *a quasi composition of regions* is designated within the undertaken research problem. The simulation performed is also utilitarian in nature, if related to the undertaken research problem. The outcome of the simulation is to help researchers in assessing how the researched properties change depending on the selected aggregation scale. Changes in the properties of statistics within *the scale problem* may result from the estimation process (different data and varying amounts of data depending on the aggregation scale). Changes may also result from determined properties that characterise the spatial data under research (e.g., spatial autocorrelation).

As *the scale problem* may be solved with a simulation, *the aggregation problem* is merely of an empirical nature. In the economic research conducted we are dealing with *the aggregation problem*, while constructing *a quasi composition of regions* at the selected aggregation scale the researcher cannot use a ready, a priori single composition of territorial units (e.g., NUTS). A problematic situation will appear when the researcher will establish a single composition of regions that is not consistent with the nature of the undertaken research problem. Solving *the aggregation problem* consists in finding an appropriate and single composition of territorial units, the use of which will make the obtained outcome sensible. In such a case, only the researcher's knowledge and scientific experience will enable him to designate regions correctly and

[17] In practice, however, arbitrariness means a random character of compositions of territorial units generated by a computer software within the zoning system or the grouping system.

avoid *the aggregation problem*. As regards a simulation, it will not provide any additional information on the matter.

While conducting an analysis it may turn out that the undertaken research problem deviates in its nature from the generally accepted composition of territorial units (e.g., NUTS).The analysis of the impact of a metropolis serves a good example. It was proved that a metropolis with its connections and impact on the environment deviate substantially from the accepted administrative division of regions. Establishing a composition of territorial units for a metropolis and other regions is challenging.

Also, we may face a situation where it is necessary to establish a definite number of areas for which there is no counterpart in the form of a readily made composition of territorial units. An example may be the creation of SGM (Standard Gross Margin) regions. Poland's joining the European Union in 2004 enforced the adjustment of statistics to the standards binding in the European Union. The division of Poland into SGM regions required homogeneous regions in the levels of agricultural development and culture. The clustering of data conducted for nine diagnostic variables allowed the territory of Poland to be divided into four agricultural SGM regions. They were given official names and were included in the annex of the Treaty on the Accession of the Republic of Poland to the European Union. The establishment of SGM regions is an example of a positive solution applied to *the aggregation problem*. Economic analyses concerning agriculture conducted for SGM regions should lead to the obtainment of correct results. In the case of making analyses for spatial data two additional problems may arise due to the non-fulfilment of *condition 2* and *condition 4*. As regards *condition 2*, it may happen that *a quasi composition of regions* will be extended by an aggregation scale for which the correctness of results within the undertaken research problem is not preserved. The results received based on the data from that aggregation scale will lead to the formulation of incorrect conclusions. This problem is referred to in the paper as *the aggregation scale interpretation problem* (ASIP).

A good example of *the aggregation scale interpretation problem* is the analysis of the unregistered unemployment rate. The unemployment phenomenon is characterized by strong spatial dependence. If we calculate the spatial autocorrelation for the unemployment rate based on the NUTS 4 level, we will obtain a strong positive spatial correlation. However, if we calculate this property based on based on the NUTS 2 level, then we will obtain a negative autocorrelation. The received result is inappropriate since the unemployment phenomenon is heterogeneous within too large regions at NUTS 2 level. The NUTS 2 level has a too high aggregation scale and reaches beyond *the quasi composition of regions*.

Another type of problem that may occur while analysing spatial data concerns *condition 4* and is referred to in the paper as *the final areal interpretation problem* (FAIP). This problem occurs when the characteristics of phenomena or dependence are designated for too large region. Then it is possible that the data will lose the preferred properties (*homogeneity* and *systematic heterogeneity*). Two situations may take place. In the first case the data possessing properties of *homogeneity* for a specific region may be characterised by either *the systematic heterogeneity* or the *unsystematic heterogeneity* if the region enlarges. In the second case, the data characterised by *the systematic heterogeneity* resulting from the enlargement of the analysed region change their properties into *the unsystematic heterogeneity*. In both cases it is necessary to decrease the size of the region under analysis in order to obtain appropriate properties of data, or to use different, better suited research tools.

The area of agricultural land may serve as an example of *the final areal interpretation problem*. We may determine the average area of agricultural land based on the data at the NUTS 4 level (districts). If we calculate the average for a single province (NUTS 2), then the data should possess the properties of homogeneity[18] and based on the average we will obtain reliable results for the agrarian structure. However, if the average is calculated for a country's territory (NUTS 0), then the average area of agricultural land will not represent any cognitive value. This results from the fact that the data on the area of agricultural land are characterised by the property of the systematic heterogeneity or the unsystematic heterogeneity for an enlarged region.

Simulation analysis

In the case of the scale problem, a simulation analysis should accompany empirical research and a specific research problem should determine the simulation assumptions. As the aggregation problem results from the researcher's mistake, the scale problem results from the data properties and the aggregation process. Therefore, it should be checked by means of a simulation, to what extent the scale problem impacts the research conducted within the undertaken research problem. This

[18] The selected province should be characterised by a similar agrarian structure within all of its districts and by a similar agricultural culture.

subchapter presents a simulation of the consideration of the scale problem within a hypothetical research problem[19].

A scale problem will be considered for the aggregation of data originating from Poland's NUTS 5 level to the NUTS 4 level. This simulation will consist in examining to what extent the basic properties of data are modified during the aggregation process. The mean and the variance calculated for simulated data will be analysed. Determining the regularities in the changes concerning the above-mentioned properties will allow the received results to be interpreted correctly. Figure 1 presents the composition of territorial units used for the purposes of the simulation – 2,497 municipalities (NUTS 5) and 379 district (NUTS 4). Both of the compositions make up the territory of Poland.

The simulation should be applied to the four basic conditions that assure the correct analysis of spatial data. Referring it to condition 1[20] will consist in treating the simulated data as a hypothetical economic category expressed in relative units[21]. Another step is designating quasi compositions of regions within condition 2. In this case, the quasi composition of regions will consist of the compositions of territorial units at two aggregation scales – of the NUTS 5 and NUTS 4 compositions. Condition 3 does not refer to simulated data, however, it may be assumed here that data will be simulated in a correct way. Next, in accordance with condition 4, a region will be determined in relation to which conclusions will be drawn. The region will be Poland's whole territory (NUTS 0), shown in Figure 1. Since data are going to be simulated, it should be additionally assumed what property will characterise them. For this specific case, it was assumed that data will be the realisation of the spatial process with the property of homogeneity[22]. That means that the internal structure

[19] The paper included a simulation of the realisation of the spatial process with a constant expected value, a variance and of the lack of spatial noise. However, there do not exist economic phenomena which would possess such properties. The spatial differentiation of phenomena depicts the existence of spatial dependence, a systematic, or unsystematic, property of heterogeneity. Therefore, the paper will undertake a hypothetical research problem in the form of a hypothetical analysis of an economic category expressed in relative units.

[20] A hypothetical research problem was accepted.

[21] Such data may simulate, for example, values of the demographic dependency ratio in the form of the number of persons at post-working age in relation to the number of persons at working age.

[22] The simulated data will be subsequent realisations of the spatial noise. The simulated data will be characterised by a constant mean, a constant variance and by lack of spatial autocorrelation. The process of spatial noise is a random field that

of the spatial process will be composed only of the property of homogeneity. Data will be generated for 2,479 municipalities and then will be aggregated to the level of 379 districts (NUTS 4). Therefore, the purpose of the simulation will be checking in the accepted quasi composition of regions, given the analysed phenomenon is characterised by the property of homogeneity, whether the mean and variance do not change under the impact of aggregation. A positive answer would mean that in the case of the mean and the variance similar results will be obtained, no matter whether they are calculated at the NUTS 5 level or at the NUTS 4 level.

Figure 1. The regions used in the simulation, 2,479 municipalities (NUTS 5) and 379 districts (NUTS 4)

Source: elaborated by the author.

The first step to be made in the simulation is to obtain data expressed in relative quantities. These data will be obtained indirectly. First, at the NUTS 5 level, an analysis of two processes of spatial noises will be made, namely, of process 1 and process 2. The process of spatial noise was accepted due to the fact that it is characterised by the property of

---

possesses the properties of stationarity in a broad sense. Additional assumptions made concern the constant value of the mean and the function of covariance in the following from: $K(u,u') = \begin{cases} \delta^2, & gdy \ \ u_i = u_j \\ 0, & gdy \ \ u_i \neq u_j \end{cases}$ (see Arbia 2006). Usually the zero expected value is accepted. This assumption was rejected in the simulations.

homogeneity[23]. The realisation of these two processes will be treated as hypothetical data expressed in the absolute quantities. Next, process 3 is obtained as the ratio of process 1 and process 2 and it will be treated as hypothetical data expressed in the relative quantities. For process 1 the mean equal to 10 and the variance equal to 1.6 will be assumed. For process 2 the assumptions are the following: the mean equal to 5 and the variance equal to 0.5. Process 1 and 2 will be generated in five variants. In the first variant the processes will not be correlated. In subsequent variants the correlation between the processes at the levels of 0.3, 0.5, 0.7 and 0.9 will be assumed. For each variant 500 realizations of process 1 and of process 2 will be simulated[24], and based on them the realizations of process 3 will be obtained[25].

Table 1. The results of the simulations of the processes for the correlation coefficient equals 0

NUTS 5	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	-0.001	-0.001	10.001	5.000	2.043	1.599	0.499	0.164
standard deviation	0.017	0.019	0.024	0.013	0.008	0.046	0.014	0.006
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	-0.001	0.516	-0.001	0.514		-0.002	0.523	
standard deviation	0.012	0.293	0.013	0.297		0.013	0.299	
NUTS 4	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	699.793	0.992	65.391	32.692	2.013	1409.7	353.19	0.047
standard deviation	5.408	0.001	0.158	0.086	0.011	14.300	4.009	0.008

[23] It must be emphasized that it is unlikely for empirical data to be characterized by a constant mean, a constant variance and by lack of spatial autocorrelation. The process of spatial noise was assumed to be the simplest process generating data. Empirical spatial data are characterised most frequently by spatial autocorrelation and systematic heterogeneity. The presented simulation should be extended by the processes possessing the mentioned properties.

[24] The realizations of the processes are simulated from a two-dimensional vector of the random field.

[25] A negative correlation between phenomena cannot be observed, therefore, this case will not be considered.

	I(X1)	p-value	I(X2)	p-value	I(X3)	p-value
mean	-0.013	0.622	-0.012	0.612	-0.006	0.538
standard deviation	0.005	0.055	0.006	0.064	0.031	0.268

(cov-covariance, cor-correlation, I(X)- Moran's I statistics)
Source: elaborated by the author.

Table 2. The results of the simulations of the processes for the correlation coefficient equals 0.3

NUTS 5	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	0.269	0.301	10.004	5.002	2.031	1.593	0.502	0.113
standard deviation	0.016	0.016	0.023	0.012	0.006	0.043	0.014	0.004
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	0.000	0.500	0.000	0.477		0.000	0.485	
standard deviation	0.013	0.298	0.011	0.281		0.013	0.297	
NUTS 4	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	702.709	0.994	65.410	32.704	2.010	1411.9	353.86	0.032
standard deviation	6.006	0.001	0.150	0.080	0.009	13.981	3.801	0.005
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	-0.013	0.623	-0.013	0.618		-0.006	0.530	
standard deviation	0.005	0.054	0.006	0.063		0.028	0.256	

Source: elaborated by the author.

Tables 1,2,3,4,5 present the results obtained from descriptive statistics for subsequent variants connected with the correlation level. Based on the simulated realizations of processes the following were calculated: the covariance and the correlation between process 1 and process 2, the means, variations and the statistics of Moran's *I* for process 1, process 2, process 3. Theses statistics were calculated both at the aggregation scale for NUTS 5 and for NUTS 4. The obtained results allowed the evaluation of the impact of the aggregation scale on the examined descriptive statistics of the processes.

In the case of the simulated realizations of process 1 and process 2 (hypothetical data expressed in absolute quantities.) the mean and the variance of processes increased together with the change of the aggregation scale into higher one. This proves the need for avoiding making analyses based on data expressed in absolute quantities whose values of examined statistics grow together with increases in the aggregation scale and those data are not spatially comparable. In the case of this kind of data, the value of the correlation grew intensely along with changes in the aggregation scale. In addition, the values of the covariance, correlation, means and of the variances of the processes reached the same levels regardless of their initial correlation levels.

Table 3. The results of the simulations of the processes for the correlation coefficient equals 0.5

NUTS 5	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	0.451	0.502	9.998	5.000	2.023	1.606	0.501	0.080
standard deviation	0.019	0.015	0.024	0.011	0.005	0.045	0.013	0.003
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	-0.001	0.526	0.000	0.499		-0.001	0.511	
standard deviation	0.012	0.278	0.013	0.289		0.013	0.294	
NUTS 4	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	705.15	0.999	65.359	32.684	2.002	1409.1	353.2	0.005
standard deviation	7.331	0.000	0.164	0.085	0.003	13.70	4.110	0.001
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	-0.012	0.611	-0.013	0.615		-0.004	0.513	
standard deviation	0.005	0.057	0.006	0.063		0.029	0.256	

Source: elaborated by the author.

Table 4. The results of the simulations of the processes for the correlation coefficient equals 0.7

NUTS 5	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	0.625	0.699	10.00	5.001	2.015	1.602	0.500	0.048

			0					
standard deviation	0.023	0.011	0.021	0.012	0.004	0.046	0.014	0.002
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	0.001	0.485	0.002	0.495		0.001	0.475	
standard deviation	0.012	0.279	0.013	0.286		0.012	0.287	
NUTS 4	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	705.154	0.999	65.359	32.684	2.002	1409.8	353.29	0.005
standard deviation	7.331	0.000	0.164	0.085	0.003	13.708	4.110	0.001
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	-0.013	0.616	-0.013	0.617		-0.003	0.505	
standard deviation	0.005	0.057	0.005	0.061		0.031	0.284	

Source: elaborated by the author.

Table 5. The results of the simulations of the processes for the correlation coefficient equals 0.9

NUTS 5	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	0.809	0.900	9.991	4.992	2.008	1.606	0.503	0.017
standard deviation	0.024	0.004	0.025	0.013	0.002	0.043	0.014	0.001
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	-0.001	0.517	-0.001	0.522		-0.001	0.513	
standard deviation	0.012	0.278	0.012	0.280		0.013	0.295	
NUTS 4	cov	cor	mean			variance		
	(X1,X2)	(X1,X2)	X1	X2	X3	X1	X2	X3
mean	705.154	0.999	65.359	32.684	2.002	1409.8	353.29	0.005
standard deviation	7.331	0.000	0.164	0.085	0.003	13.708	4.110	0.001
	I(X1)	p-value	I(X2)	p-value		I(X3)	p-value	
mean	-0.012	0.609	-0.012	0.611		0.001	0.468	
standard	0.005	0.055	0.005	0.059		0.033	0.275	

deviation						

Source: elaborated by the author.

Within the realisation of process 3 (hypothetical data expressed in relative quantities) the aggregation of data did not affect the mean value, however, it influences decreases in the value of the variance. This is a significant conclusion because if data possess the properties of the spatial noise, then, regardless of their aggregation scale, close mean values will be derived from them. However, at higher aggregation scales a lower covariance will be obtained. In the case of the dependent variable, it may be take the form of a higher R-squared for a line regression model.

The aggregation of the realization of the processes did not result in the occurrence of the autocorrelation of spatial data, which is proved by statistically insignificant Moran's *I*. That means that for data possessing the properties of the spatial noise, aggregation does not result in the occurrence of spatial autocorrelation which could impact the value of statistics under research.

Conclusions

The paper deals with the issue of *the modifiable areal unit problem* (MAUP), which is connected with the analysis of spatial data referring to irregular regions. The paper discussed the conditions that are necessary for maintaining the correctness of spatial analyses performed. The described conditions indicate the need for making the research problem a starting point for every spatial analysis. Besides, the level of the aggregation of spatial data needs to be determined based on which conclusions form analyses will be drawn and boundaries of the regions for which these conclusions are to be formulated.

Also, the paper raised the problem of the arbitrary nature of compositions of territorial units. That means that their boundaries and shapes may be created in any way. The author pointed out, however, that this arbitrariness is related to and limited by the specificity of the considered research problem. The finally accepted composition of territorial units should result from the undertaken research problem as well as from the researcher's experience.

The paper has introduced the concept of *a quasi compositions of regions*. It was defined as a set of particular compositions of territorial units for subsequent aggregation scales. Among all potential compositions of territorial units, *the quasi compositions of regions* is formed exclusively by

those which allow the analysis within the undertaken research problem to be conducted. The considerations made allowed the issue of *the modifiable areal unit problem* to be redefined. Both *the scale problem* and *the aggregation problem* were linked to the undertaken research problem and to the accepted *quasi composition of regions*. It is of great importance to the spatial analyses performed since the arbitrary acceptance of compositions of territorial units, which are excluded from the *quasi composition of regions*, leads to the formulation of incorrect conclusions. That means that the concept of *the modifiable areal unit problem* presented in subject literature is formulated inappropriately because in the case of an analysis based on empirical data it allows compositions of territorial units not included in *the quasi composition of regions*.

The redefinition of *the modifiable areal unit problem* compels the change in simulations made within *the scale problem*. The identification of the change in the properties of processes should constitute the purpose of making simulations while moving between the aggregation scales of the accepted *quasi composition of regions*. Therefore, data are generated exclusively for compositions of spatial units belonging to *a quasi composition of regions*. However, generating data for all arbitrary compositions of regions within the zoning system or the grouping system does not result in solving *the scale problem*. Quite conversely, it obscures solving the problem by gaining a wide range of incorrect values of the characteristics under examination.

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.

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.

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.

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 aggregation problem*, „Equilibrium", Volume 9, Issue 3, [in print].

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

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, [in print].

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

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.

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.

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