The modifiable areal unit problem in regional economics

Tamás Dusek

Széchenyi István University
Hungary, Győr
9026, Egyetem tér 1.
dusekt@sze.hu

Paper submitted to the 45th Congress of the European Regional Science Association
Amsterdam, Holland, August 23-27 2005

Abstract

There is a very well-known fundamental problem in spatial data analysis namely that all results of quantitative methods are potencially influenced by the way of spatial delimitation. This problem is mostly called modifiable areal unit problem (MAUP). However, beside the rich tradition in empirical spatial data analysis, the effect of MAUP on putting forward and testing a theory and the effect on model-building is an issue that is rarely investigated. The MAUP creates the need for the investigation of the connection between theories and data and the micro-macro dualism. My paper presents the epistemological background of the problem and gives illustrations of the negative consequences of ignoring them in spatial macroeconomics.

Keywords: Modifiable areal unit problem, Spatial aggregation, Purchasing power parity, Optimal currency area
JEL classification: F0, R1

Introduction

The effect of spatial aggregation, mostly called the "modifiable areal unit problem" (MAUP), consists of two related but distinctive components: the scale effect and the zoning effect. The scale effect is the variation in results, that may be obtained, when the same areal data are combined into sets of increasingly larger areal units of analysis. The zoning effect is any variation in the results, due to alternative units of analysis, where the number of units is constant.

Previous works on MAUP focused mainly on its effect on results of various descriptive statistics. However, it is useful to distinguish at least five aspects of this question:

- the effect on the results of descriptive statistics;
- the effect on the interpretation of the results;
- the practical-political effect;
- the effect on the applicability of inferential statistics;
- the effect on building theories and models.

The first three topics have an enormous literature. The scale and zoning effects on the results of descriptive statistics are illustrated through several database and numerous methods e.g. correlation (Gehlke–Biehl, 1934; Robinson, 1950; Yule–Kendall, 1950; Openshaw, 1977; Openshaw, 1984a; Holt *et al.*, 1996), regression (Openshaw, 1984b), multivariate regression (Fotheringham–Wong, 1991), spatial autocorrelation (Jelinski–Wu, 1996; Qi–Wu, 1996), spatial interaction models (Goodchild, 1979; Webber, 1980; Batty–Sikdar, 1982; Schwab–Smith, 1985; Putnam–Chung, 1989; Ubøe, 2004), location-allocation models (Hillsman–Rhoda, 1978; Bach, 1981; Current–Shilling, 1987; Fotheringham *et al.*, 1995; Francis *et al.*). Their dependence on the given zoning system should be taken into account when interpreting the results.

The practical side of the question depends on the application of the analysis. Though the results of the studies describing the historical situation or satisfying our curiosity can be deceiving and have a more restricted validity, than those of the authors. However, they can only have their practical effect, if they are used for administrative, regional development reasons or in political decision-making. This problem is most noticeable in the establishing new regional units, when it has a delimiting factor on the region's participation in the different regional development subsidies. Drawing the boundaries of the constituencies can influence the result of the elections (Taylor, 1973; Gudgin–Taylor, 1974). The method of aggregation can also critically affect the results on designating the optimal locations of new institutions (Bach, 1981; Fotheringham *et al*., 1995).

According to some optimistic opinions, development in computing has resulted in new opportunities to explore the issues of the MAUP. Sophisticated mathematical simulations were developed, in the field of synthetic spatial data generators as well (Amrhein, 1995; Reynolds, 1998). However, in the lack of certain given basic spatial units it is doubtful what the possible regular zoning effects can be compared with and what is the theoretical novelty which can not be discovered by the help of mental experiments. The vital and sometimes forgotten difference between MAUP and ecological fallacy is that in the later case there is a natural, objective and unmodifiable basic unit of analysis (the persons), that in the case of spatial analysis does not exist. What would be, for example, the basic spatial unit of the unemployment rate, population density, price level, economic growth, average yield of potato crops, the pattern of interregional trade or the other, only in spatially construability variables?

Surprisingly the study of theoretical and technical effects on statistical inferences, despite its enormous significance, is pushed into the background. The effects on putting forward a theory and on model building in regional economics is also a seldom investigated issue. It is a very unfortunate situation because many potentially dangerous problems remain undiscovered. In this paper I give an outline of the main questions and the effects of areal delimitation on the theories and methodology of regional economics.

The treatment of space in economics and in regional economics

Spatial researchers often criticise the standard imaginary spaceless world of economics. The validity of these criticisms depend on two questions. Firstly, what is the impact of the spacelessness on the validity of the theories? Secondly, is the treatment of space adequate or not from the spatial problems' point of view? The first very important general question can not be investigated here, only to the extent of its contact with the second question.

The treatment of space has many different forms:

1. Dimensionless aggregated one-point economies.
2. The network of aggregated one-point economies without transport costs.
3. The network of aggregated one-point economies with transport costs.
4. Space built by individual basic points and individual movements.
5. Other or mixed approaches.

There is a model, for example, among the mixed approaches in which not all factors of production have transport costs; one example of other approaches is the model of one dimensional linear continuous space. The first type of the above mentioned approaches is the typical but almost always implicit approach of standard micro- and macroeconomics. The second one is often used in international economics, the third one in regional economics or in the literature of "New Economic Geography". The fourth approach is also used in regional economics, for example Lösch's and Ohlin's systems were based on individuals.

Only the fourth approach is truly free from the MAUP, therefore the demarcation line is more important between the third and fourth approach than between the first and second. Great part of criticism concerning dimensionless aggregated one-point economies is valid to the network of aggregated one-point economies, too because both types of analysis also deal with spatially aggregated zoning-system dependent data.

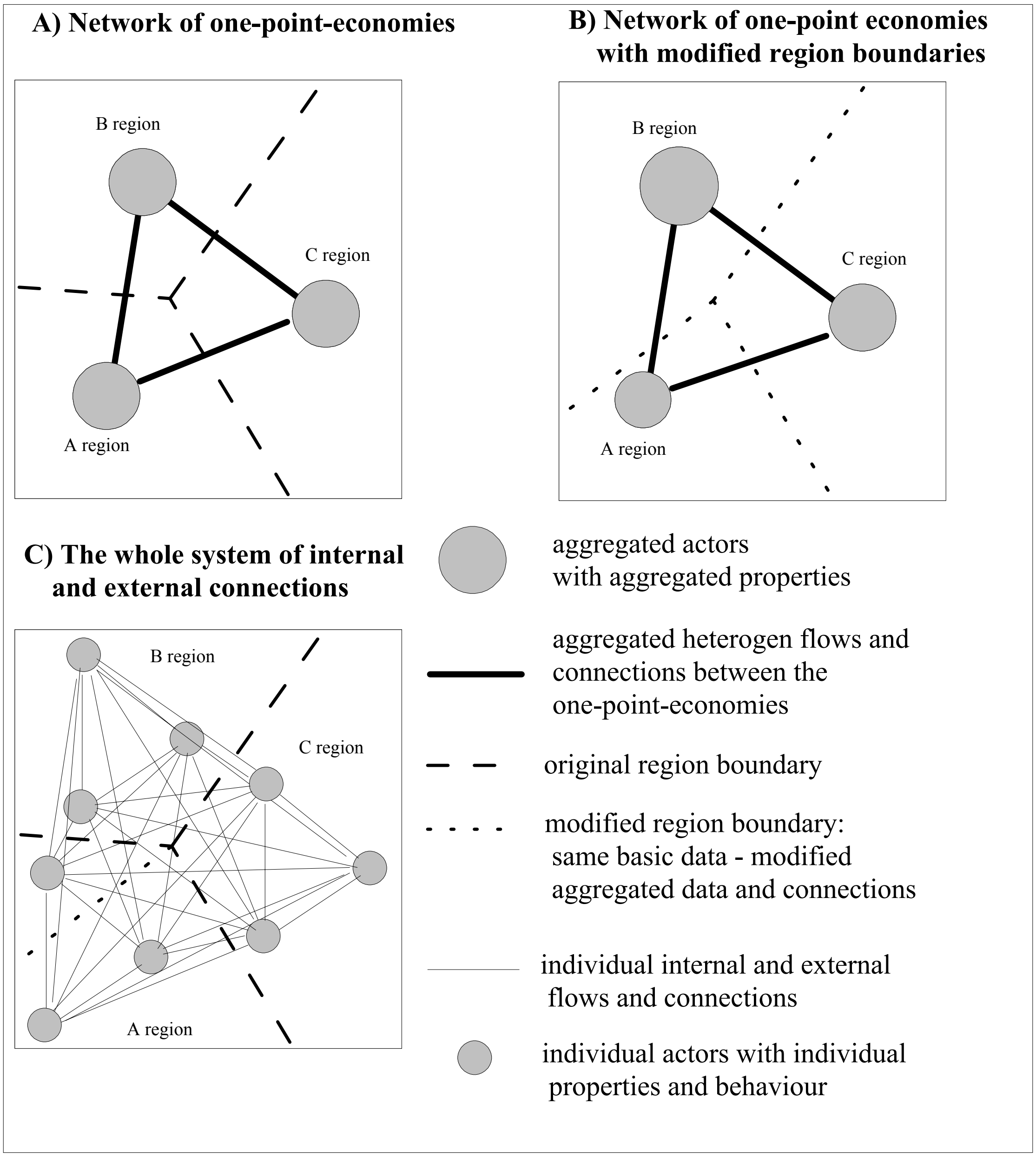
The MAUP in regional economics

Owing to the partial dependence of the areal delimitation of the statistical results of data analysis, it is only fair to ask how this influences theories that explain the connection between spatially aggregated variables. We can see such macro variables within the models of regional economics, that can only be interpreted under the form of spatial aggregation, even if they have identifiable individual basic data. Such variables are, for example, the price level and its change, total investment, unemployment, the quantity of money, rate of interest, interregional trade, average propensity to consume, aggregate demand etc.

Regions, similar to countries, merge economical actors, which are spatially, temporally, in their degree of quality, in quantity and in behaviour heterogeneous. Even if, with the modification of the zoning system, the same and in an objective sense unmodified basic data corroborated one theory, then again another, would mean it is impossible to test the various theories with the help of data; or could be formulated through a sort of non-mathematical probability (not based on exact, known distribution and on random samples). Therefore, the partial zoning system dependence of the results creates difficulties that are impossible to overcome for those who treat the data as an exclusive tool for corroboration or falsification of theories. For those who treat the data as the help for description of historical situations, and by the way illustrating theories, the zoning system dependence is not a problem but a fact that must be taken into consideration when interpreting the results.

The problem of heterogenity can be seen on Figure 1. The number of individual actors and their connections are much larger than those of the connections of aggregated regions, but these are not possible to depict. There can be legal, constitutional differences between intraregional and interregional flows, if regions are countries. In other cases, the sources of differences can mainly be traced back to the differences of distances. It is a common misconception to treat the spatial units as individual behavioural units, without their own spatial extension. I detail this problem in the next part of this study.

Figure 1 The MAUP in regional economics



Aggregation in economics

The MAUP is strongly connected to the general problem of aggregation. There are two main approaches to the general problem of aggregation in economics: a purely technical, mathematical one and a fundamental one. The main difference between them is that the first approach does not care for some tacit assumptions but for some technical issues, which are only relevant if the tacit assumptions are true. Examples of this kind of inquiry are: Theil (1954), Gupta (1969) or Stroker (1993).

The fundamental approaches are concerned with the more substantial issues, which were detailed as follows. By assuming a causal relationship between the spatially aggregated data you have to manage the problem of the qualitative difference between the basic elements. It is useful to distinguish three types of differences as well as spatial and temporal ones, namely the differences in value, behaviour and composition. Differences are defined here in value, as the different values of a variable in different basic elements, for example the different income of two people or two enterprises. Differences in behaviour means the different behaviour of basic units, which can be ignored in some cases or at group level. When this behaviour for the working of the whole system is significant, ignoring it would produce errors. For example, the distribution of demand among the different goods is of great importance (Hayek, 1984). Finally the compositional difference means that the variables (e. g. GDP, indices of price level), which were created by weighting methods and using various auxiliary assumptions, often do not have natural basic units.

The micro level explanations of macro-phenomena are problematic issues that natural sciences also have to deal with. The same macro-arrangement can go hand in hand with many different micro-arrangements. These micro-arrangements are unobservable, but it is not necessary for describing and explaining the system as a whole. This is possible, because the material components of the systems vary from each other only in pure quantitative or in quantitative expressible qualitative, but not in behavioural or in compositional characteristics. In this case the properties of system can be described with the help of statistics.

For explanation of economic-social phenomena, which are much more complex, the knowledge of relative frequency of the properties of elements of the system is not sufficient. The elements of society connect with each other in an organised way. The variables are interconnected in a very complex way and they vary simultaneously (Weaver, 1967). Therefore, knowledge of the manner in which the individual elements are connected is also needed, which can not be replaced by statistical information or by simple functions between few variables (Hayek, 1967; Weaver, 1967). "Aggregate phenomena are influenced by the way that individuals at the micro level behave and by the way that individuals interact with one another" (Fotheringham–Rogerson, 1993, p. 15). In this case full information about each element is required in order to derive practical and specific conclusions and predictions about the individual elements. "Without such specific information about the individual elements we shall be confined to (…) mere pattern predictions – predictions of some of the general attributes of the structure that will form themselves" (Hayek, 1984, p. 270).

Carl Menger saw the difficulties of aggregation clearly in 1871: "In what follows I have endeavored to reduce the complex phenomena of human economic activity to the simplest elements that can still be subjected to accurate observation, to apply to these elements the measure corresponding to their nature, and constantly adhering to this measure, to investigate the manner in which the more complex economic phenomena evolve from their elements according to definite principles" (Menger, 1994, pp. 46-47). Similarly, Hayek writes in 1931: "Neither aggregates nor averages do act upon one another, and it will never be possible to establish necessary connections of cause and effect between them as we can between individual phenomena, individual prices etc. I would even go so far as to assert that, from the very nature of economic theory, averages can never form a link in its reasoning" (Hayek, 1935, pp. 4-5).

In spite of this and other similar reflections, the connection between macro variables is often treated in the process of model building as simply a connection between micro variables. This is done when, for example, the price elasticity of export, the average propensity to consume, the capital intensity of production, the qualification of the workforce, which are categories that can be interpreted individually, are used in models. But the discussion of spatial aggregation can only be rarely found. Steiness proved in his study, in which he criticised an aggregated model about the causal relation of minority job losses and residential segregation, that because of the statistical gerrymandering, i. e. the MAUP, the model can explain both job losses and gains of minority groups as a result of segregation. This problem can be seen as a specific characteristic of each econometric model using spatially aggregated data (Steiness, 1980).

The MAUP in empirical analysis could be solved by using scale independent individual data. We can find an example of this in the use of individual locational data in the analysis of spatial concentration (Arbia, 2001; Duranton–Overman, 2002). On one hand, this is an exceptional alternative, which is impossible to attain in the majority of the analysis, because most variables in the explanation can not be interpreted at an individual level. On the other hand the problem of drawing the external boundaries of the research area remains.

Examples: the theory of optimal currency area and the doctrine of purchasing power parity

In this short survey there are only two popular theories or research territories presented, in which the MAUP and the continuous character of space were not taken into consideration and this led to long ago unrecognized problems. The theory of optimal currency areas, which was proposed by Mundell (1961), suffers from the false treatment of space and focusing on the connection between aggregated indicators without sound attention to original particles. The implicit treatment of space in OCA literature is the network of aggregated one-point economies. The discussion about the theory following Mundell's article treats a matter of detail in first line and not of the conceptual mistakes. The modifiable areal unit problem is not mentioned in the literature of OCA. Mundell's definition of region mixes the functional (factor mobility) and homogenous (uniform) elements (Mundell, 1961). The interregional flows belongs to functional elements. However, functional (nodal) regions, in the case of economics, do not have firm borders, the space divided into functional regions are continuous and it consists of overlapping regions. Further problems are not discussed here (about the conceptual economic problems see Block, 1999).

Calculation of purchasing power parity has two entirely different applications. The first is as an economic indicator, a description of local economies. This is harmless and contributes interesting information to economic history. The second application is a theoretical one, in explanation of development and connections between national price level and exchange rates. This theory can be read in every elementary textbook on international economics. The theory is entirely fallacious for the following reasons. After the tacit assumption of purchasing power parity theory, the national economies are spaceless, dimensionless points. Inside the countries the price level is constant everywhere. In reality the countries have spatial extension and the price level varied at different points within the countries and its temporal change is also different. As regards to general price level, it is only an abstraction. In fact only the individual prices exist, and one sort of general price level is extracted from the individual prices by the help of weighting, sampling and other auxiliary assumptions. There is not puzzle about this theory and "testing" of PPP. The only puzzle is merely the popularity of this theory.

The treatment of the MAUP and conclusions

The difficulties of MAUP can be solved theoretically by the reduction of macro phenomena to micro phenomena. When such a postulate is carried out in a consistent way every connection between macro phenomena can also be reduced to connections between micro phenomena. Actually the MAUP can be solved in regional economics by a clear distinction between theoretical and historical research. The results and coefficients in empirical analyses are perfectly valid descriptions of concrete historical circumstances at the respective zoning system, and it is not necessary to manipulate them with unjustified significance tests. The model building approach is not authorized for ignoring the MAUP and using spatially aggregated variables by the fact, that statistics are mostly accessible only in spatially aggregated forms. Theoretical models can not adequately be based on aggregates, which are provided by economic statistics helping historical description, without a careful account of its parts.

References

Amrhein, C. G. (1995) Searching for the elusive aggregation effect: evidence from statistical simulations. Environment and Planning A, 27, pp. 105-119.

Arbia, G. (2001) Modelling the geography of economic activities on a continuous space. Papers is Regional Science 80, pp. 411-424

Bach, L. (1981) The problem of aggregation and distance for analyses of accessibility and access opportunity in location-allocation models. Environment and Planning A, 13, pp. 955-978.

Batty, M.–Sikdar, P. K. (1982) Spatial aggregation in gravity models: 4. Generalisations and large-scale applications. Environment and Planning A, 14, pp. 795-822.

Block, W. (1999) The gold standard: a critique of Friedman, Mundell, Hayek, Greenspan. Managerial Finance, 25, pp. 15-33.

Current, J. R.–Schilling, D. A. (1987) Elimination of Source A and B Errors in p-Median Location Problems. Geographical Analysis, 19, pp. 95-110.

Duranton, G.–Overman, H. G. (2002) Testing for Localisation Using Micro-Geographic Data. CEPR Discussion Paper No. 3379

Fotheringham, A. S.–Densham, P. J.–Curtis, A. (1995) The zone definition problem in location-allocation modeling. Geographical Analysis, 27, pp. 60-77.

Fotheringham, A. S.–Rogerson, P. A. (1993) GIS and spatial analytical problems. International Journal of Geographic Information Systems, 7, pp. 3-19.

Fotheringham, A. S.–Wong, D. W. S. (1991) The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A, 23, pp. 1025-1044.

Francis, R. L.–Lowe, T. J.–Tamir, A.–Emir-Farinas, H. (2004) Aggregation Decomposition and Aggregation Guidelines for a Class of Minimax and Covering Location Models. Geographical Analysis, 36, pp. 332-349.

Gehlke, C. E.–Biehl, K. (1934) Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. Proceedings of the American Statistical Journal New Series, pp. 169-170.

Goodchild, M. F. (1979) The aggregation problem in location-allocation. Geographical Analysis, 11, pp. 240-255.

Gudgin, G.–Taylor, P. J. (1974) Electoral bias and the distribution of party voters. Transactions of the Institute of British Geographers, No. 63. pp. 53-73.

Gupta, K. L. (1969) Aggregation in economics. Rotterdam University Press

Hayek, F. A. (1935) Prices and Production. Routledge & Kegan Paul, London

Hayek, F. A. (1967) The theory of complex phenomena. In: Studies in Philosophy, Politics and Economics. Routledge and Kegan Paul, London pp. 22-42.

Hayek, F. A. (1984) The pretence of Knowledge. In: The essence of Hayek. Edited by C. Nishiyama, K. R. Leube, Hoover Institution Press, Stanford, pp. 266-277.

Hillsman, E. L.–Rhoda, R. (1978) Errors in measuring distances from populations to service centers. Annals of the regional Science Association, 12, pp. 74-88.

Holt, D.–Steel, D. G.–Tranmer M.–Wrigley, N. (1996) Aggregation and ecological effects in geographically based data. Geographical Analysis, 28, pp. 244-261.

Jelinski, D. E.–Wu, J. (1996) The modifiable areal unit problem and implications for landscape ecology. Landscape Ecology, 11, pp. 129-140.

Menger, C. (1994) Principles of economics. Libertarian Press, Grove City

Mundell, R. (1961) A theory of optimum currency areas. American Economic Review, 51, pp. 657-665.

Openshaw, S. (1977) A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. Transactions of the Institute of British Geographers, 2, pp. 459-472.

Openshaw, S. (1984a) Ecological fallacies and the analysis of areal census data. Environment and Planning A, 16, pp. 17-31.

Openshaw, S. (1984b) The modifiable areal unit problem. Concepts and Techniques in Modern Geography, Number 38, Geo Books, Norwich

Qi, Y.–Wu, J. (1996) Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices. Landscape Ecology, 11, pp. 39-49.

Putnam, S. H.–Chung, S-H. (1989) Effects of spatial system design on spatial interaction models. 1: The spatial system definition problem. Environment and Planning A, 21, pp 27-46.

Reynolds, H. D. (1998) The modifiable areal unit problem: empirical analysis by statistical simulation. Doctoral Thesis, University of Toronto

Robinson, W. S. (1950) Ecological correlations and the behavior of individuals. American Sociological Review, 15, pp. 351-356.

Schwab, M. G.–Smith, T. R. (1985) Functional invariance under spatial aggregation from continuous spatial interaction models. Geographical Analysis, 17, pp. 217-230.

Steinnes, D. N. (1980) Aggregation, gerrymandering, and spatial econometrics. Regional Science and urban Economics, 10, pp. 561-569.

Stoker, T. M. (1993) Empirical Approaches to the problem of Aggregation Over Individuals. Journal of Economic Literature, 31, pp. 1827-1874.

Taylor, P. J. (1973) Some implications of the spatial organisations of elections. Transactions of the Institute of British Geographers, No. 60. pp. 121-136.

Theil, H. (1954) Linear aggregation of economic relations. North-Holland Publishing Company, Amsterdam

Ubøe, J. (2004) Aggregation of gravity models for journeys to work. Environment and Planning A, 36, pp. 715-729.

Weaver, W. (1967) Science and Complexity. In: Science and imagination. Selected papers of Warren Weaver, Basic Books, New York, London, pp. 25-33.

Webber, M. J. (1980) A theoretical analysis of aggregation in spatial interaction models. Geographical Analysis, 12, pp. 129-141.

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