Journal of Urban Economics 67 (2010) 287–302

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? ☆

A. Briant [a,*], P.-P. Combes [b,1], M. Lafourcade [a,c]

[a] Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France
[b] University of Aix-Marseille, GREQAM, 2 rue de la Charité, 13236 Marseille Cedex 02, France
[c] University of Paris-Sud 11, ADIS, 54 Boulevard Desgranges, 92331 Sceaux Cedex, France

ARTICLE INFO

Article history:
Received 22 October 2008
Revised 21 September 2009
Available online 7 October 2009

JEL Classification:
R12
R23
C10
C43
O18
F1
C4
J3
R1

Keywords:
MAUP
Concentration
Agglomeration
Wage equations
Gravity

ABSTRACT

This paper evaluates, in the context of economic geography estimates, the magnitude of the distortions arising from the choice of a specific zoning system, which is also known as the Modifiable Areal Unit Problem (MAUP). We undertake three standard economic geography exercises (the analysis of spatial concentration, agglomeration economies, and trade determinants), using various French zoning systems differentiated according to the size and shape of their spatial units. While size might matter, especially when the dependent variable of a regression is not aggregated in the same way as the explanatory variables and/or the zoning system involves large spatial units, shape does so much less. In any case, both dimensions are of secondary importance compared to specification issues.

1. Introduction

Most empirical work in economic geography relies on scattered geo-coded data that are aggregated into discrete spatial units, such as cities or regions. However, the aggregation of spatial dots into boxes of different size and shape is not benign regarding statistical inference. The sensitivity of statistical results to the choice of a particular zoning system is known as the Modifiable Areal Unit Problem (hereafter MAUP). Surprisingly, economists paid little attention to this problem up until recently.[2] Our main objective here is to assess whether differences in results across empirical studies are really sparked by economic phenomena in the process under scrutiny, or rather just by different zoning systems. We first investigate whether changes in either the *size* (equivalently the number) of spatial units, or their *shape* (equivalently the drawing of their boundaries) alter any of the

☆ This paper was prepared as part of the PREDIT research contract 04-MT-5036, whose financial support is gratefully acknowledged. The authors are also involved in the Marie Curie Research Training Network 'TOM – Transnationality of migrants' funded by the European Commission (Contract No.: MRTN-CT-2006-035873). We thank Alain Sauvant and Micheline Travet (French Ministry of Transport) for providing us with the trade data, and Sébastien Roux for his very kind cooperation on wage data. We are indebted to Yasushi Asami, Andrew Clark, Gilles Duranton, Laurent Gobillon, Keith Head, Thierry Mayer, Daniel McMillen, Henry Overman, Matthew Turner, as well as two anonymous referees and William Strange for insightful comments. Conference and seminar participants in Kyoto, Toronto, Barcelona, Lyon and Pau also provided us with very useful feedback.
* Corresponding author.
E-mail addresses: anthony.briant@m4x.org (A. Briant), ppcombes@univmed.fr (P.-P. Combes), lafourcade@pse.ens.fr (M. Lafourcade).
URLs: http://www.vcharite.univ-mrs.fr/pp/combes/ (P.-P. Combes), http://www.pse.ens.fr/lafourcade/ (M. Lafourcade).
[1] Also affiliated with the CEPR and Paris School of Economics. Financial support from CNRS is gratefully acknowledged.

[2] Two noticeable exceptions are Holmes and Lee (in press) and Menon (2008).

0094-1190/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jue.2009.09.014

estimates that are usually computed in the economic geography literature. Second, we address the important question of whether distortions due to the MAUP are large compared to those resulting from specification changes.

Disentangling these two effects is essential for policy. For instance, much work has tried to check empirically whether agglomeration enhances economic performance at the scale of countries, European regions, U.S. states or even smaller spatial units such as U.S. counties or French employment areas. The magnitude of the estimates differs between papers, but we do not know whether this reflects zoning systems or real differences in the extent of knowledge spillovers, intermediate input linkages, and labor-pooling effects on firm productivity. The resulting economic policy prescriptions regarding cluster-formation strategies will be affected accordingly. In the same vein, a large body of literature has evaluated the degree of spatial concentration, but does not check whether the conclusion that some industries are more concentrated than others results from the chosen zoning system or from more fundamental differences in the size of agglomeration and dispersion forces across industries at different spatial scales.

This paper is based on three standard empirical questions in economic geography, although many others could have been considered.[3] We start by evaluating the degree of spatial concentration under three types of French zoning systems (administrative, grid and partly random spatial units) and by comparing the differences between concentration measures (Gini vs. Ellison and Glaeser) with those between zoning systems. We then turn to regression analysis as not only is the measure of any spatial phenomenon likely to be sensitive to the MAUP, but also its correlation with other variables. We estimate the impact of employment density on labor productivity and compare the magnitude of agglomeration economies across zoning systems and econometric specifications. Finally, we run gravity regressions. We study how changes in the size and shape of spatial units affect the elasticities of trade flows within France with respect to both distance- and information-related trade costs.

All of these empirical exercises suggest that, when spatial units remain small, changing their size only slightly alters economic geography estimates, and changing their shape matters even less. Both distortions are secondary compared to specification issues. More caution should be warranted with zoning systems involving large units, however. The MAUP is obviously less pervasive when data variability is preserved from one scale to another. When moving from dots to boxes, specific attention should be devoted to the following key points: 1 – the size of boxes in comparison with the original dots, 2 – the way data are aggregated, i.e. averaging or summation, 3 – the degree of spatial autocorrelation in the data. The MAUP is less jeopardizing when data are spatially-autocorrelated and averaged, as is the case in wage regressions. By way of contrast, the MAUP is more challenging when variables in a regression are not computed under the same aggregation process. In gravity regressions for instance, moving from one scale to another requires a summation of trade flows on the left-hand side, whereas distance is averaged on the right-hand side.

The remainder of the paper is organized as follows. Section 2 provides a simple illustration of the possible size- and shape-dependency of spatial statistical inference, along with a data simulation exercise. Section 3 lists the zoning systems for which our estimations are carried out. As a first sensitivity test, Section 4 is dedicated to the study of French spatial concentration patterns. Sections 5 and 6 investigate the extent to which changing econometric specifications and zoning systems affect the size and significance of wage and trade determinants, respectively. Section 7 concludes and suggests further lines of research.

[3] For comparison purposes, we use the same specifications as those typically found in the literature (see Combes et al., 2008b), even though we do not necessarily think that they are the most apt.

2. The modifiable areal unit problem: a quick tour

The Modifiable Areal Unit Problem is a longstanding issue for geographers. In their seminal contribution, Gehlke and Biehl (1934) were the first to emphasize that simple statistics such as correlation coefficients could vary tremendously across zoning systems. They note that, in the United States, the correlation between male juvenile delinquency and the median equivalent monthly housing rent increases monotonically with the size of spatial units. Openshaw and Taylor (1979) pursued this line of investigation and, drawing on correlations between the percentage of Republican voters and the percentage of the population over 60, standardize what they called the "Modifiable Areal Unit Problem".[4]

2.1. A simple illustration of the MAUP

Spatial statistics may vary along two dimensions: firstly, the level of aggregation, or the *size* of spatial units, and secondly, at a given spatial resolution, the drawing of their boundaries, or their *shape*. Fig. 1 illustrates these two related issues via the employment density–labor productivity relationship.

Black points display the location of skilled workers, whose individual productivity is denoted $\bar{y}$, while empty dots stand for unskilled workers, with productivity $\underline{y} < \bar{y}$. In the top figure, space is divided into four rectangles, each consisting of three skilled and two unskilled workers. The spatial distribution of workers across units is uniform and average productivity is the same across units. To illustrate the shape effect, consider the bottom-left figure. Spatial concentration emerges here, with two clusters of six high-skilled workers and two clusters of four low-skilled workers. Average productivity is higher in the former due to the spatial sorting of labor skills. Hence, agglomeration economies, defined here as the positive correlation between productivity and employment density, are zero in the first zoning system but positive in the second. We now turn to the size effect. In the bottom-right figure, we consider smaller rectangles with the same proportions as in the top figure. Spatial concentration is also found here, but the relationship between productivity and density is less marked than in the bottom-left case. Indeed, the difference in productivity between low- and high-productivity regions remains the same (except for empty boxes), whereas the density gap is higher in the bottom-right case. Hence, the extent and scope of agglomeration economies change with the size and shape of units, even though the underlying spatial information – the location and productivity of workers – remains the same.

The question we pursue in this paper is hence twofold. How much does moving from a particular zoning system to another alter the perception of an economic phenomenon? And how does this alteration vary accordingly to whether information is summed or averaged under this aggregation process? Section 2.2 provides a first clue to these questions, drawn from a simple simulation exercise.

2.2. Mean and variance distortions: a first illustration with simulated data

A number of authors have provided detailed analyses of the MAUP based on simulated data. According to Arbia (1989), both size and shape distortions are minimized (although never elimi-

[4] See Fotheringham and Wong (1991) for an extended review of the earliest MAUP contributions.