Preface

Defining what the concept of "Region" exactly means represents a central theme in the field of spatial science. Since the beginnings of regional analysis, discussions about the proper way of delimitating the concept of region itself have been present. This is probably due to the fact that we lack a universal delimitation of the appropriate spatial scale in theoretical terms. This definition fundamentally depends on what we are searching for, or what our specific objective is.

Traditionally, regional scientists have been limited in our empirical analysis to working with data collected for administrative regions. These administrative divisions were normally designed on the basis of historical and/or political factors. We have had no option but to take the information available at that scale as the input for regional modeling. Modern regional analysis, however, requires thinking at a different spatial scale. For example, endogenous growth theory focuses attention on the local level rather than national or larger regional levels. New Economic Geography, moreover, claims that most of the economic dynamics take place in local environments such as cities in which spillovers, agglomeration economies, or connections via networks are present and play a fundamental role.

The new requirements of information demanded by more recent approaches in regional science stimulated the generation of statistical information at a smaller scale. Data availability has increased enormously over recent years, together with the development of techniques to estimate new information from previous data resources. Thus, one of the most important challenges for modern regional analysis is to take advantage of this more disaggregated information in order to more precisely define the spatial scale and delimitation of our studies. The main objective of this book is the exploration of ways of defining the appropriate spatial scale and how to use or generate local data to build regional models at the local level.

We organize the book into three parts that deal with three different but related general issues, which are listed below:

1. The first one is focused on suggesting appropriate definitions of the spatial scale to use. It comprises four contributions that reflect on (1) the proper way of defining economic areas from the information collected at the local level and (2)

which spatial scale is the most suitable depending on the specific objective of the analysis. Different concepts such as local labor markets, metropolitan areas, functional areas, or analytical regions are considered and discussed in these four chapters, giving a complete idea of the different approaches and perspectives regarding the identification of local and regional dynamics.

2. The second issue deals with estimation of alternatives when the required data are not directly observable. The second section of the book presents some options to recover data at a small scale. Several approaches to estimate information at the local level or to integrate different spatial levels of data are revised and illustrated with applications. Techniques based on Entropy Econometrics, ISQ procedures, and Disequilibrium Adjustment framework, among others, are proposed. The last two chapters explore how to maximize the information contained in local data by means of dynamic panel models and spatial econometrics.

3. The third general issue is related to economic analysis and modeling from data collected at different spatial scales, with a special focus on data at the local level. The book finishes with a section that studies this issue by presenting several empirical applications with real-world examples, on which several common problems in regional and urban economics are approached from models built on spatially disaggregated data.

Building Economic Areas from Local Data

The classic contributions of Von Thünen, Lösch, and Christaller among others gave a definition of an economic region consistent with the Economic theory of land, markets, and distance costs. Since these first approaches were proposed, many new ideas have emerged in regional and urban economics. The New Economic Geography (NEG) pointed out the importance of agglomeration economies and the distinction between central and peripheral areas, thereby introducing new challenges to the way of defining regions with an economic meaning. Many researchers have developed quantitative techniques with the aim of identifying coherent local areas. Different names were given to these areas, such as *Functional Economic Area*, *Labor Market Area*, etc., but they all mean a territory that internalizes the home-to-work daily journeys of their residents.

In accordance with this approach, the first chapter of the book, written by Sforzi, starts by defining and calculating regions that satisfy the idea of internalizing the home-to-work daily journeys of their residents. To begin with, from the methodological point of view, Sforzi expands traditional approaches to the identification of metropolitan areas to introduce a new category he labels the "network approach," which explicitly recognizes that metropolitan areas are cliques of networks of cities, and can be monocentric or polycentric. He goes on to provide two sets of metropolitan areas for the case of Italy, which are identified by means of rigorous and replicable standards that can be used by other researchers in subsequent investigations.

Rubiera and Viñuela start from a definition of local units and propose a subsequent aggregation of these local units into aggregated regions. This aggregation covers all the space of a country and is organized in accordance with NEG postulates. The classification of the territory carried out by these authors is consistent with the classical point of view and also with the relevance of agglomeration and scale economies pointed out in recent contributions. This chapter applies the concept of incremental distance in order to analyze the relevance of the position of each territory with respect to the metropolis. As an empirical illustration, Rubiera and Viñuela apply this re-aggregation of basic spatial units into analytical regions to the Spanish economy and compare them with other administrative or historical regions that usually have no economic meaning and, consequently, are not appropriate for analyzing the economy from a regional/spatial perspective.

The third chapter, by Boix, Veneri, and Almenar, explores ways of aggregating units that form a metropolitan area, which are used as the reference point for the previous approach of Rubiera and Viñuela. They propose a procedure in line with Sforzi but which take into account the specific connections that emerge in a big metropolis and its influential area. The metropolitan units identified have three basic purposes: to provide a general view of the characteristics of each country's metropolitan reality, to compare the metropolitan processes of both countries, and to identify metropolitan units that could be used in subsequent analysis. This has been done by the authors comparing two functional approaches to the concept of metropolitan area: a general methodology applicable to most European Union countries based on the concept of the *Functional Urban Area*, and their proposed *Dynamic Metropolitan Areas*, which are specifically designed to deal with the particular characteristics of networking and polycentricism.

Finally, from the base of the previous chapter, Brezzi, Piacentini, and Sanchez-Serra present the results and conclusions of some recent work carried out at the OECD to develop an international methodology for measuring the socioeconomic and environmental performance of urban and metropolitan areas. In their chapter, the authors propose an international methodology for the definition of urban areas that is applied to 27 OECD countries. The methodology identifies urban areas as functional economic units, characterized by densely inhabited "urban cores" and "hinterlands" whose labor market is highly integrated with the "cores" by commuting flows. The development of a harmonized functional economic definition overcomes previous limitations related to administrative definitions by increasing cross-country comparison. The definition of urban areas that they use takes into account the possibility of polycentric development, since more cores physically separated can be included in the same urban area. They integrate information from geographical sources with population data to get a better understanding of urban forms and urbanization processes. Finally, the proposed methodology identifies for each country all urban systems with a population of at least 50,000, enabling analysis of medium-sized urban areas and not only of large metropolitan areas.

Estimation of Spatial Disaggregated Data

The problem of having the information required to define different regional units is always present. This justifies the second part of the book, which focuses on different techniques applied to estimate disaggregated data from observable aggregates.

This section starts with a chapter by Fernández-Vazquez and Garduño, where they present an application of an estimation procedure based on entropy econometrics in order to infer data on wages for Mexican municipalities using aggregated information. Mexico is a specially useful case because official data are available with a high level of disaggregation. Specifically, their objective is to estimate wages paid by industry and municipality for a group of four Mexican states. The information required to put into practice the technique proposed are data that reflect a priori beliefs about the possible wage distribution across municipalities and industries, as well as observable municipal and industry aggregates. From these two pieces of information, they apply a Cross Entropy adjustment and compare their estimates with the actual values observed in the Mexican Economic Census for 2009. Given that assuming perfectly observable information on the municipal aggregates was quite unrealistic, the standard adjustment is extended in order to consider the possibility of errors in the margins.

In the following chapter, Lemelin and Mainguy explain the so-called ISQ method and use it to estimate the GDP of the subareas that form the province of Quebec, Canada. The ISQ technique is presented as a top-down method, which consists of allocating total labor income and net income of unincorporated business (NIUB) by industry among regions. For this purpose, they use allocators constructed from fiscal data on wages and salaries and NIUB obtained from the Quebec Ministry of Revenue. For each industry, other components of value added are then distributed in proportion to the sum of total labor income and NIUB. The key ingredients in the method are a compilation of fiscal data on incomes and reliable home-to-work commuting tables by industry. As the next natural step following their estimation exercise, the authors analyze the estimates to examine the recent evolution of the geographical pattern of economic activity in Quebec.

Kim and Hewings show in their chapter an application of the Disequilibrium Adjustment Framework to small region forecasting for the metropolitan area of Chicago. This chapter highlights the dynamic nature of adjustment models as one of its key advantages and explores the possibility of further applications beyond its existing uses for empirical assessments. Specifically, the authors apply a spatial econometric version of the regional disequilibrium adjustment model for small area socioeconomic forecasting and impact analysis. Rather than using the standard adjustment model only for forecasting purposes, the chapter introduces the idea of combining it hierarchically with a regional econometric input–output model (REIM) that provides a long-term trajectory of economic growth that is not reflected by the adjustment model alone. Further, they present an application of the proposed combined framework for a small area population and employment forecasting under various scenarios.

In the same fashion as the first chapter in this second section, Bernardini-Papalia suggests the use of entropy-based estimation applied to recover geographically disaggregated data. More specifically, she proposes a generalized cross-entropy estimator that allows for modeling heterogeneity in subgroup indicators by addressing the spatial dependency problem. The approach proposed in her chapter offers a tractable framework for modeling the underlying variation in small area indicators, in particular when the data set contains outliers. The basic idea is the proposal of an estimator based on an entropy measure of information which provides an effective and flexible procedure for reconciling micro and macro data. A maximum entropy (ME) procedure gives the possibility of including out-of-sample information which can be introduced as additional constraints in the optimization program or specifying particular priors for parameters and errors. The proposed method of estimation is presented as capable of yielding disaggregate data consistent with prior information from different data sources in the absence of high-quality and detailed data. Problems of collinearity and endogeneity are also tackled without imposing strong distributional assumptions. Within this framework, the author shows how partial information at the disaggregated level can be combined with aggregated data to provide estimates of latent variables or indicators which are of interest at the small area level.

The chapter by Mayor and Patuelli estimates a dynamic panel model using a spatial filter in order to account for spatial heterogeneity and/or spatial autocorrelation in unemployment. They compare different methods for obtaining short-run unemployment forecasts in small geographical units and observe their performance between different countries. Their empirical application analyzes regional unemployment rates at the NUTS-3 level for two countries: Spain and Switzerland. Taking advantage of the strong heterogeneity in terms of population or area size across NUTS-3 regions between these two countries, they investigate the variation in the performance of different spatial econometric methods. They further evaluate the forecasting performance of two competing econometric methods: a spatial vector autoregressive (SVAR) model and a dynamic heterogeneous-coefficients panel data model based on an eigenvector-decomposition spatial filtering (SF) procedure.

Finally, the last chapter in this section, by Paez and Mur, presents a methodological contribution to the literature on estimation of spatial models from data at small scale. The authors focus on Geographically Weighted Regression (GWR) techniques and propose several modifications to the conventional specification. The standard GWR analysis is based on a global perspective, but the authors argue that spatial location matters. In their chapter, they suggest different strategies for constructing local weighting matrices that reflect the local surrounding of each observation. Specifically, they study three different but related questions: the development of a GWR test to compare local versus global estimates; the definition of the bandwidth, which they solve in a fully adaptative framework; and the nonuniqueness of the GWR estimates, which follows from the uncertainty in relation to the selection of the kernel.

Applications of Spatial Analysis with Small Area Observations

In the final section of the book, five applied studies are presented where economic analysis is carried out for functional regions and small-area data are used as inputs for the empirical analysis.

Paredes, Lufin, and Aroca contribute with the first chapter in this section. The authors present a study where they apply the concept of Functional Urban Area for quantifying the size and scope of agglomeration economies in Chile. Their hypothesis is that thicker labor markets generate a wage premium in comparison with thin labor markets. To carry out their testing strategy, a redefinition of the geographical space to identify functional thick labor markets is required, and they use spatial and social network analysis to identify functional areas. Once their functional regions are identified, three empirical issues are discussed. The first is the issue raised when the aggregated wage at regional level is used as a dependent variable. The authors add an independent variable indicating the size of the region and its coefficient identifies the wage premium. A second issue is how the literature has defined the geographical scale of a thick labor market. They approach this problem by proposing a stepwise tool for spatial delineation of thick labor markets following a combination of methods based on commuting flows among counties. The third discussion emerges when the wage differential is considered as a causal effect of thick labor market instead of other control variables such as human capital. The authors propose finding identical workers between a thick and thin labor markets using Coarsened Exact Matching (CEM) as a previous step to estimating the wage differential.

Duque, Royuela, and Noreña present a case study of Medellin (the second largest city in Colombia) using a stepwise procedure that efficiently delimits intra-urban slums. The exact delimitation of the borders of these areas has not been a matter of study for local authorities. However, a proper identification and delineation of these areas is a central issue, as it can help prevent poverty traps and crime nests. The authors argue that using administrative areas implies a simple identification but that this procedure can be severely inefficient since in many cases statistical inference based on normative regions may be strongly affected by aggregation problems. They claim that analytical regions, on the contrary, prevent several statistical problems and ensure homogeneous impacts of specific public policy actions.

López, Angulo, and Artal present an empirical application that studies the patterns of co-localization of the economic activities in the Mediterranean Axis in Spain. They base their analysis on statistics built on the concept of spatial dependence that usually characterizes spatial processes. They present an analysis different from traditional studies on cluster analysis, which usually employs official data on employment levels, number of enterprises, or business volume for geographical units such as the county, municipality, region, or country. The authors use data taken at the greatest level of spatial integration (post code level) and they identify clusters of companies dedicated to the same economic activity. Then, they propose applying to these highly disaggregated data a methodology borrowed from

epidemiology studies (Kulldorff Scantest) to improve the detection and identification of clusters.

Chasco and Le Gallo provide an analysis with a multilevel perspective using techniques capable of dealing with data at various spatial scales and supports. The authors explain that the combination of data at different levels creates two categories of problems: (1) statistical problems due to using aggregate or disaggregate data from different levels that could imply a loss of representativeness and (2) interpretation problems, which may arise if we use a wrong method of aggregation, generating an "ecological fallacy" when the level of aggregation is higher than normal, or an "atomistic fallacy" when the level of disaggregation is lower that normal. In their chapter, they propose using explanatory variables at the aggregate level that serve as "moderators" of the relationships present at the individual level. They illustrate how this procedure works with an example: a hedonic price model that measures the impact of air and noise pollution on housing prices. In this example, multilevel models could be very useful because effects on prices operate at different levels and they must be taken into consideration jointly.

Pablo-Martí and Arauzo-Carod contribute with a chapter that closes this section and the book itself. The objective of their contribution is to identify possible patterns of concentration of manufacturing and services industries in Spain. The authors divide Spain into homogeneous cells and check whether each industry follows a concentrated or dispersed pattern and whether collocation exists for pairs of industries so that clusters made by different industries can also be identified. The methodology they apply allows the overcoming of previous drawbacks in terms of data constraints, computing limitations, and border's definitions.

Some Final words

This is an applied research book about spatial analysis characterized by its dual nature, both in regard to the authors and to the contents. The list of contributors is a mixture of well-known names in regional science together with younger researchers in the field of regional and urban economics and spatial econometrics. Regarding the contents of their contributions, most of the chapters present technical or methodological contributions, but always illustrated with useful empirical applications. Armed with this background, we face two fundamental challenges for modern spatial analysis: spatial scale definition and the use of data at small scale.

Defining the Spatial Scale in Modern Regional Analysis
New Challenges from Data at Local Level
Fernández Vázquez, E.; Rubiera Morollón, F. (Eds.)