UNIVERSITY OF OVIEDO

APPLIED ECONOMICS DEPARTMENT

SURPASSING THE ADMINISTRATIVE DIVISION LIMITS ON REGIONAL ANALYSIS:

THREE ESSAYS ON URBAN AND REGIONAL ECONOMICS

ANA VIÑUELA

2010

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE ECONOMÍA APLICADA

SUPERANDO LOS LÍMITES DE LA DIVISIÓN ADMINISTRATIVA DEL TERRITORIO EN EL ANÁLISIS REGIONAL:

TRES ENSAYOS EN ECONOMÍA URBANA Y REGIONAL

Memoria que presenta Ana Viñuela para optar al grado de Doctor por la Universidad de Oviedo, con la condición de Doctor Europeo, bajo la dirección de los doctores Geoffrey Hewings y Fernando Rubiera.

Oviedo, Diciembre de 2010.

Edita: Universidad de Oviedo,
Biblioteca Universitaria, 2011
Colección Tesis Doctoral-TDR nº 77
ISBN: En trámite
D.L.: AS.00488-2011

Universidad de Oviedo

Surpassing the administrative division limits on regional analysis: Three essays on urban and regional economics

Ana Viñuela Jiménez

Depósito Legal: AS.00488-2011
http://www.tesisenred.net/ TDR-0127111-095614

Acknowledgements

To my family (for their love)

To my colleagues (for being there)

To my friends (for their sense of humor)

To my Thesis supervisors (for their incredible support)

To Begoña, Esteban and Fernando (who can fit in more than two of the above)

To Geoff (for being the glue that stuck us together)

To Nora and Andrés (for the sleepless nights)

To Alan (for many days of wine and cheese)

To all of them… thanks!

Table of Contents

List of Tables

List of Figures

List of Maps

Introduction

Agglomeration economies are one of the key concepts in modern Regional and Urban Economics analysis. As the population becomes more and more concentrated into densely-populated urban spaces, namely the cities, a series of external effects essential for understanding territorial dynamics is triggered off. All too often, however, economists forget that economic activity, like all human activity, occurs in a place - a space or a territory - and the place, with its geo-structural characteristics, conditions this activity. Despite this, there exists a large body of literature, both theoretical and empirical, devoted to studying how agglomeration economies permit us to understand economic issues such as the dynamics of growth and innovation, trends in productive specialization or processes of spatial concentration of economic activity.

If agglomeration economies are capable of explaining why economic activity locates in some places but not others, or explaining why large metropolises follow paths that are very different from those of medium-sized or small cities, it would seem natural to think that this is also a fundamental concept in order to understand the spatial dynamics of other aspects of the economy such as labour markets. Labour market theories are normally developed and tested at national level. It is generally accepted in the literature that the main conclusions at this level basically hold when aggregating to the supra-national level or disaggregating to regional and local levels. However, this is not necessarily true. Several studies have pointed out the relevance of regional disparities and their persistence over space and time. The question remains as to whether there may be some common factors that contribute to the explanation of those regional disparities in labour market outcomes. This is the underlying hypothesis running through the three essays that compose this thesis: agglomeration economies are a fundamental concept which can help in understanding the spatial differences arising in labour markets.

To develop this central hypothesis, it is necessary to move away from the traditional classification of the territory into politico-administrative regions and search for an alternative spatial classification capable of including agglomeration economies. This leads us to fundamental questions that motivate the title of this thesis. How much faith can we have in the administrative division of territory? Is it possible to consider some other more efficient way of defining the *regions*? Should this be done by constructing regions *a posteriori* following statistical aggregation methods or should the classification be proposed *a priori* based on economic theory? How can we test the usefulness of an alternative regional classification that we might propose?

These questions have inspired the first chapter of this thesis. Using disaggregated data at local level, we depart from the traditional political-administrative definition of regions and instead choose an alternative territorial classification based on economic criteria. This classification takes into consideration that in Regional and Urban Economics population *size* matters, as larger populations provide more opportunities to accede to and take advantage of agglomeration economies. Simultaneously, the *distance* to size also matters, as a territory with a small population but located close to a large metropolis can take advantage of the agglomeration economies produced in a large city without incurring the negative external effects.

With these factors in mind, we resort to a mode of aggregating territorial units which has been widely used in the literature to test aspects of localization, specialization and growth of economic activity and which has been applied in several studies to the object of our research, namely the Spanish economy. Though constructed *a priori* based on economic criteria not directly linked to labour economics nor specifically designed to study regional labour markets, in this first chapter we show that this analytical classification results in more appropriate regions – higher degree of compactness and separation in relation to several labour variables - than the administrative ones commonly used.

After evaluating the aptness of this alternative classification when carrying out analysis on local labour markets, in the second and third chapters of this thesis we address the question at the heart of this research: how are spatial differences in labour markets generated and how can agglomeration economies help explain these differences.

The aim of the second chapter is to evaluate employability using a spatial approach. There are several empirical studies addressing the issue of employability, both for Spain and other countries, and these have led to some consensus within the scientific community about the factors determining the employability of individuals. Among the variables whose influence has been demonstrated are the educational level, sex, age and family conditions. However, the fact that these individuals live and work in a particular territory is usually overlooked. Therefore, the issue arises as to how employability patterns change when we move to more disaggregated spatial levels. Will we observe discrepancies with respect to the national results, as we hinted earlier? If so, can these differences be explained by the effects of agglomeration economies? In short, can Regional and Urban Economics shed light on a topic which belongs to the sphere of Labour Economics?

Using the analytical classification described in the first chapter, in Chapter 2 we analyze the existence of spatial differences in the probability of being employed in Spain. Once the personal characteristics are controlled for, the interpretation of these differences will provide us information on the relevance of size (agglomeration economies) and distance to the metropolis (location) for the probabilities of being employed.

However, if the characteristics of the region of residence are relevant when explaining the probabilities of being employed - and this turns out to be the case - the natural reaction to this seemingly deterministic outcome would seem to be that workers can move to another type of region in search of better job opportunities. With this in mind, the objective of the third chapter is to try to gain some insights into another key aspect of labour markets: the geographical mobility of workers and in particular from the periphery to the core. Over the last 20 years Spain has experienced a significant increase of in-flows to core regions, either from the peripheral Spanish regions or from other parts of the world. Again, using the analytical regions described in the first chapter, which allow us to distinguish between core (central) and peripheral regions, this chapter explores the chain of effects that the arrival of new workers has on the core regions.

The analysis is based on an input-output migration model incorporating information of commuting flows that, has been extended to the study of the distribution of *jobs* and also *residences* across the different types of central regions (basically metropolitan areas and main cities classified by size). According to this model, agglomeration economies can also affect the distribution pattern of residences and jobs: depending on their size, cities can either be attractive places to work but not to live (high housing costs, congestion or some other negative externalities) or vice versa.

In the last section of this thesis, named *Conclusions and Extensions,* we summarize the main lessons to be drawn and briefly lay out several questions which arose during the elaboration of this research and which will constitute the main line of research of the doctoral candidate in the immediate future. The results obtained in this study encourage us to continue with this line of research, which in a certain way draws together Regional and Urban Economics and Labour Economics.

Chapter 1

An analytical regions proposal: surpassing the administrative division of territory in the study of the Spanish labour market

Objectives

To prove that a functional division of the territory based on the existence of agglomeration economies and the importance of location results in more convenient regions for the study of labour markets issues than the administrative ones commonly used (NUTS regions).

Methodology

Evaluation of the functional regions using the Theil Index and the Davies-Bouldin Validation Index to test the internal homogeneity and heterogeneity between the areas defined, with an application to labour market issues.

Synopsis

Internal homogeneity and also heterogeneity between the regions are both desirable properties for a better understanding of the local labour markets and for increasing the efficiency of any industrial policy applied at local level. However, studies of the labour markets that include a spatial dimension are commonly limited to administrative rather than appropriately-defined functional regions. Using micro data from the latest Census available, the Spanish territory can be divided into functional regions that emphasize the importance of location and agglomeration economies (size). Using both the Theil Index and the Davies-Bouldin Validation Index, this chapter shows that the classification based on economic criteria results in more convenient regions – higher degree of compactness and separation - than the administrative ones commonly used (NUTS regions).

1. An analytical regions proposal: surpassing the administrative division of territory in the study of the Spanish labour market

1.1 Introduction

Economic data is usually disaggregated according to the administrative or normative division of the territory. However, this does not necessarily make clear economic sense as these classifications are or were constructed in terms of some sort of political, administrative or historical criteria. In Spain, as in most European Union countries, the data are commonly organized according to the *Nomenclature des Unites Territoriales Statistiques*[1] into NUTS I or NUTS II regions (Autonomous Communities in the Spanish case). At best, data can be more spatially disaggregated into NUTS III regions (*Provinces* in the Spanish case).

When using data from these normative regions, we may fall into a well-known trap called the ecological fallacy (EF). This problem occurs when conclusions for aggregated data do not reflect the reality of individuals belonging to this aggregation.[2] In other words, the group or region created is not necessarily homogeneous, which is also referred to as aggregation bias in the literature. Given the availability of abundant and comparable data at European level data under the NUTS system, regional economics researchers have not tended to consider any alternative spatial aggregations. Instead, regional scientists have been devoted to formulating

[1] *Nomenclature des Unites Territoriales Statistiques* (NUTS) is the geographical system established by Eurostat for the production of regional statistics within the European Union. According to Eurostat, these "normative regions are the expression of a political will; their limits are fixed according to the task allocated to the territorial communities, to the sizes of population necessary to carry out these task efficiently and economically, or according to historical, cultural and other factors" (Eurostat, 2006).

[2] EF was first introduced by Robinson (1950) and has been studied by many other authors since then. See, for example, Richardson *et al.* (1987), Piantadosi *et al.* (1988), Greenland and Morganstern (1989) and Richardson (1992).

statistical models or estimation procedures to reduce aggregation bias. Gotway and Young (2002) provide a detailed overview of several statistical solutions which have been proposed to deal with this problem.

If access is available to micro data for smaller units, researchers could try to carry out their own regionalization procedure in order to avoid this fallacy. However, if this approach is adopted, the resulting summary values may be influenced by the choice of the aggregation or regionalization technique, by the number of regions in which territory is divided, or by other restrictions imposed on the groups. This is another source of statistical bias called the 'modifiable areal unit problem' (MAUP) (Openshaw, 1984).

Thus, a way to reduce aggregation bias, caused either by the use of existing data from normative regions or by the characteristics of the regionalization exercise carried out, is to try to identify regions showing the maximum intraregional homogeneity as well as maximum interregional heterogeneity (Fischer, 1980).

Based on this simple idea, the objective of this chapter is to evaluate a set of analytical regions constructed using economic criteria as opposed to the normative regions commonly used (NUTS I, NUTS II and NUTS III). Although this has been successfully applied to explain the location of economic activity in Spain, to date there is no empirical evidence suggesting the superiority of these functional regions over the administrative division. Thus, using micro data from the 2001 Spanish Census[3] - the last year available - we illustrate that this configuration of the territory can be very valuable in regional economic studies, especially when studying labour market issues.

[3] The Spanish Census is administered by the INE (National Statistics Institute of Spain).

The rest of the chapter is organized as follows: in the next section we discuss the meaning of *regions* and describe the aggregation criteria used for creating analytical regions with economic sense based on agglomeration economies and the importance of location. Section 3 deals with the evaluation of the normative regions versus the analytical ones when including the spatial dimension in the study of the Spanish labour markets. Internal homogeneity within the regions and heterogeneity between them are evaluated by gender, industry and level of qualification. In the last section we summarize the main conclusions of this chapter.

1.2 Surpassing the administrative regions: an analytical proposal based on economic criteria

The concept of *Region* from the economic point of view has only one basic element: economic openness. A *region* is a unit in which capital and labour move freely and goods and services are totally open to trade with other regions without any frontiers or limitations. The openness of the regions and their interaction with other regions are their main characteristics.

From this basic idea the observer can use different ways to divide a particular territory into parts or *Regions*. Three elements must be taken into account (Behrens and Thisse, 2007). First, a *Region* is part of a set in which each comprising element has some specificities which make it different from the rest. Secondly, a set of regions always involves a partition of some geographical space that contains a large number of places, with a place serving as the elementary spatial unit that we use (in our case municipalities). Thirdly, a well-known result in set theory is that there is one-to-one correspondence between the family of partitions in a set and the family of equivalence relations of the same set. An equivalence relation on a set is (i) reflexive, (ii) symmetric and (iii) transitive. These imply that (i) an object is always similar to itself; (ii) if one object is similar to another the latter is similar to the former and (iii) two objects similar to a third one are themselves similar. Based on these three basic criteria, many possible sets of regions may be defined, and as a result, depending on the point of view selected, many types of concepts of *Region* can be constructed. It would only depend on the point of view of the analyst.

In the labour economics field, data constraints have led to a situation where the researcher interested in the performance of regional or local labour markets (LLMs) usually bases her work on information about states or administrative regions, such as NUTS I, NUTS II or NUTS III in the European case. However, some attempts have been made to use other spatial aggregations of the data reflecting functional relationships between workers and jobs. Thus, the design of these local labour markets has been based on commuting patterns as a means identifying the borders of labour catchment areas. In practice, this means that at least 75% of residents work in the area and that 75% of those who work in the area also live there,[4] with a minimum size of 3,500 resident workers. The regionalization procedure is based on an algorithm originally developed by Coombes *et al.* (1986)[5] that uses a multi-stage aggregation process. This methodology has been applied by the Department of Employment in Great Britain (which defines the so-called Travel-To-Work-Areas or TTWAs), Sforzi *et al* (1997) for Italy, Papps and Newell (2002) for New Zealand, Andersen for Denmark (2002) or Watts (2004) for Australia. Casado-Diaz (2000) and Feria and Susino (2005) have applied it for the Spanish case, but due to the lack of data they compute local labour markets only for two Autonomous Communities, Valencia and Andalucía respectively. A modification of the methodology proposed in Coombes *et al.* has been applied to Italy by ISTAT-IRPET (1989), ISTAT (1997) ISTAT (2005), and also for the Spanish territory as a whole by Boix and Galletto (2006).

---

[4] See the UK National Statistics website: http://www.statistics.gov.uk/geography/ttwa.asp

[5] For a description of the previous method, see Smart (1974). For a discussion of problems that arose with that method, see Ball (1980) and Coombes and Openshaw (1982).

If the researcher chooses a pure urban and regional economics perspective (see Fujita *et al.* 1999), for the definition of a *Region* a small number of attributes should be highlighted, namely: (i) location matters, because industries (and therefore economic activity and employment) are always drawn to places best suited for commerce and interaction with markets; and (ii) size matters, because dynamic industries, or the most advanced in each epoch, are naturally drawn to large cities and places within easy reach. A corollary can be deduced from (i) and (ii), namely: (iii) proximity to size also matters. Another basic idea of regional economics is: (iv) cost matters, because without adequate size or a propitious location, places will grow if they have a clear labour cost advantage or, alternatively, an exceptional resource endowment (Polèse, 2009). The existence of agglomeration economies (size) and distance are the key factors of this definition.

In less abstract terms, the gains derived from large-scale production and from the positive externalities associated with size lead to the concentration of economic activity in central locations from which the largest possible market is accessible. Transportation costs constrain this concentration behaviour, but the strength of this limitation depends on the consumption characteristics of the activity. Those activities that require intense personal interaction between consumers and producers (which includes many services) and/or are consumed daily or very frequently will display quasi-equal distributions over space. In contrast, activities that are tradable over broader distances, not requiring proximity to the point of consumption, and/or are demanded less frequently will concentrate their production in a limited number of central locations. As distance costs fall and trade increases, larger concentrations should normally grow in size. A shift in the national economy towards agglomeration-sensitive goods and services (and away from, say, agriculture) also favours the growth of larger concentrations (see Parr, 2002).

As large concentrations grow, diseconomies naturally appear, producing an expulsion effect for some activities. Wages and land prices are in part a function of city size. Wage-sensitive and space-extensive activities will be pushed out by what is sometimes called the “crowding-out effect” of rising wages and land prices in large metropolitan areas.[6] This crowding-out effect will most notably be felt by medium-technology manufacturing - which has less need for the highly skilled labour in large cities (Henderson and Thisse, 1997) - and also by wholesaling and distribution, which are extensive consumers of space, giving rise in turn to the growth of smaller cities.

On the other hand, the agglomeration economies associated with urban concentration lead to firms within the same industry benefitting through lower recruitment and training costs (shared labour-force), knowledge spillovers, lower industry-specific information costs, and increased competition (Rosenthal and Strange 2001, Beardsell and Henderson 1999, Porter 1990). The increasing size of the metropolis makes certain infrastructures possible, such as international airports, post-graduate universities, research hospitals, etc. The recent literature stresses the positive link between productivity and the presence of a diversified, highly-qualified and versatile labour pool (Duranton and Puga, 2002; Glaeser, 1998; Glaeser *et al.*, 1995). As highlighted by Hall (1998), Eaton and Eckstein (1997) and Castells (1976), large metropolises stimulate the exchange of knowledge, and the link between urban agglomeration and economic growth has been explored by Polèse (2005). Activities that are characterized by the need for high creativity and innovation will in general choose to locate in or near to major metropolitan areas (Desmet and Fafchamps, 2005).

It is reasonable to infer that the trade-off between the positive and negative effects that push economic activities towards large cities or drive them out should give rise to an economic landscape characterised by

[6] There is a large literature on the effects of agglomeration economies on wage behaviour: see, among others, Partridge et al (2008, 2009) or Head and Mayer (2006). However, this is outside the scope of our research.

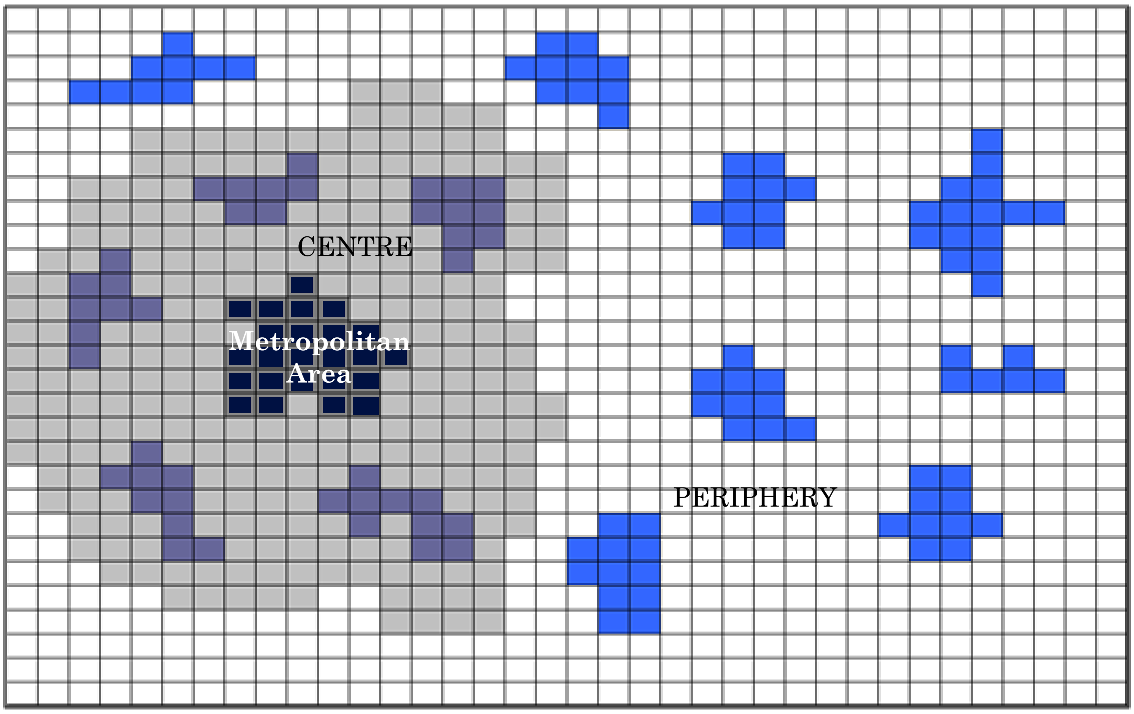
regularities in industrial and employment location patterns based on the size of and distance from some other (larger) cities (see Redding and Venables, 2004). This provides the conceptual foundation for the urban economics types of region proposed in Coffey and Polèse (1988), Polèse and Champagne (1999), and Shearmur and Polèse (2007) for Canada and its posterior application to the Spanish case by Polèse et a (2007).

Though originally designed for explaining the location of economic activity and economic growth, this analytical classification is potentially very useful for labour market analysis. In practical terms, the spatial statistical units (either census divisions, counties, municipalities, *länders*, etc.) that constitute the national economic space are aggregated based on the population size and distance criteria. Thus, these new analytical regions created are groupings of analogous statistical units, classified by size and by distance (to the nearest metropolitan area).

Figure 1.1 presents a schematic representation for an idealized national space economy. The reader will undoubtedly note the resemblance with the classic idealized economic landscapes of Christaller, Lösch, and Von Thünen, all of which posit one metropolis or marketplace at the centre. Thus, Figure 1.1 represents one big metropolis at the centre, but also four smaller "central" urban areas of different population sizes around it, as well as other "central" rural areas (these areas, either urban or rural, are close to the metropolis). Another four analogous size classes represent the "peripheral" urban areas, which are located at some distance from the metropolis and surrounded by their corresponding rural areas. It is implicitly assumed that urban areas are distributed in accordance with the rank-size rule.[7]

[7] This was proposed by George Zipf (Zipf, 1949) and consists of identifying a statistical relationship between the size of cities and their position in the city ranking. The rule has been tested for several countries and time periods, with prominent empirical studies including Gabaix (1999) and Ades and Glaeser (1995).

Figure 1.1. Schematic Representation of the Classification of Spatial Units.



Key
Metropolitan Area
Central Urban
Central Rural
Peripheral Urban
Peripheral Rural

Source: Own elaboration based on Polèse et al. (2007).

Following this classification, in relation to size there are three main types of areas:

Metropolitan areas: metropolitan areas of more than five hundred thousand inhabitants. These metropolitan areas include the city and its surrounding area of influence. They are *ad hoc* specifications.

Urban areas: urban agglomeration areas with more than ten thousand inhabitants.

Rural areas (RA): less than ten thousand inhabitants, including all areas that are not urban but which may contain towns.

A parallel distinction, based on proximity to major metropolitan areas, is applied to all non-metropolitan urban areas:

> *Central areas* (CA): all areas within approximately one hour's drive of a *metropolitan* area.
>
> *Peripheral areas* (PA): all areas situated more than one hour's drive from metropolitan areas.

The one hour's drive criterion takes into account several factors such as road conditions (e.g., highway or not), the spatial limits of metropolitan areas, and the distinctive characteristics of the area being classified. Thus, as illustrated on Figure 1.1, central areas do not necessarily form perfect rings around metropolitan areas. The one-hour threshold, also used in other applications, has been found to be very robust and a good indicator of the range within which spatial interaction with the metropolis remains fairly easy, especially for face-to-face relationships related to the consumption of high-order services.

1.3 The Spanish case: analytical regions *versus* administrative regions

Over the last three decades Spain has moved gradually from a strongly centralised public administration system to a highly decentralised model.[8]

The State of Autonomous Communities, included in the 1975 Constitution, has transformed Spain *de facto* into a federal system. The degree of decentralization in public spending in Spain is not far from that of Germany, which is one of the most decentralized governments in the European Union (Toboso, 2005). Most of the new responsibilities of the

[8] A complete review of all these changes are the objective of a special issue of *Revista de Economía* (2003, number 811) which summarizes the social, political and economic transformations in Spain over the previous 25 years.

developing Spanish welfare state have become responsibility of *regional* or *local* governments instead of central government.

Thus, the Autonomous Communities (NUTS II regions) are the reference for territorial and urban planning, health, education, active labour market policies and many other important policies. In the literature, the efficiency and ability of the Autonomous Communities in terms of better identifying regional and local needs is not an issue.[9] That is, debate among economists is not concerned with the *efficiency* of the system, but rather with its *equality,* and most studies are therefore concerned with its model of financing.[10] In this context, any alternative (and perhaps better) functional regional classification is hardly considered when designing or implementing economic policies where identifying the local particularities is a plus.

Spain is divided into seventeen Autonomous Communities (NUTS II regions), some of which include several provinces (NUTS III) for a national total number of 50 provinces.[11] Each province is in turn divided into several municipalities, ranging from 34 (Las Palmas) to 371 (Burgos). Furthermore, the seventeen Autonomous Communities are also aggregated into seven administrative regions (NUTS I regions), which have no real internal meaning and are only used for comparative purposes with some other European member-states. In 2001 Spain had 8,106 municipalities. The Spanish Census gives population and employment figures (by age, gender, industry and level of qualification and occupation) for each municipality.

In the presence of agglomeration economies, economic activity and therefore employment tend to be unevenly spatially distributed, instead exhibiting positive spatial dependence. These disparities and spatial dependence can be observed at regional level either using the NUTS I, NUTS II or NUTS III regions.

---

[9] See, for example, Sevilla (2005).

[10] As an example, see the monograph of the *Instituto de Estudios Fiscales* coordinated by Salinas and Álvarez (2003).

[11] Ceuta and Melilla are excluded from the study.

For example, in the year 2001 the Spanish employment rates[12] for the NUTS I regions ranged from 50.67% (South) to 62.72% (Madrid). At the NUTS II regional level, the figures ranged from 49.44% (Andalucía) to 63.65% (Cataluña), while for the NUTS III regions, the province with the lowest employment rate was Cádiz (43.45%) and the one with the highest was Girona, with 65.98% employment rate. To get a better idea of the magnitude of these disparities, Table 1.1 and Table 1.1 show the average and the standard deviation for NUTS I, NUTS II and NUTS III regional employment rates.

Table 1.1. NUTS I regional average employment rates.

NUTS I. Average Employment Rates					
Below Average			**Above Average**		
	Average	**Standard Deviation**		**Average**	**Standard Deviation**
South	50.67%	0.0985	North-East	60.54%	0.0743
North-West	55.03%	0.0663	East	61.93%	0.0668
Centre	55.29%	0.1004	Madrid	62.72%	0.0515
Canary Islands	56.33%	0.0671			
TOTAL	**Average**	**57.77%**	**Standard Deviation**		**0.099**

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

The wide disparities observed at NUTS I, NUTS II and NUTS III level reflect little or no consideration for the meaningfulness of these administrative or normative regions, which are the ones commonly used when studying labour market topics. If data at more disaggregated levels were available, alternative types of regions (i.e., analytical regions) could be used and the aggregation method chosen (regionalization procedure) should be of major interest to researchers and policy makers.

[12] Although EUROSTAT provides figures for employment rates for the NUTS I, NUTS II and NUTS III regions, for comparison purposes with the analytical regions under evaluation these figures have been calculated from the corresponding aggregation of the municipal data. Moreover, when using the 2001 Census from the Spanish National Institute of Statistics (INE), employment rates can only be estimated with the *potential* active population, i.e., all the population aged 16 to 64, either active or non active.

When grouping data, lots of very well-known regionalization algorithms for spatial aggregation can be used.[13] The purpose of these regionalization algorithms is basically to generate regions that are internally coherent but clearly different from each other. These algorithms minimize the objective function value in such a way that the intraregional heterogeneity is as low as possible. The methods assume certain prior knowledge about the relevant variable (or variables) for aggregation, the number of regions to be designed or the existence of contiguity constraints, so in this sense they are *supervised*.

[13] For a review of the literature on regionalization methods, see Fischer (1980), Murtagh (1985), Gordon (1996; 1999) and Duque *et al.* (2006).

Table 1.2. NUTS II and NUTS III regional average employment rates.

NUTS II (CC.AA.) Average Employment Rates					
Below Average			**Above Average**		
	Average	**Standard Deviation**		**Average**	**Standard Deviation**
Andalusia	49.44%	0.0987	**Murcia**	58.27%	0.0428
Extremadura	50.00%	0.0998	**Basque Country**	58.90%	0.0612
Asturias	52.33%	0.0667	**C. Valenciana**	59.18%	0.0680
Cantabria	55.78%	0.0535	**Aragón**	61.81%	0.0787
Galicia	55.95%	0.0693	**Balearic Islands**	62.60%	0.0468
Canary Islands	56.33%	0.0671	**Rioja**	62.66%	0.0792
Castilla-Mancha	56.56%	0.0891	**Madrid**	62.72%	0.0515
Castilla-Leon	56.65%	0.0949	**Navarra**	63.18%	0.0642
			Cataluña	63.65%	0.0538
Total	**Average=**	**57.77%**	**Standard Deviation=**		**0.099**

NUTS III (Provinces) Average Employment Rates					
Below Average			**Above Average**		
	Average	**Standard Deviation**		**Average**	**Standard Deviation**
Cádiz	43.55%	0.0635	**Alicante**	57.87%	0.0796
Badajoz	47.40%	0.0747	**Murcia**	58.27%	0.0428
Granada	48.38%	0.1063	**Toledo**	58.94%	0.0570
Jaén	48.73%	0.0975	**Valencia**	58.95%	0.0493
Huelva	48.90%	0.0712	**Lugo**	58.98%	0.0611
Sevilla	49.04%	0.0732	**Almería**	60.91%	0.0885
Córdoba	49.67%	0.0787	**Burgos**	61.29%	0.0832
Málaga	51.38%	0.1061	**Teruel**	61.40%	0.1027
Asturias	52.33%	0.0667	**Guadalajara**	61.47%	0.1077
Ciudad Real	52.83%	0.0586	**Zaragoza**	61.49%	0.0653
Ourense	53.36%	0.0690	**Álava**	61.88%	0.0603
León	53.88%	0.0847	**Segovia**	61.91%	0.0769
Salamanca	54.23%	0.0718	**Guipúzcoa**	62.06%	0.0513
Cáceres	54.26%	0.0920	**Tarragona**	62.32%	0.0510
Zamora	54.32%	0.0769	**Balearic Islands**	62.60%	0.0468
Albacete	55.12%	0.0640	**Rioja**	62.66%	0.0792
Coruña	55.42%	0.0494	**Madrid**	62.72%	0.0515
Sta. Cruz Tenerife	55.65%	0.0567	**Navarra**	63.18%	0.0642
Valladolid	55.77%	0.0598	**Barcelona**	63.38%	0.0469
Cantabria	55.78%	0.0535	**Huesca**	63.48%	0.0616
Vizcaya	56.23%	0.0568	**Castellón**	64.22%	0.0647
Palencia	56.27%	0.0681	**Soria**	65.79%	0.0965
Ávila	56.36%	0.0721	**Lleida**	65.96%	0.0530
Pontevedra	56.39%	0.0655	**Girona**	65.98%	0.0579
Palmas	56.93%	0.0696			
Cuenca	57.35%	0.0803			
Total	**Average**	**57.77%**	**Standard Deviation**		**0.099**

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

However, as explained in the previous section, the regionalization method proposed by Polèse *et al.* (2007) is not based on any mathematical algorithms minimizing differences or on a pure labour economics basis (commuting patterns), but on regional and urban economics criteria based on location, size and proximity to size.

Applying the original classification to the particular characteristics of the Spanish case,[14] the classification would be as follows:

Metropolitan areas: The same 500,000 threshold for Canada is used in the Spanish case to define metropolitan areas. However, given the special demographics of Spanish cities, metropolitan areas are subdivided into two classes. The first, MA1, includes metropolitan areas with more than two and a half million inhabitants. The second, MA2, refers to metropolitan areas with a population of between 500,000 and 2,500,000 inhabitants.

Urban areas (UA1 and UA2): The urban agglomerations with more than ten thousand inhabitants are grouped into two classes. The first, UA1, includes all areas with more than 100,000 inhabitants and less than 500,000; the second, UA2, all urban areas with populations between 10,000 and 100,000 inhabitants.

Rural areas (RA): all areas that are not urban, which may contain towns, but with less than ten thousand inhabitants in 2001.

[14] A strict application of the original definition of "metropolitan area" (areas with more than 500,000 inhabitants) in the Spanish case generates two clear sub-groups: Madrid and Barcelona Metropolitan Areas (which both concentrate more than 4 million inhabitants), and the rest of the Spanish metropolitan areas (with populations less than 1.5 million in all cases). Similarly, two categories of *urban* regions are created (less than or more than 100,000 inhabitants). For more details, see Appendix I.

The hour's drive criterion also holds in the Spanish case:

Central areas (CA): all areas within approximately one hour's drive of a *metropolitan* area (either MA1 or MA2).

Peripheral areas (PA): all areas situated more than one hour's drive from metropolitan areas (either MA1 or MA2).

Thus, taking into account the particular demographic structure and city size characteristics, in the Spanish case eight types of regions were created. Table 1.3 presents a summary of the eight types of regions based on the size and distance criteria into which the 8,106 Spanish municipalities are aggregated.

In line with this classification, in Map 1.1 we can see the resulting division of Spanish territory, an unfamiliar and novel landscape where these "types of regions" do not necessarily share borders (there is no contiguity constraint) or belong to the same politico-administrative region, i.e. a NUTS III region (*province)* or NUTS II region (Autonomous Community).[15]

[15] As Fischer (1980) points out, "The main distinguishing characteristic between regions and regional types is the spatial contiguity of the spatial basic units."

Table 1.3. Territorial classification by size and position. Application to the Spanish case.

Metropolitan areas (MA) of more than 2,500,000 inhabitants ([1])	**MA1**	
Metropolitan areas of between 500,001 and 2,500,000 inhabitants ([2])	**MA2**	
	Central Urban Areas (no more than a one hour drive from a MA)	Peripheral urban areas (more than a one hour drive from a MA)
Urban areas of between 100,001 and 500,000 inhabitants ([3])	**CUA1**	**PUA1**
Urban areas of between 50,001 and 100,000 inhabitants	**CUA2**	**PUA2**
Rural areas, less than 50,000 inhabitants	**CRA**	**PRA**

***Notes**:*

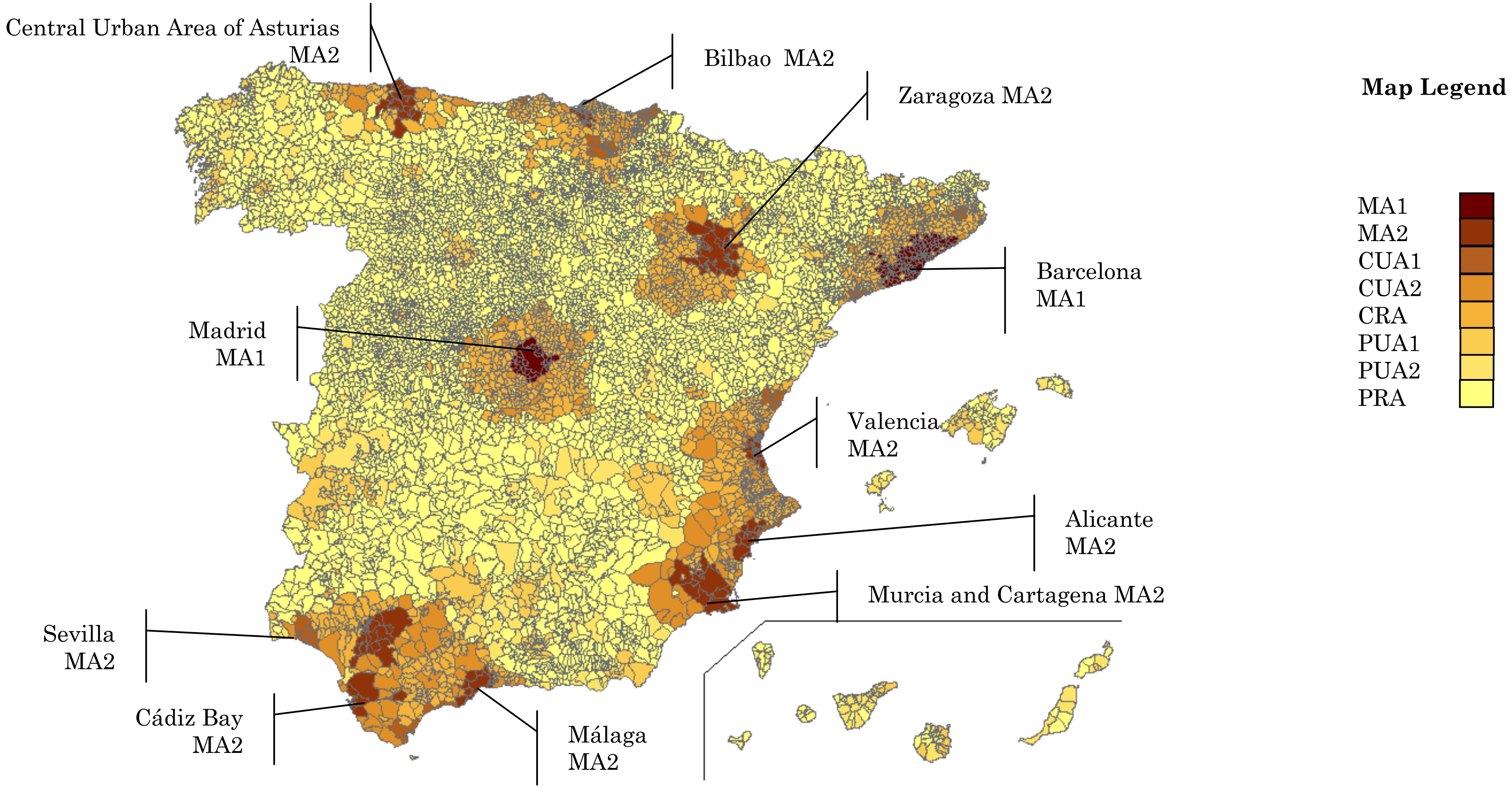
([1]) Metropolitan Areas of Madrid and Barcelona.

([2]) Metropolitan Areas of Alicante, Bilbao, Cadiz Bay, the Central Urban Area of Asturias, Malaga, Murcia and Cartagena Conurbation, Seville, Valencia and Zaragoza.

([3]) There are more than 200 municipalities that can be classified as "urban" and "central", with the most important ones being Castellon, Girona, Huelva, Malaga, San Sebastian, Santander-Torrelavega, Tarragona, Vitoria and some of their surrounding municipalities.

Source: Polèse et al. (2007).

Map 1.1. Spanish Territory into the Eight Analytical Types of Regions.



***Source**: Own elaboration based on Polèse et al. (2007).*

The employment rates using this analytical classification are shown in Table 1.4. It should be remarked that internal dispersion seems to be stronger the higher the number of municipalities (or observations) included in the region, which suggest the existence of some scale effect. Some of the municipalities are extremely small.[16] The municipalities with less than 50,000 inhabitants are included in the regions called Central Rural Areas (CRA) and Peripheral Rural Areas (PRA) which, despite concentrating only 7.1% and 11.8% respectively of the total population, represent 20.9% and 65.6% of the number of Spanish municipalities respectively.

Table 1.4. Average employment rates and standard deviations by analytical regions.

	Number of Municipalities included	Average employment rate	Standard Deviation
Central regions			
Metropolitan Areas 1 (MA1)	192	62.79%	0.0365
Metropolitan Areas (MA2)	202	54.18%	0.0757
Central Urban Areas 1 (CUA1)	74	58.72%	0.0785
Central Urban Areas 2 (CUA2)	171	58.39%	0.0699
Central Rural Areas (CRA)	1,694	58.71%	0.0901
Peripheral regions			
Peripheral Urban Areas 1 (PRA1)	259	55.91%	0.0837
Peripheral Urban Areas 2 (PUA2)	193	56.65%	0.0698
Peripheral Rural Areas (PRA)	5,321	55.37%	0.1061
Total	**8,106**	**57.77%**	**0.0999**

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

[16] In 2001 there were two municipalities - Salcedillo (Teruel) and Illán de Vacas (Toledo) - with only 7 inhabitants. Almost 12% of the municipalities have less than 100 inhabitants. The figure increases to 26% for less than 200 inhabitants.

Although this chapter does not deal directly with the labour economics literature on local labour markets, i.e., the analytical areas were not constructed using any sort of commuting criteria between municipalities, it can be easily demonstrated that the travel-to-work commuting patterns of these analytical regions generated under the size and proximity criteria also fit into the definition of a local labour market area. This is an additional positive feature of the analytical regions as it is agreed that the LLMAs are the ideal geographical areas for the implementation of any regional industrial policy or for reporting disaggregated labour figures (Ball, 1980).

Table 1.5 shows commuting patterns for the eight regions under analysis. The metropolitan areas and the regions that include the bigger municipalities (cities with more than 100,000 inhabitants and their surrounding areas or influence)[17] would also strictly fulfil the 75% threshold for commuting patterns set in the labour economics literature, i.e., they are self-contained local labour markets or travel-to-work areas from both the demand and the supply side.

Table 1.5. Commuting Patterns for the Analytical Regions.

Analytical Regions	MA1	MA2	CUA1	CUA2	CRA	PUA1	PUA2	PRA
Percentage of people working in the residence region	91.2	89.5	84.3	69.6	59.3	87.7	77.7	65.8
Percentage of people living in the working region	88.1	84.6	75.4	69.2	68.6	76.2	79.5	83.9

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

[17] As suggested by Papps and Newell (2002), "The metropolitan area system is not designed to deal with rural areas." (p. 6)

However, in theory an optimal region should fulfill at least one of two principles: *internal homogeneity*, whereby individual regions should be as homogeneous in the attribute space as possible, and *external separation*, whereby different regions should be as far apart in the attribute space as possible (Fischer, 1980). Taking these two principles into account, in the next section we will evaluate the analytical regions and the administrative ones using two different validation indexes.

1.4 Evaluation of the analytical classification

1.4.1 Evaluation criteria

The regionalization procedure used for designing the analytical regions is not based on any statistical algorithm or method. However, when evaluating these analytical regions by comparing them to other administrative divisions, the same criteria proposed for clustering evaluation and selection of an optimal clustering scheme applies: i) *Compactness*, i.e., the members of each region should be as close to each other as possible; and ii) *Separation*, i.e., the regions themselves should be different from each other. There are different approaches to measuring the distance between two regions: distance between the closest members of the regions (single linkage), distance between the most distant members (complete-linkage) or distance between the centres of the clusters (comparisons of centroids) (Berry and Linoff, 1997).

Applied to local labour markets, the first criterion (internal homogeneity) implies the existence of very well-defined labour markets where the municipalities included share common characteristics, problems and diagnosis. This compactness is a desirable feature for any active labour policy designed to be implemented at local level. As for the second criteria (dissimilarity between regions), the existence of differences between regions in practical terms implies the spatial customization of policies, i.e., policymakers should be very aware of the economic characteristics of each region in order to try to identify its particular problems. The success of a policy in a certain region does not guarantee its immediate success in another.

1.4.2 Hypotheses

To evaluate the relative performance of the administrative regions (NUTS I, NUTS II and NUTS III) versus the analytical ones designed under economic criteria, we will test the following hypotheses:

H1: under the existence of agglomeration economies and economies of scale, the analytical regions proposed are better for describing the *employment distribution patterns*, either total or by gender, in Spanish territory than any other administrative division.

H2: the analytical regions capture the patterns of *distribution of employment by industry* better than the administrative regions commonly used.

H3: the *spatial distribution of employment by type of job performed and level of qualification* can be better explained using the analytical regions than any other type of administrative division.

To test these three hypotheses, first we will use the well-known Theil inequality index commonly applied to the distribution of income and wealth (Theil, 1967). The index can be decomposed into the sum of a *between* and a *within* component.[18] Its *within* component will be useful to quantify the intraregional homogeneity of the regions when dealing with the spatial distribution of employment. Thus, given the characteristics of Theil's index, if the internal homogeneity of the regions increases (a decrease of the within component), this necessarily implies that the heterogeneity between regions increases (a rise of the *between* component).

[18] For more details about the calculation of the Theil Index, see Appendix II.

In reality, this is not necessarily the case: a given region may be very well-defined internally but exhibit no particular differences from another (the reverse also applies). In order to include both properties simultaneously, we can use techniques specifically designed for clustering validation. Thus, in a second stage we test the three hypotheses again with the Davies-Bouldin Validity Index (Davies and Bouldin, 1979).[19] Given the characteristics of our database and the administrative and analytical classifications of the regions (non-hierarchical crisp clusters with different numbers of regions), the Davies-Bouldin index is the most appropriate clustering validation technique.[20] This index estimates the average similarity between each region and the most similar one to it and exhibits no trends with respect to the number of regions. Small values of the index are indicative of the presence of compact and well-separated regions.

1.5 Evaluation with the Theil Index

The first hypothesis we are testing is related to the regional distribution of employment and the distribution of employment by gender.

Table 1.6 shows the Theil index and its components (total and by gender) when the 8,106 Spanish municipalities are aggregated into administrative regions – NUTS I (7 regions), NUTS II (17 Autonomous Communities), NUTS III (50 Provinces) - and into the analytical regions (8 regions).

[19] For more details about the calculation of the Davies-Bouldin index, see Appendix III.

[20] For a good review of the main clustering validation techniques, see Halkidi *et al.* (2001).

Table 1.6. Decomposition of the Theil inequality index by gender.

		Administrative Regions			Analytical Regions
		NUTS III (PROV) (50 Regions)	NUTS II (CCAA) (17 Regions)	NUTS I (7 Regions)	MA1 to PRA (8 Regions)
Total Population					
Theil's Index	*Between*	0.6230	0.4888	0.2401	1.2498
2.4326	*Within*	1.8096	1.9437	2.1925	1.1828
Male					
Theil's Index	*Between*	0.5964	0.4646	0.2304	1.1937
2.3112	*Within*	1.7149	1.8468	2.0808	1.1175
Female					
Theil's Index	*Between*	0.6713	0.5327	0.2573	1.3455
2.6434	*Within*	1.9721	2.1103	2.3861	1.2979

Source: Authors' calculations based on 2001 Spanish Census (INE. 2007).

In spite of the scale effect - i.e., everything else equal, intraregional inequality drops with the number of regions - the *within* component for the eight analytical regions is clearly lower than for any of the NUTS regions. In other words, the classification proposed shows a higher degree of internal homogeneity in the distribution of employment so that the regional labour markets generated under the size and distance criteria are more integrated or coherent (even by gender) than any other political-administrative division of the territory.

The second hypothesis under scrutiny is related to the spatial patterns of distribution of employment by industry. The 2001 Spanish Census offers employment figures for 16 types of industries.[21] The results for the Theil index (total and decomposed) according to the industrial classification used in the Census (Table 1.7) show a higher level of internal homogeneity within the analytical regions for all industries except "Agriculture, hunting and forestry activities and fishing" and "Extractive Industries". One simple explanation is that these particular activities are

[21] "Agriculture, Hunting and Forestry Activities" and "Fishing" have been aggregated, so results are shown for only 15 types of industries.

necessarily linked to the physical location of the natural resources at hand - land, forests, rivers or mines - and therefore the chances to *choose* the geographical location of employment or for employment to move *freely* are very limited. In other words, the distribution of these activities does not depend on agglomeration economies but on the location of the natural resources.

For testing the third hypothesis, i.e., the analytical regions are more homogeneous when dealing with different levels of qualification in the labour markets, the classification available at the 2001 Census describes nine groups aggregated by type of work and level of qualification. In this special classification, qualification is understood as the capacity to carry out the tasks which comprise any given job. Therefore, it includes two different facets: level of qualification and specialization within this level qualification. In order to work with a more standard classification, we aggregate these nine groups into “High-Qualified Occupations”, “Medium-Qualified Occupations” and “Low-Qualified Occupations”.[22] The results of the Theil index and its components by level of qualification[23] are shown in Table 1.8.

[22] For details about the nine categories and their aggregation, see Appendix IV.

[23] See Appendix V for the Theil Index results for the nine categories.

Table 1.7. Decomposition of the Theil inequality index by type of industry.

		Administrative Regions			Analytical Regions
		NUTS III (PROV) (50 Regions)	NUTS II (CCAA) (17 Regions)	NUTS I (7 Regions)	MA1 to PRA (8 Regions)
Agriculture. hunting and forestry activities and fishing					
Theil's Index	Between	0.5421	0.4842	0.4005	0.4605
1.2976	Within	0.7555	0.8134	0.8971	0.8371
Extractive Industries					
Theil's Index	Between	1.3393	0.9010	0.6745	0.6457
3.0814	Within	1.7421	2.1803	2.4068	2.4356
Manufacturing					
Theil's Index	Between	0.6453	0.4501	0.3674	1.2409
2.3362	Within	1.6909	1.8861	1.9688	1.0953
Production and Distribution of Energy					
Theil's Index	Between	0.6367	0.5014	0.4287	1.2291
2.6969	Within	2.0602	2.1955	2.2682	1.4678
Construction					
Theil's Index	Between	0.5751	0.4620	0.4008	1.0312
2.0300	Within	1.4548	1.5679	1.6292	0.9988
Minorsalers; Repairs					
Theil's Index	Between	0.6465	0.5082	0.4522	1.3422
2.5604	Within	1.9139	2.0522	2.1082	1.2182
Hotels and Restaurants					
Theil's Index	Between	0.7466	0.6213	0.5418	1.2225
2.5406	Within	1.7941	1.9193	1.9988	1.3182
Transportation, Storage and Communications					
Theil's Index	Between	0.8367	0.6794	0.6217	1.5005
2.9777	Within	2.1410	2.2983	2.3560	1.4772
Financial Intermediation					
Theil's Index	Between	0.8580	0.7099	0.6594	1.6692
3.4331	Within	2.5751	2.7232	2.7738	1.7639
Real State, Rental and Business Services					
Theil's Index	Between	0.9213	0.7451	0.6893	1.7464
3.4331	Within	2.5118	2.6880	2.7438	1.6868
Public Administration and Defense					
Theil's Index	Between	0.6190	0.5128	0.4655	1.3232
2.7992	Within	2.1803	2.2864	2.3337	1.4760
Education					
Theil's Index	Between	0.6177	0.4881	0.4250	1.4343
2.9176	Within	2.3000	2.4296	2.4927	1.4833
Health and Veterinary Activities					
Theil's Index	Between	0.6091	0.4711	0.4160	1.5005
3.0352	Within	2.4261	2.5642	2.6192	1.5347
Other social activies and servicies for households					
Theil's Index	Between	0.7727	0.6229	0.5673	1.4776
2.9100	Within	2.1373	2.2871	2.3427	1.4324
Household's Activities					
Theil's Index	Between	0.9173	0.7598	0.6925	1.5667
3.3296	Within	2.4123	2.5698	2.6371	1.7629

Source: Authors' calculations based on 2001 Spanish Census (INE. 2007).

Table 1.8. Decomposition of the Theil inequality index by level of qualification and occupation.

		Administrative Regions			Analytical Regions
		NUTS III (PROV) (50 Regions)	NUTS II (CCAA) (17 Regions)	NUTS I (7 Regions)	MA1 to PRA (5 Regions)
High Qualified Occupations					
Theil's Index	*Between*	0.6820	0.5421	0.4874	1.4522
2.9777	*Within*	2.2957	2.4356	2.4903	1.5255
Medium Qualified Occupations					
Theil's Index	*Between*	0.6081	0.4725	0.4174	1.1963
2.3306	*Within*	1.7225	1.8581	1.9132	1.1343
Low Qualified Occupations					
Theil's Index	*Between*	0.6168	0.4866	0.4263	1.2032
2.3215	*Within*	1.7047	1.8349	1.8952	1.1183

Source: Authors' calculations based on 2001 Spanish Census (INE. 2007).

For all levels of qualification, the within component is clearly lower than any other administrative division. The differences are slightly more pronounced for the high qualified jobs, where agglomeration economies might play a more effective role.

1.6 Evaluation with the Davies-Bouldin Index

Regarding the first hypothesis - the regional distribution of employment and the distribution of employment by gender - the figures in Table 1.9, using the Davies-Bouldin index, also show better results (lower value of the index) for the analytical regions than for any other type of normative division.

Table 1.9. Davies-Bouldin inequality index by gender.

	Administrative Regions			Analytical Regions
	NUTS III (PROV) (50 Regions)	NUTS II (CCAA) (17 Regions)	NUTS I (7 Regions)	MA1 to PRA (8 Regions)
Total	579.96	3313.01	57.70	36.33
Male	232.58	518.72	40.33	33.22
Female	525.03	189.63	1914.03	20.36

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

Surprisingly enough, when taking into account internal homogeneity and also heterogeneity between regions (the main advantage of the DB index), NUTS I regions have better results than the NUTS II or NUTS III regions. This proves that despite being made up of Autonomous Communities with a higher degree of internal heterogeneity, the NUTS I division "artificially created" for comparison purposes within the European Union at least manages to divide the Spanish territory into seven large areas which are clearly differentiated and reasonably homogeneous: North-West, East, North-East, Madrid, Centre, South and Canary Islands. In any case, however, the results for the eight analytical regions are better.

As expected, we can talk about one labour market for men and another one for women. These are two different labour markets that show a higher level of homogeneity – and heterogeneity between them - when studied separately.

When testing the second hypotheses, the results for the Davies-Bouldin validation index (Table 1.10) confirm a more homogeneous distribution of employment by type industry within the analytical regions (than the administrative regions) and also a higher level of differentiation between the analytical regions.

Table 1.10. Davies-Bouldin inequality index by industry.

	Administrative Regions			Analytical Regions
	NUTS III (PROV) (50 Regions)	NUTS II (CCAA) (17 Regions)	NUTS I (7 Regions)	MA1 to PRA (8 Regions)
Agriculture. hunting and forestry activities and fishing	88.29	105.12	42.08	123.23
Extractive Industries	361.34	59.34	53.41	98.12
Manufacturing	390.49	100.55	25.60	547.64
Production and Distribution of Energy	383.65	420.25	54.29	44.44
Construction	245.92	131.27	23.29	15.21
Minorsalers; Repairs	328.13	2140.30	69.43	23.55
Hotels and Restaurants	450.71	527.21	188.20	20.90
Transportation. Storage and Communications	340.48	159.81	116.33	23.47
Financial Intermediation	857.83	839.29	60.31	10.92
Real State. Rental and Business Services	375.78	107.10	33.13	9.81
Public Administration and Defense	866.40	101.17	20.18	29.09
Education	789.15	117.58	34.01	25.13
Health and Veterinary Activities	433.19	29.45	37.13	29.68
Other social activities and services for households	520.02	437.47	158.37	13.76
Household's Activities	497.93	354.12	29.10	12.67

Source: Authors' calculations based on 2001 Spanish Census (INE. 2007)

For the third hypothesis tested, namely the suitability of the analytical regions for the study of the spatial patterns of distribution of employment by level of qualification and occupation, the results from Table 1.11 confirm those from the Theil index. The analytical regions are a better option for any regional analysis studying labour market issues by level of qualification or occupation.

Table 1.11. Davies-Bouldin inequality index by occupation.

	Administrative Regions			Analytical Regions
	NUTS III (PROV) (50 Regions)	NUTS II (CCAA) (17 Regions)	NUTS I (7 Regions)	MA1 to PRA (8 Regions)
High Skilled Occupations	149.06	503.67	64.94	19.55
Medium Skilled Occupations	681.13	167.34	99.87	28.18
Low Skilled Occupations	330.02	402.27	64.77	33.68

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

1.7 Conclusions

To date, analysis of the spatial dimension of Spanish labour market has been limited to administrative, rather than appropriately-defined functional, geographic units. Alternative divisions of the territory based on the existence of agglomeration economies and the importance of geography have been used in the literature to understand the location of economic activity (Polèse *et al.*, 2007). However, their robustness against the administrative ones commonly used has to date not been evaluated. The objective of this chapter is to prove those functional regions defined under such economic criteria provide better defined regions – more compactness and separation - than the administrative ones commonly used to carry out labour market studies at sub-national level.

Using micro data from the last Spanish Census available, the functional and administrative regions are evaluated using the Theil index and the Davies-Bouldin Validation index. Applied to employment (by gender, industry and level of qualification and occupation), both indexes show better results for the analytical regions than for any of the ordinary administrative ones (NUTS I, II or III regions). The analytical classification generates areas where the distribution of employment is more homogeneous within and more heterogeneous between the regions. Agglomeration economies and distance (to the metropolis) seem to be relevant for understanding the patterns of distribution of employment, either by gender, by industry or by level of qualification and occupation. In practice this means a clearer way for identifying local labour markets and explaining their differences and similarities.

In the light of these results, this chapter suggests that, subject to the availability of data, this alternative classification could be considered when carrying out labour economics studies that include a *spatial* dimension. The use of this classification can offer a better understanding of the job opportunities, location of industries, concentration of unemployment, occupations and many other labour related topics. Surpassing the administrative division of the territory, this classification manages to have explanatory power in spatial labour economics topics while including geo-economic characteristics as relevant as location and agglomeration economies.

Chapter 2

An analysis of urban size and territorial location effects on employment probabilities: the Spanish case

Objectives

Using the regional classification described in Chapter 1, the purpose of this chapter is to analyze the relevance of urban size and location on the probability of being employed in the Spanish economy

Methodology

Logit model applied to each type of functional region and differentiating between male and female.

Synopsis

The probability of being employed varies depending on several factors. Many of these are related to personal characteristics such as educational level, age, gender or number and age of children. Nevertheless, other factors may be relevant, in particular the geographical environment. Using the classification described in Chapter 1, the spatial patterns of employment distribution are analysed. Our results show some relevant differences between the functional regions defined: municipalities with similar sizes and located at similar distance from a metropolis but belonging to different Autonomous Communities or provinces share similar employability patterns.

2. An Analysis of Urban Size and Territorial Location Effects on Employment Probabilities: the Spanish Case

2.1 Introduction

Most economic theories of labour markets have been developed and tested at the national level and it is generally accepted in the literature that the main conclusions at this level basically hold when aggregating to the supra-national level or disaggregating to regional and local levels. Nevertheless, it seems obvious that the greater the degree of spatial disaggregation, the higher the possibilities of uncovering different behaviours. This might occur because, at a very local level, the economies could be more specialised and more susceptible to specific local policies or geographical factors. This means that some of the generally-accepted conclusions in labour economics, as well as in other fields of economic analysis, may not hold when applied to a certain local area. This in turn might help in understanding why many local and regional policies do not work as well as was expected or, in certain cases, even harmful. A study of regional or local characteristics in terms of geographical factors or industrial structure is necessary before applying general recipes.

In this chapter, we are interested in identifying the existence of certain types of regularities across space. It is frequently thought that the differences between local and general behaviours are completely explained by local specificities that may not be relevant for other areas. However, it is of interest to identify regularities that may apply to some other particular cases, with all the necessary care that should be taken with inter-cultural comparisons.

We centre our research on the probability of being employed for different types of workers, checking whether the most successful profiles are independent of the spatial region for which the analysis applies. This is especially relevant for two reasons:

First, most of the conclusions about which profile will be more successful and which type of employee will have greater problems to find a job come from empirical approaches that, because of the lack of data at local level, normally use data at a national level.

Secondly, the focus of many active employment policies is on relevant factors that explain the higher chances to be employed at national level on the assumption that similar factors are relevant at the local or regional level. However, in many cases, the national determinants may not be really appropriate at local level.

The main problem when carrying out this type of analysis is that when studying local data, the probabilities of getting lost in the peculiarities of each regional or local area are high. On the other hand, if we aggregate the local data according to the usual regional classifications, i.e., politico-administrative regions that may not be relevant from a purely economic point of view, we may not be able to achieve useful conclusions. The key question, therefore, is how to classify and aggregate the local data in order to be able to draw general conclusions. The solution is to use a classification that is relevant from a regional and urban economics point of view, as the analytical regions described in Chapter 1. Thus, the employability analysis proposed in this chapter will be applied to each of those eight regions into which Spanish territory is divided and where no mention whatsoever is made of the politico-administrative regional frontiers commonly used. The interpretation of the differences will provide us information on the relevance of size (*agglomeration economies)* and distance to the metropolis (*location)* for the probabilities of being employed.

The rest of the chapter is developed as follows. In the next section we review the main literature and results on employability and the different types of regional data used therein. A brief description of the micro-data used for this chapter is provided in Section 3. The main hypothesis and the empirical model applied to the study of employability by type of region can be found in Section 4. The more relevant results and regional regularities are summarised in Section 5, and the main conclusions are presented in the final section.

2.2 What Do We Know about Employability at Regional Level?

One of the targets of any government is to achieve a high, sustained level of employment. Therefore, the study of the determining factors of employment is an area of great interest for many economists. Regional economics examines the interaction between employment and the economy in general, exploring some of the sub-national characteristics that may result in significant differences in levels and growth rates of employment across space.

While very little has been done on the spatial dimension in labour market outcomes, there are plenty of studies that analyze the groups that suffer the highest risk of being unemployed. These typically use as their main sources the Labour Force Survey elaborated by individual countries or in the European case, the Living Conditions Survey or the European Household Panel, all of which are representative statistics at national or regional levels (NUTS II or NUTS III division).

The scarcity of disaggregated information beyond the politico-administrative regions limits our possibilities for performing any analysis where the size of the geographical units used could be taken into account. Sometimes, this inconvenience is bridged when the data source yields information about the *condition* of the unit (metropolitan, urban, rural), as is the case of the National Longitudinal Surveys or the Panel Study of Income Dynamics, both in the United States (used among others by Glaeser and Maré, 2001). On other occasions, administrative registered data are used, which includes disaggregated information in geographical terms but presents some problems related to the fact of not having been elaborated for research purposes (Alonso and Del Rio, 2007; Alonso *et al*., 2009).

The particularities of the Spanish labour market make it a very interesting case study. For a long time the aggregate unemployment rate in Spain was the highest of the European countries, leading to a great number of studies on the explanatory factors for this (see, for example, Bentolila and Blanchard, 1990; Blanchard and Jimeno, 1995; Dolado and Jimeno, 1997, among others). In addition to the persistence of unemployment, another special characteristic of the Spanish case is the existence of regional disparities that may indeed reflect the existence of regional employment markets, i.e. differentiated spatial behaviours in response to changes in labour activity (Decressin and Fatás, 1995; Jimeno and Bentolila, 1998; Bande *et al*., 2008). For example, López-Bazo *et al*. (2005) conclude that in the 1980s such differences were explained mainly by the industrial mix and wages, while in the 1990s the differences across *provinces* (NUTS III) in amenities explain the regional dispersion of the employment rates.

Overall, the conclusions of these and other studies highlight the specificity of regional labour markets. The question remains as to whether there may be some common factors that contribute to the explanation of the behaviour of these markets, such as for example the size of the regional units being considered or their spatial location. One of the main contributions of this study is that, using disaggregated data at local level, we depart from the traditional political-administrative definition of regions (NUTS II or III) and instead use a new territorial classification of municipalities which is more economically meaningful (in the sense that has been explained in the previous section). Our objective is to use these alternative *types of regions* to analyze the differences in the probability of being employed in Spain.

2.3 Database: the Spanish Census

One of the major problems of this new approach is the availability of suitable data at local level in order to aggregate it into the alternative types of regions. In the Spanish case, one of the few databases that includes individual information on the main variables needed to study employability[24] as well as information at municipal level is the 2001 Spanish Census, administered by the INE (National Statistics Institute of Spain). Although there are partial updatings every three years, the complete database is only available every ten years. The last two full Spanish censuses available are for 1991 and 2001. For the purpose of this study we shall use the data from 2001.

The database is based on 5% of the total population living in Spain in 2001 according to the Census. As our purpose is to study the influence of the type of region in which a person lives (metropolitan area, urban vs. rural, central vs. peripheral) on the chances of being employed, from the initial sample all individuals under the age of 16 and over 65 were removed. This leaves us with a database that includes 1,374,612 working-age individuals, 7.1% of which are foreigners. For each person we have the following information: age, gender, highest level of studies achieved, marital status (single/married/separated-divorcee/ widow-widower), labour situation (employed/ unemployed/ inactive), number of children above/below the age of four, and place of birth (domestic/foreign). In our sample, 53.9% of the individuals were working, 8.9% were searching for a job, and the total active

[24] The variables will be discussed in more detail in the next section.

population was 863,786. The remainder (36.9%) are individuals not economically active.

As regards the spatial units, the Census includes information on the politico-administrative regions where the individuals lived in 2001 (seventeen Autonomous Communities or NUTS II regions, and fifty provinces or NUTS III regions). The great advantage of the 2001 Spanish Census is that it also includes information on the *municipalities* into which provinces are divided. The number of municipalities in the provinces ranges from 35 to 370, with a total figure of 8,106 municipalities in 2001. Thus, for each individual it is possible to know the municipality of residence and therefore the *type of region* where that person lived (MA1, MA2, CUA1, CUA2, CRA, PUA1, PUA2 or PRA). According to our sample, in 2001 38.5% of the active population had their residence in a metropolitan area (MA1 and MA2), 24.6% lived in central areas (CUA1, CUA2 and CRA) and the remainder (36.9%) lived in the periphery.

Table 2.1 shows more details on the distribution of the sample (active population) by type of region and also by gender, age, marital status, educational level, place of birth and parenthood. While foreigners and individuals with university studies tend to concentrate in the metropolitan areas, individuals with no studies or basic studies are located mainly in small and peripheral regions.

Table 2.1. Descriptive statistics of the variables.

	MA1		MA2		CUA1		CUA2		CRA		PUA1		PUA2		PRA		TOTAL	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
EDUCATIONAL LEVEL																		
No Studies	0.070	0.255	0.072	0.259	0.061	0.240	0.089	0.285	0.083	0.276	0.065	0.246	0.100	0.300	0.113	0.317	0.082	0.275
Basic Studies	0.432	0.495	0.480	0.500	0.494	0.500	0.538	0.499	0.565	0.496	0.469	0.499	0.550	0.497	0.603	0.489	0.508	0.500
High School	0.166	0.372	0.144	0.352	0.136	0.343	0.126	0.332	0.114	0.318	0.159	0.365	0.128	0.334	0.103	0.303	0.138	0.345
Technical Training	0.119	0.324	0.124	0.330	0.139	0.346	0.115	0.319	0.120	0.325	0.113	0.317	0.094	0.292	0.087	0.282	0.112	0.316
University Studies	0.213	0.410	0.179	0.383	0.170	0.375	0.132	0.339	0.118	0.323	0.194	0.396	0.128	0.334	0.094	0.292	0.159	0.366
GENDER																		
Male	0.492	0.500	0.494	0.500	0.494	0.500	0.505	0.500	0.519	0.500	0.491	0.500	0.504	0.500	0.523	0.499	0.502	0.500
Female	0.508	0.500	0.506	0.500	0.506	0.500	0.495	0.500	0.481	0.500	0.509	0.500	0.496	0.500	0.477	0.499	0.498	0.500
AGE																		
<24	0.180	0.384	0.194	0.395	0.189	0.392	0.198	0.398	0.180	0.384	0.196	0.397	0.201	0.401	0.183	0.387	0.190	0.392
25-29	0.135	0.342	0.128	0.334	0.130	0.336	0.128	0.334	0.123	0.328	0.128	0.334	0.125	0.331	0.117	0.322	0.127	0.333
30-44	0.347	0.476	0.348	0.476	0.347	0.476	0.357	0.479	0.364	0.481	0.347	0.476	0.362	0.481	0.354	0.478	0.352	0.478
>45	0.338	0.473	0.330	0.470	0.334	0.472	0.317	0.465	0.333	0.471	0.329	0.470	0.312	0.463	0.346	0.476	0.331	0.471
PLACE OF BIRTH																		
Local	0.902	0.297	0.942	0.233	0.932	0.252	0.926	0.262	0.940	0.237	0.935	0.246	0.928	0.258	0.946	0.226	0.929	0.256
Foreigner	0.098	0.297	0.058	0.233	0.068	0.252	0.074	0.262	0.060	0.237	0.065	0.246	0.072	0.258	0.054	0.226	0.071	0.256
MARITAL STATUS																		
Single	0.407	0.491	0.399	0.490	0.401	0.490	0.369	0.483	0.367	0.482	0.416	0.493	0.382	0.486	0.385	0.487	0.393	0.488
Married	0.525	0.499	0.538	0.499	0.534	0.499	0.574	0.494	0.581	0.493	0.518	0.500	0.564	0.496	0.573	0.495	0.548	0.498
Separated/Divorcee	0.048	0.213	0.042	0.200	0.045	0.206	0.038	0.190	0.033	0.179	0.046	0.209	0.033	0.180	0.022	0.148	0.039	0.193
Widow	0.019	0.138	0.021	0.145	0.021	0.142	0.019	0.137	0.019	0.136	0.021	0.142	0.020	0.141	0.020	0.139	0.020	0.140
CHILDREN																		
With children	0.156	0.363	0.126	0.332	0.135	0.342	0.143	0.350	0.152	0.359	0.115	0.319	0.123	0.329	0.129	0.336	0.136	0.343
No children	0.844	0.363	0.874	0.332	0.865	0.342	0.857	0.350	0.848	0.359	0.885	0.319	0.877	0.329	0.871	0.336	0.864	0.343
With children under 4	0.877	0.328	0.877	0.328	0.874	0.332	0.858	0.350	0.861	0.346	0.881	0.324	0.863	0.344	0.878	0.327	0.872	0.334
No children under 4	0.123	0.328	0.123	0.328	0.126	0.332	0.142	0.350	0.139	0.346	0.119	0.324	0.137	0.344	0.122	0.327	0.128	0.334
With children >4	0.231	0.422	0.194	0.395	0.208	0.406	0.221	0.415	0.230	0.421	0.182	0.386	0.195	0.396	0.191	0.393	0.207	0.405
No children > 4	0.769	0.422	0.806	0.395	0.792	0.406	0.779	0.415	0.770	0.421	0.818	0.386	0.805	0.396	0.809	0.393	0.793	0.405
N	**187.282**		**141.157**		**38.396**		**100.942**		**77.034**		**98.104**		**101.803**		**119.068**		**863.786**	

2.4 Hypothesis and Method

Our objective is to study the probability of being employed as a function of the specific profile of each individual, and in particular of the type of region where that person resides. To do so, we use a logit model in which the dependent variable is the probability of being employed, *p*, which takes value 1 if the individual is employed and value 0 if the individual is unemployed.

The model is specified as follows:

$$y_t = \frac{1}{1+e^{-(X'_t\beta+\delta t+\varepsilon_t)}} \tag{2.1}$$

with linear transformation:

$$log\,it(y_t) = ln\frac{y_t}{1-y_t} = X'_t\beta + \delta t + \varepsilon_t \tag{2.2}$$

The independent variables, X_i, are those commonly used in studies of individuals' probabilities of being employed: age (AGE); highest educational level attained (*EDL*); marital status (MS); a variable capturing whether the individual was born outside of Spain (FOR); variables capturing whether the individual has children under or above the age of 4 (CHBA4 and CHAA4 respectively); and variables indicating the type of Region where he/she lives (MA1, MA2, CUA1, CUA2, CRA, PUA1, PUA2, PRA).[25]

Given the well-know differences in behaviour of men and women in the labour market, we have estimated separate models for each. As we will see, the results show significant differences in their employability probabilities and also differences according to the type of region.

[25] See Appendix VI for details.

Specifically, our interest is to test the following hypotheses:

H1: the probability of being employed increases with size

This hypothesis suggests that the concentration of economic activity and the agglomeration economies generated in the largest metropolitan areas increase the opportunities for matching labour supply and demand and therefore the probabilities of being employed, whatever the level of studies or any other characteristics of the worker. This increase should be clearer in those activities which are highly sensitive to agglomeration economies which tend to demand more qualified or creative jobs.

However, this hypothesis does not mean that only larger cities can benefit from agglomeration economies. In keeping with the idea of Phelps *el al.* (2001) regarding borrowed size and with empirical results from both Canada (Polèse and Shearmur, 2004) and Spain (Polèse *et al.*, 2007, and Rubiera, 2006), it is the areas within *and also close to* larger cities that will benefit from these advantages. In other words, proximity to size also matters.

H2: the probability of being employed diminishes with distance from the metropolis

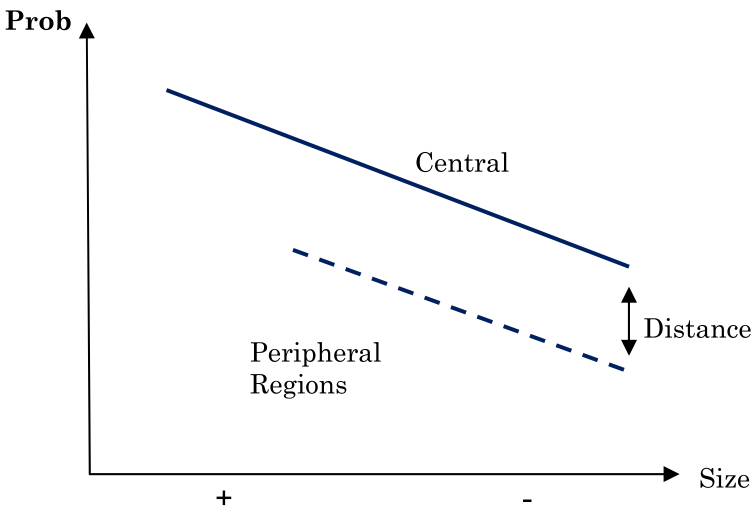
The economic structure and growth of an area are clearly affected by its proximity to a large metropolis. Areas close to the metropolis (i.e., "central" areas as opposed to "peripheral" areas) can benefit from agglomeration economies linked to the concentration of firms belonging to the same industry or from the mere concentration of firms in the metropolitan areas. Such benefits include, among others, knowledge spillovers, a pool of skilled labour, the diffusion of innovations, or the existence of infrastructure such as international airports and highways depending upon a large local market.[26]

[26] There are agglomeration economies linked to the co-location of many firms within the same industry (localization economies). These economies may be related to a shared labour-

However, the price of land and the higher salaries usually paid in the biggest cities expel some spatially-intensive activities less sensitive to agglomeration economies, especially industrial activities, to other locations that nevertheless remain close to the metropolis (main market). For this reason, urban and rural areas located close to big cities are expected to show higher chances of being employed for individuals with technical training.

These two hypotheses are graphically illustrated in Figure 2.1.

Figure 2.1. Probabilities of Being Employed by Type of Region: Size and Distance Effect.



force, knowledge *spillovers*, rapid diffusion of innovations, and stimulation due to the competition between firms (see Rosenthal and Strange, 2001; Porter, 1990; Beardsell and Henderson, 1999). There are also economies linked to the co-location of many diverse activities (urbanization economies). Infrastructure such as international airports and highways depend upon a large local market, as do schools, universities and cultural activities. In addition, the presence of a diversity of economic sectors may stimulate the cross-over of ideas, leading to innovations or even to new economic activities (see Jacobs, 1984; Quigley, 1998). For more details about the different types of *agglomeration economies* see Phelps and Ozawa (2003).

Thus, regardless of the rest of explanatory variables, when ordering the types of regions in descending order of size (MA1, MA2, CUA1 and PUA1, CUA2 and PUA2, CRA and PRA), these two hypotheses imply: i) a downward-sloping curve showing lower probabilities of being employed with the size of the region, and ii) the existence of a gap between the "central" type of regions and the "peripheral" ones with the same size.

2.5 Main Results

In this section, we present the main results of the econometric model proposed for the eight types of regions in which the Spanish territory is divided. Separate estimations are made for men and women, as is common in the literature. The basic econometric results are presented in Table 2.2 and Table 2.3. The signs and values of the coefficients for the level of studies and the place of birth are the ones expected at the 1% level of significance, while age does not show a clear pattern by type of region.

Table 2.2. Determinants of Employment by Type of Region for Men.

	MA1			MA2			CUA1			CUA2			CRA			PUA1			PUA2			PRA		
	Coef.		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**
EDUCATIONAL LEVE																								
No Studies	-0.717	***	0.055	-1.019	***	0.054	-0.907	***	0.117	-1.050	***	0.071	-1.285	***	0.085	-0.990	***	0.065	-1.104	***	0.063	-1.193	***	0.065
Basic Studies	-0.243	***	0.037	-0.352	***	0.038	-0.197	**	0.081	-0.320	***	0.056	-0.251	***	0.070	-0.278	***	0.045	-0.270	***	0.051	-0.429	***	0.055
Technical Training	0.201	***	0.048	0.200	***	0.048	0.375	***	0.101	0.300	***	0.074	0.415	***	0.090	0.360	***	0.060	0.270	***	0.070	0.083		0.071
University Studies	0.226	***	0.045	0.320	***	0.049	0.538	***	0.109	0.160	**	0.076	0.275	***	0.098	0.468	***	0.059	0.264	***	0.071	-0.012		0.077
ref: High School																								
AGE																								
< 25	-0.463	***	0.039	-0.306	***	0.036	-0.275	***	0.077	-0.385	***	0.051	-0.483	***	0.063	-0.344	***	0.047	-0.286	***	0.047	-0.390	***	0.045
30-44	-0.096	**	0.039	-0.020		0.036	0.130	*	0.078	-0.055		0.052	0.073		0.064	-0.064		0.046	0.013		0.047	0.151	***	0.043
> 45	-0.339	***	0.047	-0.041		0.046	0.004		0.099	-0.169	***	0.063	-0.060		0.074	-0.026		0.058	-0.048		0.057	-0.007		0.050
ref: 25-30																								
PLACE OF BIRTH																								
Foreigner	-0.563	***	0.040	-0.417	***	0.051	-0.327	***	0.103	-0.446	***	0.062	-0.440	***	0.082	-0.340	***	0.060	-0.495	***	0.055	-0.491	***	0.058
MARITAL STATUS																								
Married	1.242	***	0.036	1.237	***	0.035	1.209	***	0.076	1.101	***	0.048	0.736	***	0.055	1.234	***	0.043	1.053	***	0.042	0.716	***	0.037
Separated/Divorce	0.526	***	0.084	0.448	***	0.091	0.627	***	0.208	0.197		0.121	0.272	*	0.158	0.452	***	0.107	0.324	***	0.124	0.112		0.138
Widow	0.776	***	0.173	0.863	***	0.174	0.979	**	0.380	0.686	***	0.237	0.465	*	0.263	0.804	***	0.215	0.876	***	0.241	0.553	***	0.198
ref: Single																								
CHILDREN																								
Children < 4	0.058		0.045	0.008		0.042	-0.086		0.091	-0.063		0.054	0.050		0.065	0.132	**	0.055	-0.030		0.049	-0.005		0.047
Children > 5	-0.066	**	0.033	-0.158	***	0.035	-0.170	**	0.074	-0.046		0.045	0.037		0.052	-0.037		0.044	0.003		0.042	0.114	***	0.039
Constant	2.076	***	0.051	1.644	***	0.053	1.760	***	0.110	2.138	***	0.074	2.245	***	0.091	1.568	***	0.065	1.819	***	0.068	2.056	***	0.070

: significance at 10%; **: significance at 5%; *: significance at 1%.*

Table 2.3. Determinants of Employment by Type of Region for Women.

	MA1			MA2			CUA1			CUA2			CRA			PUA1			PUA2			PRA		
	Coef.		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**	**Coef.**		**S.E.**
EDUCATIONAL LEVEL																								
No Studies	-0.597	***	0.056	-0.705	***	0.061	-0.711	***	0.136	-0.866	***	0.075	-1.678	***	0.083	-0.707	***	0.073	-0.817	***	0.068	-1.137	***	0.065
Basic Studies	-0.463	***	0.033	-0.488	***	0.036	-0.460	***	0.079	-0.526	***	0.049	-0.606	***	0.061	-0.421	***	0.043	-0.301	***	0.046	-0.453	***	0.048
Technical Training	-0.104	**	0.041	-0.123	***	0.042	-0.073		0.091	-0.146	**	0.058	-0.093		0.073	-0.115	**	0.051	-0.108	*	0.057	-0.044		0.059
University Studies	0.405	***	0.038	0.396	***	0.042	0.430	***	0.093	0.339	***	0.060	0.265	***	0.076	0.321	***	0.049	0.401	***	0.057	0.197	***	0.058
ref: High School																								
AGE																								
< 25	-0.316	***	0.040	-0.346	***	0.037	-0.349	***	0.078	-0.231	***	0.047	-0.309	***	0.060	-0.311	***	0.046	-0.236	***	0.046	-0.379	***	0.044
30-44	0.053		0.035	0.206	***	0.034	0.153	**	0.072	0.142	***	0.044	0.222	***	0.053	0.221	***	0.042	0.219	***	0.042	0.267	***	0.040
> 45	0.210	***	0.041	0.532	***	0.043	0.446	***	0.090	0.424	***	0.056	0.520	***	0.065	0.556	***	0.052	0.571	***	0.054	0.672	***	0.050
ref: 25-30																								
PLACE OF BIRTH																								
Foreigner	-0.098	**	0.038	-0.212	***	0.049	-0.169	*	0.096	-0.113	*	0.060	-0.143	*	0.080	-0.211	***	0.055	-0.147	**	0.056	-0.096		0.061
MARITAL STATUS																								
Married	0.068	**	0.031	0.271	***	0.032	0.363	***	0.067	0.400	***	0.042	0.090	*	0.051	0.290	***	0.038	0.368	***	0.040	0.178	***	0.038
Separated/Divorce	0.167	***	0.051	0.178	***	0.052	0.241	**	0.109	0.138	**	0.070	0.073		0.094	0.217	***	0.061	0.153	**	0.070	-0.023		0.082
Widow	0.607	***	0.104	0.588	***	0.103	0.922	***	0.257	0.912	***	0.157	0.433	***	0.163	0.581	***	0.124	0.767	***	0.139	0.674	***	0.140
ref: Single																								
CHILDREN																								
Children < 4	-0.476	***	0.032	-0.438	***	0.034	-0.474	***	0.070	-0.569	***	0.041	-0.477	***	0.048	-0.415	***	0.041	-0.404	***	0.040	-0.495	***	0.040
Children > 5	-0.178	***	0.028	-0.152	***	0.030	0.027		0.061	-0.062	*	0.037	-0.067		0.043	-0.119	***	0.038	0.019		0.037	-0.020		0.036
Constant	1.979	***	0.044	1.281	***	0.048	1.491	***	0.101	1.570	***	0.063	1.842	***	0.078	1.322	***	0.058	1.206	***	0.061	1.355	***	0.061

: significance at 10%; **: significance at 5%; *: significance at 1%.*

However, our interest is not in analysing the determinants of employment but the *probability* of being employed and how this probability is affected by the urban size and the position of the type of region where the individuals live. Thus, from these results we calculate the probabilities of being employed for the 640 different plausible profiles of a worker (four classes of civil status, four groups of age, five different level of studies, being a national or a foreigner, having or not having children above the age of four, and having or not having children below the age of four). For each profile there are different probabilities depending on the gender and also on the type of region.

Given the variety of profiles and therefore of results, several comparisons could be made such as the employability of a foreigner versus a national, or a man versus a woman (other things being equal). We could also analyse whether marital status has any effect on the chances of being employed or whether technical training studies are a better way to get into the labour market than holding a University degree. However, the aim of this chapter is to explore the existence of any location patterns of employment according to a definition of regions that includes *size* and *location* characteristics. In particular, our hypotheses are that given the existence of *agglomeration economies*, the probabilities of being employed are expected to increase with *size* and to decrease with *distance* from the metropolis. Therefore we shall choose one profile and then vary only the region of residence of the individuals. In Figure 2.2, we have represented the probabilities of being employed for one of the most common profiles: a non-foreign married individual with no children, aged between 30 and 44 and with university studies.

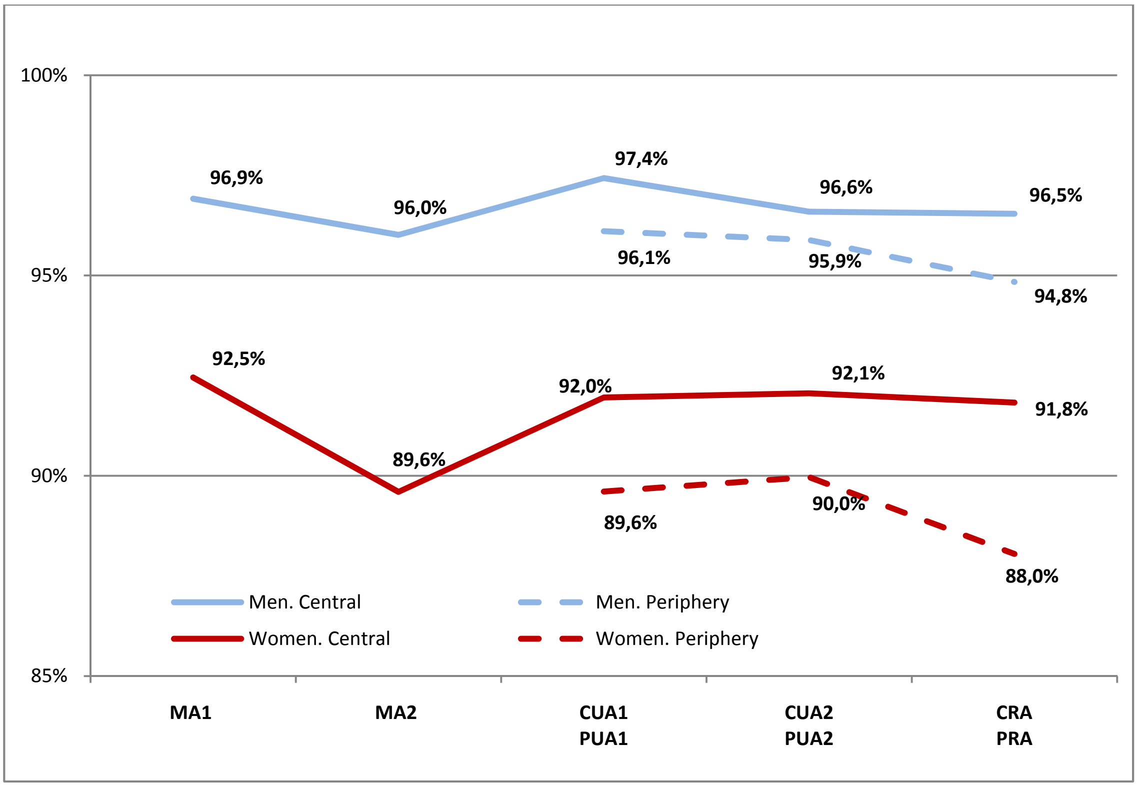
As expected, there is a significant gap between central and peripheral areas in terms of the chances of being employed, confirming the *distance hypothesis* (higher concentration of employment in the central regions), at least for jobs requiring higher qualifications. However, in no case does the gap exceed 2% for men or 3% for women.

As regards the *size hypothesis*, the positive relationship between employment and size is confirmed but it does not seem to be very strong for this particular profile (individual with university studies). In the largest metropolitan areas (MA1), a man with university studies has an almost 97% chance of being employed, while in a *rural* area the probability falls to 96.5% (central) and 94.8% (peripheral). A similar pattern can be observed for woman but with a difference in levels, a very well-known characteristic in the Spanish labour market.[27]

[27] Although it has decreased in recent years, in Spain the employment gap between men and woman is around 15 points.

Figure 2.2. Probability of being employed by gender and region.

(Profile: married individual, age 30-44, non-foreigner, no children)



Of particular note are the metropolitan areas with less than 500,000 inhabitants (MA2), where there seems to be a higher concentration of unemployment for people with a University degree. We might ask whether the exceptional behaviour of the MA2 is exclusive to individuals with higher levels of studies. In the following graphs (Figure 2.3) we can observe that regardless of the level of education, the chances of being employed in metropolitan areas 2 are always lower than in its *bigger* or *smaller* sisters, namely the metropolitan areas 1 (MA1) and central urban areas 1 (CUA1). There are two plausible explanations. First, these areas have long suffered from unemployment as they are areas specialised mainly in traditional industries that have undergone intense de-industrialization in recent decades. Second, in terms of size these metropolitan areas might not be large enough so as to fully take advantage of the benefits associated with agglomeration economies or to compete against the bigger metropolitan areas, while at the same time they suffer the negative effects (wage and land price increase, congestion etc.) associated with large metropolises.

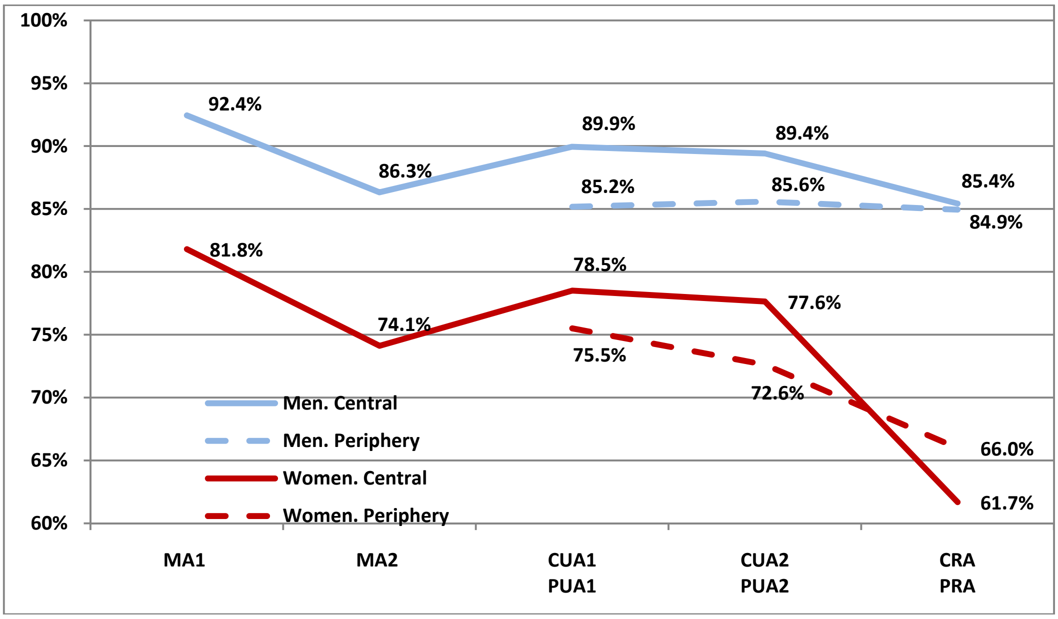
Nevertheless, as high value-added activities, and therefore highly qualified jobs, are expected to be concentrated in the central areas, this first overview based on the profile of an individual with university studies may be biased. Obviously, the level of studies should have a differential impact in any comparison between central areas and peripheral ones. This effect can be observed in Figure 2.3, where we represent the probability of being employed by type of region for the same profile as the one presented above (married, age 30-44, non-foreigner, no children), but modifying his/her level of studies. As expected, the higher the level of studies, the higher the probability of being employed in any type of region. This result holds for both women and for men.

It is worth pointing out that the probability of being employed for somebody holding a university degree versus an individual with technical training is slightly higher in metropolitan areas (MA1 and MA2) and urban areas of a certain size, either central or peripheral (CUA1 and PUA1). However, the contrary is true for the least populated regions (those below 100,000 inhabitants, i.e. Urban Areas 2 and Rural Areas), regardless of whether they are considered central or peripheral. In other words, when searching for a job at national level, the chances of being employed for somebody with technical studies are higher than the equivalent for an individual with university degree in *small* regions, either central or peripheral.

Figure 2.3. Probability of being employed by region, gender and level of studies.

(Typology: married individual, age 30-44, non-foreigner, no kids)

No studies



Basic studies

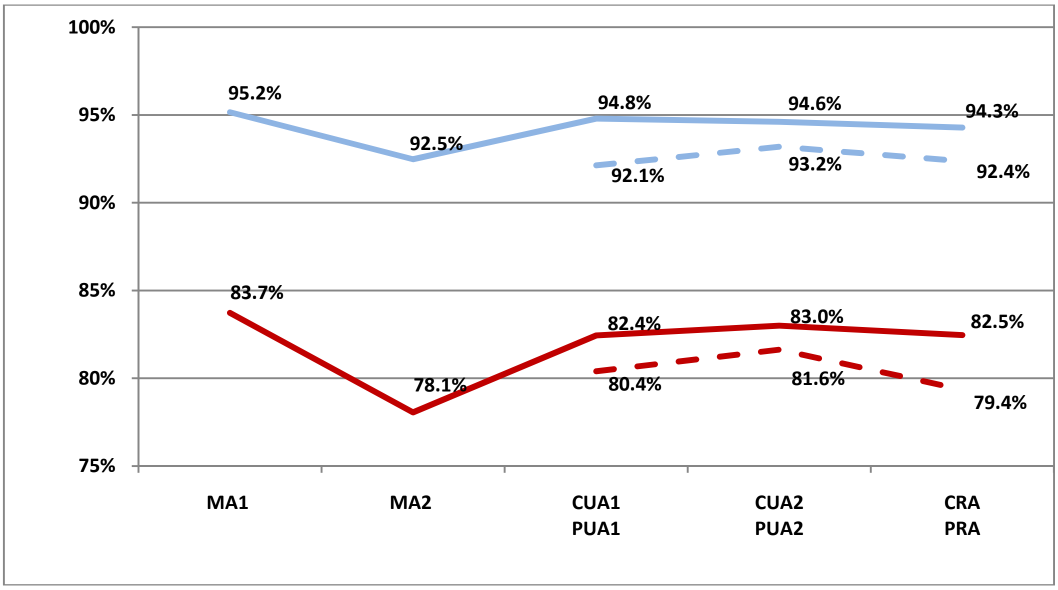
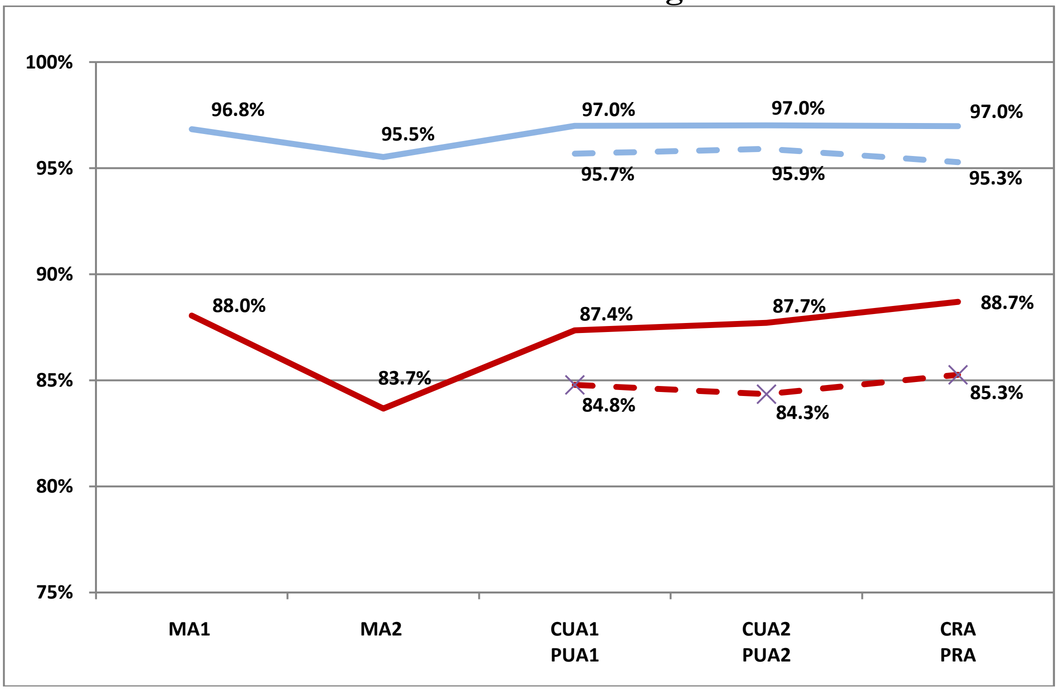


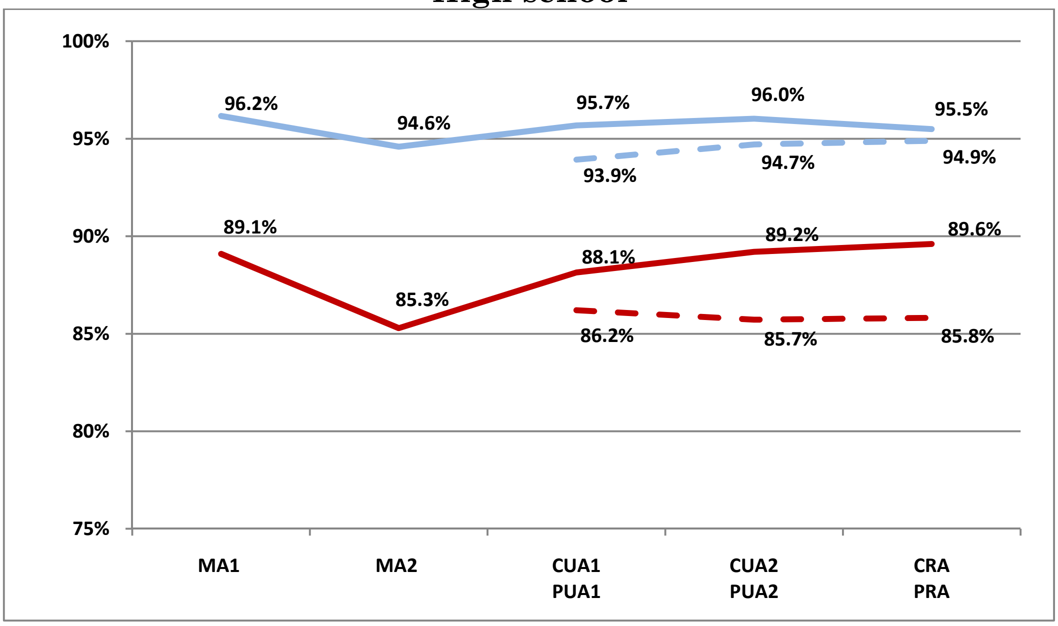
Figure 2.3b. Probability of being employed by region, gender and level of studies.

(Typology: married individual, age 30-44, non-foreigner, no kids)

Technical training



High school



On the other hand, the expected negative relationship between employment and the size of the region where the individual lives seems to be stronger for those with no studies or basic studies. In other words, the size of the region does not seem to have a great impact on employability unless non-qualified jobs are sought; in this case, a person has more chances to find a job in larger rather than smaller regions. This might suggest a higher concentration of non-qualified workers in small and medium-sized regions, which increases the competition for the few non-qualified jobs available. In Table 2.4 we present the concentration index of the level of education for each type of region. As we can see, the previous suggestion is confirmed by the data.[28]

Table 2.4. Index of Spatial Concentration of Educational Levels by Region.

	CENTRAL REGIONS					PERIPHERAL REGIONS		
	MA1	**MA2**	**CUA1**	**CUA2**	**CRA**	**PUA1**	**PUA2**	**PRA**
No Studies	0.85	0.88	0.75	1.08	1.01	0.79	1.22	1.38
Basic Studies	0.85	0.94	0.97	1.06	1.11	0.92	1.08	1.19
High School	1.20	1.05	0.99	0.91	0.83	1.15	0.93	0.74
Technical Training	1.06	1.11	1.23	1.02	1.07	1.01	0.84	0.78
University Studies	1.34	1.12	1.06	0.83	0.74	1.22	0.80	0.59

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

By type of region, we can observe that for all levels of studies the differences between the probabilities of being employed are much less pronounced in metropolitan areas 1 (MA1) than in the remaining regions. In MA1 the gap between a male with university studies and a male with no studies barely exceeds 4 points (96.9% and 92.4% respectively), while in the remaining smaller-sized regions the gap always exceeds 10%. This result confirms the importance of agglomeration economies, as there is a higher

[28] Note that for individuals with University studies, the probabilities of being employed increase with the size of the type of region where they live, even given their higher level of concentration in those areas.

concentration of jobs (of any type) in the largest region considered (metropolitan areas with more than 2.5 million inhabitants).

The employability rate between UA1 and UA2 (urban areas with populations between 100,001 and 500,000 and 50,001 and 100,000 respectively), located in either a central or a peripheral location, does not vary significantly and does not show a clear tendency. This appears to suggest that for those regions of insufficient size to be considered a "metropolitan area", size is not as relevant as its central-peripheral location. In other words, for an individual living in a medium-sized type of region, the chances of getting a job seem to be affected by the distance of the region to a metropolis but not by its relative size. As expected, the *distance hypothesis* (the chances of being employed are much higher the closer the region where the individual lives is to a metropolitan area) is confirmed by the results. This holds true for all levels of studies.

For rural areas (less than 50,000 inhabitants), both size and distance seem to be relevant when evaluating employability, i.e., regardless of the level of studies, there are fewer chances of being employed in rural areas as opposed to urban ones and in peripheral rural areas as opposed to central ones.

Finally, some remarks are in order on the contribution of this chapter with respect to the spatial differences in employment rates. Researchers do not deny the importance of territory and the role played by the regions. However, using the politico-administrative division it is difficult to explain why the rates are different (industrial distribution of economic activity, culture, etc.) or to determine which policies can be effective at local-regional level.

Our results underline the relevance for employment of the size and location of the region where individuals live. Using these (economic) regions the differences across space can be explained by the existence of agglomeration economies and location. In this sense, similar people living in the same (politico-administrative) region could have different probabilities of being employed depending on the municipality in which they live (rural or urban, central or peripheral).

Our results suggest that any government trying to implement a successful employment policy at national or even local level should design several policies addressed to different social groups as well as geographical areas other than the administratively-defined regions used to date. In other words, as well as it is accepted that employment policies for women have necessarily to differ from those designed for men, or that measures to promote hiring young people and first-time job seekers are not the same as those for adults over 45, a new regional approach taking into account characteristics as the size and location of the regions should become part of the policy agenda.

2.6 Conclusions

It is generally accepted that the conclusions of the economic studies carried out at either national or regional levels can be applied at a local level. However, this is not necessarily true. In this chapter, we have presented a spatial analysis of employment at local level where, among other factors, the demographic and geographical characteristics can and do affect the results.

Applying the novel regional classification suggested in Polèse *et al.* (2007) for the Spanish case, the empirical results support the hypothesis that the *size* –in terms of population - and the *location* – in terms of distance to a metropolis - are explanatory variables for the probability of being employed. In other words, employment depends on the personal characteristics of the individuals (level of education, age, sex, etc.) but also on the *type of region* (as defined in Polèse *et al.* 2007) where they live.

Assessing the importance of location, our results show a significant gap in the chances of being employed between "central" and "peripheral" types of regions, i.e., the closer the region is to the metropolis, the higher the concentration of economic activity and therefore employment. Likewise, the expected negative relationship between employment and the *size* of the region where the individual lives is confirmed, and this seems to be stronger for non-skilled individuals than for people with university studies.

In terms of employability, the largest Spanish metropolitan areas (MA1) seem to be enjoying the full benefits of agglomeration economies while the smaller metropolitan areas (MA2) seem to be suffering their negative effects. Likewise, for urban areas that cannot be considered "metropolitan areas" (UA1 and UA2), size does not seem to be as important as their central-peripheral location. In rural areas (less than 50,000 inhabitants), both size and distance seem to be relevant determinants of employability. In other words, regardless of the level of studies, there are fewer chances of being employed in rural areas as opposed to urban areas and in peripheral rural areas as opposed to central rural areas.

Recognizing the importance of these spatially differentiated results would have a significant impact on current policy discussions, shifting the focus from general solutions to more spatially customized ones where size and location are considered. As well as differences in ages, gender or industrial structure are taken into account in the design of employment policies (at national or local level), our results suggest that an additional *spatial dimension* that somehow includes the size and location of the local area where the person lives should be considered.

Chapter 3

From the periphery to the core: direct and indirect effects of the migration of labour

Objectives

Measuring the direct and indirect effects caused by the arrival of workers from the periphery to the core; measuring the displacement of population and jobs among the rest of the regions and identifying some patterns of distribution of jobs and residence.

Methodology

Multiregional input-output model of migrations that includes commuting.

Synopsis

In the last 20 years Spain has experienced a significant increase of internal mobility, in particular from the peripheral to the core regions, as well as an extraordinary inflow of workers coming from all around the world, who also tend to settle down in the core regions. Using the analytical regions described in the Chapter 1., which allow us to distinguish between core (central) and peripheral regions, this chapter considers both migration and commuting flows and explores the chain of effects that the arrival of new workers have in the regions located in central places.

3. From the periphery to the core: direct and indirect effects of the migration of labour

3.1 Introduction

In Spain internal migrations – either inter or intraregional - have increased sharply since the 1980s. The recipients of these internal migrations are mainly central urban provinces. Simultaneously, there has been an extraordinary increase in the arrival of population from abroad, especially from Southern American and Northern African countries, which has shown a clear preference for settling down in highly-populated, urban central provinces.[29]

However, it could be argued that it is not the *province* itself that is the attracting force, but the *city* and its surrounding areas. In other words, the metropolitan areas and the big cities – the core - might be the ones offering new job opportunities, mostly service sector jobs, prompting the arrival of new workers (Bover and Velilla 1999). At the same time, the arrival of new workers may also have a negative impact, as agglomeration diseconomies (e.g. congestion, pollution) or transportation and housing costs begin to rise with a city growth thereby pushing out residents that might not consider them as attractive places to live in any more (Glaeser, 1998).

The objective of this chapter is to analyze the direct and indirect effects - in terms of internal migration and commuting flows - that the arrival of new workers either coming from abroad or from a peripheral region have on the so-called "*central* Spain". In this sense, the analytical classification described in detail in Chapter 1 captures the size effect,

[29] For details, see Appendixes VII and VIII. The data come from the *Census of Residential Variations* and the *Historical Population Series*, both of which are annual databases administered by the Spanish National Institute of Statistics (INE).

allowing us to distinguish different cities' sizes, and divides the territory into core and peripheral regions. From the assumption that foreign immigration affects internal mobility by displacing population, this research aims to show the potential of input-output methodology for studying interregional migrations under this general framework of interconnected migration systems. It is important to note that the methodology proposed in this paper is not an attempt to investigate the causal relationships between internal migrations and some explanatory variables. Instead, we suggest a type of analysis that shares many common points with the so-called Garin-Lowry models. Originally proposed by Lowry (1964) and Garin (1966), these models have been frequently used to explain the allocation of population and labour among different locations from an initial push of basic employment (see for example, McGill, 1977; Batty, 1983, Guldman and Wan, 1998; Jun, 2005).

The next section explains the fundamentals of a multiregional model that studies the effects of new arrivals to the core from the periphery by means of a multipliers matrix. The proposed methodology is based on the input-output migration model developed in Fernández-Vázquez *et al*. (2010). Section 3 explains how the model can be extended to the study of the distribution of jobs and residences across the types of regions by incorporating information of commuting flows. In section 4, categories for the central regions are suggested through an indicator that combines the two models described in the previous sections. Section 5 presents an empirical application of this methodology for Spain using the microdata from the most recent Census available. Finally, section 6 presents the concluding remarks of the chapter.

3.2 Core and periphery: a new approach for the Spanish case

> "The core regions of a country are those regions which are economically and politically dominant; they contain the principal cities of the country and have traditionally experienced high rates of net in-migration from the other, less urbanized, peripheral regions" (Vining and Pallone, 1982, pp. 340).

In principle, identifying a region as *core* or *peripheral* should not be controversial. However, as Vining and Pallone (1982) already pointed out, there are "a very large number of regionalization schemes available to any researcher on regional population trends" and it is impossible to prove whether a thesis holds for all such plausible schemes. Constrained by the availability of regional data, the researcher usually works with a classification based on the conventional politico-administrative boundaries.

In the Spanish case, the functional classification described in detail in Chapter 1 results on eight regional types, where the first five - Metropolitan Areas (MA1 and MA2), Central Urban Areas (CUA1 and CUA2), and Central Rural Areas (CRA) - represent the *core* or central regions while the last three - Peripheral Urban Areas (PUA1 and PUA2) and Peripheral Rural Areas (PRA) - would constitute the *periphery*.[30]

Based on this classification, Table 3.1 shows figures for year 2008 of the population concentration and incoming migration to the core either from the peripheral Spanish regions or from abroad. Central regions concentrate 64.40% of the total population and also take most of the population growth pressure. Thus, 69.13% of immigrants decide to live in the core, and internally, the central regions received 645,655 individuals from peripheral Spanish regions whereas only 320,941 left.

[30] Note that this classification would be also useful for evaluating the rural-urban patterns of migration and their effect on the urban type of regions. However, that is not the purpose of this chapter, which is focused on the relations between the core and the periphery.

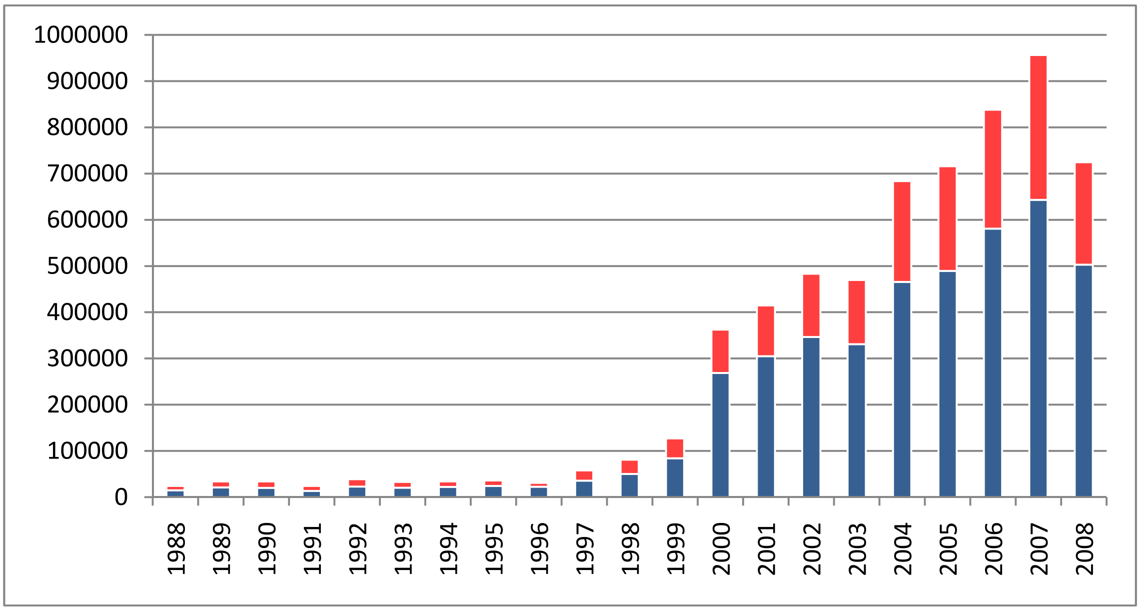
Table 3.1. Spanish Core and Periphery: Immigration and in-migration.

	Number of municipalities	Pop.	Internal Inmigration	Net Internal Inmigration	Immigration from abroad
CORE (Central Regions) Areas no more than one hour drive from a MA					
	2,333	29,629,417	645,655	324,714	495,944
MA1 Metropolitan Areas with more than 2,500,000 inhabitants					
Madrid MA	162	5,473,262	138,354	54,612	118,458
Barcelona MA	30	4,905,570	152,826	121,272	121,272
MA2 Metropolitan Areas with more than 500,000 inhabitants					
Valencia MA	47	1,568,630	30,241	14,699	24,104
Seville MA	27	1,342,145	13,944	7,925	9,927
Malaga MA	15	931,769	18,207	12,723	15,053
Murcia&Cartagena	14	908,752	17,643	10,370	14,524
Bilbao MA	34	905,577	10,688	5,185	8,226
Central Asturias MA	18	860,276	11,189	6,840	7,830
Zaragoza MA	28	758,818	16,926	11,755	13,591
Alicante MA	13	712,098	12,095	11,329	9,793
Cadiz MA	6	630,826	4,347	1,808	2,322
CUA1 Central Urban Areas of between 100,001 and 500,000 inhabitants					
	74	2,536,748	60,748	26,287	43,901
CUA2 Central Urban Areas of between 50,001 and 100,000 inhabitants					
	171	4,393,200	97,832	45,130	68,233
CRA Rural areas, less than 50,000 inhabitants					
	1,694	3,701,746	60,615	57,757	45,757
PERIPHERY (Peripheral regions) Areas more than one hour drive from a MA					
	5,772	16,372,626	-645,655	-320,941	221,392
TOTAL	**8,106**	**46,002,043**	**0**	**0**	**717,336**

Source: Authors' calculations based on Historical Population Series and Census of Residential Variations (2008).

The intensity of these inflows of population to the core, both from abroad and from the peripheral regions, has been growing since the beginning of the 1990s (Figure 3.1 and Figure 3.2).

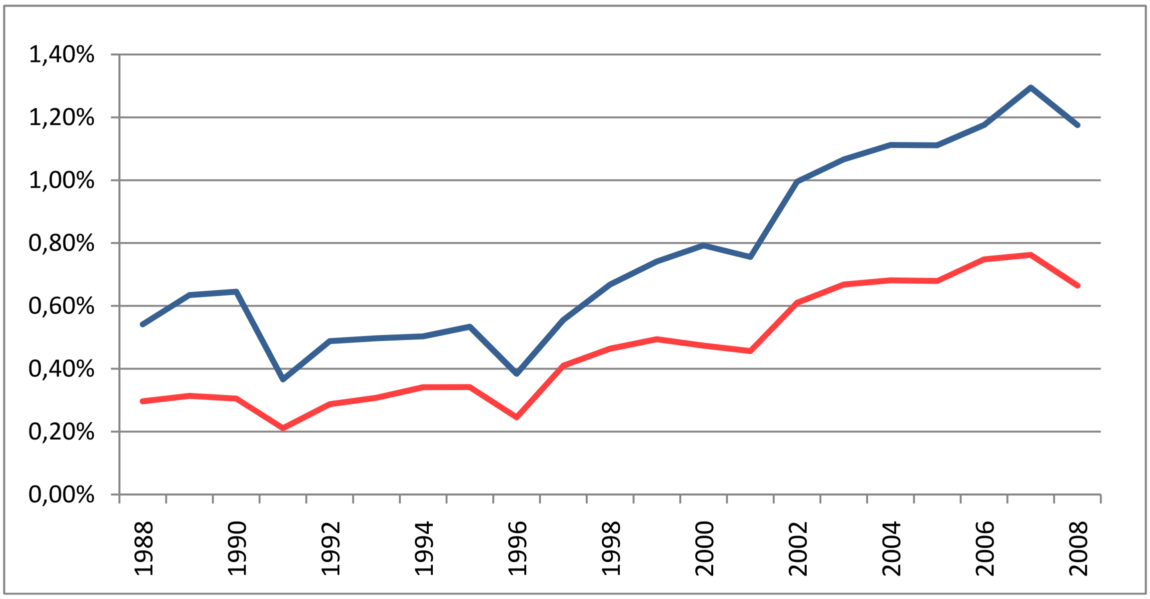
Figure 3.1. Immigration Movements to the Core and the Periphery.



Key	Immigrations to the core (absolute figures) Immigrations to the periphery (absolute figures)

Source: Authors' calculations based on Census of Residential Variations (1988 to 2008).

Figure 3.2. In-migration movements to the Core and the Periphery.



Key	In-migrations to the core (% population from the periphery) In-migrations to the periphery (% population from the core)

Source: Authors' calculations based on Historical Population Series and Census of Residential Variations (2008).

The effects that these arrivals generate on the migration patterns in the core itself can be quantified with a multiregional model based on Fernández-Vázquez *et al.* (2010). This model, which later on will be extended to include commuting flows, is explained in the next section.

3.3 The arrival of workers to the central locations modeled with an input-output approach

The point of departure for the analysis will be to consider a set of N geographical units that experience the spatial re-allocation of some part of their workforce in a given period of time. For the sake of simplicity, in this chapter we will specifically consider a set of N central regions (CR) that together constitute the core, where some interregional migration flows are observed. In this situation, the following table reflects the migration flows among these central regions (CR):

Table 3.2. Matrix of migrations among N central regions.

	CR_1	CR_2	…	CR_N
CR_1	m_{11}	m_{12}	…	m_{1N}
CR_2	m_{21}	m_{22}	…	m_{2N}
…	…	…	…	…
CR_N	m_{1N}	m_{2N}	…	m_{NN}

where a typical element m_{ij} denotes the number of workers that migrate from central region i to another central region j. For any central region j considered, the net inflow of migrants (nm_j) received will be:[31]

$$nm_j = [ic_j + p_j] - [oc_j + a_j] \tag{3.1}$$

[31] The traditional matrix algebraic notation is applied: bold uppercases denote matrices, bold lowercases (column) vectors and italic lowercases scalars. A prime indicates transpose.

Equation (3.1) defines the net inflow of workers for each type of central region. In other words, this variable is defined as the difference between the arrival of new workers - consisting of the in-migration coming from other central regions (ic_j) plus the workforce coming from the periphery (p_j), i.e., either peripheral regions or from another country - and the outflows of people given by the out-migration to other regions located into the core (oc_j) as well as the out-migration to non-central regions (a_j). Note from Table 3.2 that it is possible to obtain ic_j as the column sum $\sum_{i=1}^{N} m_{ij}$. Likewise, the out-migrations from region i to other central regions (oc_i) can be obtained as the row sum $\sum_{j=1}^{N} m_{ij}$. Conversely, the migration flows to and from non-central regions (p_j and a_i, respectively) do not appear in Table 3.2 but they can be included in it jointly with net migration (nm_j) in order to construct a new table that fulfils the demographic identity (3.1), and where the row sums equal the column sums. With this purpose in mind, let us define a N×1 vector $\boldsymbol{x}$, where a typical element x_j shows the inflows arriving to central region j. Note that the elements of this vector can be defined as the sum $x_j = ic_j + p_j$, or alternatively as $x_j = nm_j + [oc_j + a_j]$, from equation (3.1).

This equivalence holds also when considering the whole vector:

$$\boldsymbol{x}' = \boldsymbol{ic}' + \boldsymbol{p}' \tag{3.2a}$$

$$\boldsymbol{x} = \boldsymbol{nm} + [\boldsymbol{oc} + \boldsymbol{a}] \tag{3.2b}$$

Consequently, Table 3.2 can be modified in the following way in order to verify that the row and column sums are both equal to vector **x**.

Table 3.3. Matrix of migrations among N central regions and the periphery.

	CR1	**CR2**	**...**	**CRN**	***oc***	***a***	***nm***	***x***
CR1	m_{11}	m_{12}	...	m_{1N}	$oc_1 = \sum_{j=1}^{N} m_{1j}$	a_1	nm_1	x_1
CR2	m_{21}	m_{22}	...	m_{2N}	$oc_2 = \sum_{j=1}^{N} m_{2j}$	a_2	nm_2	x_2
...	...	...	...	...	...	...	...	...
CRN	m_{1N}	m_{2N}	...	m_{NN}	$oc_N = \sum_{j=1}^{N} m_{Nj}$	a_N	nm_N	x_N
ic'	$ic_1 = \sum_{i=1}^{N} m_{i1}$	$ic_{12} = \sum_{i=1}^{N} m_{i2}$	...	$ic_N = \sum_{i=1}^{N} m_{iN}$				
p'	p_1	p_2	...	p_N				
x'	x_1	x_2	...	x_N				

Some assumptions are required in order to explain $\boldsymbol{x'}$ through an input-output model. One basic assumption is that the arrival of new workers from the periphery (vector $\boldsymbol{p'}$) is something exogenous to the set of N central regions analyzed. Additionally, from Table 3.3 we will define a coefficient $b_{ij} = \frac{m_{ij}}{x_i}$. These b_{ij} coefficients measure the number of workers that move from central region i to central region j relative to the total number of workers received in i (including those coming from peripheral regions). If, for instance, $b_{ij} = 0.15$, this would imply that for each 100 workers received in region i, this region "pushes" 15 to region j. The coefficients b_{ij} are thus a sort of "rate of displacement" that defines the behaviour of central region i when it receives new workers. The basic idea behind these b_{ij} is that the arrival of these new workers to one central region N produces a type of "diseconomy" (e.g.: rises in housing prices, traffic congestion, etc.) that encourages some of the workers living there to migrate.

Another essential assumption of the model is that the b_{ij} coefficients are assumed to be fixed in the short run. The matrix $\boldsymbol{B}$ contains all the b_{ij} coefficients for the N regions included in the model:

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & . & b_{1N} \\ b_{21} & b_{22} & \vdots & b_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ b_{N1} & b_{N2} & . & b_{NN} \end{bmatrix} \quad (3.3)$$

As a result, the vector of workers coming from other central regions ($\boldsymbol{ic}'$) can be expressed as:

$$\boldsymbol{ic}' = \boldsymbol{x}'\boldsymbol{B} = [x_1 \quad \cdots \quad x_N] \begin{bmatrix} b_{11} & b_{12} & . & b_{1N} \\ b_{21} & b_{22} & \vdots & b_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ b_{N1} & b_{N2} & . & b_{NN} \end{bmatrix} = [ic_1 \quad \cdots \quad ic_N] \quad (3.4)$$

and equation (2a) can be rewritten in terms of $\boldsymbol{B}$ as:

$$\boldsymbol{x}' = \boldsymbol{ic}' + \boldsymbol{p}' = \boldsymbol{x}'\boldsymbol{B} + \boldsymbol{p}' \quad (3.5)$$

Suppose that the group of N central regions receives in a given period a vector of new workers $\boldsymbol{p}'$. However, this initial inflow of $\boldsymbol{p}'$ new workers will "push" some of the workforce out to another one of the N central regions. This will generate a new round of movements in the system of N central locations equal to $\boldsymbol{p}'\boldsymbol{B}$, which will further displace a part of the workforce equal to $\boldsymbol{p}'\boldsymbol{BB}$, and so on. The expression that describes the overall process of obtaining the new vector of incoming workers $\boldsymbol{x}'$ is:

$$\boldsymbol{x}' = \boldsymbol{p}' + \boldsymbol{p}'\boldsymbol{B} + \boldsymbol{p}'\boldsymbol{B}^2 + \boldsymbol{p}'\boldsymbol{B}^3 + \cdots = \boldsymbol{p}'[\boldsymbol{I} + \boldsymbol{B} + \boldsymbol{B}^2 + \boldsymbol{B}^3 + \cdots] \quad (3.6)$$

Where $\mathbf{I}$ is the identity matrix. Under certain mathematical conditions, (3.6) can be written as:

$$\boldsymbol{x}' = \boldsymbol{p}\,[\boldsymbol{I} - \boldsymbol{B}]^{-1} \tag{3.7}$$

Equation (3.7) explains how the arrival of new workers to the central regions ($\boldsymbol{x}'$) changes due to variations in the vector of workers coming from peripheral regions ($\boldsymbol{p}'$). The idea underlying equations (3.6) and (3.7) is that any increase in the movement of workers from the periphery to a region situated in the core, apart from the direct impact that it has on this specific region, generates a set of indirect effects on the entire system of N central regions that turns out to be larger than the initial shock.[32]

In the framework defined by the previous model, the elements of the matrix $[\boldsymbol{I} - \boldsymbol{B}]^{-1}$ play a crucial role. The structure of this matrix is:

$$[\boldsymbol{I} - \boldsymbol{B}]^{-1} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdot & \beta_{1N} \\ \beta_{21} & \beta_{22} & \vdots & \beta_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ \beta_{N1} & \beta_{N2} & \cdot & \beta_{NN} \end{bmatrix} \tag{3.8}$$

where the element β_{ij} shows the variation in the number of workers that arrive to the central region j due to the arrival of one additional worker to the central region i.

This means that β_{ij} can be interpreted as an approximation to the following derivative:

$$\beta_{ij} = \frac{dx_j}{dp_i} \tag{3.9}$$

---

[32] Readers accustomed to the input-output literature will easily see the analogy of this proposal with the so-called Gosh input-output model. For an excellent review of the properties and interpretation of this model, see Dietzenbacher (1997).

It is important to note that even if there are no direct migration flows between regions i and j, β_{ij} might still be different from zero given that it also measures the indirect effects. For example, the migration from the periphery of workers to central region i displaces part of the workforce to the central region h, and consequently some workers from h move to the central region j.

3.4 The spatial allocation of economic activity: the role of commuting

The previous model (3.7) explains how workers allocate and reallocate their place of residence across the system of N central regions (vector $\boldsymbol{x}$) given the initial stimulus of new entries of workers from the periphery. It is important to highlight that it models the choice of residence of the workers, which may not be the same place where they have their jobs. This is a significant difference because while it is possible that a region i displaces workers to another region j as a consequence of the diseconomies produced by additional dwellers which turn region i into a less attractive place to live in, these workers might keep their jobs in the original region i because it is still attractive to work there. In other words, the model allows the possibility of commuting to be included given that the place of residence and job location might not be in the same region.

In a fashion similar to the previous section, the commuting flows between the N central regions can be represented in the following table (Table 3.4), where the regions of origin are displayed by rows and the destinations are shown by columns.

Table 3.4. Matrix of commuting flows among the N central regions.

	CR1	**CR2**	**…**	**CRN**	$\boldsymbol{x}$
CR1	f_{11}	f_{12}	…	f_{1N}	x_1
CR2	f_{21}	f_{22}	…	f_{2N}	x_2
…	…	…	…	…	…
CRN	f_{1N}	f_{2N}	…	f_{NN}	x_N
$\boldsymbol{l}$	l_1	l_2	…	l_N	

f_{ij} stands for the flow of workers in a given period of time that arrived to live in the central region i but commute to the central region j. The main diagonal elements represent the workers that live and work in the same region. Note that $\sum_{j=1}^{N} f_{ij} = x_i$ and that $\sum_{i=1}^{N} f_{ij} = l_j$, where l_j is the total number of jobs occupied by the vector of workers $\boldsymbol{x}$ that are allocated in region j.

For the sake of convenience, we will work with proportions of commuters instead of working directly with flows. These proportions will be defined as $c_{ij} = \frac{f_{ij}}{x_i}$ and measure the fraction of workers who live in central region i but work in central region j. For example, c_{ij} = 0.25 means that 25% of the workers who migrated to region i commute to region j. If $\boldsymbol{C}$ denotes the $N \times N$ matrix of proportions c_{ij}, it is straightforward to see that:

$$\boldsymbol{l}' = \boldsymbol{x}'\boldsymbol{C} \tag{3.10}$$

Equation (3.10) links the entries of workers that live in the system of N central regions ($\boldsymbol{x}'$) with the distribution of their jobs across the same N regions ($\boldsymbol{l}'$). Note that this equation is simply a mathematical expression that relates the place of residence to the location of the jobs. However, by combining equations (3.7) and (3.10) we can construct the following model that explains the spatial allocation of the new jobs from the exogenous shock produced by the arrival of workers from the periphery to the core ($\boldsymbol{p}'$):

$$\boldsymbol{l}' = \boldsymbol{x}'\boldsymbol{C} = \boldsymbol{p}'[\boldsymbol{I} - \boldsymbol{B}]^{-1}\boldsymbol{C} \qquad (3.11)$$

Equation (3.11) models the spatial location across the core of the economic activity (new jobs) generated as a consequence of new workers coming to live in the set of central regions. The idea is that workers from the peripheral regions ($\boldsymbol{p}'$) arrive to the central regions, which produces a sequence of indirect effects through migrations - matrix $[\boldsymbol{I} - \boldsymbol{B}]^{-1}$ - that subsequently implies a specific distribution of jobs - matrix $\boldsymbol{C}$. The whole sequence of multiplier effects on the generation of jobs is given by the product of the matrices $[\boldsymbol{I} - \boldsymbol{B}]^{-1}$ and $\boldsymbol{C}$. Letting $\boldsymbol{M}^* = [\boldsymbol{I} - \boldsymbol{B}]^{-1}\boldsymbol{C}$, equation (3.11) can be written as:

$$\boldsymbol{l}' = \boldsymbol{p}'\boldsymbol{M}^* \qquad (3.12)$$

where:

$$\boldsymbol{M}^* = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdot & \mu_{1N} \\ \mu_{21} & \mu_{22} & \vdots & \mu_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ \mu_{N1} & \mu_{N2} & \cdot & \mu_{NN} \end{bmatrix} \qquad (3.13)$$

and a typical element μ_{ij} shows the variation in the number of jobs in the central region j given by the arrival of one additional worker to central region i.

Note that these elements are given by the sums $\sum_{h=1}^{N} \beta_{ih} c_{hj}$, so that in more detail we have:

$$\mu_{ij} = \sum_{h=1}^{N} \beta_{ih} c_{hj} = \sum_{h=1}^{N} \frac{dx_h}{dp_i} \frac{f_{hj}}{x_h} = \frac{dx_1}{dp_i} \frac{f_{1j}}{x_1} + \frac{dx_2}{dp_i} \frac{f_{2j}}{x_2} + \cdots + \frac{dx_N}{dp_i} \frac{f_{Nj}}{x_N} \quad (3.14)$$

The idea underlying the μ_{ij} elements is that they comprise a two-stage process: the entries of workers from the periphery to region i generate a variation - through the whole round of indirect effects captured in β_{ih} - in the inflows of labour to region h, a proportion $c_{hj} = \frac{f_{hj}}{x_h}$ of which are going to commute to region j. When considered together and summed across all the regions h, μ_{ij} can be taken as an approximation to the derivative:

$$\mu_{ij} = \frac{dl_j}{dp_i} \quad (3.15)$$

We may be interested in studying the capability of each region of getting the jobs that are generated by the entrance of new workers in the central regions. In other words, we could be interested in estimating how many jobs will locate to region j when all the central regions receive one additional worker coming from the periphery. Note that this number can be obtained by the sum $\mu_{\cdot j} = \sum_{i=1}^{N} \mu_{ij}$ and it can be written as:

$$\mu_{\cdot j} = \sum_{i=1}^{N} \mu_{ij} = \sum_{i=1}^{N} \frac{dl_j}{dp_i} = \frac{dl_j}{dp_1} + \frac{dl_j}{dp_2} + \cdots + \frac{dl_j}{dp_N} \quad (3.16)$$

Relatively large values of $\mu_{\cdot j}$ would indicate that region j attracts comparatively more jobs than the average region when workers coming from the periphery migrate to the core.

3.5 The net effect: "resident absorbing" or "job attracting" regions

The elements of matrices (3.8) and (3.13) highlight the effect that the arrivals of workers from the periphery (vector $\boldsymbol{p}$) have both on the residential (vector $\boldsymbol{x}$) and job location (vector $\boldsymbol{l}$) patterns. It seems logical to expect that among the set of N central regions some of them experience a relatively larger effect on one of these patterns. For example, the new workers coming from the periphery might decide to live in some specific location i but commute to a different location because there are better opportunities for work there. This would result in a large effect on x_i but a small effect on l_i. Conversely, in a given region i we might observe a huge generation of jobs as a consequence of the exogenous shock of vector $\boldsymbol{p}$, but the number of workers who have their residence there might be small (because of high prices of housing, say) and consequently we would have large effects on l_i but smaller ones on x_i.

From this simple idea, we can define a measure of "net demand of commuters" for a region i as the difference $d_i = (l_i - x_i)$, which compares the number of jobs that are located in that region with the number of workers that reside there. If this difference is positive, it means that region i would require commuters from other areas in order to fill the jobs that are not taken by the local workers. In other words, this would be a signal that would indicate that region i is attractive for working in but not for living in. The opposite would apply if this difference were negative.

If we compute this difference for the whole set of N central regions, we would have:

$$\boldsymbol{d}' = \boldsymbol{l}' - \boldsymbol{x}', \tag{3.17}$$

And taking into account equations (3.11) and (3.7):

$$\boldsymbol{d}' = \boldsymbol{l}' - \boldsymbol{x}' = \boldsymbol{p}'[\boldsymbol{I} - \boldsymbol{B}]^{-1}\boldsymbol{C} - \boldsymbol{p}'[\boldsymbol{I} - \boldsymbol{B}]^{-1} = \boldsymbol{p}'[\boldsymbol{I} - \boldsymbol{B}]^{-1}[\boldsymbol{C} - \boldsymbol{I}] \tag{3.18}$$

If we denote with the matrix obtained by the product $[\boldsymbol{I} - \boldsymbol{B}]^{-1}[\boldsymbol{C} - \boldsymbol{I}]$ by $\boldsymbol{\Delta}$, the previous equation can be written as:

$$\boldsymbol{d}' = \boldsymbol{p}'\boldsymbol{\Delta} \tag{3.19}$$

where:

$$\boldsymbol{\Delta} = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdot & \delta_{1N} \\ \delta_{21} & \delta_{22} & \vdots & \delta_{2N} \\ \cdot & \cdot & \vdots & \cdot \\ \delta_{N1} & \delta_{N2} & \cdot & \delta_{NN} \end{bmatrix} \tag{3.20}$$

and a typical element δ_{ij} shows the variation in the requirement of commuters in central region j produced by one additional worker migrating from the periphery to central region i. Considering that each element δ_{ij} comes from the product $[\boldsymbol{I} - \boldsymbol{B}]^{-1}[\boldsymbol{C} - \boldsymbol{I}]$, its expression is given by:

$$\delta_{ij} = \sum_{h \neq j}^{N} \beta_{ih} c_{hj} + \beta_{ij}(c_{jj} - 1) \tag{3.21}$$

Note that the element δ_{ij} will be positive if:

$$\sum_{h \neq j}^{N} \beta_{ih} c_{hj} > \beta_{ij}\left(\frac{x_j - f_{jj}}{x_j}\right) \tag{3.22}$$

i.e., if the arrival of new workers from the periphery to region i produces an increase in the commuters to j larger than the increase in the number of workers that commute from j $(x_j - f_{jj})$. On the contrary, the variation in p_i could cause an increase in the workers that have their residence in another region h and commute to work in j that is smaller than the growth in number of workers living in j but commuting to somewhere else. In such a case, δ_{ij} would be negative, which would indicate that the rise in immigrant workers from the periphery arriving to region i produces an increase in residents rather than in jobs in j. If, on the other hand, δ_{ij} is positive, the additional workers coming from the periphery to region i would increase the jobs located in j to a greater extent than the workers residing there.

From these δ_{ij} multipliers it is possible to classify the regions into two different types: those that are capable of attracting comparatively more jobs than residents ("job attracting") when new workers enter in the system of central regions, and those where the opposite happens ("resident absorbing"). This information can be obtained by the sum $\delta_{.j} = \sum_{i=1}^{N} \delta_{ij}$.

In general terms, the following vector contains these sums for all the regions:

$$\boldsymbol{\delta}' = \boldsymbol{e}'\boldsymbol{\Delta} \tag{3.23}$$

When an element $\delta_{.j}$ of the vector $\boldsymbol{\delta}'$ is positive, this indicates that the region j can be classified as "job attracting". If, on the contrary, $\delta_{.j}$ is negative then region j would be "resident absorbing".

3.6 Effects of the in-migration and immigration to the core in Spain: an empirical application for 1991-2001

This section applies the aforementioned classification and methodology to the construction of a model of migration and commuting flows for Spain between 1991 and 2001 using the data from the last National Census. As explained above, in Spain internal mobility has traditionally not been very high, but it has recently experienced a considerable increase together with a remarkable rise in the reception of immigrants.

The data required for the model have been obtained taking a sample of microdata extracted from the most recent *Censo de Población y Viviendas* - Population and Housing Census (PHC) - compiled by the Spanish National Statistical Institute for 2001. The sample comprises approximately 5% of the whole census, corresponding to a sample size of around two million people. The information contained in the survey includes the part of population who were working in 2001, the municipality where they were working and the municipality where they were living. Moreover, information about the municipality where they lived in 1991 is also available.

After aggregating those municipalities into the suggested classification of analytical regions, we can construct a matrix of interregional migration flows - similar to Table 3.3 - for Spain between 1991 and 2001.

Table 3.5. Migration flows among the five types of central regions and from the periphery. Spain 1991-2001.

	MA1	MA2	CUA1	CUA2	CRA
MA1	351	811	360	4,940	5,337
MA2	969	657	465	2,596	1,825
CUA1	272	332	95	585	941
CUA2	1,177	1,195	501	348	2,100
CRA	989	827	455	1,632	203
$\boldsymbol{p}'$	10,656	4,397	1,533	4,405	2,251
$\boldsymbol{x}'$	14,414	8,219	3,409	14,506	12,657

Source: own computation of the authors from the sample of microdata of the 2001 Spanish Census (INE, 2007).

For the sake of simplicity, we have just focused on the data required to build a 5×5 table of interregional migrations among the locations classified as central regions - which contains the m_{ij} elements - and a 1×5 vector of immigrants from the peripheral locations ($\boldsymbol{p}'$). The column sums of the m_{ij} elements plus the values of vector $\boldsymbol{p}'$ are equal to the vector of workers that from 1991 to 2001 migrated to one of these regions ($\boldsymbol{x}'$). This table illustrates that during the period 1991-2001 the set of central regions received 53,205 workers, of which 29,963 came from a region in the core - corresponding to the sum $\sum_{i=1}^{5}\sum_{j=i}^{5} m_{ij}$ – with the remaining 23,242 coming from a peripheral location corresponding to the sum of the elements of vector $\boldsymbol{p}'$.

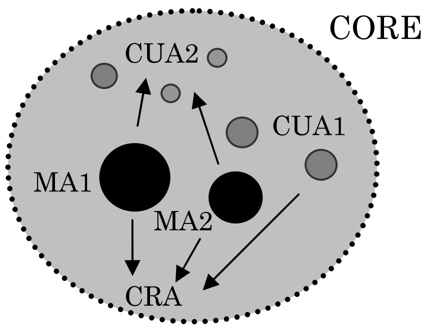
In order to obtain the inverse $[\boldsymbol{I}-\boldsymbol{B}]^{-1}$, the migration flows presented in Table 3.5 have been divided by the vector of total inflows $\boldsymbol{x}'$ to compute the coefficient matrix $\boldsymbol{B}$. The β_{ij} multipliers quantify how many workers the region i displaces directly and indirectly to region j as a consequence of the arrival of workforce from the periphery to region i. If we obtain the sums $\beta_{i\cdot} = \sum_{j \neq i}^{N} \beta_{ij}$, excluding the elements of the main diagonal to avoid the effects induced on themselves, we will have an indicator that measures the amount of workers displaced from region i to other central regions. Table 3.6 shows these indicators.

Table 3.6. β multipliers.

	$\boldsymbol{\beta_{i\cdot}}$
MA1	1.3623
MA2	1.3757
CUA1	1.2154
CUA2	0.6172
CRA	0.5576

The figures in Table 3.6 show that, roughly speaking, the greater the size, the larger the number of workers that each type of location displaces to other regions. In the largest metropolises, the arrival of 1,000 workers from the periphery would provoke more than 1,300 reallocations of residence to other regions over a period of ten years. This volume decreases with size down to less than one half for the case of regions including small but central municipalities (CRA). The next figure shows the direction of the five most important β_{ij} multipliers, those larger than the third quartile.

Figure 3.3. Displacement effect (β multipliers).



Source: Own elaboration based on Polese et al. (2007).

The effects described in this figure are in line with the $\beta_{i.}$ indicators of the previous table: the biggest cities MA1, MA2 and CUA1 are pushing workers to live in other locations presumably because in the former there are some diseconomies that encourage some workers to migrate to smaller cities.

Besides the data on residence mobility between 1991 and 2001 the PHC also informs about the commuting patterns in 2001, as for each person included in the sample both the municipality of residence and the location of the job were registered. This allows us to obtain a matrix that contains the commuting trips among the central regions, with the same structure as Table 3.4.

Table 3.7. Commuting flows among the five types of central regions. Spain 1991-2001.

	MA1	MA2	CUA1	CUA2	CRA	x
MA1	13,210	34	14	701	455	14,414
MA2	57	7,207	105	524	326	8,219
CUA1	23	43	2,869	210	264	3,409
CUA2	2,818	1,563	453	8,391	1,281	14,506
CRA	3,434	1,182	673	1,746	5,622	12,657
ι'	19,542	10,029	4,114	11,572	7,948	

Source: own computation of the authors from the sample of microdata of the 2001 Spanish Census (INE 2007)

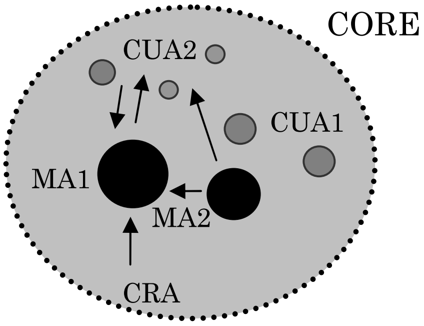
From the data contained in Table 3.7, the proportions of commuters defined as $c_{ij} = \frac{f_{ij}}{x_i}$ is computed to obtain the matrix $\boldsymbol{C}$. This matrix will be used to replicate equation (11) for our study case and compute the multiplier matrix $\boldsymbol{M}^* = [\boldsymbol{I} - \boldsymbol{B}]^{-1}\boldsymbol{C}$, where a typical element μ_{ij} quantifies the variation in the jobs located in j generated by the arrival of one additional worker to region i. We now focus our analysis on the measurement of the number of jobs located in region j when all the regions of the core receive one additional worker coming from the periphery. Again, we exclude the main-diagonal elements of $\boldsymbol{M}^*$. Table 3.8 shows the resulting sums $\mu_{\cdot j} = \sum_{i \neq j}^{N} \mu_{ij}$.

Table 3.8. μ multipliers.

	$\boldsymbol{\mu_{\cdot j}}$
MA1	1.7478
MA2	1.0414
CUA1	0.4985
CUA2	1.6324
CRA	1.2497

The results in Table 3.8 indicate that the largest cities (MA1) and the smaller ones (CUA2 and CRA) are the locations that manage to retain the jobs distributed across the central regions when they receive immigrant workers from the periphery. For example, if all the central regions - excluding the MA1 itself - received 1,000 workers from the periphery, then after all the rounds of redistribution of residence and taking into account the commuting patterns between the regions, 1,748 jobs would be allocated to the MA1 region. Note that in this case there is not a clear correlation with size. The biggest metropolitan areas (MA1) are the ones retaining most of the jobs, followed by small central cities less than 100,000 inhabitants (CUA2). Similarly to the previous example, Figure 3.4 depicts the direction of the five most important μ_{ij} multipliers.

Figure 3.4. Job creation (μ multipliers).



Source: Own elaboration based on Polèse et al. (2007).

This figure shows that the biggest cities (MA1 type) are those that manage to attract most of the jobs filled by the migrating labour force. The successive rounds of reallocation of workers' residences together with the commuting patterns lead to a spatial distribution of jobs where the largest metropolises retain the greatest part, and the workers who reside in small-scale municipalities such as UCA2 or CRA (which receive the biggest indirect effects in terms of migration - see Figure 3.2) commute to work in those MA1 cities.

All in all, the general picture is that the immigration of labour from the periphery to the core pushes workers from living in big cities to residing in the smaller locations situated in the core. However, commuting permits the large metropolises to keep the jobs filled with these workers. To gain more insight into the different effects on the distribution of jobs and residences across the five types of regions, we focus on vector $\boldsymbol{d}$. The basic insight is that over the entire period 1991-2001 the central regions have received workers from either a region in the centre or from a peripheral location (Table 3.5) and while in the aggregate the number of workers should be equal to the number of jobs, the distribution of these two variables is different depending on the type of region.[33]

Therefore, we have computed the matrix $\boldsymbol{\Delta}$ composed of the δ_{ij} coefficients that quantify the variation in the requirement of commuters in the central region j produced by one additional worker migrating from the periphery to the central region i. According to the interpretation given in the previous section, the locations that can be classified as net "job attracting" because they manage to attract comparatively more jobs than residents will be characterized by having a positive value of the sum $\delta_{.j} = \sum_{i \neq j}^{N} \delta_{ij}$. In the opposite case, if $\delta_{.j}$ is negative then the region j could be labelled as "resident absorbing". Again, we exclude from this sum the elements of the main diagonal of matrix $\boldsymbol{\Delta}$. The coefficients are presented in Table 3.9.

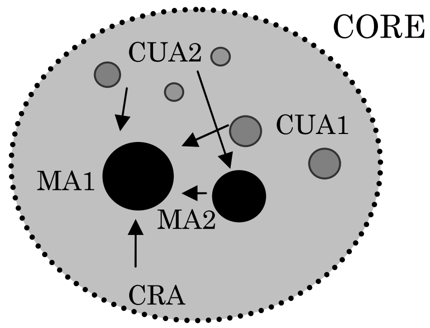
Table 3.9. δ coefficients.

	$\boldsymbol{\delta_{.j}}$
MA1	1.058
MA2	0.425
CUA1	0.181
CUA2	-0.088
CRA	-0.535

[33] It should be remembered that these numbers refer to a sample size of approximately 5% of the total population.

The figures and signs in Table 3.9 suggest a very high correlation between the size of the cities and their classification as "job attracting" or "resident absorbing". Taking as an example the case of the MA1 cities, if all the central regions - excluding the MA1-type itself - received 1,000 workers from the periphery, the final result would be that they would require 1,058 commuters from other central regions because they attract 1,058 more jobs than residents. A similar result, though on a smaller scale, can be observed for the cases of cities classified as MA2 and CUA1. On the other hand, the smallest municipalities, which have been grouped as CUA2 and CRA, can be considered as "resident absorbing". Again, we can get a richer description of the situation if we represent graphically the direction of the most important (larger than the third quartile) δ_{ij} elements.

Figure 3.5. Job attracting regions (δ coefficients).



Source: Own elaboration based on Polèse et al. (2007).

This figure shows that the MA1-type cities (the biggest ones) are the job-winners in the process triggered by the migration of labour from the periphery. The entrance of workers from the periphery to any other central regions causes a rise in the jobs located in the MA1 cities far larger than the increase in their resident-workers. Figure 3.5 also suggests a size-based hierarchy for this phenomenon given that the next type of urban regions in the size ranking - MA2 - also attract more jobs that residents (mainly from the CUA2 cities).[34]

3.7 Conclusions

Spain has experienced over the last two decades an intense arrival of immigrants and in-migrants to its central regions. The arrival of population has effects on the recipient regions through internal migrations and/or commuting to some areas that might be more attractive.

Using a novel definition of core and periphery based on the analytical regions proposed by Polèse *et al.* (2007), and extending the input-output model suggested on Fernández-Vázquez *et al.* (2010) to include the possibility of commuting, this chapter has assessed the effects that the arrival of new workers (either nationals or foreigners) have in the core. Using the last available Census, estimations for Spain show that the arrival of in- and im-migration to the core generates a set of direct and indirect effects induced by the redistribution of population among other regions.

The arrival of workers from the periphery to the core provokes reallocations of residence in all cases. However, the intensity of these reallocations increases with size, which proves the existence of agglomeration diseconomies associated with big cities. At the same time, when the possibility of commuting is considered the arrival of workers from the periphery to the core generates both the reallocation of jobs (economic activity) and also of residences. The larger cities are the ones pushing out more residents to other areas while keeping most of the jobs. In other words, they are becoming attractive areas to work in (economies of agglomeration), but not to live in (high housing costs, congestion or some other negative externalities). The opposite is true for the smaller cities, which are resident-

[34] Appendix IX reports the three multiplier matrices obtained from the data.

absorbing but not job-attracting. Thus, the distribution pattern of residences proves to be different to the distribution pattern of jobs.

Even more, these results highlight the idea that the effects of the arrival of population are not only felt in the recipient region/city but might also generate comparatively far larger effects on other regions in terms of internal migration and commuting flows.

Conclusions and Extensions

The concept of Region is one of the elements which has differentiated Regional Economics from other fields of Applied Economics. In spite of this, however, researchers in this field of economic analysis have not paid a great deal of attention to this concept. All too often, Regions have been identified with the politico-administrative units into which nation-states have divided their territory and for which statistical information is widely available. However, a Region can be defined in many different ways. Frameworks which define regions according to analytical/theoretical criteria provide greater scope for applied studies and permit a more complete interpretation of the results contained therein.

In this research we have proposed a concept of Region which goes beyond the administrative division of territory. Our regional aggregation has been based on agglomeration economies, one of the fundamental concepts in the fields of Economic Geography and Urban and Regional Economics. In accordance with the work of Polèse *et al.* (2007), the territory has been classified into analytical regions which take into account the size of the population and the distance from the main urban areas. In doing so, we achieve an aggregation which corresponds with the differences in agglomeration economies across space. However, their robustness in comparison with the administrative units commonly used has – to date - not been evaluated. The objective of the first chapter of this thesis was to prove that the functional regions defined under such economic criteria provide better defined regions – in terms of greater compactness and separation - than the administrative ones commonly used to carry out labour market studies at sub-national level.

Using micro data from the last available Spanish Census, the functional and administrative regions are evaluated using the Theil index and the Davies-Bouldin Validation index. Applied to employment (by gender, industry and level of qualification and occupation), both indexes show better results for the analytical regions than for any of the ordinary administrative ones (NUTS I, II or III regions). In other words, the analytical classification generates areas where the distribution of employment is more homogeneous within and more heterogeneous between the regions. Agglomeration economies and distance (to the metropolis) seem to be relevant for understanding the patterns of distribution of employment, either by gender, by industry or by level of qualification and occupation. In practice, this provides a clearer way for identifying local labour markets and explaining their differences and similarities. In light of the results from the first chapter, we suggest the use of this alternative classification –subject, of course, to the availability of data - when carrying out Labour Economics studies that include a *spatial* dimension.

The following chapters have provided two applications of this analytical division of the territory to Labour Economics issues: the factors affecting the probability of being employed (Chapter 2) and the effects that labour mobility and commuting have on the central regions (Chapter 3).

In the second chapter, we presented a spatial analysis of employment at local level where, among other factors, the demographic and geographical characteristics can and do affect the outcome. The empirical results support the hypothesis that *size* – in terms of population - and *location* – in terms of distance to a metropolis - are explanatory variables for the probability of being employed. In other words, employment depends not only on the personal characteristics of the individuals (level of education, age, sex, etc.) but also on the type of analytical region – as defined in Chapter 1 - where they live.

Regarding the importance of location, our results show a significant gap in the chances of being employed between "central" and "peripheral" types of regions, i.e., the closer the region is to the metropolis, the higher the concentration of economic activity and therefore employment. Likewise, the expected negative relationship between employment and the *size* of the region where the individual lives is confirmed, and this seems to be stronger for non-skilled individuals than for people with university studies.

In terms of employability, the largest Spanish metropolitan areas (MA1) seem to be enjoying the full benefits of agglomeration economies while the smaller metropolitan areas (MA2) seem to be suffering their negative effects. Likewise, for urban areas that cannot be considered “metropolitan areas” (UA1 and UA2), size does not seem to be as important as their central-peripheral location. In rural areas (less than 50,000 inhabitants), both size and distance seem to be relevant determinants of employability. In other words, regardless of the level of studies, there are fewer chances of being employed in rural areas as opposed to urban areas and in peripheral rural areas as opposed to central rural areas.

Recognizing the importance of these spatially differentiated results should have a significant impact on current policy discussions, shifting the focus from general solutions to more spatially customized ones where size and location are considered. Just as differences in age, gender or industrial structure are taken into account in the design of employment policies (at national or local level), these results suggest that an additional *spatial dimension* that somehow includes the size and location of the local area where the person lives should be considered.

Some important migration policy implications can also be derived when using these alternative functional regions to analyze the direct and indirect effects that the arrival of workers has in the core regions. Spain has experienced over the last two decades an intense arrival of both immigrants and in-migrants to its central regions, and as a consequence (though not exclusively) of these inflows, we can observe internal migrations and/or commuting to some areas that might be more attractive.

Using the last available Census, the estimations for Spain of an input-output multi-regional model that includes the possibility of commuting show that the arrival of in- and im-migration to the core generates a set of effects induced by the redistribution of population among other regions. The arrival of workers from the periphery to the core provokes reallocations of residence in all cases (displacement effect). However, the intensity of these reallocations increases with size, which shows the existence of some agglomeration diseconomies associated with big cities. When the possibility of commuting is also considered, the arrival of workers from the periphery to the core generates the reallocation of both jobs (economic activity) and residences. The larger cities are the ones pushing out more residents to other areas, while keeping most of the jobs. In other words, they are becoming attractive areas to work in, but not to live in (due to, among other reasons, high housing costs, congestion or other negative externalities). The oppposite is true for the smaller cities, which are attractive for residing in but for working in. The distributional pattern of residences proves to be different to the distributional pattern of jobs.

These results highlight the idea that the effects of the arrival of population are not only felt by the recipient region/city but may generate comparatively far larger effects on other regions in the form of internal migration and commuting flows, something that policy makers should bear in mind.

To conclude, surpassing the administrative division of the territory, this classification manages to have explanatory power in spatial Labour Economics topics while including relevant geo-economic characteristics such as location and agglomeration economies. The use of this classification has proved to offer a better understanding of the patterns of distribution of employment (by gender, by industry or by level of qualification and occupation), job opportunities, and of the probabilities of being employed depending on the level of qualification or the degree of attractiveness of a region for working or living purposes. Some other questions spatially related to the performance of regional labour markets remain unanswered.

Future lines of research include the application of this classification to the study of labour economic issues such as the determinants of unemployment, inter-industrial labour mobility or the existence of over-qualification taking into account spatial factors (i.e. the type of analytical region where the potential worker lives) which are usually ignored. Agglomeration economies and distance play an important role in the location of economic activity, and therefore should affect the labour outcomes once the worker has decided to live in certain type of region. Obviously, such a decision does not have to be permanent, and workers can move in order to improve their labour opportunities. Therefore, a further possible question of relevance is the internal migration decisions between and within analytical regions. That is, can certain regularities be observed? Are people moving from peripheral regions to central or metropolitan areas or the other way round? Are internal migrations better explained in terms of size, i.e. in terms as counter-urbanization or urbanization? Are these movements linked to job opportunity decisions? Can we observe any differences according to their level of qualification? Even more, workers can be employed in a certain type of region but live in another, i.e., we could observe migration on a daily basis (commuting). Is one type of analytical region attracting workers or attracting residents? Do people tend to live and work in the same type of region? Could the analytical division be improved in order to specifically include the commuting criteria used in the local labour markets literature? We believe that these questions provide a fascinating and important future research agenda.

Resumen y conclusiones en castellano

Las economías de aglomeración son uno de los conceptos clave en Economía Regional y Urbana. A medida que los individuos se concentran en áreas urbanas densamente pobladas, principalmente las ciudades, se desencadenan una serie de efectos fundamentales para entender las dinámicas territoriales.

Al igual que las economías de aglomeración nos ayudan a comprender cuestiones económicas relacionadas con el crecimiento e innovación regional, las tendencias de especialización productiva o los procesos de concentración espacial de la actividad económica, también deberían contribuir a explicar las dinámicas espaciales que se observan en otras disciplinas, como por ejemplo la Economía Laboral.

Las teorías sobre los mercados de trabajo normalmente se desarrollan y contrastan a nivel nacional; sin embargo, sus principales conclusiones no suelen corroborarse a nivel supranacional o desagregando a nivel regional. Si nos centramos en los mercados de trabajo *locales* o *regionales*, los estudios suelen subrayar la importancia de las disparidades regionales existentes y su persistencia a lo largo del espacio y del tiempo. Pero la cuestión es si pudiese existir algún factor o factores comunes que contribuyan a explicar esas disparidades regionales.

La hipótesis de partida, presente a lo largo de los tres capítulos que componen esta tesis, es las economías de aglomeración podrían ser un factor clave. Para desarrollar esta hipótesis es necesario apartarse de la clasificación tradicional del territorio en regiones político-administrativas y buscar una clasificación espacial alternativa capaz de incluir las economías de aglomeración. Utilizando datos desagregados a nivel local, en el primer capítulo de esta tesis elegimos una clasificación territorial alternativa basada en el tamaño (economías de aglomeración) y localización (centro versus periferia). Aunque construida *a priori* de acuerdo con criterios económicos no vinculados directamente con Economía Laboral ni especialmente diseñada para el estudio de mercados de trabajo regionales, en este primer capítulo mostramos que esta clasificación analítica proporciona regiones más adecuadas –un mayor nivel de homogeneidad interna y de heterogeneidad en relación con diversas variables laborales-, que las regiones administrativas normalmente utilizadas.

Tras evaluar la idoneidad de esta clasificación alternativa para el análisis de los mercados de trabajo locales, en el segundo capítulo evaluamos la empleabilidad intentando identificar la existencia de patrones espaciales. Entre los factores que determinan la probabilidad de estar empleado se encuentran el nivel educativo, el sexo, la edad y las condiciones familiares; sin embargo, el lugar en que los individuos viven y trabajan suele ser, por norma general, ignorado. Nuestros resultados muestran que el tamaño y la localización del lugar en que el individuo vive son determinantes para comprender su empleabilidad.

Sin embargo, la reacción natural ante este hecho sería que los trabajadores migrasen hacia otro tipo de región con el fin de tener mayores oportunidades de empleo. Así, el objetivo del tercer capítulo es estudiar otro aspecto clave de los mercados de trabajo: la movilidad geográfica de los trabajadores, en particular desde la periferia hacia el centro.

A lo largo de los últimos 20 años España ha experimentado un significativo aumento de flujos de entrada a las regiones centrales, bien desde las regiones periféricas españolas o desde otras partes del mundo. De nuevo, haciendo uso de las regiones analíticas descritas en el primer capítulo, con las que podemos distinguir entre regiones centrales y periféricas, este capítulo examina la cadena de efectos que la llegada de nuevos trabajadores tiene en el centro. El análisis está basado en un modelo input-output de migración que, incorporando información sobre flujos de *commuting*, es ampliado para estudiar la distribución de *empleos* y de *residentes* a lo largo de los diferentes tipos de regiones centrales (básicamente áreas metropolitanas y grandes ciudades clasificadas por tamaño). Según este modelo, las economías de aglomeración también influyen sobre el patrón de distribución de residencias y de empleos: dependiendo de su tamaño, las ciudades pueden ser sitios atractivos para trabajar pero no para vivir (elevado precio de la vivienda, problemas de congestión u otras externalidades negativas) o viceversa.

Para finalizar, superando la división administrativa del territorio, la clasificación analítica utilizada tiene poder explicativo en temas de Economía Laboral donde está presente la dimensión espacial. El empleo de esta clasificación parece ofrecer una mejor comprensión de los patrones de distribución del empleo, de las oportunidades de empleo, la probabilidad de estar empleado o del grado de atractivo de una región con fines de residencia o de trabajo. Otras muchas cuestiones espaciales relacionadas esperan respuesta. Los resultados obtenidos en este estudio nos animan a continuar con esta línea de investigación, que en cierto sentido sirve de punto de encuentro entre la Economía Regional y Urbana con la Economía Laboral.

Bibliography

Ades, A. and Glaeser, E. (1995): “Trade and circuses: explaining urban giants”, *The Quarterly Journal of Economics*, 60, pp. 195-228.

Alonso, O. and Del Río, C. (2007): “The geographical concentration of unemployment: A male-female comparison in Spain”, *Regional Studies*, 42 (3), pp. 401-412.

Alonso, O., Del Río, C. and Toharia, L. (2009): “Un análisis espacial del desempleo por municipios”, *Revista de Economía Aplicada*, 49 (XXVII), pp. 47-80.

Andersen, A.K. (2002): “Are Commuting Areas Relevant for the Delimitation of Administrative Regions in Denmark?”, *Regional Studies*, 36 (8), pp 833 – 844.

Ball, R.M. (1980): “The use and definition of Travel-to-Work Areas in Great Britain: Some problems”, *Regional Studies*, 14 (2), pp. 125-139.

Bande, R., Fernández, M. and Montuenga, V. (2008): “Regional unemployment in Spain: Disparities, business cycle and wage setting”, *Labour Economics*, 15 (5), pp. 885-914.

Batty, M. (1983): “Linear urban models”. *Papers in Regional Science*, 53, pp. 1-25.

Beardsell, M. and Henderson, V. (1999): “Spatial evolution of the computer industry in the USA”, *European Economic Review*, 43, pp. 431-56.

Behrens, K. and Thisse, J.F. (2007): “Regional Economics: A New Economic Geography Perspective”, *Regional Science and Urban Economics*, 37, pp. 457-465.

Bentolila, S. and Blanchard, O.J. (1990): “Spanish unemployment”, *Economic Policy*, 10, pp. 233-281.

Berry M.J. and Linoff, G. (1997), *Data Mining Techniques for Marketing, Sales and Customer Support,* John Wiley and Sons.

Blanchard, O. J. and Jimeno, J. F. (1995): "Structural Unemployment: Spain versus Portugal", *American Economic Review*, 85 (2), pp. 277-292.

Boix, R. and Galletto, V. (2006), "Sistemas industriales de trabajo y distritos industriales marshallianos en España", *Economia Industrial*, 359, pp. 165–184.

Bover, O. and Velilla, P. (1999): "Migrations in Spain: Historical Background and Current Trends". Banco de España. Working Paper Nº 9909.

Casado-Diaz, J.M. (2000): "Local Labour Market Areas in Spain: A Case Study", *Regional Studies*, 34 (9), pp. 843-856.

Castells, M. (1976): *La Cuestión Urbana*, Siglo XXI Editores, México D.F.

CNIG (2001): *Mapas digitales de los municipios españoles*, Centro Nacional de Información Geográfica, Ministerio de Fomento.

Coffey, W.J. and Polèse, M. (1988): "La Transformation de L´espace Économique Canadien: ¿Assistons-nous à un Mouvement Centre-Péripherie?", *Revue d'économie régionale et urbaine*, 1.

Coombes, M.G. and Openshaw, S. (1982): "The Use and Definition of Travel-to-Work Areas in Great Britain: Some Comments", *Regional Studies*, 16 (2), pp. 141-149.

Coombes, M.G., Green, A.E. and Openshaw, S. (1986): "An Efficient Algorithm to Generate Official Statistical Reporting Areas: The Case of the 1984 Travel-to-Work Areas Revision in Britain", *Journal of the Operational Research Society*, 37 (10), pp. 943-953.

Davies, D.L. and Bouldin, D.W. (1979): "A cluster separation measure". *IEEE Transactions* on *Pattern Analysis and Machine Intelligence*, 1, pp. 224–227.

Decressin, J. and Fatás, A. (1995): "Regional labour market dynamics in Europe", *European Economic Review*, 39, pp. 1627-1655.

Desmet, K. and Fafchamps, M. (2005): "Changes in the spatial concentration of employment across US countries: a sectoral analysis (1972-2000)", *Journal of Economic Geography*, 5, pp. 261-284.

Dolado, J.J. and Jimeno, J.F. (1997): "The causes of Spanish unemployment: A structural VAR approach", *European Economic Review*, 41 (7), pp. 1281-1307

Duque, J.C., Artís, M. and Ramos, R. (2006): "The ecological fallacy in a times series context: evidence from Spanish regional unemployment rates", *Journal of Geographical Systems,* 8 (4), pp. 391-410.

Duranton, G. and Puga, D. (2002): "Diversity Specialization in Cities: Why, Where and Does it Matter", McCann, P. (Ed.): *Industrial Localization Economics*, pp. 151-186, Cheltenham-Edward Elgar.

Eaton, J. and Eckstein, Z. (1997): "Cities and growth: Theory and evidence from France and Japan", *Regional Science and Urban Economics,* 27(4-5), pp. 443-74.

Eurostat (2006): *Nomenclature of territorial units for statistics—NUTS. Statistical regions of Europe*, http://europa.eu.int/comm/eurostat/ramon/nuts/home_regions_en.html

Feria, J.M. and Susino, J. (Coord.), (2005): *Movilidad por razón de Trabajo en Andalucía*. Sevilla: Instituto de Estadística de Andalucía, Estadísticas Demográficas, 2005. 177 p

Fernández-Vázquez, E., García-Muñiz, A. S. and Ramos-Carvajal, C. (2010): "The impact of immigration on interregional migrations: an input-output analysis with an application for Spain"; *Annals of Regional Science*, (forthcoming). Online publication date: 11-Dec-2009.

Fischer, M.M. (1980): “Regional taxonomy: A comparison of some hierarchic and non-hierarchic strategies”, *Regional Science and Urban Economics*, 10, pp. 503-537.

Fujita M., Krugman, P. and Venables, A. (1999): *The Spatial Economy: Cities, Regions and International Trade*, Cambridge, MA, MIT Press.

Gabaix, X. (1999): “Zipf´s law for cities: an explanation”, *The Quarterly Journal of Economics*, pp. 739-767.

Garin. R.A. (1966): “A matrix formulation of the Lowry model for intra-metropolitan activity location”, *Journal of the American Institute of Planners*, 32, pp. 361-364.

Glaeser E. and Maré, D. (2001): “Cities and skills”, *Journal of Labour Economics,* 19, pp. 316–342.

Glaeser, E.L., Scheinkman, J. and Shleifer, A. (1995): “Economic Growth in a Cross-Section of Cities”, *Journal of Monetary Economics*, 36, pp.117-143.

Glaeser, E.L. (1998): “Are Cities Dying?”, *Journal of Economics Perspectives*, 12 (2), pp. 139-160.

Gordon, A.D. (1996): “A survey of constrained classification”, *Computational Statistics &Data Analysis*, 21, pp. 17-29.

Gordon, A.D. (1999): *Classification,* (2nd ed.). Monographs on statistics and applied probability, Chapman & Hall.

Gotway C.A., Young L.J. (2002): “Combining incompatible spatial data”. *Journal American Statistics Association*, 97, pp. 632–648.

Greenland S. and Morganstern H. (1989): “Ecological bias, confounding, and effect modification”. *International Journal Epidemiology*, 18, pp. 269–274.

Guldman, J.M. and Wang, F. (1998): "Population and employment density function revisited: a spatial interaction approach". *Papers in Regional Science*, 77 (2), pp. 189-211.

Halkidi, M., Y. Batistakis and Vazirgiannis, M. (2001): "On Clustering Validation Techniques", *Journal of Intelligent Information Systems*, 17 (2/3), pp. 107-145.

Hall, P. (1999): *Cities and civilization*, Phoenix Giant, London.

Head, K. and Mayer, T. (2006): "Regional wage and employment responses to market potential in the EU", *Regional Science and Urban Economics,* 36, pp. 573-594.

Henderson, J.V. and Thisse, J.F. (1997): "On the Strategic Community Development," *CEPR Discussion Papers* 1550, C.E.P.R. Discussion Papers.

INE (1991): *Censo de Población, 1991*, Instituto Nacional de Estadística (available online at www.ine.es).

INE (2007): *Censo de Población y Viviendas, 2001*, Instituto Nacional de Estadística, Madrid (available online at www.ine.es).

INE (2009): *Encuesta de Variaciones Residenciales*, series 1998-2008. Instituto Nacional de Estadística, Madrid (available online at www.ine.es).

ISTAT (1997), I sistema locali del lavoro 1991, *Argomenti*, núm. 10, Istituto Nazionale di Statistica.

ISTAT (2005), *I sistema locali del lavoro. Censimento 2001. Dati definitivi*, Istituto Nazionale di Statistica [available online at www.istat.it].

ISTAT-IRPET (1989), *I mercati locali del lavoro in Italia*, Franco Angeli Libri, Istituto Centrale di Statistica –Istituto Regionale per la Programmazione Economica della Toscaza, Milano.

Jacobs, J. (1984): *Cities and the wealth of nations*, Vintage, New York.

Jimeno, J.F. and Bentolila, S. (1998): “Regional unemployment persistence: Spain 1976–94”, *Labour Economics*, 5, pp. 25–51.

Jun. M.J. (2005): “Forecasting urban land-use demand using a metropolitan input-output model”, *Environment and Planning A*, 37, pp. 1311-1328.

López-Bazo, E., del Barrio, T. and Artís, M. (2005): “Geographical distribution of unemployment in Spain”, *Regional Studies*, 39 (3), pp. 305-318.

Lowry I.S. (1964): *A model of metropolis*. Rand Corporation, Santa Monica, California.

McGill, S.M. (1997): “The Lowry model as an input-output model and its extension to incorporate full intersectoral relations”. *Regional Studies*, 11 (5), pp. 337-354.

Murtagh, F. (1985): “A Survey of Algorithms for Contiguity-constrained Clustering and Related Problems”. *The Computer Journal*, 28 (1), pp. 82-88.

Openshaw, S. (1984): “The modifiable areal unit problem”. *Concepts and techniques in modern geography*, 38, GeoBooks, Norwich

Papps, K.L. and Newell, J.O. (2002): “Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data”, IZA Discussion Paper 443.

Parr, J. (2002): “Agglomeration economies: ambiguities and confusions”, *Environment and Planning A*, 34, pp. 717-731.

Partridge, M.D., Rickman, D.S., Ali, K. and Olfert, M.R. (2008): “Employment growth in the American urban hierarchy: long live distance”, *The B.E. Journal of Macroeconomics,* (Contributions), 8(1), pp 1-36.

Partridge, M.D., Rickman, D.S., Ali, K. and Olfert, M.R. (2009): "Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy", *Journal of International Economics,* 78 (1), pp. 126-140.

Phelps, N.A. and Ozawa, T. (2003): "Contrasts in agglomeration: proto-industrial, industrial and post-industrial forms compared", *Progress in Human Geography,* 27, pp. 583-604.

Phelps, N., Fallon, R. and Willians, C. (2001): "Small firms, borrowed size and the urban-rural shift", *Regional Studies*, 53 (7), pp. 613-624.

Piantadosi S., Byar D.P., Green S.B. (1988): "The ecological fallacy", *American Journal of Epidemiology*, 127, pp. 893–904.

Polèse, M. (2005): "Cities and National Economic Growth: A Reappraisal", *Urban Studies*, 42 (8), pp. 1429-1451.

Polèse, M. (2009): *The Wealth and the Poverty of Regions: Why Cities Matters*, University of Chicago Press, Chicago.

Polèse, M. and Champagne, E. (1999): "Location matters: Comparing the distribution of economic activity in the Mexican and Canadian urban systems", *International Journal Science Review*, 22 (1), pp. 102-132.

Polèse, M., and Shearmur, R. (2004): "Is Distance Really Dead? Comparing the Industrial Location Patterns over Time in Canada", *International Regional Science Review*, 27 (4), pp. 1-27.

Polèse M., Shearmur, R. and Rubiera, F. (2007): "Observing regularities in location patters. An analysis of the spatial distribution of economic activity in Spain", *European Urban & Regional Studies*, 14 (2), pp. 157-180.

Porter, M. (1990): *The competitive advantage of nations*, Free Press, New York.

Quigley, J.M. (1998): "Urban Diversity and Economic Growth". *Journal of Economic Perspectives*, 12 (2), pp. 127-138.

Redding, S. and Venables, A., (2004): "Economic geography and international inequality", *Journal of International Economics*, 62(1), pp. 53-82.

Richardson S. (1992): "Statistical methods for geographical correlation studies", in Elliot P, Cuzick J., English D., Stern R. (eds.): *Geographical and environmental epidemiology: methods for small area studies,* Oxford University Press, New York, pp. 181–204.

Richardson S., Stucker L. and Hemon D. (1987): "Comparison of relative risks obtained in ecological and individual studies: some methodological considerations". *International Journal of Epidemiology*, 16, pp. 111–120.

Robinson W.S. (1950): "Ecological correlations and the behaviour of individuals", *American. Sociological Review,* 15, pp. 351-357.

Rosenthal, S.S. and Strange C.W. (2001): "The determinants of agglomeration", *Journal of Urban Economics*, 50 (2), pp. 191–229.

Rubiera, F. (2006): *Ciudades, Crecimiento y Especialización Territorial. Dinámicas Espaciales de Concentración del Empleo y la Población en España*, Consejo Económico y Social. Madrid (España).

Salinas, J. and Álvarez, S. (Coord.) (2003): *El Gasto público en la Democracia.* Estudios en el XXV aniversario de la Constitución Española de 1978. Instituto de Estudios Fiscales. Madrid.

Sevilla, J. (2005): *Financiación autonómica.* Problemas y propuestas, Madrid, exlibris ediciones.

Sforzi F., Openshaw S. and Wymer C. (1997) Le procedura di identificazione dei sistemi locali del lavoro [The procedure to identify local labour market area], in Sforzi F. (Ed.) *I sistemi locali del lavoro 1991*, pp. 235–242. ISTAT, Rome.

Shearmur, R. and Polèse, M. (2007): "Do Local Factors Explain Local Employment Growth?: Evidence from Canada, 1971-2001", *Regional Studies*, Vol. 41.4, pp. 453–471, June 2007.

Smart, M.W. (1974): "Labour market areas: uses and definitions". *Progress in Planning*, 2, pp. 239-353.

Theil, H. (1967): *Economics and information theory*, North-Holland, Amsterdam.

Toboso, F. (2005): "Algunas claves organizativas del Estado de las Autonomias Español y sus consecuencias". *Revista Asturiana de Economia*, 32, pp. 43-66.

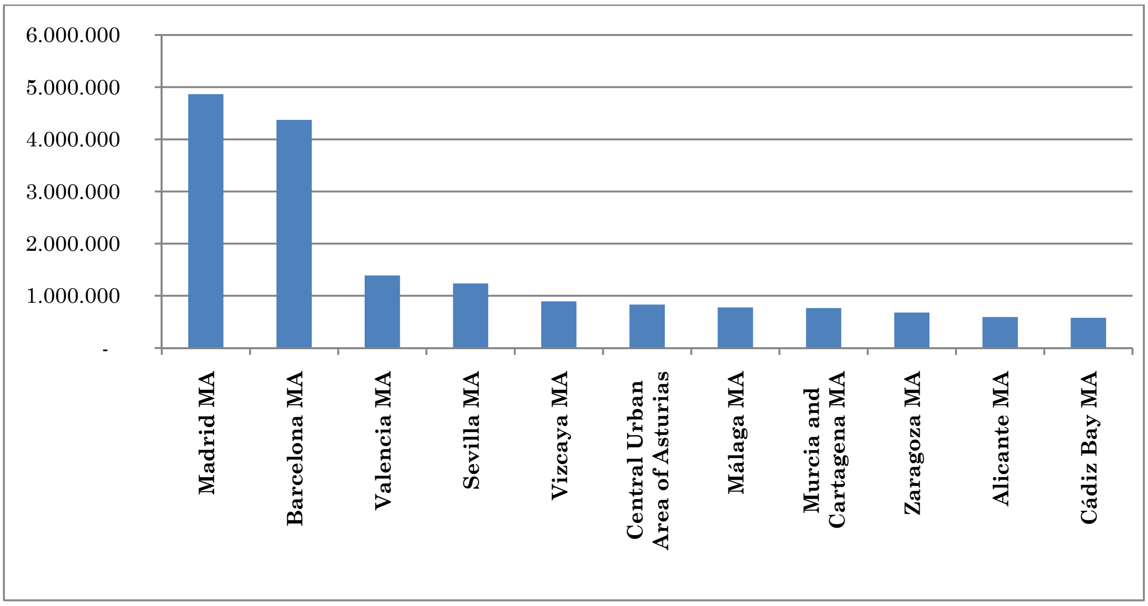
Vining, D.R. and Pallone, R. (1982): "Migration between core and peripheral regions: a description and tentative explanation of the patterns in 22 countries", *Geoforum*, 13 (4), pp. 339-410.

Watts, M. (2004): "Local labour markets in New South Wales: fact or fiction?", *Centre of full employment and equity working paper*, no. 04-12, November [http://e1.newcastle.edu.au /coffee/pubs/wp/2004/04-12.pdf].

Zipf, G. (1949): *Human behaviour and the principle of least effort*. Addison-Wesley, Reading.

Appendixes

Appendix I. Population of the Spanish Metropolitan Areas and the other Analytical Regions



SPANISH ANALYTICAL REGION	Total Population	Number of municipalities
MA1		
Madrid MA	4,866,821	30
Barcelona MA	4,372,091	162
MA2		
Valencia MA	1,389,585	47
Sevilla MA	1,237,066	27
Vizcaya MA	895,086	34
Central Urban Area of Asturias	832,843	18
Málaga MA	776,744	15
Murcia and Cartagena MA	766,222	14
Zaragoza MA	679,721	28
Alicante MA	592,230	13
Cádiz Bay MA	577,756	6
CUA1	2,163,392	74
CUA2	3,669,212	171
CRA	3,014,919	1,694
PUA1	5,587,002	259
PUA2	3,970,633	193
PRA	5,318,132	5,321
Total	**40,709,455**	**8,106**

Appendix II. Decomposition of the Theil Inequality Index

Applying the Theil inequality index (Theil 1967) to employment, the formula is computed as follows:

$$T = \sum_{m=1}^{n} \frac{PopEmp_{municip}}{PopEmp_{Esp}} \log\left(\frac{PopEmp_{municip} / PopEmp_{Esp}}{1/n} \right)$$

where n is the number of municipalities considered (8,106), $PopEmp_{municip}$ is the population employed in municipality m, and $PopEmp_{Esp}$ represents the Spanish working population.

The Theil index can be completely and perfectly decomposed into a *between-group* component (Tg) and a *within-group* component (Tw). Intraregional homogeneity can be therefore quantified by the within-group component. Thus:

$$T = Tg + Tw$$

with

$$Tg = \sum_{r=1}^{R} \frac{PopEmp_r}{PopEmp_{Esp}} \log\left(\frac{PopEmp_r / PopEmp_{Esp}}{n_r / n} \right)$$

$$Tw = \sum_{r=1}^{R} \frac{PopEmp_r}{PopEmp_{Esp}} \sum_{m=1}^{n} \frac{PopEmp_{municip}}{PopEmp_r} \log\left(\frac{PopEmp_{municip} / PopEmp_r}{1/n_r} \right)$$

where r indexes regions, with n_r representing the number of municipalities in region r and $PopEmp_r$ the population employed in the region r to which the municipality belongs.

As the within component quantifies the heterogeneity between the individuals of a region, small values indicate the existence of a high degree of internal homogeneity.

Appendix III. The Davies-Bouldin Validation Index

This index (Davies and Bouldin, 1979) is a function of the ratio of the sum of within-region scatter to between-region separation, and is defined as:

$$DB_m = \frac{1}{m}\sum_{i=1}^{m} R_i$$

where $R_i = max_{j=1,\ldots\ldots,mj \neq i} R_{ij}, i = 1,\ldots., m$ and m is the number of regions in which the Spanish territory is divided for each classification.

Then, the similarity index R_{ij} between region i (R_i) and region j (R_j) is defined as:

$$R_{ij} = \frac{S_i + S_j}{d_{ij}}$$

where S_i is a measure of dispersion of R_i and $d(C_i, C_j) \equiv d_{ij}$ the dissimilarity between two regions. The index R_{ij} satisfies the following:

1. $R_{ij} \geq 0$
2. $R_{ij} = R_{ji}$
3. If $S_i = 0$ and $S_j = 0$, then $R_{ij} = 0$
4. if $S_j > S_k$ and $d_{ij} = d_{ik}$, then $R_{ij} > R_{ik}$
5. if $S_j = S_k$ and $d_{ij} < d_{ik}$ then $R_{ij} < R_{ik}$.

The dissimilarity between region R_i and region R_j, in a l-dimensional space is defined as:

$$d_{ij} = \left\| \overline{EmpPop_i} - \overline{EmpPop_j} \right\| = \sqrt{\sum_{k=1}^{l} \left| \overline{EmpPop_{ik}} - \overline{EmpPop_{jk}} \right|^2}$$

and the dispersion of a region R_i is defined as:

$$S_i = \sqrt{\frac{1}{n_i}\sum_{x \in R_i} \left\| EmpPop - \overline{EmpPop_i} \right\|^2}$$

As DB_m is the average similarity between each region and its most similar one, small values of DB_m are indicative of the presence of compact and well-separated regions. The DB_m index exhibits no trends with respect to the number of regions.

Appendix IV. Type of job and level of qualification according to the 2001 Census classification and broad aggregation

CN01 CLASSIFICATION (2001 Spanish Census)	**AGGREGATION**
Business and public administration management/managers	*Highly-skilled personnel: directors, highly-qualified professionals and skilled technical staff*
Technical staff and scientific and intellectual professionals	
Technical and professional support staff	*Medium-skilled personnel: technical staff and administrative support staff*
Administrative staff	
Skilled workers in agriculture and fisheries	
Artisans and qualified workers in manufacturing, construction and mining excluding machine operators and installation workers	
Hostelry, security and retail sales workers	*Low-skilled personnel: tertiary workers and industrial, artisanal and agricultural operators/labourers*
Machine operators, installation workers and fitters.	
Unskilled workers	

Appendix V. Decomposition of the Theil inequality index by level of qualification and occupation (9 categories)

		Administrative Regions			Analytical Regions
		NUTS III (PROV) (50 Regions)	NUTS II (CCAA) (17 Regions)	NUTS I (7 Regions)	MA1 to PRA (8 Regions)
Business and public administration management/managers					
Theil's Index	*Between*	1.2665	1.2534	0.4419	0.6264
2.5227	*Within*	1.2563	1.2693	2.0809	1.8963
Technical staff and scientific and intellectual professionals					
Theil's Index	*Between*	1.5881	1.5795	0.5218	0.7260
3.2942	*Within*	1.7060	1.7146	2.7723	2.5682
Technical and professional support staff					
Theil's Index	*Between*	1.5722	1.5577	0.5548	0.7867
3.0469	*Within*	1.4747	1.4892	2.4921	2.2602
Hostelry, security and retail sales workers					
Theil's Index	*Between*	1.3473	1.3360	0.4774	0.6759
2.6168	*Within*	1.2695	1.2808	2.1394	1.9409
Skilled workers in agriculture and fisheries					
Theil's Index	*Between*	0.3152	0.3113	0.2496	0.3604
1.0321	*Within*	0.7170	0.7209	0.7826	0.6717
Artisans and qualified workers in manufacturing, construction and mining excluding machine operators and installation workers					
Theil's Index	*Between*	1.1107	1.0911	0.3619	0.5742
2.0863	*Within*	0.9756	0.9952	1.7243	1.5121
Machine operators, installation workers and fitters					
Theil's Index	*Between*	1.1173	1.0859	0.3254	0.5464
2.1345	*Within*	1.0172	1.0487	1.8092	1.5882
Unskilled workers					
Theil's Index	*Between*	1.1214	1.1103	0.4990	0.6771
2.3019	*Within*	1.1805	1.1916	1.8028	1.6248

Source: Authors' calculations based on 2001 Spanish Census (INE, 2007).

Appendix VI. Variables and Data Used in the Empirical Approach

Variables		**Database**
EMPYN	Employed: Yes/No	Dummy variable that provides information about an individual's labour situation: employed *vs* non employed. This variable is constructed using the labour data included in the 2001 Spanish Census (2007). The Census is administered by the Spanish National Institute of Statistics, INE.
MA1, MA2	Metropolitan areas: more than 500,000 inhabitants	See Table 1. The distances to a metropolitan area are calculated using the digital maps of the CNIG (Spanish National Centre for Geographical Research). Information from the National Government Ministry of Infrastructures and Public Works was used for the delimitation of the metropolitan areas.
CUA1, CUA2	Urban (more than 10.000 but less than 500.000 inhabitants) Central Areas	
PUA1, PUA2	Urban (more than 50.000 but less than 500.000 inhabitants) Peripheral Areas	
CRA	Rural Central areas	
PRA	Rural Peripheral areas	
AGE	Age	For each individual included in the Spanish Census there is information about the year of birth. We aggregate the variable AGE into 4 groups: less than 25, between 25 and 29, between 30 and 44, and more than 44 years of age.
MS	Marital Status	The 2001 Census distinguishes between 5 categories: single, married, widow, separated and divorced. For our purposes we aggregated the separated and divorced individuals.
EDL	Educational level	There are 10 different categories for this variable in the 2001 Spanish Census. For our purposes we aggregate them in 5 categories: without studies, basic studies, secondary education, vocational training and university education
FORE	Foreigner: Yes/No	Dummy variable that establishes if an individual is a foreigner or was born in Spain. The 2001 Census includes information about the country of birth for each individual.
CHBA4	Children under age of 4: Yes/No	Dummy variable reflecting the existence of children under the age of 4.
CHAA4	Children above age of 4: Yes /No	Dummy variable reflecting the existence of children above the age of 4.
Spatial unit of the analysis: Spanish municipalities		

Appendix VII. Spanish total population and population inflows from abroad (1988-2008)

	Population Inflows from abroad	Spanish Total Population	Percentage
1988	24,352	39,092,185	0.06%
1989	33,883	39,415,407	0.09%
1990	33,913	39,752,852	0.09%
1991	24,301	38,745,531	0.06%
1992	38,825	39,001,640	0.10%
1993	33,003	39,656,225	0.08%
1994	34,071	40,088,586	0.08%
1995	36,031	40,319,957	0.09%
1996	30,791	39,719,693	0.08%
1997	57,776	n.a.	
1998	81,140	39,719,990	0.20%
1999	127,208	40,071,101	0.32%
2000	362,331	40,357,873	0.90%
2001	414,700	40,971,951	1.01%
2002	483,174	41,692,162	1.16%
2003	469,831	42,573,283	1.10%
2004	683,784	43,053,851	1.59%
2005	715,860	43,966,572	1.63%
2006	838,093	44,564,112	1.88%
2007	956,563	45,049,712	2.12%
2008	724,383	46,002,043	1.57%

Source: Authors' calculations based on Historical Population Series and Census of Residential Variations. 1988-2008.

Appendix VIII. Population inflows from abroad by province of destination (2008)

Province	Immigrant Population and Percentage				
Madrid	135,157	18.66%	**Cádiz**	6,097	0.84%
Barcelona	128,843	17.79%	**Huelva**	5,383	0.74%
Valencia	37,292	5.15%	**Ciudad Real**	5,352	0.74%
Alicante	34,880	4.82%	**Rioja**	5,216	0.72%
Málaga	28,347	3.91%	**Burgos**	4,746	0.66%
Baleares	25,723	3.55%	**Albacete**	4,723	0.65%
Murcia	23,996	3.31%	**Álava**	4,389	0.61%
Almería	20,547	2.84%	**Valladolid**	4,373	0.60%
Girona	20,276	2.80%	**Guadalajara**	4,207	0.58%
Palmas (Las)	18,359	2.53%	**Cuenca**	3,983	0.55%
Zaragoza	17,865	2.47%	**Córdoba**	3,933	0.54%
Sta. Cruz de Tenerife	17,141	2.37%	**León**	3,778	0.52%
Tarragona	17,032	2.35%	**Salamanca**	3,622	0.50%
Sevilla	12,274	1.69%	**Huesca**	3,330	0.46%
Vizcaya	11,106	1.53%	**Badajoz**	3,325	0.46%
Lleida	10,462	1.44%	**Jaén**	3,067	0.42%
Toledo	9,862	1.36%	**Lugo**	3,019	0.42%
Navarra	9,528	1.32%	**Ourense**	2,970	0.41%
Asturias	9,465	1.31%	**Teruel**	2,853	0.39%
Coruña (A)	8,813	1.22%	**Segovia**	2,266	0.31%
Castellón	8,375	1.16%	**Ávila**	1,915	0.26%
Granada	8,359	1.15%	**Cáceres**	1,738	0.24%
Guipúzcoa	8,129	1.12%	**Palencia**	1,308	0.18%
Pontevedra	7,363	1.02%	**Soria**	1,278	0.18%
Cantabria	7,061	0.97%	**Zamora**	1,257	0.17%
TOTAL				**724,383**	**100%**

Source: Authors' calculations based on Historical Population Series and Census of Residential Variations. 2008.

Appendix IX. Matrices of multipliers for Spain (1991-2001)

Table IX. 1. Matrix of multipliers $[\boldsymbol{I}-\boldsymbol{B}]^{-1}$.

	MA1	MA2	CUA1	CUA2	CRA
MA1	1.1427	0.1676	0.0796	0.5448	0.5703
MA2	0.2376	1.1964	0.1128	0.5529	0.4724
CUA1	0.1798	0.1917	1.0751	0.3760	0.4679
CUA2	0.1409	0.1392	0.0633	1.1664	0.2738
CRA	0.1316	0.1180	0.0614	0.2466	1.1459

Table IX. 2. Matrix of multipliers $\boldsymbol{M}^{*}=[\boldsymbol{I}-\boldsymbol{B}]^{-1}\boldsymbol{C}$.

	MA1	MA2	CUA1	CUA2	CRA
MA1	1.3095	0.2626	0.1175	0.4650	0.3503
MA2	0.4624	1.1548	0.1529	0.4798	0.3223
CUA1	0.3734	0.2663	0.9440	0.3692	0.3376
CUA2	0.4314	0.2744	0.1062	0.7321	0.2395
CRA	0.4806	0.2381	0.1219	0.3184	0.5444

Table IX. 3. Matrix of multipliers $\boldsymbol{\Delta}=[\boldsymbol{I}-\boldsymbol{B}]^{-1}[\boldsymbol{C}-\boldsymbol{I}]$.

	MA1	MA2	CUA1	CUA2	CRA
MA1	0.1668	0.0950	0.0380	-0.0798	-0.2200
MA2	0.2248	-0.0416	0.0400	-0.0731	-0.1501
CUA1	0.1935	0.0746	-0.1311	-0.0068	-0.1303
CUA2	0.2905	0.1352	0.0429	-0.4343	-0.0343
CRA	0.3490	0.1201	0.0605	0.0718	-0.6016