

Labour Market Areas and cluster subdivision for Spain 2011

Lucas Martínez-Bernabéu¹, José Manuel Casado-Díaz¹

¹ University of Alicante, Alicante, Spain

Received: 21 April 2022/Accepted: 20 June 2022

Abstract. Labour market areas (LMAs) are a type of functional areas that provide the appropriate framework for spatially-explicit analyses of the labour market to guide policymaking in this field. This resource presents a data base with the most recent definition of LMAs for Spain using a state-of-the-art methodology. All defined LMAs are highly self-contained in terms of travel-to-work flows and exceed a minimum size compatible with the usual requirements in terms of confidentiality and representativeness, allowing the recoding of microdata geographic referencing and therefore a much greater granularity compared to what is usual for this type of data. It can be downloaded from <https://rua.ua.es/dspace/handle/10045/121236>.

1 Introduction

Labour market areas (LMAs) are functional regions beyond the administrative boundaries, defined for purposes of compiling, reporting, and evaluating employment, unemployment, workforce availability, and related topics. An LMA is an economically integrated territory where the majority of people live and work. The boundaries of these regions are designed to ensure self-containment in terms of commuting flows; they embrace territorial units that present a high interaction among them, and by doing so, LMAs also tend to be self-contained in related socioeconomic activities such as housing, shopping, and leisure.

LMAs are potentially the most appropriate spatial units for sub-national economic policy (Brandmueller et al. 2017, Coombes et al. 2012, Stabler et al. 1996). The sensitivity of statistical econometric models with respect to different levels of spatial aggregation, the so-called ‘modifiable areal unit problem’, is a well-known issue (Openshaw, Taylor 1979). When the spatial econometric analysis is performed over well-defined LMAs, the spillover effects between regions are more consistent and less intense, thus easier to handle with spatial econometric techniques or safer to ignore (Osland, Thorsen 2013, Stimson et al. 2016, 2018). Another advantage over using administrative regions is that LMA boundaries can be updated periodically to reflect changing patterns in socioeconomic relations, using up-to-date data on commuting flows (Eurostat 2020, Martínez-Bernabéu, Casado-Díaz 2021). Defining these regions consistently in all countries allows for international statistical comparability of many socioeconomic aspects (OECD 2002, 2020, Halás et al. 2019), as sought by the Global Statistical Geospatial Framework in its third principle: common geographies for dissemination of statistics (United Nations 2019). The advantages of policy relevance, ease of use in econometric analysis, practical updatability, and international comparability have prompted numerous European countries to devise official LMA definitions (Eurostat 2020, 2021).

Spain does not have an official regionalization of its territory into LMAs or other forms of functional regions based on commuting flows. In the statistical analysis of the Spanish socioeconomic phenomena, the Administration and researchers most often use municipalities (LAU2), provinces (NUTS3), or autonomous communities (NUTS2) as the spatial framework. These administrative regions were historically or arbitrarily defined in the past two centuries, and do not generally represent the spatial labour and socioeconomic reality of Spain.

2 Background

The data on commuting flows from the latest census (2011) was made available to researchers (under contract) in 2014. Using this data, [Martínez-Bernabéu, Casado-Díaz \(2016\)](#) developed a set of functional, highly self-contained, LMAs whose population exceeds 20,000 inhabitants in both 2001 and 2011. The minimum of 20,000 inhabitants reflects the specific threshold that, as happens in other countries, is imposed by the Spanish National Statistics Office (Instituto Nacional de Estadística) in order to comply with confidentiality regulations and sample representativeness in microdata dissemination. In this way, it is possible to recode the geographic references of the 2001 and 2021 census microdata (and potentially also of other data from statistical operations with large samples), so that the geographic reference to the place of residence, work, origin (in the case of migrants), among others, becomes a functional unit instead of an administrative one. In the Spanish case, for example, the geographical reference in census microdata is the municipality only when it has more than 20,000 inhabitants; otherwise, only information on the province and the size range of the municipality of residence is given, which greatly limits the analyses, especially in less densely populated settings. LMAs are further subdivided into clusters of at least 20,000 inhabitants (in both 2001 and 2011) to enhance geographical detail in microdata analysis when finer-grained resolution is more important than functional self-containment.

The method used to define these LMAs was a grouping evolutionary algorithm (GEA) previously applied to the 2001 Spanish Census by [Martínez-Bernabéu et al. \(2012\)](#). This optimization algorithm maximises the overall internal cohesion (integration) of LMAs by iteratively making stochastic changes in the partitioning of municipalities into LMAs and selecting the ones with the highest cohesion, subject to restrictions of minimum levels of labour force self-containment and population size of the LMAs. This method proved itself capable of identifying more LMAs with similar average self-containment than other methods like the well-established Travel-To-Work Areas method ([Coombes, Bond 2008](#)). It was also successfully applied to Chile ([Casado-Díaz et al. 2017](#)), and the results were shared with the academic community ([Rowe et al. 2017](#)).

The above-mentioned proposal of LMAs for Spain 2011 was further improved with an enhanced methodology in [Martínez-Bernabéu, Casado-Díaz \(2021\)](#). It uses a combination of heuristics along with GEA to implement a three-step optimization process. In this process, the input matrix of commuting flows is first compacted in a simple hierarchical clustering algorithm that groups together municipalities with enough inter-dependence in terms of the ratio of residents that commute between areas. Then, GEA performs the regionalisation of the compacted matrix into LMAs. Finally, a greedy heuristic further optimizes the resulting LMAs in terms of local indicators on self-containment and commuting interaction between each municipality and the neighbouring LMAs, allowing the compacted municipalities to be reallocated independently. The results showed significant improvements in terms of dependence between municipalities and their LMA and greatly reduced the appearance of arguably inconsistent regions, all in all making the final phase of adjustments significantly easier and faster. The present resource paper publishes the geography resulting from this improved methodology and makes it available to researchers and practitioners.

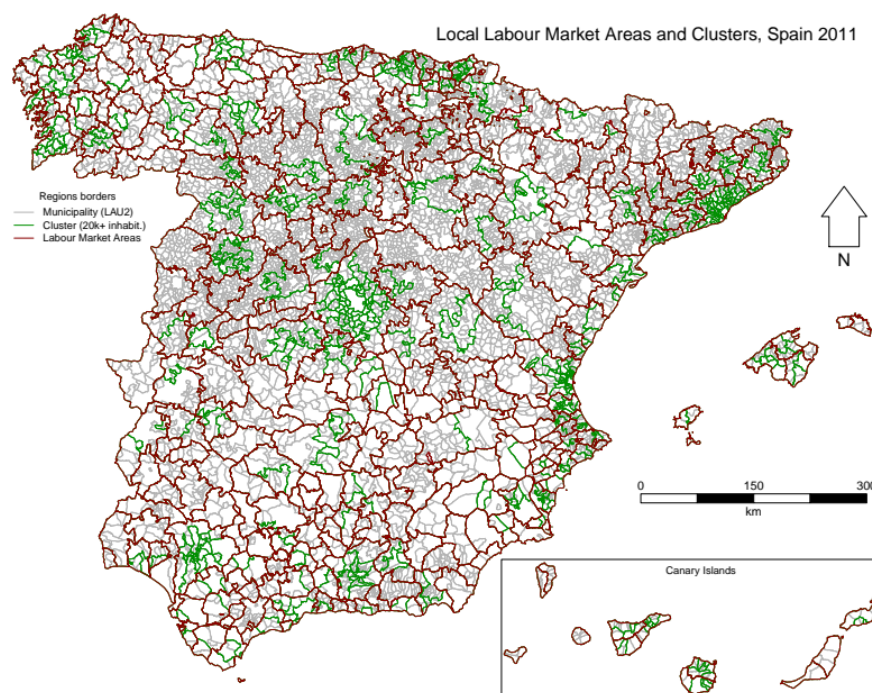


Figure 1: LMA and cluster borders

3 Description of the resource

The data set includes 287 LMAs, further subdivided into 822 clusters, providing full coverage of the Spanish territory (Figure 1). The linked resource contains the correspondence between municipalities (the basic territorial units considered here), cluster and LMA, in CSV and XLS format. The XLS file also contains statistics on area, inhabitants, working residents, jobs, and supply- and demand-side self-containment for individual LMAs and clusters. The boundaries of municipalities, clusters, and LMAs are provided in Shapefile format. It also includes a PDF file with a plotted map of the LMA and cluster boundaries.

These LMAs were defined to maximise internal commuting interaction while reaching a minimum self-containment of 75% for populations of at least 20,000 inhabitants (except for the case of the island El Hierro in the Canary Islands, whose whole population is 9922), linearly relaxed to 66.7% for populations over 100,000 inhabitants. These functional areas exhibit high degrees of internal cohesion and external self-containment in terms of travel-to-work flows and, accordingly, are an appropriate geography for spatially-explicit studies that benefit from reduced spillover effects between regions.

The clusters within each LMA were defined with the priority of maximising geographical detail (number of separate regions). When several groupings with the same number of regions were possible, the one that maximised internal cohesion was chosen. These regions cannot be considered as functional areas due to their limited level of self-containment. However, they allow researchers to perform microdata spatial analyses that benefit from finer-grained geographical resolutions (median area is 291 km² for clusters and 1343 km² for LMAs). Although lower levels of self-containment imply larger spatial spillovers, this effect is nuanced by the fact that such spillovers are more easily handled because they are mostly limited to other clusters within the same LMA.

4 Potential applications

A definition of LMAs for Spain entails an opportunity to better understanding the structure and dynamics of labour markets and their effects on other socioeconomic phenomena,

particularly since its use reduces the impact of the MAUP issue when performing spatially-explicit analyses. The literature provides numerous and varied examples of analyses that benefit from using functional instead of administrative regions. As an illustration of such studies, [Kosfeld et al. \(2006\)](#) investigate regional convergence of labour productivity and income per capita in Germany using LMAs to avoid distortion due to commuter flows between administrative districts. [Autor et al. \(2015\)](#) analyse the simultaneous impacts of technology and trade on US employment levels and job composition, juxtaposing their effects across local labour markets, over time, between sectors and occupations, and among workers of different education, age and sex categories, using commuting cones that approximate local labour markets as the appropriate spatial unit of analysis. In the UK, [Jacob et al. \(2019\)](#) study the impacts of increased commuting on well-being, considering the characteristics of the local labour market and differentiated by gender. [Coile \(2021\)](#) also use US commuting zones to analyse the employment rates at older ages. [Stimson et al. \(2018\)](#) measure the endogenous performance of regional markets across Australia, and recommend to focus the efforts of policy-makers on region-specific development policies. [Eliasson et al. \(2021\)](#) study geographical disparities in human capital and income in Finland and Sweden. In Chile, [Carriel et al. \(2022\)](#) identify and explain the determinants of long-distance commuting among different sectors, worker skills and levels of education, and [Goya \(2022\)](#) analyse the Marshallian and Jacobian externalities in creative industries across Chile, both works supporting the use of LMAs in the analyses. [Vallone, Chasco \(2020\)](#) evaluate, also in Chile, the influence of spatial proximity on the evolution of cities to detect regional differences in their spatio-temporal dynamics, and they use LMAs as a robustness check.

In addition to the advantages associated with the usual uses of LMAs as the reference unit for the study of labour phenomena at the local level, it should be added that in this specific case they have been defined in such a way that they allow, as has been pointed out, the recoding of the geographical references in the census microdata (in this case those corresponding to two waves, those of 2001 and 2011). This is a substantial improvement for two reasons. On the one hand, it makes it possible to incorporate contextual variables (associated with characteristics of the LMA of residence, for example) into the individual records, thus substantially enriching the information available for each individual when carrying out analyses based on microeconomic models. On the other hand, when the unit of analysis is a territorial unit instead of the individual, the availability of this recodified microdata makes it possible to define a large number of variables that can be calculated as the average value of the variables in the individuals who reside (or work, for example) in the territory analysed or by calculating the relative weight of a specific characteristic (or an ad hoc combination of them) in that territory, substantially enlarging the range of territorial characteristics available when the information is disseminated at the level of municipalities (LAU 2), without the loss of spatial resolution that implies using provinces (NUTS 3).

Acknowledgement

This work was supported by grants PID2020-114896RB-I00 (Agencia Estatal de Investigación, AEI / 10.13039/501100011033; Spanish State Plan for Scientific and Technical Research and Innovation 2017-2020) and AICO/2021/062 (Generalitat Valenciana, Department of Innovation, Universities, Science and Digital Society; R&D&I Programme of the Comunidad Valenciana).

References

- Autor DH, Dorn D, Hanson GH (2015) Untangling trade and technology: Evidence from local labour markets. *The Economic Journal* 125: 621–646. [CrossRef](#)
- Brandmueller T, Schäfer G, Ekkehard P, Müller O, Angelova-Tosheva V (2017) Territorial indicators for policy purposes: NUTS regions and beyond. *Regional Statistics* 7: 78–89. [CrossRef](#)
- Carriel V, Lufin M, Pérez-Trujillo M (2022) Do workers negative self-select when they commute? Evidence for the Chilean case of long-distance commuting. *The Annals of Regional Science*: 1–25. [CrossRef](#)
- Casado-Díaz JM, Martínez-Bernabéu L, Rowe F (2017) An evolutionary approach to the delimitation of labour market areas: An empirical application for Chile. *Spatial Economic Analysis* 12: 379–403. [CrossRef](#)
- Coile C (2021) The geography of retirement. National bureau of economic research, w29433). https://www.nber.org/system/files/working_papers/w29433/w29433.pdf
- Coombes M, Casado-Díaz JM, Martínez-Bernabéu L, Carausu F (2012) Study on comparable labour market areas – final research report. Eurostat-framework contract, (6001), 2008.001-2009.065. https://ec.europa.eu/eurostat/cros/content/study-comparable-labour-market-areas_en
- Coombes MG, Bond S (2008) Travel-to-work areas: The 2007 review. London: Office for national statistics, 2008. https://www.ncl.ac.uk/media/wwwnclacuk/curds/files/TTWA_report.pdf
- Eliasson K, Haapanen M, Westerlund O (2021) Patterns of inter-and intra-regional differences in human capital and earnings: Evidence from Finland and Sweden 1987-2015. *Applied Geography* 135: 102539. [CrossRef](#)
- Eurostat (2020) European harmonised labour market areas: Methodology on functional geographies with potential. Luxembourg: Publications office of the European Union. [CrossRef](#)
- Eurostat (2021) Applying the degree of urbanisation: A methodological manual to define cities, towns and rural areas for international comparisons. Luxembourg: Publications office of the European Union. [CrossRef](#)
- Goya D (2022) Marshallian and jacobian externalities in creative industries. Inter-american development bank. IDB-WP-1263. [CrossRef](#)
- Halás M, Klapka P, Hurbánek P, Bleha B, Péntes J, Pálóczi G (2019) A definition of relevant functional regions for international comparisons: The case of Central Europe. *Area* 51: 489–499. [CrossRef](#)
- Jacob N, Munford L, Rice N, Roberts J (2019) The disutility of commuting? The effect of gender and local labor markets. *Regional Science and Urban Economics* 77: 264–275. [CrossRef](#)
- Kosfeld R, Eckey HF, Dreger C (2006) Regional productivity and income convergence in the unified Germany, 1992-2000. *Regional Studies* 40: 755–767. [CrossRef](#)
- Martínez-Bernabéu L, Casado-Díaz J (2016) Delineating zones to increase geographical detail in individual response data files: An application to the Spanish 2011 census of population. *Moravian Geographical Reports* 24: 26–36. [CrossRef](#)
- Martínez-Bernabéu L, Casado-Díaz JM (2021) Automatic delimitation of labour market areas based on multi-criteria optimisation: The case of Spain 2011. *Environment and Planning B*. [CrossRef](#)

- Martínez-Bernabéu L, Flórez-Revuelta F, Casado-Díaz JM (2012) Grouping genetic operators for the delineation of functional areas based on spatial interaction. *Expert Systems with Applications* 39: 6754–6766. [CrossRef](#)
- OECD (2002) Redefining territories: The functional regions. OECD regional development studies, OECD publishing, Paris. [CrossRef](#)
- OECD (2020) Delineating functional areas in all territories. OECD territorial reviews, OECD publishing, Paris. [CrossRef](#)
- Openshaw S, Taylor PJ (1979) A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In: Wrigley N (ed), *Statistical applications in spatial sciences*. Pion, London, 127–144
- Osland L, Thorsen I (2013) Spatial impacts, local labour market characteristics and housing prices. *Urban Studies* 50: 2063–2083. [CrossRef](#)
- Rowe F, Casado-Díaz JM, Martínez-Bernabéu L (2017) Functional labour market areas for Chile. *REGION* 4: R7–R9. [CrossRef](#)
- Stabler JC, Olfert MR, Greuel JB (1996) Spatial labor markets and the rural labor force. *Growth and Change* 27: 206–230. [CrossRef](#)
- Stimson RJ, Flanagan M, Mitchell W, Shyy TK, Baum S (2018) Modelling endogenous employment performance across Australia's functional economic regions over the decade 2001 to 2011. *The Australasian Journal of Regional Studies* 24: 3–34
- Stimson RJ, Mitchell W, Flanagan M, Baum S, Shyy TK (2016) Demarcating functional economic regions across Australia differentiated by work participation categories. *The Australasian Journal of Regional Studies* 22: 27–57
- United Nations (2019) The global statistical geospatial framework. Department of economic and social affairs, United Nations, New York. https://unstats.un.org/unsd/-statcom/51st-session/documents/The_GSGF-E.pdf
- Vallone A, Chasco C (2020) Spatiotemporal methods for analysis of urban system dynamics: An application to Chile. *The Annals of Regional Science* 64: 421–454. [CrossRef](#)

