A Bayesian Interpolation Method to Estimate Per Capita GDP at the Sub-Regional Level: Local Labour Markets in Spain

Domenica Panzera and Ana Viñuela

Abstract Although economic data is increasingly available for almost any topic, there is still a dearth of data at very high levels of spatial disaggregation. GDP data (total or per capita) is available for the European NUTS II regions and the Spanish National Statistics Office (INE) regularly publishes such information at the NUTS III level (which, in the Spanish case, corresponds to provinces). In this research, we move toward a higher level of spatial disaggregation. We estimate the per capita GDP for the 804 Local Labor Market Areas (LLMs) into which the Spanish territory can be divided by applying the Bayesian Interpolation Method (BIM) introduced by Palma and Benedetti (Geograph Sys 5:199–220, 1998), and considering spatial dependence between observations. Before proceeding with the estimation, we test the methodology by estimating per capita GDP for provinces (NUTS III level), as if such information is only known at a more aggregated level (NUTS II regions) and compare the results to actual provincial values. We then derive per capita GDP values for LLMs given the observed data at the provincial level. The results obtained reveal a high level of internal heterogeneity in GDP per capita within the Spanish administrative regions, and highlight the importance of agglomeration economies and relative location.

Keywords Spatially disaggregated data • Bayesian interpolation method • Local labor market areas

D. Panzera (✉)
Department of Economic Studies, University "G. d'Annunzio" of Chieti-Pescara, Chieti, Italy
e-mail: domenicapanzera@yahoo.it

A. Viñuela
Applied Economics Department, University of Oviedo, Oviedo, Spain
e-mail: avinuela@uniovi.es

J.-C. Thill (ed.), *Spatial Analysis and Location Modeling in Urban and Regional Systems*, Advances in Geographic Information Science,
https://doi.org/10.1007/978-3-642-37896-6_2

Introduction

A significant volume of research has explored the underlying causes of income disparities across sub-national regions and the concomitant dispersion in economic outcomes. Nowhere is this more apparent than in Spain. When investigating these issues, regional economists are constrained by the use of data which, at their most disaggregate, are available at NUTSIII level,[1] which corresponds to provinces in the Spanish case. However, given the internal heterogeneity within the regions at various scales, the challenge becomes one of trying to estimate income values for regional systems based on economic rather than politico-administrative criteria.

Differences in economic outcomes within administrative regions reflect differences in productivity, employment levels and wages (Moretti 2011). While in Spain there are employment data available at the local (municipality) level – which can be used by industry level or in total to demonstrate the internal heterogeneity of Spanish regions – still, no data exists on GDP (or productivity) at local levels.

Yet the Local Labor Markets (LLMs) present themselves as a more suitable *economic* unit of analysis. This is based upon two ideas: (1) that workers and firms interact primarily in Local Labor Markets, the size of which are much smaller than that of the national market; and (2) that few people move from one market to another (Armstrong and Taylor 1993; Bartik 1996; Hughes and McCormick 1994; Topel 1986). In most countries, perhaps only a relatively small number of large LLMs (cities) account for most of the country's output. However, economic researchers still face the problem of a lack of GDP figures, either in total or per capita, at this level of disaggregation.

The purpose of this paper is to estimate per capita GDP for the Local Labor Markets (hereafter LLMs) into which the Spanish territory is divided.

In section "Disaggregating data spatially: The *local* unit of analysis and methodologies", we briefly explain the choice of the LLM as the appropriate sub-regional unit for estimating per capita GDP and the problem of disaggregating data from the regional level in the general framework of areal interpolation. Section "The Bayesian Interpolation Method" presents the proposed methodology – the Bayesian Interpolation Method (hereafter, BIM) – to disaggregate per capita GDP. The solution to this problem requires formulating a hypothesis regarding the probability distribution of the *original* process and exploits information about the *spatial dependence* between observations, as well as auxiliary information observed at the sub-regional level. Prior to estimating per capita GDP for the Spanish LLMs, in section "Estimating Per Capita GDP for Spanish Provinces" the performance of the BIM is tested when disaggregating per capita GDP for *Autonomous Communities* (NUTS II divisions) into per capita GDP for *Provinces* (NUTS III divisions), while treating the latter as unknown. In section "Estimating Per Capita GDP for Spanish

[1]The Nomenclature of Territorial Units for Statistics (NUTS) was established by Eurostat in order to provide a single uniform delineation of territorial units for the production of regional statistics which is based upon the administrative divisions applied in the Member States.

LLMs", the results of the per capita GDP estimated at LLM level using the BIM are presented and discussed. Finally, section "Conclusions and future research" suggests future extensions of this work and a research agenda for this topic.

Disaggregating Data Spatially: The *Local* Unit of Analysis and Methodologies

Local Unit of Analysis: Local Labor Markets

What is the proper unit of analysis for estimation of per capita GDP at the local level? This question is closely related to the definition of an *economic region* in the regional economics literature. The hallmark of an economic region is the existence of an especially high degree of interdependence among individual incomes within the area (Parr 2008).

A number of researchers in the United States during the 1960s (Fox and Kumar 1965), and subsequently in Europe during the 1970s (Smart 1974), designed quantitative techniques or regionalization schemes for identifying such regions. Different names have been applied to these areas, such as Functional Economic Areas and Labor Market Areas (and, as used here, Local Labor Markets or LLMs), but they all referred to a region with an increasingly interwoven and internally interactive area that encompasses the home-to-work daily journeys of its residents.

Originally designed to set the limits of a *city*, these *regions* reflect functional relationships between workers and jobs.[2] With an urban center in the core and fringe municipalities surrounding this core, the LLMs have formed increasingly interwoven and internally interactive areas, in their own right, over time. At the present time in Europe, jobs are increasingly created in the fringe areas, but people prefer to live in the city centers, resulting in two-way journey-to-work flows. Based upon commuting patterns, LLMs demarcate the borders of labor catchment areas. In practice, this means that at least 75% of residents work in the LLM and that 75% of those who work in the LLM also live there. To qualify as an LLM, an area must have a minimum of 3500 residents that work in the area.

The regionalization procedure to establish the borders of an LLM is based upon an algorithm originally developed by Coombes et al. (1986). Starting with the municipal administrative unit and combining the data on the resident employed population, total employed population and journey-to-work commuting patterns, boundaries of the LLMs are defined through a multi-stage aggregation process.[3] This methodology, slightly modified to meet the specific characteristics of countries,

[2]For a discussion on the increasing tendency to treat the region as the city and a city as a region in urban and regional analysis, see Parr (2008).

[3]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).

has been applied in several European countries. The EU Department of Employment, Social Affairs & Inclusion has defined the LLMs (so-called Travel-To-Work-Areas or TTWAs) for Great Britain, for Italy (Sforzi et al. 1997), for Denmark (Andersen 2002), and for Spain (Boix and Galleto 2006).

Based on the idea that workers and firms interact mainly within their respective LLMs, these economic regions may be appropriate spatial units of analysis for studying topics such as: the underlying causes of spatial income disparities across a country; the existence of agglomeration economies and diseconomies and their effect on location decisions, economic structure and growth between larger and smaller cities (Behrens and Thisse 2007); or the relationship between urban agglomeration and productivity (Melo et al. 2009). However, these types of studies or analyses cannot be carried out for most European countries, due to the lack of per capita GDP figures at a further level of disaggregation than NUTS III regions.

Approaches to the Lack of Spatially Disaggregated Data

The European System of National and Regional Accounts (ESA 95) provides its users with a consistent and reliable quantitative description of the economic structure of the European Member States and their regions. The European regional accounts are closely connected to the National Accounts and are based upon the transactions that occur within a specific NUTS-level region.[4] Some attempts have been made by the European Union to provide economic indicators (including GDP figures) for some European cities, as well. The Urban Audit Project provides comparable statistics collected every three years for 321 cities in the 27 countries of the EU, along with 36 additional cities in Norway, Switzerland and Turkey.[5]

In Spain, the National Statistics Institute (INE), which is responsible for the elaboration of the national and regional accounts, provides GDP figures for the NUTS II regions (*Autonomous Communities)* and also for the NUTS III regions (*Provinces*). However, GDP figures at the next level of spatial disaggregation, such as LLMs, are not available.

Situations in which observable regional data are decomposed into their respective smaller spatial units can be viewed as a special case of areal interpolation. Areal interpolation refers to the process of estimating one or more variables for a set of *target* zones, based upon the known values in the set of *source* zones. Target units can be either finer-scale, or misaligned, with respect to the source units.

---

[4]For a discussion on the regionalization method implemented by Eurostat for deriving the regional accounts or conceptual difficulties that constrain the economic variables available at NUTS level – such as the existence of many productive activities crossing regional boundaries, units of production operating in different regions, inter-firm transfers or people living in one region, but working in another, see Eurostat, System of Regional Accounts, ESA1995, (http://circa.europa.eu/irc/dsis/nfaccount/info/data/ESA95/en/titelen.htm)

[5]For more details see http://www.urbanaudit.org/

A number of solutions for deriving data not available at the desired scale of spatial resolution have been proposed in the regional economics literature.[6] Applied to disaggregate Spanish per capita GDP for Autonomous Communities, and treating actual provincial per capita GDP (at the NUTS III level) as not observed, Peeters and Chasco (2006) compared two different approaches to disaggregate information from aggregated data: the classical OLS (ordinary least squares) model and a new Generalized Cross-Entropy (GCE) model. The latter involves the assumption of spatial heterogeneity across observations, and yields predictions of the unobserved that "are *superior* in terms of accuracy".[7] Another recent contribution dealing with the prediction of disaggregated regional data has been provided by Polasek et al. (2010). In this work, the authors extended the Chow-Lin method (Chow and Lin 1971), originally proposed to predict high-frequency time series data from related series, by its application to regional cross-sectional data. For this purpose, the authors specified a spatial autoregressive model (SAR) for the disaggregated data, thereby introducing the spatial dependence effect in the analysis. The spatial dependence parameter is estimated by running the SAR model on the aggregated data; GLS estimators for the parameters associated with regressors observed at the disaggregated level are derived and the prediction of the unobserved dependent variable is carried out.

In the present study, the Spatial Chow-Lin procedure is developed within a Bayesian framework, and it is applied to predict the GDP per capita for Spanish Provinces (assumed to be unobserved), given available indicators at the provincial level and the aggregated data observed for Autonomus Communities. Taking the actual values of provincial GDP as a benchmark, in this empirical application the authors find a significant improvement in the estimates when moving from the classical OLS to the spatial autoregressive specification.

In the next section, we propose the use of the BIM to spatially disaggregate per capita GDP using Spanish data. Bearing in mind that the aim is to estimate per capita GDP for the LLMs defined in section "Local unit of analysis: Local labor markets", in section "Estimating Per Capita GDP for Spanish Provinces" we will test the proposed methodology by disaggregating Autonomous Communities' per capita GDP into Provincial per capita GDP, treating the latter as an unobserved variable. Using the same methodology, in section "Estimating Per Capita GDP for Spanish LLMs" we will estimate per capita GDP for a higher level of disaggregation (LLMs) and verify the existence of intraregional disparities in economic welfare.

[6]For a review on this topic see e.g. Flowerdew and Green (1992), and for a general framework on the areal interpolation of socio-economic data see Goodchild et al. (1992).

[7]In their GCE approach, Peeters and Chasco (2006) develop the prediction model considering the aggregated data to be a weighted geometric mean of the disaggregated data and incorporate the spatial heterogeneity effect through the introduction of unit-specific coefficients for provinces. Unlike the GCE model, the BIM that will be proposed in this work focuses on the spatial dependence effect, which is incorporated into the covariance structure of the data generating process, allowing for the identifcation of a model characterized by fewer parameters according to the parsimony principle.

The Bayesian Interpolation Method

The problem of deriving data that are not available at the desired scale of interest is here formalized by assuming the areal data as a realization of a spatial stochastic process or random field.

Consider a set of n areal units $\{\omega_i, i = 1, 2, \ldots, n\}$ which form a partition Ω over a geographical domain, and denote by Z a variable of interest. The set of random variables $\{Z(\omega_i),\ \ i = 1, \ldots, n\}$, indexed by their locations, defines a random field. We refer to the set $\{Z(\omega_i),\ \ i = 1, \ldots, n\}$ as the *original process*.

Consider a partition Ω^* of the same geographical domain in $m < n$ areal units $\left\{\omega_j^*,\ j = 1, \ldots, m\right\}$ obtained by grouping the areal units ω_i, $i = 1, \ldots, n$, into larger areas. Changing the support of Z creates a new variable Z^*, and data collected on Z^* can be viewed as a realization of the *derived* or *aggregated process* $\left\{Z^*\left(\omega_j^*\right),\ j = 1, \ldots, m\right\}$.

Now assume that data on the variable of interest are available with reference to the partition Ω^*, while we are interested in the level of disaggregation corresponding to the partition Ω. In the described framework, the areal data conversion problem consists of reconstructing the realizations of the original process by starting with the knowledge, based upon data, of the derived process.

As a possible solution to the areal data conversion problem we propose the Bayesian Interpolation Method (BIM) introduced by Palma and Benedetti (1998). According to the BIM the joint distribution of the random vector $\boldsymbol{Z} = (Z(\omega_1), \ldots, Z(\omega_n))'$ is assumed to be a multivariate normal distribution, and $\boldsymbol{Z}$ is expressed in an additive form, that is:

$$\boldsymbol{Z} = \boldsymbol{S} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{S} = (S(\omega_1), S(\omega_2), \ldots, S(\omega_n))'$ refers to the variable of interest at the n sites ω_i, and $\boldsymbol{\varepsilon}$ is a random vector of error terms.

As an additional assumption, the random vectors $\boldsymbol{S}$ and $\boldsymbol{\varepsilon}$ are modeled through the Conditional Autoregressive (CAR) specification (Besag 1974). CAR models satisfy a Markov property in the space where the value of a random variable in a region, given the observations in all the other regions, depends only on the observations in neighboring regions. Assuming CAR models for the random vectors $\boldsymbol{S}$ and $\boldsymbol{\varepsilon}$ does not entail any loss of generality, since any Gaussian process on a finite set of sites can be modeled according to this specification (Ripley 1981, p.90).

The CAR model for $\boldsymbol{S}$ is specified by the set of full conditional distributions:

$$S(\omega_i) \mid S(\omega_{-i}) \sim N\left(T_i + \rho \sum_{j=1}^{n} w_{ij}\left(S\left(\omega_j\right) - T_j\right), \sigma_i^2\right) \qquad i,j = 1, \ldots, n \tag{2}$$

where $\omega_{-i} = \{\omega_j : j \neq i, j = 1, \ldots, n\}$, T_i is the expected value of $S(\omega_i)$, σ_i^2 is its conditional variance, ρ is a scalar parameter measuring the spatial autocorrelation, and w_{ij} is the (ij)-th entry of a $n \times n$ spatial weights matrix $\boldsymbol{W}$, with $w_{ii} = 0$ for all i.

Under some specified regularity conditions (Besag 1974), the full conditional distributions in (2) uniquely determine the following joint distribution:

$$\boldsymbol{S} \sim MVN\left(\boldsymbol{T}, \boldsymbol{\Sigma}_{\boldsymbol{S}}\right) \tag{3}$$

where $\boldsymbol{T} = (T_1, \ldots, T_n)'$ and $\boldsymbol{\Sigma}_{\boldsymbol{S}}$ is the covariance matrix expressed as:

$$\boldsymbol{\Sigma}_{S} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{M} \tag{4}$$

with $\boldsymbol{I}$ denoting a n-dimensional identity matrix and $\boldsymbol{M} = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right)$.

The matrix $\boldsymbol{W}$ in (4) is commonly specified by normalizing a proximity matrix $\boldsymbol{C}$, whose elements c_{ij} are defined as:

$$c_{ij} = \begin{cases} 1 \text{ if } \omega_j \in N(i), \\ 0 \text{ if } i = j, \\ 0 \text{ otherwise} \end{cases} \quad i, j = 1, \ldots, n.$$

with $N(i)$ denoting the set of neighbors of the areal unit ω_i identified according to any proximity criterion. Let $c_{i+} = \sum_{j=1}^{n} c_{ij}$ denote the cardinality of $N(i)$, for $i = 1, \ldots, n$. Constructing the n-dimensional diagonal matrix $\boldsymbol{D} = \text{diag}\ (c_{1+}, \ldots, c_{n+})$, the matrix $\boldsymbol{W}$ can be expressed as follows:

$$\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{C}. \tag{5}$$

The number of neighbors can vary for the areal units. However, since the CAR specification requires the symmetry condition expressed by $\sigma_j^2 w_{ij} = \sigma_i^2 w_{ji}$, for all $i, j = 1, \ldots, n$, when the weighting scheme in (5) is used, the conditional variances have to be inversely proportional to c_{i+}, for $i = 1, \ldots, n$ (Clayton and Berardinelli 1992; Wall 2004). Then, by setting $\sigma_i^2 = \sigma_S^2 / c_{i+}$, for $i = 1, \ldots, n$, where σ_S^2 is a scalar parameter measuring the overall variability of $\boldsymbol{S}$, the covariance matrix in (4) becomes:

$$\boldsymbol{\Sigma}_S = \sigma_S^2(\boldsymbol{D} - \rho\boldsymbol{C})^{-1} \tag{6}$$

where the matrices $\boldsymbol{D}$ and $\boldsymbol{C}$ are defined as above. The specification of the covariance matrix in (6) characterizes the specification that is referred to as CAR(ρ, σ^2) by Gelfand and Vounatsou (2003), and a sufficient condition for its positive definiteness is $|\rho| < 1$.

Similar considerations hold for the random vector $\boldsymbol{\varepsilon}$ which is modeled through a zero-centered CAR, so that its joint distribution is given by:

$$\boldsymbol{\varepsilon} \sim MVN\left(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right) \tag{7}$$

where 0 denotes a $n \times 1$ vector of zeros and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma_\varepsilon^2(\boldsymbol{D} - \rho_\varepsilon\boldsymbol{C})^{-1}$, with σ_ε^2 and ρ_ε denoting scalar parameters.

From the formulated assumptions it follows that the joint distribution of the original process can be expressed as:

$$\boldsymbol{Z} \sim MVN\left(\boldsymbol{S}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\right) \tag{8}$$

where $\boldsymbol{S}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ are specified as above.

The transformation of the n-dimensional random vector $\boldsymbol{Z} = (Z(\omega_i), \ldots, Z(\omega_n))'$, related to the original or disaggregated process, into the m-dimensional random vector $\boldsymbol{Z}^* = \left(Z^*\left(\omega_1^*\right), \ldots, Z^*\left(\omega_m^*\right)\right)'$, related to the derived or aggregated process, can be formalized by introducing a linear operator $\boldsymbol{G}$. The transformation operator $\boldsymbol{G}$ is constructed as a $m \times n$ matrix whose elements can be specified according to any sum or averaging operations, so that:

$$\boldsymbol{Z}^* = \boldsymbol{G}\boldsymbol{Z} = \boldsymbol{G}(\boldsymbol{S} + \boldsymbol{\varepsilon}) = \boldsymbol{G}\boldsymbol{S} + \boldsymbol{G}\boldsymbol{\varepsilon}.$$

Since the observed aggregated data derive from the unobserved disaggregated data through the operator $\boldsymbol{G}$, Palma and Benedetti (1998) give a solution to the areal data conversion problem based on identifying the posterior probability distribution of $\boldsymbol{S}|\boldsymbol{Z}^*$. This posterior probability distribution can be derived by the Bayes' rule as follows:

$$P\left(\boldsymbol{S}|\boldsymbol{Z}^*\right) \propto P(\boldsymbol{S})\, P\left(\boldsymbol{Z}^*|\boldsymbol{S}\right) \tag{9}$$

where $P(\boldsymbol{S})$ denotes the prior probability distribution of $\boldsymbol{S}$, and $P(\mathbf{Z}^* | \mathbf{S})$ is its likelihood on the basis of the observed data.

Based upon the assumption in (3), $\boldsymbol{S}$ has a multivariate normal distribution. Furthermore, the conditional distribution of $\boldsymbol{Z}^*|\boldsymbol{S}$ can be derived from the distribution of $\boldsymbol{Z}$ as follows (Anderson 1958, p. 26):

$$Z^* \mid S \sim MVN\left(GS, G\Sigma_\varepsilon G'\right).$$

From the Gaussian nature of the distributions of $\boldsymbol{S}$ and $\boldsymbol{Z}^*|\boldsymbol{S}$, it follows that $\boldsymbol{S}|\boldsymbol{Z}^*$ also has a multivariate normal distribution.

Under the additional hypothesis of a known covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ the Bayesian approach described leads us to the following result (Pilz 1991):

$$\boldsymbol{S} \mid \boldsymbol{Z}^* \backsim MVN\left(\widehat{\boldsymbol{S}}, \boldsymbol{V}_{\widehat{S}}\right)$$

where $\widehat{\mathbf{S}}$ and $\mathbf{V}_{\widehat{S}}$ are BIM estimators specified as follows:

$$\boldsymbol{V}_{\widehat{S}} = \left[\boldsymbol{G}'\left(\boldsymbol{G}\frac{(\boldsymbol{D} - \rho_\varepsilon \boldsymbol{C})}{\sigma_\varepsilon^2}\boldsymbol{G}'\right)^{-1}\boldsymbol{G} + \frac{(\boldsymbol{D} - \rho \boldsymbol{C})}{\sigma_S^2}\right]^{-1} \tag{10}$$

$$\widehat{\boldsymbol{S}} = \boldsymbol{V}_{\widehat{S}}\left[\frac{(D - \rho C)}{\sigma_S^2}\boldsymbol{T} + \boldsymbol{G}'\left(\boldsymbol{G}\frac{(D - \rho_\varepsilon C)}{\sigma_\varepsilon^2}\boldsymbol{G}'\right)^{-1}\boldsymbol{Z}^*\right] \tag{11}$$

Any inference on the original process can be carried out by the posterior distribution with parameters defined by (10) and (11). Point estimates for $\boldsymbol{S}$ can be obtained using $\widehat{\boldsymbol{S}}$, which is its *Maximum A Posterior* (MAP) estimate, as shown in Benedetti and Palma (1994). Confidence intervals and hypothesis tests can be performed in the standard way using multivariate normal distributions.

Additional issues are raised by the pycnophylactic, or mass preserving, property which consists of finding an estimate of $\boldsymbol{S}$ such that, by applying the transformation operator $\boldsymbol{G}$, the observed data are again obtained (Tobler 1979). In order to preserve this property, the posterior distribution of $\boldsymbol{S}|\boldsymbol{Z}^*$ is conditioned to the linear constraint $\boldsymbol{GS} = \boldsymbol{Z}^*$, so that the constrained version of the BIM estimators is obtained as follows:

$$\widetilde{\boldsymbol{S}} = \widehat{\boldsymbol{S}} + \boldsymbol{V}_{\widehat{S}}\boldsymbol{G}'\left[\boldsymbol{G}\boldsymbol{V}_{\widehat{S}}\boldsymbol{G}'\right]^{-1}\left(\boldsymbol{Z}^* - \boldsymbol{G}\widehat{\boldsymbol{S}}\right) \tag{12}$$

$$\boldsymbol{V}_{\widetilde{S}} = \boldsymbol{V}_{\widehat{S}} - \boldsymbol{V}_{\widehat{S}}\boldsymbol{G}'\left[\boldsymbol{G}\boldsymbol{V}_{\widehat{S}}\boldsymbol{G}'\right]^{-1}\boldsymbol{G}\boldsymbol{V}_{\widehat{S}} \tag{13}$$

Point estimates of $\boldsymbol{S}$ can be obtained by using $\widetilde{\boldsymbol{S}}$; confidence intervals and hypothesis tests can be performed in the standard way using multivariate normal distributions.

Estimating Per Capita GDP for Spanish Provinces

Before proceeding to the estimation of per capita GDP for LLMs, the performance of the method is tested by disaggregating per capita GDP for NUTS II regions (17 Spanish *Autonomous Communities*) into NUTS III regions (50 *Provinces*), while treating the latter as not observed. Both figures are published by the National Statistics Institute. The availability of actual data at provincial level provides a benchmark for evaluation of the estimates and methodology.

According to the BIM, per capita GDP values for the 50 Spanish Provinces can be derived by applying the procedure described in section "The Bayesian Interpolation Method", to disaggregate data observed for Autonomous Communities. Specifically, the estimates for per capita GDP at provincial level are computed as the BIM estimates, given in Eqs. 11, 12. The computation of the unconstrained and constrained BIM estimates requires specifiying the prior mean vector $\boldsymbol{T}$ and the covariance matrix $\boldsymbol{\Sigma}_S$, as well as the aggregation matrix $\boldsymbol{G}$.

In the presence of covariates observed at provincial level, a natural way to express the prior mean of $\boldsymbol{S}$ is:

$$\mathbf{T} = \mathbf{X}\widehat{\boldsymbol{\beta}} \tag{14}$$

Table 1 Regression coefficient estimates for data at autonomous community (NUTS II) level

Coefficients	Model 1		Model 2		Model 3		Model 4	
Constant	−19.194	***	−30.913	***	−23.466	***	−47.533	***
(standard errors)	(3.764)		(6.716)		(4.777)		(9.667)	
Employed population	89.108	***			102.479	***		
(p-values)	(9.312)				(13.208)			
Active population			102.380	***			143.490	***
(p-values)			(14.417)				(22.747)	
Foreign population					−30.286		−67.900	*
(p-values)					(21.817)		(30.973)	
R^2	0.8592		0.7707		0.8763		0.8293	

Significance levels: '***' 0.001, '**' 0.01, '*' 0.05

where $\mathbf{X}$ is a $n \times k$ matrix which includes $k-1$ explanatory variables and an initial column of ones, and $\widehat{\boldsymbol{\beta}}$ is a $k \times 1$ vector of estimated regression parameters, including an intercept term.

There is great availability of socioeconomic variables at the provincial level that can be included as explanatory variables of the variable under study (per capita GDP). However, our final goal is to reconstruct per capita GDP figures for LLMs, and only a few socioeconomic variables -such as total population, population by age and gender, foreign population, active population, and employed population- are available for the Spanish LLMs. Therefore, to ensure coherence with the analysis carried out later at the LLM level, only some of these socioeconomic variables are included in our analysis.

The explanatory variables used to estimate the prior mean vector come from the 2001 Population and Household Census published and administrated by the INE, and the regression coefficients associated with these variables are estimated based upon data observed for the Autonomous Communities.

Table 1 shows results for different regression models estimated with the Autonomous Communities data. The coefficients associated with the explanatory variables included in Models 1 and 2 are both positive and highly significant, providing evidence of the contribution of the employed and active population in explaining per capita GDP. The coefficient associated with the foreign population is negative and not significant for Model 3 and negative and significant at the 0.05 level for Model 4. Comparing Models 1 and 2 we can see that the best fit is reported for Model 1, which exhibits a higher R^2.

The coefficients estimated for Model 1 are then used in the specification of the prior mean vector $\boldsymbol{T}$. The regression residuals' variance is used to estimate the scalar parameter σ_S^2 included in the covariance matrix $\boldsymbol{\Sigma}_S$. Three different values of the autocorrelation parameter, namely $\rho = 0.06$, $\rho = 0.2$ and $\rho = 0.8$, are in turn assumed. The proximity matrix $\boldsymbol{C}$ is constructed according to the 5 nearest neighbors criterion, and is conveniently symmetrized.[8]

[8]Estimations were also run using different k values for the C matrix. No significant differences were found.

The aggregation matrix $\boldsymbol{G}$ expresses the aggregation of NUTS III (*Provinces*) into NUTS II (*Autonomous Communities*). The entries of $\boldsymbol{G}$ are specified as follows:

$$g_{ji} = h_{ji}\frac{P_i}{P_j} \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m \tag{15}$$

where P_i and P_j denote the Total Population in the *i-th* Province and in the *j-th* Autonomous Community respectively, and h_{ji} takes the value 1 if the *i-th* Province belongs to the *j-th* Autonomous Community, and 0 otherwise. The ratio $\frac{P_i}{P_j}$ is needed so that *per capita* values for Provinces can be aggregated into per capita values for Autonomus Communities.

The BIM estimates are computed in their unconstrained version – Eq. (11) – and constrained version – Eq. (12). As explained above, the constrained solutions preserve the pycnophylactic property (which ensures that the values estimated for the target zones after the aggregation yield a total equal to the total of the values observed for the source zones); the pycnophylactic property is not considered in the unconstrained version of the BIM estimates. Table 2 shows the evaluation of the accuracy of the estimates – that is, the comparison between the estimated and the actual values – using indices from the forecasting literature such as as the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE).[9]

Table 2 Accuracy of constrained and unconstrained per capita GDP estimates for provinces (prior mean specified by Model 1)

ρ	Type	Correlation	RMSE	MAE	MAPE
0.06	*Unconstrained*	0.89107	1.68583	1.39997	0.09493
0.06	*Constrained*	0.94021	1.25207	0.85685	0.05952
0.2	*Unconstrained*	0.89107	1.68583	1.39997	0.09493
0.2	*Constrained*	0.93966	1.25697	0.85770	0.05961
0.8	*Unconstrained*	0.89107	1.68582	1.39997	0.09493
0.8	*Constrained*	0.92973	1.36126	0.92428	0.06611

[9]The considered accuracy measures are computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right|$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right|$$

As expected, the constrained estimates perform better than the unconstrained versions (the calculated values of the accuracy measures are lower). The constrained solutions also exhibit a higher correlation with actual data. The values of the correlation coefficient for the RMSE, MAE and MAPE appear to be essentially independent of the hypothesis regarding ρ, which reveals the robustness of the estimates with respect to the autocorrelation parameter.

Results of the estimated per capita GDP for Provinces are reported in Table 3. The results assuming Model 1 and Model 2 in the specification of the prior mean and variance are compared with the actual values. There is an improvement in the estimates when the prior mean is specified according to Model 1, which is confirmed by the computed accuracy measures.

The estimated per capita GDP values suggest that the BIM performs well in converting data from the NUTS II to the NUTS III level. This testing exercise validates the application of the BIM to higher levels of spatial disaggregation. Results for the estimation of per capita GDP for the Spanish LLMs are presented in the following section.

Estimating Per Capita GDP for Spanish LLMs

The 806 LLMs into which the Spanish territory is divided (Table 4) show high disparities in size, in terms of both population and number of municipalities. The largest LLMs in terms of population are the Madrid and Barcelona LLMs, which comprise 20.51% of the total Spanish population. More than 85% of the LLMs can be considered rural (i.e., less than 50,000 inhabitants), but only 23.23% of the population live in these rural LLMs, which are typically located in inaccessible/poorly-connected areas characterized by hilly terrain -such as those in the northern part of the country.[10]

In this section, the BIM is applied to disaggregate per capita GDP data for the 50 Spanish provinces in order to determine per capita GDP for the 804 LLMs.[11] Data on per capita GDP for the 804 Spanish LLMs are computed as the constrained BIM estimates given in (12). The computation of the BIM estimates requires specifying the prior mean vector and the covariance matrix of $\boldsymbol{S}$ as well as the aggregation matrix $\boldsymbol{G}$.

where y_i and $\widehat{y_i}$ denote real and estimated values respectively. RMSE and MAE can range from 0 to $+\infty$ and are negatively oriented values, so that lower values of these indices means better estimates.

[10] As expected, the delimitation of LLMs based upon commuting criteria does not match the boundaries of the administrative regions; consequently, there are a few cases in which municipalities included in the same LLM belong to different NUTS III regions. As will be explained below, this fact has been taken into account when defining the aggregation matrix.

[11] Ceuta and Melilla LLMs are not considered in the analysis due to their particular status of autonomous cities.

Table 3 Disaggregation of per capita GDP (EUR) for Spanish provinces

	Actual values	Prior mean specified by Model 1			Prior mean specified by Model 2		
		$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$	$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$
Álava	23,164	22,003	21,876	21,023	21,550	21,312	19,901
Albacete	12,530	12,994	13,013	12,906	14,036	14,068	14,000
Alicante	15,017	15,684	15,666	15,291	15,398	15,264	14,311
Almería	15,584	18,975	18,983	18,688	17,925	17,969	17,928
Ávila	13,346	13,131	13,085	12,457	12,003	11,869	10,788
Badajoz	10,710	9281	9222	8620	9759	9796	10,115
Balears	20,875	20,875	20,875	20,875	20,875	20,875	20,875
Barcelona	20,212	20,487	20,468	20,303	20,969	20,963	20,929
Burgos	18,787	18,759	18,702	18,310	18,661	18,558	17,932
Cáceres	10,805	13,131	13,226	14,207	12,353	12,292	11,773
Cádiz	12,594	9333	9235	8414	11,556	11,469	10,927
Castellón	19,039	19,779	19,686	19,031	17,779	17,642	16,768
Ciudad Real	13,277	10,780	10,748	10,465	10,739	10,690	10,423
Córdoba	10,923	10,960	10,883	10,332	13,154	13,136	13,087
Coruña, A	13,327	12,359	12,381	12,512	12,840	12,802	12,789
Cuenca	12,843	12,407	12,386	12,091	10,383	10,314	9997
Girona	20,832	21,382	21,399	21,481	20,458	20,548	21,187
Granada	11,119	10,793	10,883	11,586	11,535	11,672	12,503
Guadalajara	15,671	17,388	17,350	17,170	16,886	16,873	16,904
Guipúzcoa	21,649	21,970	21,964	21,756	21,262	21,257	21,143
Huelva	13,057	11,448	11,494	12,293	12,969	12,892	12,190
Huesca	17,688	17,607	17,647	17,827	15,419	15,524	16,244
Jaén	10,849	9941	9992	10,135	9074	9098	9498
León	13,957	13,566	13,601	13,890	13,549	13,658	14,438
Lleida	20,816	20,038	20,099	20,577	17,352	17,562	18,921
Rioja, La	18,881	18,881	18,881	18,881	18,881	18,881	18,881
Lugo	12,591	13,961	13,937	14,037	12,321	12,432	13,347
Madrid	22,377	22,377	22,377	22,377	22,377	22,377	22,377
Málaga	12,683	14,226	14,247	14,215	13,197	13,264	13,823
Murcia	13,901	13,901	13,901	13,901	13,901	13,901	13,901
Navarra	21,349	21,349	21,349	21,349	21,349	21,349	21,349
Ourense	12,095	10,927	10,907	10,911	9558	9563	9813
Asturias	14,160	14,160	14,160	14,160	14,160	14,160	14,160
Palencia	15,850	15,324	15,359	15,429	15,229	15,184	14,986
Palmas	17,014	16,804	16,752	16,513	16,852	16,755	16,314
Pontevedra	13,195	14,263	14,254	14,054	14,836	14,837	14,398
Salamanca	13,940	13,789	13,679	12,957	14,672	14,533	13,558

(continued)

Table 3 (continued)

	Actual values	Prior mean specified by Model 1			Prior mean specified by Model 2		
		$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$	$\rho = 0.06$	$\rho = 0.2$	$\rho = 0.8$
Sta. Cruz de Tenerife	15,729	15,961	16,018	16,281	15,908	16,014	16,500
Cantabria	15,979	15,979	15,979	15,979	15,979	15,979	15,979
Segovia	16,914	17,135	17,120	16,859	15,121	15,144	14,999
Sevilla	12,580	13,392	13,396	13,684	11,645	11,592	11,205
Soria	17,202	19,069	19,144	19,952	16,350	16,140	14,724
Tarragona	21,406	19,189	19,287	20,229	17,833	17,673	16,544
Teruel	18,090	14,691	14,766	15,627	11,261	11,127	10,220
Toledo	13,110	14,645	14,681	15,165	14,867	14,918	15,306
Valencia	16,070	15,468	15,500	15,891	16,094	16,212	17,031
Valladolid	17,143	17,613	17,631	17,899	19,036	19,104	19,647
Vizcaya	19,355	19,458	19,494	19,835	19,996	20,060	20,487
Zamora	12,030	11,792	11,933	12,754	10,528	10,725	11,950
Zaragoza	17,698	18,235	18,214	18,040	19,278	19,273	19,240
Correlation		0.940	0.940	0.930	0.891	0.888	0.856
RMSE		1.252	1.257	1.361	1.706	1.728	1.966
MAE		0.857	0.858	0.924	1.180	1.201	1.407
MAPE		5.95%	5.96%	6.61%	7.87%	7.99%	9.19%

Source: 2001 Spanish Regional Accounts, INE; and 2001 Population and Household Census, INE (2007)

Table 4 Distribution of LLMs by population size (2001)

Names or numbers of LLMs of this size		Number of municipalities	% of total population
2,500,000 ≤ Inhabitants	Madrid	153	20.51%
	Barcelona	151	
500,000 ≤ Inhabitants <2,500,000	Valencia	52	16.49%
	Sevilla	39	
	Bilbao	59	
	Malaga	20	
	Zaragoza	96	
	Palmas de Gran Canaria	15	
	Sabadell	17	
	Sta. Cruz Tenerife	17	
100,000 ≤ Inhabitants <500,000	60	2102	31.20%
50,000 ≤ Inhabitants <100,000	50	666	8.56%
Inhabitants <50,000	686	4822	23.23%
Total population: 40,393,173	806	8106	40,847,371
	LLMs	municipalities	Inhabitants

Source: 2001 Population and Household Census, INE (2007)

Table 5 Economic and socio-economic variables for the 804 Spanish LLMs classified by population size

LLM population size	% active population	% employed population	% foreign population
2,500,000 ≤ Inhabitants	15.63%	16.04%	33.03%
500,000 ≤ Inhabitants <2,500,000	18.07%	17.67%	10.92%
100,000 ≤ Inhabitants <500,000	33.61%	33.75%	22.48%
50,000 ≤ Inhabitants <100,000	9.23%	9.18%	13.07%
Inhabitants <50,000	23.45%	23.36%	20.49%
Total population: 40,393,173 inhabitants	17,499,182	14,976,512	1,572,013

Source: 2001 Spanish Population and Household Census, INE (2007)

Table 6 Regression coefficient estimates for data at provincial level

Coefficients	Model 1		Model 2		Model 3		Model 4	
Constant	−13.902	***	−21.785	***	−15.912	***	−23.328	***
(standard errors)	(2.169)		(4.497)		(2.519)		(5.611)	
Employed population	75.975	***			82.705	***		
(standard errors)	(5.586)				(6.995)			
Active population			83.083	***			87.109	***
(standard errors)			(9.914)				(13.201)	
Foreign population					−20.515		−9.169	
(standard errors)					(13.157)		(19.640)	
R^2	0.794		0.594		0.804		0.596	

Significance levels: '***' 0.001, '**' 0.01, '*' 0.05

The prior mean vector is defined as in (14). For explanatory variables used to define the prior mean vector, we consider employed population, active population and foreign population. Data on these variables are available at the LLM level, and come from the 2001 Population and Household Census (INE 2007). Data on active population, employed population and foreign population for LLMs grouped by LLM population size are summarized in Table 5.

The coefficients associated with the explanatory variables are estimated based upon provincial data. Results for different specifications of the regression model estimated with per capita provincial data are displayed in Table 6. The regression residuals' variance is used to estimate the scalar parameter σ_S^2.

Given the expected relationship between labor variables (employment and active population) and economic activity, Models 1 and 2 show high coefficients of determination. A higher $R^2 = 0.794$ is reported for Model 1, which only considers the Employed Population in explaining per capita GDP. The regression coefficient associated with this independent variable is positive and highly significant, as is the coefficient associated with the Active Population in Model 2. The results related to Models 3 and 4 reveal that coefficients associated with the Foreign Population are negative but not significant. Therefore, in order to define the prior mean vector $\boldsymbol{T}$

we will focus on Model 1, which appears perform best[12] at the provincial as well as at the Autonomous Community level (see section "Estimating Per Capita GDP for Spanish Provinces").

In the conversion of provincial data to the LLM level, some important issues have to be considered. First, given that the disaggregate (LLM-level) data are not available, the accuracy of any predicted value simply cannot be verified. Secondly, as mentioned in section "Conclusions and future research", several LLMs go beyond the provincial limit, so that some part of the *i-th* LLM can belong to a given province *j* and rest to one or more other provinces (see Fig. 1).

In order to take into account this feature, the elements of the aggregation matrix ***G***, which expresses the LLMs' aggregation into provinces, will be constructed according to a weighted sum. A weight h_{ji} is assigned to each LLM i belonging to the province j according to the following scheme:

$$h_{ji} = \begin{cases} 1 \text{ if the LLM } i \text{ } b\text{elongs only to the province } j \\ k_{ji}\text{if the LLM } i \text{ belongs to the province } j \\ 0 \text{ otherwise} \end{cases}$$

where k_{ji} is a real value chosen to be proportional to the active population in the municipalities included in the *i-th* LLM, and such that $0 < k_{ji} < 1$ and $\sum_{j=1}^{m} k_{ji} = 1$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The entries of the aggregation matrix ***G*** are therefore specified as follows:

$$g_{ji} = h_{ji}\frac{P_i}{P_j} \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m \tag{16}$$

where P_i and P_j denote the total population in the *i-th* LLM and in the *j-th* province, respectively, in order to take into account the aggregation of *per capita* figures.

The BIM estimates are then carried out in their constrained version on the basis of the above specified prior mean vector, covariance and aggregation matrices. Different values for the spatial autocorrelation parameter, namely $\rho = 0.06$, $\rho = 0.2$ and $\rho = 0.8$, are considered. The proximity matrix ***C*** is specified according to the k nearest neighbours method ($k = 5$), and is conveniently symmetrized.

The obtained BIM estimates defines the per capita GDP for LLMS. The results for each of the 804 LLMs, obtained by setting different values for ρ, can be easily

[12]Regressions with additional variables from the *Anuario Economico de España* (La Caixa 2011), with data available at local level, were also estimated: commercial activities, number of banks, number of fixed telephone lines, vehicles available, etc. However, the coefficients for these variables were not significant and the models showed a goodness of fit inferior to the models displayed in Table 3.

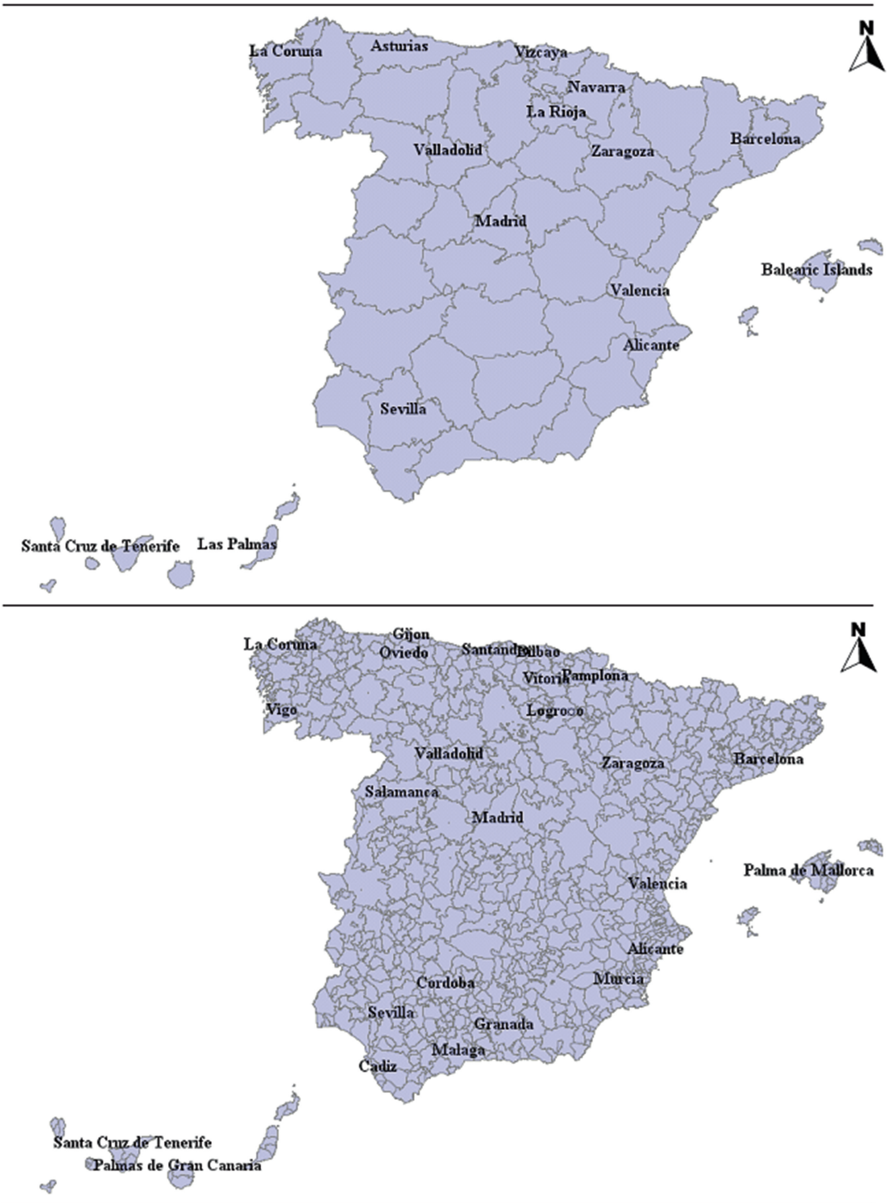

Fig. 1 Provinces and LLMs in Spain

Table 7 Estimated per capita GDP (EUR) for Spanish LLMs classified by population size

Population size	Population	Estimated per capita GDP		
		$\rho = 0.8$	$\rho = 0.2$	$\rho = 0.06$
Madrid LLM	*5,287,342*	*22,488*	*22,487*	*22,487*
Barcelona LLM	*3,027,950*	*19,422*	*19,701*	*19,733*
2,500,000 ≤ Inhabitants		21,412	21,480	21,486
500,000 ≤ Inhabitants <2,500,000		15,940	15,847	15,837
100,000 ≤ Inhabitants <500,000		16,192	16,173	16,170
50,000 ≤ Inhabitants <100,000		16,304	16,299	16,304
Inhabitants <50,000		14,338	14,328	14,327

Source: 2001 Spanish Population and Household Census, INE (2007)

visualized in a map or aggregated in a table according to size[13] as displayed in Table 7.

The results appear to confirm the existence of internal spatial disparities in per capita GDP between the metropolitan areas – LLMs comprised of a core city and surrounding municipalities with a total population greater than 2.5 million inhabitants – and LLMs in the rest of the country, comprised of small and medium size cities and rural areas. The larger the size (quantified by population) of the LLM area, as defined in Table 7 above, the higher its estimated per capita GDP. This suggests the importance of agglomeration economies. Besides the positive effects derived from large-scale production and positive externalities associated with size, agglomeration economies associated with urban concentration lead to lower recruitment and training costs and increased, knowledge spillovers and competition (Porter 1990; Beardsell and Henderson 1999; Rosenthal and Strange 2004). In other words, large metropolises stimulate the exchange of knowledge, productivity, and economic growth, and therefore economic welfare. In his hierarchy of central places, Christaller (1966) states that not all central regions or cities are equal – there are higher-order centers, with a greater concentration of economic activity, as well as other lower-order places.

Another relevant factor in regional studies is the importance of distance, and the manner in which transportation costs can affect the business location, level of economic activity, employment, employment rates and also per capita GDP. As suggested by Polèse and Shearmur (2004) for Canada and by Polèse et al. (2006) for Spain, not only the larger cities benefit from agglomeration economies but areas close to them share in these benefits as well. For a graphical demonstration of this related to the present study, see Fig. 2, where estimated per capita GDP values (with $\rho = 0.06$) above and below the Spanish average are depicted for provinces and also for LLMs.

[13]For each LLM we have the estimated per capita GDP and also actual population, so we can calculate their estimated GDP and an average per capita GDP figure for different size categories of LLMs.

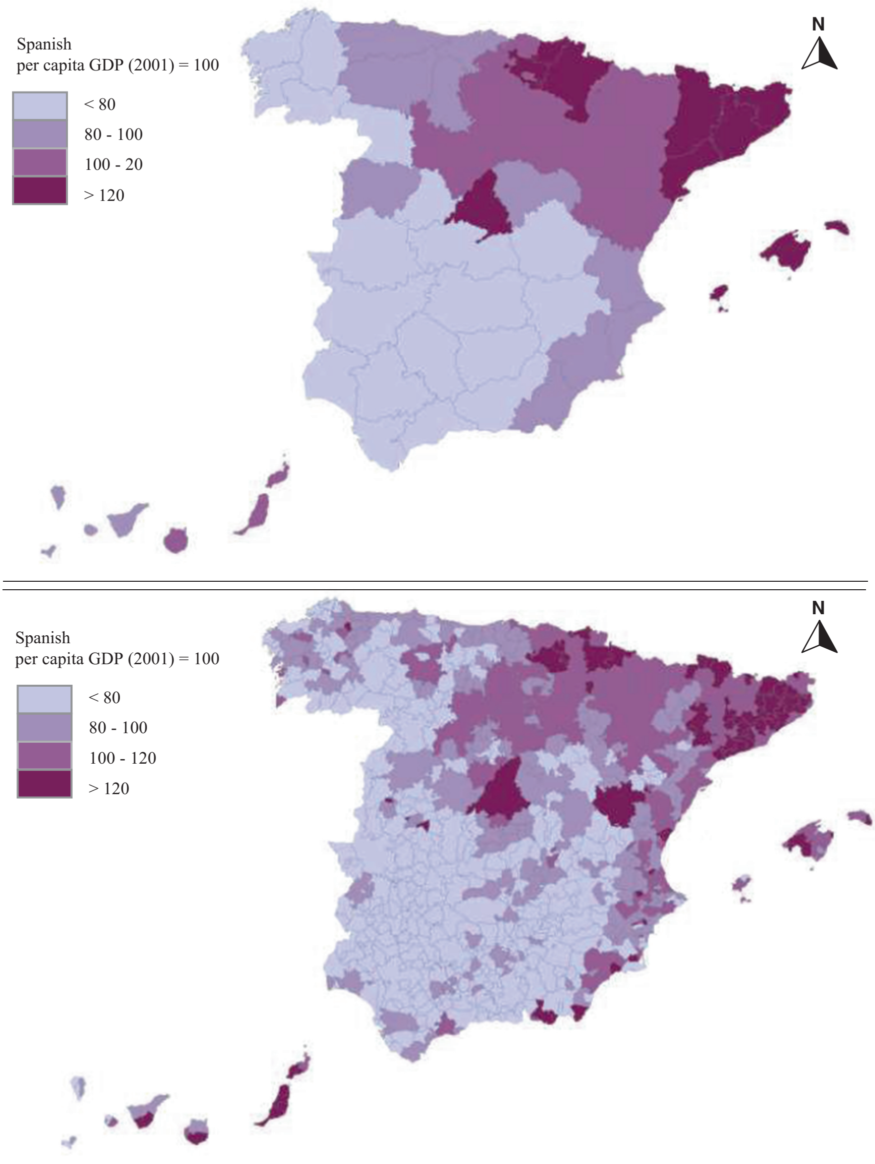

Fig. 2 *Actual* per capita GDP (provinces) and *estimated* per capita GDP for LLMs ($\rho = 0.06$)

As Polèse (2009) assessed, relative location matters. The location of a local labor market must be considered not only in regard to the national urban system, but also in the context of international connections. Proximity to international borders with important trade flows could be relevant, and this appears to be confirmed in

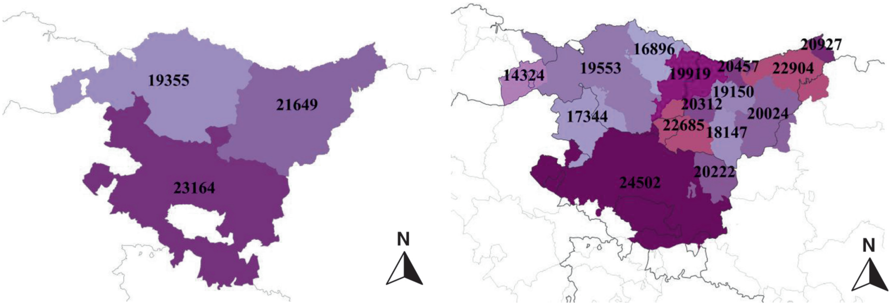

Fig. 3 *Actual* per capita GDP (provinces) and *estimated* per capita GDP for Basque LLMs ($\rho = 0.06$)

the Spanish case. In Spain, there is a northeastern concentration of the highest per capita GDP values, both by provinces and by LLM. However, by working with disaggregated data, we reveal the internal spatial disparities and the importance of the largest cities, as well as the dichotomy between the rural and the urban areas. As an example, Fig. 3 shows detailed results for LLMs that conform the Basque Country Autonomous Community (NUTS II), whose average per capita GDP was EUR 20,618 according to the 2001 Spanish Regional Accounts (published by the INE).

Divided into three provinces or NUTS III regions (Alava, Guipuzcoa and Vizcaya), the Basque Country is a highly industrialized and urbanized autonomous community with a strategic location beside the French border. In the Basque Country, there are severe intra-regional disparities in terms of per capita GDP that cannot be clearly observed at the provincial level. The severity of intra-regional differences in GDP per capita that is readily apparent at the LLM level (Fig. 3) appears to be far less pronounced when aggregated at the province level – where the GDP per capita differences for Alava, Guipuzcoa, and Vizcaya (EUR 23,164, 21,649 and 19,355, respectively) do not tell the full story of the region's disparity in terms of this measure. Previously, researchers could not verify or test the provincial-level results, as GDP per capita figures were not available at a highly disaggregate level. Therefore, the availability of this LLM-level dataset opens the door to these and many other questions related to, for instance, the impact of regional policies at the local level or the weight of cities in regional economies.

Conclusions and Future Research

Regional economists use administrative units as a proxy of the *region* in their empirical analyses, either due to the lack of alternatives or the impossibility of having a region consistent with the theoretical assumptions of regional economics.

Such a basic economic variable as per capita GDP is only available in Spain at the NUTS III level, which corresponds to provinces. However, we believe that the ideal unit of analysis is the further disaggregated LLM. Defined by the daily commuting patterns of its residents, the larger LLMs (larger cities and their surrounding areas) are presumably the national engines of growth, where most of the national GDP is concentrated. However, no data on GDP – either total or per capita – are available at the LLM level of disaggregation.

In this paper, we have estimated per capita GDP for Spanish LLMs using the Bayesian Interpolation Method. As expected, our results confirm the existence of wide disparities in per capita GDP within the administrative regions – the units commonly used in any regional study of this type – and the importance of agglomeration economies and relative location.

The proposed method was also applied to estimate per capita GDP for provinces (NUTS III level) using the data available for Autonomous Communities (NUTS II regions). Comparisons with the actual GDP per capita data available for provinces have revealed good performance by the proposed method, thereby validating its application to estimating per capita GDP at a higher level of spatial disaggregation (i.e., LLMs).

Although preliminary, the results obtained lead us to believe that this line of research should be explored in more detail. Directions for future research could include using additional auxiliary variables to explain per capita GDP, considering different specification of the neighborhood structure for LLMs or applying this methodology to some other geographical scenarios.

Acknowledgments This work is a result of the efforts of G.J.D. Hewings (University of Illinois at Urbana-Champaign) in bringing people together to prove that one plus one is greater than two. To him we give our infinite gratitude. The authors would also like to thank R. Benedetti (University of Pescara) for his useful comments and explanations on technical issues, and P. Postiglione (University of Pescara) for his suggestions.

A preliminary version of this chapter was presented at the RSAI 2012 (Timisorara, Romania) and also at a seminar of the Regional Economic Analysis Laboratory (REGIOlab) of the University of Oviedo. The authors would like to thank participants at both events for the comments and suggestions received. Finally, the chapter has greatly benefitted from the comments of the anonymous referees.

References

Andersen, A. K. (2002). Are commuting areas relevant for the delimitation of administrative regions in Denmark? *Regional Studies, 36*(8), 833–844.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. New York: John Wiley.

Armstrong, H., & Taylor, J. (1993). *Regional economics*. London: Harvester Wheatsheaf.

Ball, R. M. (1980). The use and definition of travel-to-work areas in great Britain: Some problems. *Regional Studies, 14*(2), 125–139.

Bartik, T. J. (1996). The distributional effects of local labor demand and industrial mix: Estimates using individual panel data. *Journal of Urban Economics, 40*, 150–178.

Beardsell, M., & Henderson, V. (1999). Spatial evolution of the computer industry in the USA. *European Economic Review, 43*(2), 431–456.

Behrens, K., & Thisse, J. F. (2007). Regional economics: A new economic geography perspective. *Regional Science and Urban Economics, 37*, 457–465.

Benedetti, R., & Palma, R. (1994). Markov random field-based image subsampling method. *Journal of Applied Statistics, 21*(5), 495–509.

Besag, J. (1974). Spatial iteraction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B, 36*, 192–225.

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

Caixa, L. (2011). *Anuario Económico de España*. Barcelona: Servicio de Estudios de la Caja de Ahorros y Pensiones de Barcelona.

Chow, G. C., & Lin, A. (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *The Review of Economics and Statistics, 53*(4), 372–375.

Christaller, W. (1966). *Central places in southern Germany: Translated from Die zentralen Orte in Süddeutschland*. Englewood Cliffs: Prentice Hall.

Clayton, D., & Berardinelli, L. (1992). Bayesian methods for mapping disease risk. In P. Elliott, J. Cuzick, D. English, & R. Stern (Eds.), *Geographical and environmental epidemiology: Methods for small-area studies* (pp. 205–220). Oxford: Oxford University Press.

Coombes, M. G., & Openshaw, S. (1982). The use and definition of travel to work areas in Great Britain: Some comments. *Regional Studies, 16*, 141–149.

Coombes, M. G., Green, A. E., & 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), 943–953.

Flowerdew, R., & Green, M. (1992). Developments in areal interpolation methods and GIS. *The Annals of Regional Science, 26*, 67–78.

Fox, K. A., & Kumar, T. K. (1965). The functional economic area: Delineation and implications for economic analysis and policy. *Papers in Regional Science, 15*(1), 57–85.

Gelfand, A. E., & Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics, 4*(1), 11–25.

Goodchild, M. F., Anselin, L., & Deichmann, U. (1992). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A, 25*, 383–397.

Hughes, G., & McCormick, B. (1994). Did migration in the 1980s narrow the north-south divide? *Economica, 61*, 509–527.

INE. (2007). 2001 population and household census (*Censo de Población y Viviendas*, 2001), Instituto Nacional de Estadística, Madrid. Available online at www.ine.es.

INE, Spanish Regional Accounts. (2001). *Contabilidad Regional.* Instituto Nacional de Estadistica, Madrid. Available online at www.ine.es.

Melo, P. C., Graham, D. J., & Nolan, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics, 39*(3), 332–342.

Moretti, E. (2011). Chapter 14 – Local labor markets. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics*, *Volume 4, Part B*. Amsterdam: Elsevier.

Palma, D., & Benedetti, R. (1998). A transformational view of spatial data analysis. *Geographical System, 5*, 199–220.

Parr, J. (2008). Cities and regions: Problems and potentials. *Environment and Planning A, 40*, 3009–3026.

Peeters, L., & Chasco, C. (2006). Ecological inference and spatial heterogeneity: An entropy-based distributionally weighted regression approach. *Papers in Regional Science, 85*(2), 257–276.

Pilz, J. (1991). *Bayesian estimation and experimental design in linear regression models*. New York: John Wiley.

Polasek, W., Llano, C., & Sellner, R. (2010). Bayesian methods for completing data in spatial models. *Review of Economic Analysis, 2*, 194–214.

Polèse, M. (2009). *The wealth and the poverty of regions: Why cities matters*. Chicago: University of Chicago Press.

Polèse, M., & Shearmur, R. (2004). Is distance really dead? Comparing the industrial location patterns over time in Canada. *International Regional Science Review, 27*(4), 1–27.

Polèse, M., Shearmur, R., & Rubiera, F. (2006). *Observing regularities in location patterns. An analysis of the spatial distribution of economic activity in Spain*. Montreal: INRS-Internal Document.

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

Ripley, B. D. (1981). *Spatial statistics*. New York: John Wiley.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics, 4*, 2119–2171.

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

Smart, M. W. (1974). Labour market areas: Uses and definitions. *Progress in Planning, 2*, 239–353.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*, 519–530.

Topel, R. H. (1986). Local labor markets. *The Journal of Political Economy, 94*(3), S111–S143.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference, 121*, 311–324.

http://www.springer.com/978-3-642-37895-9

Spatial Analysis and Location Modeling in Urban and
Regional Systems
Thill, J.-C. (Ed.)
2018, VI, 387 p. 111 illus., 76 illus. in color., Hardcover
ISBN: 978-3-642-37895-9