Entropy **2010**, *12*, 2171-2185; doi:10.3390/e12102171

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Incorporating Spatial Structures in Ecological Inference: An Information Theory Approach

Rosa Bernardini Papalia

Department of Statistics, University of Bologna, Via Belle Arti 41, Bologna, Italy;
E-Mail: rossella.bernardini@unibo.it

Received: 30 August 2010; in revised form: 12 October 2010 / Accepted: 12 October 2010 / Published: 14 October 2010

Abstract: This paper introduces an Information Theory-based method for modeling economic aggregates and estimating their sub-group (sub-area) decomposition when no individual or sub-group data are available. This method offers a flexible framework for modeling the underlying variation in sub-group indicators, by addressing the spatial dependency problem. A basic ecological inference problem, which allows for spatial heterogeneity and dependence, is presented with the aim of first estimating the model at the aggregate level, and then of employing the estimated coefficients to obtain the sub-group level indicators.

Keywords: generalized cross entropy estimation; ecological inference; spatial heterogeneity

PACS Codes: 89.70.-a, 89.65.Gh.

---

1. Introduction

This paper introduces an Information Theory (IT)-based method for modeling economic aggregates and estimating their sub-group (sub-area) decomposition when no individual or sub-group data are available. The proposed approach offers a tractable framework for modeling the underlying variation in sub-group indicators. A basic ecological inference problem which allows for spatial heterogeneity and dependence is presented with the aim of estimating the model at the aggregate level. The estimated coefficients are then employed to obtain the sub-group level indicators.

The latent sub-group indicators may be treated as random coefficients or modeled as a parametric function in the unit level model in which the observed aggregate is regressed on the explanatory variables both at the group and sub-group level.

By taking as a point of departure the approach presented in Johnston and Pattie [1] in Judge, Miller and Cho [2], and in Bernardini Papalia [3–7], the basic idea is to introduce an estimator based on an entropy measure of information which provides an effective and flexible procedure for reconciling micro and macro data. The maximum entropy (ME) procedures (Golan, Judge and Robinson, [8]; Golan, Judge and Miller, [9,10]; Golan, [11]) allow the possibility of taking into account out-of-sample information which can be introduced as additional constraints in the optimization program or by specifying particular priors for parameters and errors. A unique optimum solution can also be achieved if there are more parameters to be estimated than available moment conditions and the problem is ill-posed. If there exists additional non-sample information from theory and/or empirical evidence, over that contained in the consistency and adding-up constraints, for the unknown probabilities, it may be introduced in the form of known probabilities, by means of the cross-entropy formalism (Shannon, [12]; Jaynes, [13]; Kullback, [14]; Levine, [15]).

The paper is structured as follows. In Section 2.1 an introduction to the typical ecological inference problems is presented. Sections 2.2 and 2.3 introduce two alternative approaches to ecological modeling that account for spatial heterogeneity and spatial dependence problems, respectively. Section 3 provides the formulation of the proposed information theoretic approaches incorporating both spatial heterogeneity and dependence. In Section 4, the IT-based disaggregation procedure is applied to Italian data. Finally, the last section provides concluding remarks and outlines some direction for further research.

2. Ecological Inference Assuming Heterogeneity and Dependence across Space

2.1. Theoretical Framework of Ecological Inference

Ecological inference (EI) is the process of drawing conclusions about individual level behaviour from aggregate data, when no individual data are available. Situations where the only available data are aggregated at a level other than the level of interest are quite common in many application fields. This is the typical setting for Ecological Inference (Freedman, Klein, Ostland and Roberts, [16]; Schuessler, [17]; King, Rosen and Tanner, [18]), for Cross-level Inference (Achen and Shively, [19]; Cho, [20]) for Small Area Estimation (Rao, [21]), or for disaggregation methods (Barker and Pesaran, [22]). The basic idea is that in order to study the behavior of the individuals (or sub-groups of individuals), a microeconomic analysis ought to be carried out using fairly localized individual data, and data which are aggregated by areal units may be used in order to investigate the behavior of the individuals comprising those units. In this paper we specifically refer to the process of drawing conclusions about individual level behaviour from aggregate data, when no individual data are available or when individual data are incomplete.

In this inferential context, one problem is that many different possible relationships at the individual (or subgroup) level can generate the same observations at the aggregate (or group) level (King, [23]). In the absence of individual (or subgroup) level measurement (in the form of survey data), such

information need to be inferred. Estimates of the disaggregated values for the variable of interest can be inferred from aggregate data by using appropriate statistical techniques. However, in many situations, given that micro-data of interest are not available, the accuracy of any predicted value cannot be verified.

The traditional approach to ecological inference is based on the homogeneity across space hypothesis which assumes constancy of parameters across the disaggregate spatial units. This assumption is rarely tenable, since the aggregation process usually generates macro-level observations across which the parameters describing individuals may vary (Cho, [20]). It is recognized that observations at an aggregate level of analysis do not necessarily provide useful information about lower levels of analysis, particularly when spatial heterogeneity is present. Moreover, the objective of recovering disaggregate information from aggregate data may produce "ill-posed" or "undetermined" inverse problems given that there are more unknowns than data points. In EI it is also important to deal with the "modifiable area unit problem" which refers to (i) the scale effect or aggregation effects, and (ii) the grouping effect or zoning effect. In the first case the resulting aggregation bias may produce different results when data (or individuals) are grouped into increasingly larger areal units. In the second case, the resulting specification bias is connected to the variability in results due to alternative formulations of the areal units leading to differences in unit shape at the same or similar scales and arises when there is a non linear relationship that is not properly accounted for in the specification of the aggregated model.

Spatial structures are generally associated to: (i) *Absolute location* effects that refer to the impact—for each unit—of being located at a particular point in space, and to (ii) *Relative location* effects that consider relevant the position of an unit relative to other units.

The UNOBSERVED SPATIAL HETEROGENEITY (absolute location effects) can be introduced by assuming: (i) slope heterogeneity across spatial units, implying that parameters are not homogeneous over space but vary over different geographical locations; (ii) the presence of cross-sectional correlation due to the presence of some common immeasurable or omitted factors.

The SPATIAL DEPENDENCE (relative location effects) is traditionally introduced by incorporating: a spatial autoregressive process in the error term, and/or a spatially lagged dependent variable. A Spatial Error Model specification assumes that the spatial autocorrelation is modeled by a spatial autoregressive process in the error terms. It follows that: spatial effects are assumed to be identical within each unit, but all the units are still interacting spatially through a spatial weight matrix. The presence of spatial dependence is then associated with random shocks (due to the joint effect of misspecification, omitted variables, and spatial autocorrelation). In alternative, a Spatial Autoregressive Model specification, (Spatial Lag Model) assumes that all spatial dependence effects are captured by the lagged term. The spatial autocorrelation is then modeled by including a spatially lagged dependent variable. Global and local measures of spatial autocorrelation are computed to determine whether the data exhibit spatial dependence and a series of test statistics based on the Lagrange Multiplier (LM) or Rao Score (RS) principle are used to determine whether the variables in the model sufficiently capture the spatial dependence in the data. If the variables do not fully model the dependence, the diagnostics indicate whether the researcher should estimate a model with a spatially lagged dependent variable, a spatially lagged error term, or both. The LM/RS principle can also be

extended to more complex spatial alternatives, such as higher order processes, spatial error components and direct representation models, and to probit models. Paralleling and complementing the theoretical motivation may represent a useful guide for modelling the spatial dependence.

The substantive implications of properly modelling spatial heterogeneity and dependence are linked with methodological implications. The estimation of models incorporating spatial heterogeneity poses both identification and collinearity problems (due to the correlation between unobserved heterogeneity—individual specific effects—and explanatory variables). Regarding spatial dependency. The estimation of spatial lag and spatial error models poses: identification problems, endogeneity/collinearity problems; and the incidental parameter problem. As a consequence: the standard estimation procedures can produce biased parameter estimates, unbiased but inefficient parameter estimates, biased estimates of the SE(s). Consistent estimates may be obtained for both of these specifications using: a combination of Feasible Generalized Least Squares (F-GLS) and Maximum Likelihood estimation (Anselin, [24]; Smirnov and Anselin, [25]; Kelejian and Prucha, [26]; Lee, [27,28]; Bell and Bockstael, [29]), two-stage estimation procedures (Zellner, [30]), the moment conditions for GMM estimation derived by Honoré and Hu [31].

Developing estimation methods designed to improve both ecological inference and small area inference, by combining aggregate and individual-level survey data even in the presence of spatial dependency, would seem to be an important and interesting topic for research.

Hence the distinctive features of this present paper are: (i) the formulation of a generic methodological disaggregation framework within which to estimate sub group/ sub regional indicators in the presence of spatial structures; (ii) the formulation of an informational-theoretical approach to the estimation of economic aggregates or their sub-group/area decomposition, when no individual or sub-group data are available, by assuming heterogeneity across space and/or spatial dependence; (iii) the empirical application of the said approach to real data. This estimation approach presents the advantage of producing consistent estimates in small samples, in the presence of incomplete micro-level data as well as in the presence of collinearity and endogeneity problems, without imposing strong distributional assumptions.

As a first task, a functional relationship between the variable to be disaggregated and a set of variables/indicators at area level is specified by combining different macro and micro data sources. The model at the aggregate level is then estimated and the sub-group level indicators are obtained by employing these parameter estimates. The latent sub-group indicators may be treated *as random coefficients* in a regression model. Alternatively, the parameters of an initial model can be modeled as functions of a set of covariates in an expansion equation, producing a combined model (Spatial Expansion model). In the latter case, the inclusion of the error term in the expansion equation induces heteroskedasticity. Finally, different model specifications extended to include spatial effects are also introduced with the aim of testing the hypothesis of: (i) parameters homogeneity/heterogeneity; (ii) uniform/varying spatial dependence.

2.2. Ecological Inference and Heterogeneity across Space

Econometric studies have proposed several solutions to the problem of incorporating spatial structures when analyzing economic phenomena. When *heterogeneity in parameters* is assumed by

means of Spatial random coefficients models, a Feasible Generalized Least Squares (FGLS) estimation approach is generally used to deal with the induced heteroskedasticity. It is also possible to assume that a parameter is homogeneous within spatial subsets of data and heterogeneous across these subsets and to specify a Spatial switching regression model which refers to a discrete spatial heterogeneity in parameters. By using Spatial Expansion and Geographically Weighted regression models it is possible to introduce a continuous spatial heterogeneity in parameters. In the Spatial Expansion model, the parameters of an initial model are modeled as functions of a set of covariates in an expansion equation, producing a combined model. The inclusion of the error term in the expansion equation induces heteroskedasticity which is modeled via FGLS. A Geographically Weighted Regression (GWR) approach to Ecological Inference can provide a straightforward solution to the problems associated with extreme spatial heterogeneity and autocorrelation (Calvo and Escolar, [32]). This technique extends the traditional regression model, by allowing the estimated coefficients to vary from location to location and is an alternative to the more common spatial weight matrix approaches used for exploring extreme spatial heterogeneity. This method employs distance weights to give more weight in the calculation of the spatially varying parameters and includes all observations in an area that is calibrated by different bandwidths. It represents a good solution to model the spatial structure when second and third-order contiguity entered into the equations.

The assumption of *heterogeneity in functional forms* produces biased standard error estimates, while the assumption of *heterogeneity in error variance* produces Heteroskedasticity of error term, unbiased but inefficient OLS parameter estimates and biased standard error estimates.

The approach we follow when incorporating heterogeneity across space in ecological inference firstly focuses on the problem of decomposing aggregate indicators for various sub groups/regions of a population, by introducing unknown individual-specific effects into the model specification. The advantage is to test possible determinants of the variation in the underlying subgroup indicators.

We start by defining the aggregate indicator for group/region i, $\boldsymbol{y}_i$, as a weighted *geometric* mean of the latent sub group/region indicator $\mathbf{y}_{ij}$ in group/region i: $\mathbf{y}_i = \prod_{j=1}^{J_i} \left(\mathbf{y}_{ij}\right)^{\boldsymbol{\theta}_{ij}}$, that is:

$$\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\ln \mathbf{y}_{ij}\right) \boldsymbol{\theta}_{ij} \tag{1}$$

where $\mathbf{y}_{ij}$ is the indicator of the jth sub group/region in group/region i, $\boldsymbol{\theta}_{ij}$ is the weight of sub group/region j in i, with $\sum_{j=1} \boldsymbol{\theta}_{ij} = 1$, and where $I = 1,..,N$ denotes the groups/regions and $j = 1,..,J_i$ denotes the number of sub groups/regions in i.

The sub-regional indicators are not observed, but the $\mathbf{y}_i$'s and $\boldsymbol{\theta}_{ij}$'s are. In addition, by introducing an observed vector of explanatory variables for group/region i, $\boldsymbol{x}_i$, an observed vector of explanatory variables for sub group/ sub region j in group/region i, $\boldsymbol{z}_{ij}$, the latent sub-group indicators (values) are specified in a multiplicative form as follows:

$$\mathbf{y}_{ij} = \boldsymbol{\alpha}_{ij} \prod_{k=1}^{K} \mathbf{z}_{ij,k}^{\boldsymbol{\beta}_{ij,k}} \prod_{h=1}^{H} \mathbf{x}_{i,h}^{\boldsymbol{\gamma}_{ij,h}} e^{\boldsymbol{\varepsilon}_{ij}} \tag{2}$$

where $\mathbf{z}_{ij,k}$ (k = 1, K) are the covariates observed at the level of sub group/ sub region *j* within the group/region *i*, $\mathbf{x}_{i,h}$ (h = 1,..H) are the covariates observed only at the level of group/region *i*, $\boldsymbol{\alpha}_{ij}$ are unobserved fixed effects, and $\boldsymbol{\varepsilon}_{ij}$ are error terms.

By substituting Equation (2) into Equation (1), we can obtain the following model:

$$\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\ln \boldsymbol{\alpha}_{ij} + \sum_{k=1}^{K} \boldsymbol{\beta}_{ij,k} \ln \mathbf{z}_{ij,k} + \sum_{h=1}^{H} \boldsymbol{\gamma}_{ij,h} \ln \mathbf{x}_{i,h} + \boldsymbol{\varepsilon}_{ij} \right) \boldsymbol{\theta}_{ij}$$

or

$$\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\ln \boldsymbol{\alpha}_{ij} + \sum_{k=1}^{K} \boldsymbol{\beta}_{ij,k} \ln \mathbf{z}_{ij,k} + \sum_{h=1}^{H} \boldsymbol{\gamma}_{ij,h} \ln \mathbf{x}_{i,h} \right) \boldsymbol{\theta}_{ij} + \mathbf{u}_i \tag{3}$$

where $\mathbf{u}_i = \sum_{j=1}^{J_i} \boldsymbol{\varepsilon}_{ij} \boldsymbol{\theta}_{ij}$ is a "composite" error term, which is heteroskedastic.

This model implies some kind of weighted regression, capturing "distributional effects" by using data on weights for each group/region. It is important to point out that we assume: (i) unit specific coefficients for the sub groups/regions (parameter heterogeneity); (ii) a parametric specification of the unobserved spatial effects (spatial heterogeneity) through $\boldsymbol{\varepsilon}_{ij}$'s, which can be positive or negative.

Using the estimated coefficients in Equation (3) we can obtain estimates of the unobserved or latent sub-regional indicators as follows:

$$\hat{\mathbf{y}}_{ij} = \hat{\boldsymbol{\alpha}}_{ij} \prod_{k=1}^{K} \mathbf{z}_{ij,k}^{\hat{\boldsymbol{\beta}}_{ij,k}} \prod_{h=1}^{H} \mathbf{x}_{i,h}^{\hat{\boldsymbol{\gamma}}_{ij,h}} e^{\hat{\boldsymbol{\varepsilon}}_{ij}} \tag{4}$$

2.3. Ecological Inference and Spatial Dependence

As discussed previously, spatial dependency may occur when the responses to particular variables are inherently different across space. On the one hand, spatial dependence may be produced by the diffusion of behaviour between neighbouring units. Alternatively, neighbouring units may share similar behaviours due simply to the units' independent adoptions of the behaviour. If so, the spatial dependence observed in data does not reflect a truly spatial process, but merely the geographic clustering of the sources of the behaviour in question. Such dependence can be termed attributional dependence, as neighbouring units have shared attributes that produce the clustering of behaviours. Clearly, determining which process is producing spatial dependence is critical to our substantive understanding of the behaviour of interest.

As proxies for the ignorance of the sources of spatial dependence, statistically significant parameters on dummy variables for geographic areas merely indicate that behaviours differ for units in these particular areas in contrast to the reference category. Such an approach cannot indicate whether the spatial dependence is consistent with diffusion or with the spatial clustering of the behaviour's sources. Spatial diffusion occurs because units' behaviour is directly influenced by the behaviour of "neighbouring units." This diffusion effect corresponds to a positive and significant parameter on a spatially lagged dependent variable capturing the direct influence between neighbours. In the diffusion case, neighbors influence the behavior of their neighbors and vice versa. Conversely, the geographic clustering of the sources of the behaviour implies an alternative specification

If one is unable to fully model the sources of spatial dependence in the data generating process (DGP), the spatial dependence in the error terms between neighboring locations is assumed. This spatial error dependence can be modeled via a spatially lagged error term. It is also possible to hypothesize that spatial dependence is produced both by the diffusion and by the independent adoption of behaviors by neighbors. This joint spatial dependence can be modeled by incorporating both a spatially lagged dependent variable and a spatial error term, with proper identifying restrictions imposed.

In the case of attributional dependence, errors for neighboring observations exhibit simultaneous dependence. The simultaneous multidimensional nature of spatial dependence leads to implications for inference that are distinct from the time series case. The simultaneous error dependence produce a non-zero covariance between the error terms at all locations with a covariance declining as the order of contiguity increases (Anselin, [24]). This induces heteroskedasticity in the errors which must be accounted for in estimation. When there is spatial autocorrelation in cross-sectional data, constraints must be imposed on the covariances between observations and generally, there is insufficient information in cross-sectional data to estimate each of these covariances.

In the estimation context of sub groups indicators, assuming the model specification (3), in order to take into account the correlation between neighbouring groups/regions (areas), we adopt two alternative spatial model specifications: the spatial Lag and spatial error models.

When the spatial autocorrelation is modeled by a SPATIAL LAG MODEL, SPATIAL AUTOREGRESSIVE MODEL—SAR MODEL, the previous model (3) can be generalized by introducing a spatial-lag term ($\ln \mathbf{wy}_i$) into the model. The resulting latent sub-group indicators (values) are specified in a multiplicative form as follows:

$$\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\ln \boldsymbol{\alpha}_{ij} + \sum_{k=1}^{K} \boldsymbol{\beta}_{ij,k} \ln \mathbf{z}_{ij,k} + \sum_{h=1}^{H} \boldsymbol{\gamma}_{ij,h} \ln \mathbf{x}_{i,h} + \rho \ln \mathbf{wy}_i + \boldsymbol{\varepsilon}_{ij} \right) \boldsymbol{\theta}_{ij} \tag{5}$$

where ρ is a spatial lag coefficient (the parameter associated to the spatially lagged dependent variable, $\ln \mathbf{wy}$), $\mathbf{w}$ is a proximity matrix of order N.

The definition of neighbors for each observation via a spatial weights matrix is a critical decision in modeling spatial autocorrelation. In empirical applications, it is common practice to derive spatial weights from the location and spatial arrangements of observation by means of a geographic information system. In this case, units are defined 'neighbors' when they are within a given distance of each other, ie w_{ij}=1 for $d_{ij} \leq \delta$ and i≠j, where d_{ij} is the distance function chosen, and δ is the critical cut-off value. More specifically, a spatial weights matrix **w*** is defined as follows:

$$w_{ij}^{*} = \begin{cases} 0 & \text{if} \quad i = j \\ 1 & \text{if} \quad d_{ij} \leq \delta, i \neq j \\ 0 & \text{if} \quad d_{ij} > \delta, i \neq j \end{cases} \tag{6}$$

and the elements of the row-standardized spatial weights matrix **w** (with elements of a row sum to one) result:

$$w_{ij} = \frac{w_{ij}^{*}}{\sum_{j=1}^{N} w_{ij}^{*}}, \qquad i, j = 1, .., N. \tag{7}$$

The SAR model assumes that all spatial dependence effects are captured by the lagged term by showing how the performance of the dependent variable impacts all the other (neighbor) groups/regions through the spatial transformation.

In alternative, by assuming a spatial dependence is the error structure (in terms of a first order spatial autoregressive process), the resulting SPATIAL ERROR MODEL (SEM) specification relative to model (3) is derived as follows:

$$\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\ln \boldsymbol{\alpha}_{ij} + \sum_{k=1}^{K} \boldsymbol{\beta}_{ij,k} \ln \mathbf{z}_{ij,k} + \sum_{h=1}^{H} \boldsymbol{\gamma}_{ij,h} \ln \mathbf{x}_{i,h} + \left(\lambda \mathbf{w} \boldsymbol{\varepsilon}_{ij} + \boldsymbol{\tau}_{ij} \right) \right) \boldsymbol{\theta}_{ij} \tag{8}$$

where λ is a spatial autoregressive coefficient, $\mathbf{w}$ is a proximity matrix of order N, as previously defined, and $\boldsymbol{\tau}_{ij}$ are the usual stochastic error terms.

The Spatial Error Model leaves unchanged the systematic component and assumes spatially autocorrelated errors. In this respect, it is observed how a random shock in a group/region affects performances in that group/region and additionally impacts all the other groups/regions through the spatial transformation. This model specification measures the joint effect of misspecification, omitted variables, and spatial autocorrelation.

3. An Information Theoretic Approach

The application of Maximum Entropy methods and Information Theoretic techniques has been explored within the context of ecological inference (Johnston and Pattie, [1]; Judge, Miller and Cho, [2]). The first use of entropy-maximizing models concerned the application of gravity models and transportation flows. Recently, applications of Information Theoretic methods have focused on the analysis of spatial patterns of voting at the individual level (King, Rosen, and Tanner, [18])

However, the present study extends the IT approach to the case of Ecological Inference incorporating Spatial Dependence. Past studies have given little weight to the role of spatial effects in ecological inference analysis, and so this present study is going to introduce a basic framework for EI in the presence of spatial heterogeneity and dependence. It also deals with the specification of models that explicitly control for spatial effects, interpretation and IT-based formulation.

An Information Theoretic technique (Golan, Judge, and Miller, [9]; Bernardini Papalia, [5–7]) is suggested as an adequate solution in the present context since it provides an effective and flexible procedure for reconciling micro and macro data and for addressing problems related to spatial structures.

Information theoretic estimation methods present some useful advantages over classical estimation techniques (as Generalized Least Squares, GLS) that refer to the possibility:

to reformulate “ill-posed” or “under-determined” problems into “well-posed” problem,
to allow for the estimation of each individual parameter *directly*;
to deal with the problem of collinearity, simultaneity, and endogeneity arising in spatial models;
to take into account out-of-sample information which can be introduced as additional constraints in the optimization program or by specifying particular priors for parameters and errors.

In the present work we have embraced the Information Theoretic (IT)—Generalized Cross Entropy (GCE) philosophy and adopted the Kulback-Liebler information-divergence measure (Kullback, [14]).

Implementation of these methods requires that the parameters and errors of the model in Equations (5) and (8) are specified as linear combinations of some predetermined and discrete support values and unknown probabilities (weights). Thus, all coefficients $\boldsymbol{\alpha}_{ij},\boldsymbol{\beta}_{ij},\boldsymbol{\gamma}_{ij},\rho,\lambda$ and unknown errors ε_{ij}, $\boldsymbol{\tau}_{ij}$ in Equations 5 and 8, are reparameterized and expressed in terms of proper probabilities. For each parameter, a set of M support points (with 2 ≤M<∞) has been chosen: $\mathbf{s}_{\alpha}=\left(\mathbf{s}_1^{\alpha},..,\mathbf{s}_M^{\alpha}\right)'$, $\mathbf{s}_{\beta}=\left(\mathbf{s}_1^{\beta},..,\mathbf{s}_M^{\beta}\right)'$, $\mathbf{s}_{\gamma}=\left(\mathbf{s}_1^{\gamma},..,\mathbf{s}_M^{\gamma}\right)'$, $\mathbf{s}_{\rho}=\left(\mathbf{s}_1^{\rho},..,\mathbf{s}_M^{\rho}\right)',\mathbf{s}_{\lambda}=\left(\mathbf{s}_1^{\lambda},..,\mathbf{s}_M^{\lambda}\right)'$, and the corresponding unknown probabilities defined on these support spaces $\mathbf{p}_{\alpha,ij}=\left(\mathbf{p}_{ij,1}^{\alpha},..,\mathbf{p}_{ij,M}^{\alpha}\right)'$, $\mathbf{p}_{\beta,ij}=\left(\mathbf{p}_{ij,1}^{\beta},..,\mathbf{p}_{ij,M}^{\beta}\right)'$, $\mathbf{p}_{\gamma,ij}=\left(\mathbf{p}_{ij,1}^{\gamma},..,\mathbf{p}_{ij,M}^{\gamma}\right)'$, $\mathbf{p}_{\rho,ij}=\left(\mathbf{p}_{ij,1}^{\rho},..,\mathbf{p}_{ij,M}^{\rho}\right)$, $\mathbf{p}_{\lambda,ij}=\left(\mathbf{p}_{ij,1}^{\lambda},..,\mathbf{p}_{ij,M}^{\lambda}\right)$. Similarly, the errors ε_{ij}, $\boldsymbol{\tau}_{ij}$, are treated as unknowns, and a set of R support points $\mathbf{s}_{\varepsilon}=\left(\mathbf{s}_1^{\varepsilon},..,\mathbf{s}_R^{\varepsilon}\right)'$, $\mathbf{s}_{\tau}=\left(\mathbf{s}_1^{\tau},..,\mathbf{s}_R^{\tau}\right)'$, has been chosen, with 2 ≤ j < ∞ with reference to the unknown probabilities $\mathbf{p}_{\varepsilon,ij}=\left(\mathbf{p}_{ij,1}^{\varepsilon},..,\mathbf{p}_{ij,R}^{\varepsilon}\right)',\mathbf{p}_{\tau,ij}=\left(\mathbf{p}_{ij,1}^{\tau},..,\mathbf{p}_{ij,R}^{\tau}\right)'$.

For the sake of simplicity, the above support spaces are constructed as discrete, bounded entities. It is possible to construct unbounded and continuous supports within the same framework (Golan, Judge and Miller, [9]).

The support points are chosen on the basis of *a priori* information as discussed in Golan, Judge and Miller [9]. However, such knowledge is not always available, and symmetric parameter supports around zero are generally used in the presence of scarce prior information about each parameter. With regard to errors, in most cases where the underlying distribution is unknown, one conservative way of choosing the error supports $\mathbf{s}_{\varepsilon},\mathbf{s}_{\tau}$, as recommended by Golan, Judge and Miller, [9], is to employ the "three-sigma rule" established by Pukelsheim.

Under the GCE framework, the full distribution of each parameter and of each error (within their support spaces) is simultaneously estimated under minimal distributional assumptions. More specifically, the parameters $\boldsymbol{\alpha}_{ij},\boldsymbol{\beta}_{ij},\boldsymbol{\gamma}_{ij},\rho,\lambda$ and errors ε_{ij}, τ_{ij} are reparameterized as:

$$\boldsymbol{\alpha}_{ij}=\mathbf{s}_{\alpha}'\mathbf{p}_{\alpha,ij},\quad \boldsymbol{\beta}_{ij}=\mathbf{s}_{\beta}'\mathbf{p}_{\beta,ij},\quad \boldsymbol{\gamma}_{ij}=\mathbf{s}_{\gamma}'\mathbf{p}_{\gamma,ij},\quad \rho=\mathbf{s}_{\rho}'\mathbf{p}_{\rho},\quad \lambda=\mathbf{s}_{\lambda}'\mathbf{p}_{\lambda}\quad \boldsymbol{\varepsilon}_{ij}=\mathbf{s}_{\varepsilon}'\mathbf{p}_{\varepsilon,ij},\quad \tau_{ij}=\mathbf{s}_{\tau}'\mathbf{p}_{\tau,ij} \tag{9}$$

with support vectors for parameters $\boldsymbol{\alpha}_{ij},\boldsymbol{\beta}_{ij},\boldsymbol{\gamma}_{ij},\rho,\lambda$ and errors ε_{ij}, τ_{ij} given by:

$$\begin{aligned}&\mathbf{s}_{\alpha}=\left(\mathbf{s}_1^{\alpha},..,\mathbf{s}_M^{\alpha}\right)',\quad \mathbf{s}_{\beta}=\left(\mathbf{s}_1^{\beta},..,\mathbf{s}_M^{\beta}\right)',\quad \mathbf{s}_{\gamma}=\left(\mathbf{s}_1^{\gamma},..,\mathbf{s}_M^{\gamma}\right)',\quad \mathbf{s}_{\rho}=\left(\mathbf{s}_1^{\rho},..,\mathbf{s}_M^{\rho}\right)',\quad \mathbf{s}_{\lambda}=\left(\mathbf{s}_1^{\lambda},..,\mathbf{s}_M^{\lambda}\right)'\\&\mathbf{s}_{\varepsilon}=\left(\mathbf{s}_1^{\varepsilon},..,\mathbf{s}_R^{\varepsilon}\right)',\quad \mathbf{s}_{\tau}=\left(\mathbf{s}_1^{\tau},..,\mathbf{s}_R^{\tau}\right)\end{aligned} \tag{10}$$

and corresponding unknown probabilities given by:

$$\begin{aligned}&\mathbf{p}_{\alpha,ij}=\left(\mathbf{p}_{ij,1}^{\alpha},..,\mathbf{p}_{ij,M}^{\alpha}\right)',\quad \mathbf{p}_{\beta,ij}=\left(\mathbf{p}_{ij,1}^{\beta},..,\mathbf{p}_{ij,M}^{\beta}\right)',\quad \mathbf{p}_{\gamma,ij}=\left(\mathbf{p}_{ij,1}^{\gamma},..,\mathbf{p}_{ij,M}^{\gamma}\right)',\quad \mathbf{p}_{\rho,ij}=\left(\mathbf{p}_{ij,1}^{\rho},..,\mathbf{p}_{ij,M}^{\rho}\right)\\&\mathbf{p}_{\lambda,ij}=\left(\mathbf{p}_{ij,1}^{\lambda},..,\mathbf{p}_{ij,M}^{\lambda}\right)',\quad \mathbf{p}_{\varepsilon,ij}=\left(\mathbf{p}_{ij,1}^{\varepsilon},..,\mathbf{p}_{ij,R}^{\varepsilon}\right)',\mathbf{p}_{\tau,ij}=\left(\mathbf{p}_{ij,1}^{\tau},..,\mathbf{p}_{ij,R}^{\tau}\right)'\end{aligned} \tag{11}$$

with $M,R \geq 2$.

In addition, prior information reflecting subjective information or any other sample and pre-sample information is introduced by specifying the priors for all parameters and errors: $\widetilde{\mathbf{p}}_{\alpha,ij},\widetilde{\mathbf{p}}_{\beta,ij},\widetilde{\mathbf{p}}_{\gamma,ij},\widetilde{\mathbf{p}}_{\rho,ij},\widetilde{\mathbf{p}}_{\varepsilon,ij},\widetilde{\mathbf{p}}_{\tau,ij}$. These priors may come from prior data, theory, and/or other experiments.

The GCE optimization problem for the ecological spatial model corresponding to Equation (5) can be reformulated by minimizing the following objective function H(.) as follows:

$$H = \sum_i \sum_j \left(\mathbf{p}_{\alpha,ij}\right)' \ln\left(\frac{\mathbf{p}_{\alpha,ij}}{\tilde{\mathbf{p}}_{\alpha,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\beta,ij}\right)' \ln\left(\frac{\mathbf{p}_{\beta,ij}}{\tilde{\mathbf{p}}_{\beta,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\gamma,ij}\right)' \ln\left(\frac{\mathbf{p}_{\gamma,ij}}{\tilde{\mathbf{p}}_{\gamma,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\rho,ij}\right)' \ln\left(\frac{\mathbf{p}_{\rho,ij}}{\tilde{\mathbf{p}}_{\rho,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\varepsilon,ij}\right)' \ln\left(\frac{\mathbf{p}_{\varepsilon,ij}}{\tilde{\mathbf{p}}_{\varepsilon,ij}}\right) \tag{12}$$

subject to:

(i) data consistency conditions:

$$\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\mathbf{s}_\alpha' \mathbf{p}_{\alpha,ij} + \sum_{k=1}^{K} \left(\mathbf{s}_\beta' \mathbf{p}_{\beta,ij}\right) \ln \mathbf{z}_{ij,k} + \sum_{h=1}^{H} \left(\mathbf{s}_\gamma' \mathbf{p}_{\gamma,ij}\right) \ln \mathbf{x}_{i,h} + \left(\mathbf{s}_\rho' \mathbf{p}_{\rho,ij}\right) \ln \mathbf{wy}_i + \left(\mathbf{s}_\varepsilon' \mathbf{p}_{\varepsilon,ij}\right) \right) \boldsymbol{\theta}_{ij} \tag{13}$$

(ii) adding-up constraints for probabilities.

$$\sum \mathbf{p}_{\alpha,ij} = \sum \mathbf{p}_{\beta,ij} = \sum \mathbf{p}_{\gamma,ij} = \sum \mathbf{p}_{\rho,ij} = \sum \mathbf{p}_{\varepsilon,ij} = 1 \quad \forall i, j$$

Analogously, the GCE optimization problem for the ecological spatial model corresponding to Equation (8) can be reformulated by minimizing the following objective function H(.) as follows:

$$H = \sum_i \sum_j \left(\mathbf{p}_{\alpha,ij}\right)' \ln\left(\frac{\mathbf{p}_{\alpha,ij}}{\tilde{\mathbf{p}}_{\alpha,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\beta,ij}\right)' \ln\left(\frac{\mathbf{p}_{\beta,ij}}{\tilde{\mathbf{p}}_{\beta,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\gamma,ij}\right)' \ln\left(\frac{\mathbf{p}_{\gamma,ij}}{\tilde{\mathbf{p}}_{\gamma,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\lambda,ij}\right)' \ln\left(\frac{\mathbf{p}_{\lambda,ij}}{\tilde{\mathbf{p}}_{\lambda,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\varepsilon,ij}\right)' \ln\left(\frac{\mathbf{p}_{\varepsilon,ij}}{\tilde{\mathbf{p}}_{\varepsilon,ij}}\right) + \sum_i \sum_j \left(\mathbf{p}_{\tau,ij}\right)' \ln\left(\frac{\mathbf{p}_{\tau,ij}}{\tilde{\mathbf{p}}_{\tau,ij}}\right) \tag{14}$$

subject to:

(i) data consistency conditions:

$$\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\mathbf{s}_\alpha' \mathbf{p}_{\alpha,ij} + \sum_{k=1}^{K} \left(\mathbf{s}_\beta' \mathbf{p}_{\beta,ij}\right) \ln \mathbf{z}_{ij,k} + \sum_{h=1}^{H} \left(\mathbf{s}_\gamma' \mathbf{p}_{\gamma,ij}\right) \ln \mathbf{x}_{i,h} + \left(\mathbf{s}_\lambda' \mathbf{p}_{\lambda,ij}\right) w\left(\mathbf{s}_\varepsilon' \mathbf{p}_{\varepsilon,ij}\right) + \left(\mathbf{s}_\tau' \mathbf{p}_{\tau,ij}\right) \right) \boldsymbol{\theta}_{ij} \tag{15}$$

(ii) adding-up constraints for probabilities:

$$\sum \mathbf{p}_{\alpha,ij} = \sum \mathbf{p}_{\beta,ij} = \sum \mathbf{p}_{\gamma,ij} = \sum \mathbf{p}_{\lambda,ij} = \sum \mathbf{p}_{\tau,ij} = 1 \quad \forall i, j$$

The optimal solutions depend on the prior information, the data and a normalization factor. If the priors are specified such that each choice is equally likely to be selected (uniform distributions), then the GCE solution reduces to the GME one. As with the GME estimator, numerical optimization techniques should be used to obtain the GCE solution.

4. An Empirical Application

The IT formulation presented in Section 2.2 is applied to an Italian data set provided by the Italian Institute of Statistics (ISTAT), the *Osservatorio Brevetti Unioncamere*, and the EPO (European Patent Office), and refers to the nine provinces within the Emilia Romagna region for the year 2005. The objective of the analysis is to disaggregate the value-added of Emilia Romagna's provinces into six

macro-sectors: Agriculture; Industry, Construction; Commerce, Transport, Hotels/Restaurants Services, Telecommunication; Financial Intermediation Services; Other Services. The total value-added at the sub-regional level of the aforesaid 9 provinces is estimated assuming that the available information at sub regional level pertains to: (i) the total value-added of each province; (ii) the sectors' shares of the total number of workers; while at the regional level, the total value-added for each sector within Emilia Romagna is a known quantity that can be regarded as a fixed regional total.

The Information Theoretic approach may be used to yield the most uninformed distribution in keeping with the observed sample data, while minimal assumptions are made regarding the underlying distribution generating the data. As priors for parameters and errors distributions we consider uniform distributions, then the GCE solution reduces to the GME one. The basic formulation assumes that: (i) the GME estimates of the value-added of Emilia Romagna's nine provinces, disaggregated by sector, are consistent with the total value-added observed at the regional level; (ii) the value-added of the provinces are measured with error. By introducing the baseline statistical model, $\ln \mathbf{y}_i = \sum_{j=1}^{J_i} \left(\ln \boldsymbol{\alpha}_{ij} + \sum_{h=1}^{H} \boldsymbol{\gamma}_{ij,h} \ln \mathbf{x}_{i,h} \right) \boldsymbol{\theta}_{ij} + \mathbf{u}_i$, we estimate the total value-added for each sector at the level of ER's provinces, by employing all available information, that is: sub-area (provincial) level information about certain H explanatory variables $\mathbf{x}_{i,h}$ (in logs), that traditionally refer to measures of the main primary input (employment rate, ER; real capital stock, RC, as well as other measures of innovation, INN), but also refer to measures of certain exogenous or endogenous sources of spatial externalities (population density, POPD, agglomeration spillovers, spillovers connected to specialization and to diversity). A sector's share of the total number of workers here is used for $\boldsymbol{\theta}_{ij}$; together with the total value-added (in log) for each sector within the Emilia Romagna region $\ln \mathbf{y}_i$.

Our choice of model is motivated by a set of diagnostic tests which have been performed with reference to a pooled OLS estimator for a model specification without spatial variables. Spatial heterogeneity is confirmed by spatial correlation tests on GME residuals for a non spatial model specification. Specifically, Moran's I and Geary's C tests accept the null hypothesis of global spatial independence (0.131; p-value: 0.024 for the former; 0.881; p-value: 0.096 for the latter). Alternative specifications, related to spatial LAG model and spatial Error model have been the objective of a preliminary analysis; in all cases spatial lag as well as spatial autoregressive coefficients are not significant. In our analysis, the weight matrix to model spatial dependence is computed by means of the distance from Bologna, where the critical cut-off value is given by the first quartile.

We choose symmetric parameter supports around zero, given that we have very little prior information about each parameter, and M = 5 support points for each parameter, since estimation is not improved by choosing more than about five support points. We choose j = 3 support points for each error, and we specify error supports according to Pukelsheim's "Three Sigma Rule". The estimation procedure is implemented using the GAMS software and a nonlinear solver, CONOPT2.

The results (see Table 1) seem to be relatively robust in terms of estimate signs and magnitude; the GME parameter estimates do not vary a great deal as parameter supports are modified. The choice of support vectors for the parameters, within the intervals (−100, 100) and (−20, 20), has a negligible effect on the coefficients. The asymptotic standard errors are calculated using the method proposed by Golan, Judge and Miller [9].

Table 1. Estimates of the *value added* of Emilia Romagna's provinces disaggregated by six sectors for the year 2005.

Provinces (Emilia Romagna)	**Agriculture**	**Industry**	**Construction**	**Transport, Hotels, Telecom.**	**Financial Services**	**Other Services**
PIACENZA	227.4573	1568.474	366.4849	1522.572	1747.364	1183.74
PARMA	277.8251	3317.61	692.2093	2278.186	3007.466	1704.97
REGGIO NELL' EMILIA	314.3605	4576.695	917.9048	2351.606	3288.713	1552.768
MODENA	362.4585	6376.475	1076.482	3439.923	4597.192	2428.439
BOLOGNA	375.1509	7001.803	1361.174	6264.928	8143.771	5113.127
FERRARA	402.0568	1641.202	522.8528	1658.157	2074.427	1445.826
RAVENNA	341.7594	2037.979	592.185	2176.184	2394.936	1617.214
FORLI-CESENA	307.3608	2395.311	609.7457	2168.655	2435.76	1623.719
RIMINI	141.9309	1048.246	410.3823	2219.982	1991.04	1294.397
Explanatory variables: $\mathbf{x}_{i,h}$ (in logs)	*GME estimates:* $\hat{\boldsymbol{\gamma}}_{i,h}$					
Population Density: POPD	0.11*					
Employment Rate: ER	0.19*					
Per Capita Innovation Activities: INN	0.08*					
Real Capital Stock: RC	0.22**					

* 5% significant level, **1% significant level; fixed effects are included; mean values of the nine estimated varying parameters.

The distribution of the *value-added* of Emilia Romagna's nine provinces, disaggregated by sector for 2005, seems to be quite heterogeneous. Our analysis validates the hypothesis of spatial heterogeneity in slope coefficients ($\boldsymbol{\gamma}_{j,h}$) across the provinces, as well as the contribution made by the explanatory variable consisting in the share of the total number of workers operating in each sector *h* and located in province *i*.

In order to evaluate the accuracy of the disaggregation results, we computed the prediction errors as the difference between observed and predicted value-added for each province and sector; the root mean squared error (RMSE) is defined as: $RMSE = \sqrt{\frac{1}{\sum_i J_i}\sum_j^{J_i}\sum_i^{N}\left(y_{ij} - \hat{y}_{ij}\right)^2}$, and the mean absolute mean squared error (MAPE) is defined as: $RMSE = \sqrt{\frac{1}{\sum_i J_i}\sum_j^{J_i}\sum_i^{N}\left|y_{ij} - \hat{y}_{ij}\right| / y_{ij} \times 100}$.

The RMSE values show a reasonable level of precision, since the mean value is equal to 0.007 with a standard deviation of 0.004. In particular, we observed the 1.7% reduction in the MAPE compared to the OLS estimates.

5. Conclusions

In this paper we have tackled the problem of providing reliable estimates of a target variable in a set of small geographical areas, by exploring spatially heterogeneous relationships at the disaggregate level. Controlling for spatial effects means introducing models whereby the assumption is that values

in adjacent geographic locations are linked to each other by means of some form of underlying spatial relationship.

In certain cases, in order to account for spatial dependency we need to grasp the spatial variations in the regression coefficients, since empirical predictions based on global parameters may be biased, and thus misrepresent local behavior. This is particularly problematic in the case of regional analysis, where locally representative regression coefficients are required for micro-level policy decisions to be taken.

We have discussed the importance of taking into account individually- and spatially-correlated group level variations, and we have recommended the use of Information Theoretic-based methods for the estimation of variables within the small groups in question.

The Information Theoretic-based formulations could be a useful means of including spatial and inter-temporal features in analyses of micro-level behavior, and of providing an effective, flexible way of reconciling micro and macro data. An unique optimum solution may be obtained even if there are more parameters to be estimated than available moment conditions and the problem is ill-posed. Additional non-sample information from theory and/or empirical evidence is introduced in the form of known probabilities by means of the cross-entropy formalism. This procedure is capable of producing consistent estimates in small samples, in the presence of incomplete micro-level data as well as in the presence of problems of collinearity and endogeneity in the individual local models, without imposing strong distributional assumptions.

The Information Theoretic formulation has been employed in relation to an Italian data set, in order to compute the value-added of Emilia Romagna's nine provinces, per sector, by formulating a suitable set of constraints for the optimization problem in the presence of errors in the aggregates at the sub-area level. The results show that this approach provides a flexible, powerful data-disaggregation method, since it enables us to: (i) consider prior knowledge introduced by adding linear and nonlinear inequality constraints, errors in equations, and error in variables; (ii) allow for the efficient use of information from a variety of sources; (iii) reconcile data at different levels of aggregation within a coherent framework.

In this work we have only considered models with continuous response variables. Further work should be done in order to explore IT methods by considering (i) area/group level count outcomes and rates data with spatial structures; and (ii) temporal dependence.

Acknowledgements

This paper has been completed within the BLU-ETS project "Blue-Enterprises and Trade Statistics", a small or medium-scale focused research project funded by the Seventh Framework Programme of the European Commission, FP7-COOPERATION-SSH (Cooperation Work Programme: Socio-Economic Sciences and the Humanities).

References

1. Johnston, R.; Pattie, C. Ecological inference and entropy-maximizing: An alternative estimation procedure for split-ticket voting. *Polit. Anal.* **2000**, *8*, 333–345.

2. Judge, G.; Miller, D.; Cho, W.K.T. An information theoretic approach to ecological inference. In *Ecological Inference: New Methodological Strategies*; King, G., Rosen, O., Tanner, M.A., Eds.; Cambridge University Press: Cambridge, UK, 2004; pp. 162–187.
3. Bernardini Papalia, R. Modeling mixed spatial processes and spatio-temporal dynamics in information-theoretic frameworks. In *COMPSTAT 2006: Proceedings in Computational Statistics*; Rizzi, A., Vichi, M., Eds.; Physica-Verlag: Heidelberg, Germany, 2006; pp. 1483–1491.
4. Bernardini Papalia, R. Combining Incomplete Information and Learning in Microeconometric Models. In *Correlated Data Modeling*; Gregori, D., Mackenzie, G., Friedl, H., Corradetti, R., Eds.; Franco Angeli: Milano, Italy, 2007; pp. 137–143.
5. Bernardini Papalia, R. A Composite Generalized Cross Entropy Formulation in Small Samples Estimation. *Econom. Rev.* **2008**, *27*, 596–609.
6. Bernardini Papalia, R. Analyzing Trade Dynamics from Incomplete Data in Spatial Regional Models: A Maximum Entropy Approach. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*; Springer-Verlag: New York, NY, USA.
7. Bernardini Papalia, R. Data Disaggregation Procedures within a Maximum Entropy Framework. *J. Appl. Statist.* **2010**, in press.
8. Golan, A.; Judge, G.; Robinson, S. Recovering Information from Incomplete or Partial Multisectoral Economic Data. *Rev. Econom. Statist.* **1994**, *76*, 541–549.
9. Golan, A.; Judge, G.; Miller, D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*; Wiley: New York, NY, USA, 1996.
10. Golan, A.; Judge, G.; Miller, D. The Maximum Entropy Approach to Estimation and Inference: An Overview. In *Advances in Econometrics. Applying Maximum Entropy to Econometric Problems*; Fomby, T.B., Hill, R.C., Eds.; Jai Press Ltd.: London, UK, 1997; pp. 3–24.
11. Golan, A. Information and Entropy Econometrics: A Review and Synthesis. In *Foundations and Trends® in Econometrics*; Now Publishers: Hanover, MA, USA, 2008; Volume 2, pp. 1–145.
12. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Techn. J.* **1948**, *27*, 379–423.
13. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
14. Kullback, J. *Information Theory and Statistics*; Wiley: New York, NY, USA, 1959.
15. Levine, R.D. An information theoretical approach to inversion problems. *J. Phys. A* **1980**, *13*, 91–108.
16. Freedman, D.A.; Klein, S.P.; Ostland, M.; Roberts, M.R. Review of "A solution to the ecological inference problem". *J. Am. Statist. Assoc.* **1998**, *93*, 1518–1522.
17. Schuessler, A.A. Ecological inference. *PNAS* **1999**, *96*, 10578–10581.
18. King, G., Rosen, O. and Tanner, M.A. *Ecological Inference: New Methodological Strategies*; Cambridge University Press: Cambridge, UK, 2004; pp. 162–187.
19. Achen, C.H.; Shively, W.P. *Cross-level Inference*; Chicago University Press: Chicago, IL, USA, 1995.
20. Cho, W.K.T. Latent groups and cross-level inferences. *Elect. Stud*. **2001**, *20*, 243–263.
21. Rao, J.N.K. *Small Area Estimation*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2003.

22. Barker, T.; Pesaran, M.H. An introduction in disaggregation in econometric modelling. In *Disaggregation in Econometric Modelling*; Barker, T., Pesaran, H., Eds.; Routledge: London, UK, 1990; pp. 1–14.
23. King, G. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*; Princeton University Press: Princeton, NJ, USA, 1997.
24. Smirnov, O.; Anselin, L. Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Comput. Statist. Data Anal.* **2001**, *35*, 301–319.
25. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer: Boston, MA, USA, 1988.
26. Kelejian, H.H.; Prucha, I.R. A generalized moments estimator for the autoregressive parameter in a spatial model. *Int. Econ. Rev.* **1999**, *40*, 509–533.
27. Lee, L.F. *Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Econometric Models: I. Spatial Autoregressive Processes*; Ohio State University: Columbus, OH, USA, 2001.
28. Lee, L.F. *Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Econometric Models: II. Mixed Regressive, Spatial Autoregressive Processes*; Ohio State University: Columbus, OH, USA, 2001.
29. Bell, K.P.; Bockstael, N.E. Applying the generalized-moments estimation approach to spatial problems involving microlevel data. *Rev. Econ. Statist.* **2000**, *82*, 72–82.
30. Zellner, A. A Bayesian Method of Moments (BMOM): Theory and Applications. *Adv. Econome.* **1997**, *12*, 85–105.
31. Honoré, B.E.; Hu, L. Estimation of cross sectional and panel data censored regression models with endogeneity. *J. Econom.* **2004**, *122*, 293–316.
32. Calvo, E.; Escolar, M. The local voter: A Geographically Weighted Approach to Ecological Inference. *Am. J. Polit. Sci.* **2003**, *47*, 189–204.