Rosa Bernardini Papalia [1]
Esteban Fernandez-Vazquez [2]

PREDICTING INDICATORS AT SMALL SCALE USING ENTROPY ECONOMETRICS

Abstract

Statistical information for empirical analysis is very frequently available at a higher level of aggregation than it would be desired. Economic and social indicators by income classes, for example, are not always available for cross-country comparisons, and this problem aggravates when the geographical area of interest is sub-national (regions). In this paper we propose entropy-based methodologies that use all available information at each level of aggregation even if it is incomplete. This type of estimators have been studied before in the field of Ecological Inference. This research is related to a classical problem in geographical analysis called to modifiable area unit problem, where spatial data disaggregation may give inaccurate results due to spatial heterogeneity in the explanatory variables. An empirical application to Spanish data is also presented.

Keywords

Disaggregated regional data, distributionally weighted regression, generalized cross entropy.

1. Introduction

One relatively frequent limitation for empirical economics is the lack of data available at an appropriate spatial scale. Although the target, in principle, would be to work at a smaller geographical scale, the non-availability of geographically disaggregated information usually limits the conclusions of the analysis at an aggregate level. There is a growing need to produce economic and social indicators at a disaggregate geographic scale and this kind of information has become a focus of recent academic enquiry and planning policy concerns. In this paper we propose entropy-based methodologies that use all available information at each level of aggregation even if it is incomplete. This type of estimators have been studied before in the field of Ecological Inference (EI) (see Judge et al., 2004; Fernandez-Vazquez et al., 2013, Peteers and Chasco, 2006; Bernardini Papalia, 2013, Bernardini Papalia et al., 2013). Generally speaking, EI is the process of estimating disaggregated information from data reported at aggregate level. The foundations of EI were introduced in the seminal works by Duncan and Davis (1953) and Goodman (1953), whose techniques were the most prominent in the field for more than forty years, although the work of King (1997) supposed a substantial development by proposing a methodology that reconciled and extended previously adopted approaches.

[1] Department of Statistical Sciences, University of Bologna, Italy, e-mail: rossella.bernardini@unibo.it
[2] University of Oviedo, Spain, e-mail: evazquez@uniovi.es

Within the set of techniques used for EI problems, the estimation procedures based on entropy econometrics are gaining weight. Recent applications can be found in Judge et al. (2004), Peeters and Chasco (2006) or Bernardini Papalia (2010). This research is related to a classical problem in geographical analysis called to modifiable area unit problem, where spatial data disaggregation may give inaccurate results due to spatial heterogeneity in the explanatory variables. On this background, our proposal is to address within an IT framework the research question of how to exploit all the available aggregate information in order to improve the estimation of disaggregated economic/social indicators. In such a situation, we propose to approach the EI problem by using distributional data to estimate a weighted regression that will be estimated by Generalized Cross Entropy. The proposed estimators present the advantages to produce disaggregated indicators by balancing the costs and errors of the disaggregation for a study area of interest. The methods also account for spatial effects of data autocorrelation and heterogeneity. Autocorrelation is where certain variables included in the model as determinants are related in space, and hence violate traditional statistical independence assumptions, and heterogeneity is where the associationss between variables change across space. The proposed estimators present the advantages to produce disaggregated indicators by balancing the costs and errors of the disaggregation for a study area of interest. The methods also account for spatial effects of data autocorrelation and heterogeneity. Autocorrelation is where certain variables included in the model as determinants are related in space, and hence violate traditional statistical independence assumptions, and heterogeneity is where the associationss between variables change across space. The paper is divided into five further sections. The next section presents the estimation of disaggregated data in terms of a Distributionally Weighted Regression (DWR). The use of entropy econometrics in the context of DWR estimators that account for parameter heterogeneity is presented in section three. Section four evaluates the performance of this type of estimators by means of a numerical simulation under several scenarios. Section four presents an empirical application with Spanish data. The last section presents the main conclusions and possible further lines of research.

2. Distributionally weighted regression: an overview

Consider a geographical area that can be divided into $i = 1,..,T$ smaller spatial sub-areas. Further to this geographical division, suppose that there is another dimension on which we would like to observe some variable or indicator. Consider that this second dimension is the classification into $j = 1, \dots, K$ different classes (for example, population classified by income, age, etc.). The objective of the estimation problem would be to recover the values of the variable disaggregated by sub-areas and by classes, from aggregate information.

We start by paying attention to some indicator of interest which is observable at the level of the $i = 1, \dots, T$ geographical areas, $y_{i\cdot}$ In the context of a DWR, $y_{i\cdot}$is usually defined as a weighted sum of the latent indicators y_{ij}. i.e.:

$$y_{i\cdot} = \sum_{j=1}^{K} y_{ij}\theta_{ij}\,;\ \forall i = 1, \dots, T \qquad (1)$$

where θ_{ij} stands for the observable weights given to class *j* in area *i* defined as population shares:

$$\theta_{ij} = N_{ij}/N_{i\cdot} \qquad (2)$$

Being $N_{i\cdot}=\sum_{j=1}^{K}N_{ij}$. In DWR, the relations between the (latent) disaggregated indicators y_{ij} and the (observable) aggregates $y_{i\cdot}$ are only contained in equation (1). However, other possible relations between disaggregate and aggregate information could be observable as well. Sometimes, aggregate indicators across the i geographical areas for each one of the j classes are available as well and they could be incorporated to the estimation process. Consider the aggregate indicator $y_{\cdot j}$ defined as:

$$y_{\cdot j}=\sum_{i=1}^{T}y_{ij}\eta_{ij}\,;\ \forall j=1,\dots,K \tag{3}$$

where η_{ij} stands for the observable weights given to the area i in class j defined now as population shares:

$$\eta_{ij}=N_{ij}/N_{\cdot j} \tag{4}$$

where $N_{\cdot j}=\sum_{i=1}^{T}N_{ij}$. Note that the additional information considered here are just the K aggregates defined in equation (3), since the weights η_{ij} are, by definition, observable if the θ_{ij} weights are observable too.

Next, the values for the unobservable indicators y_{ij} are modeled as functions of H observable explanatory variables for the class j in each area i ($\boldsymbol{X}_{ij}$, which can include a specific intercept for each class j in area i) and R aggregate covariates observable at the level of the $i=1,\dots,T$ geographical areas ($\boldsymbol{Z}_i$). Assuming a linear relation between the indicator of interest and the covariates, but without loss of generality, this function is defined as:

$$y_{ij}=\sum_{ij,h=1}^{H}\beta_{ij,h}x_{ij,h}+\sum_{i,r=1}^{R}\gamma_{i,r}z_{i,r}+\varepsilon_{ij}, \tag{5}$$

where $\boldsymbol{\beta}_{ij}$ and $\boldsymbol{\gamma}_i$ are the vectors with the parameters to be estimated and ε_{ij} is a residual.[3]

3. Entropy econometrics

The estimation of DWR models like (5) can be based on the use of Entropy Econometrics (EE) for estimating linear models. Generally speaking, EE techniques are used to recover unknown probability distributions of random variables that can take M different known values. The estimate $\widetilde{\boldsymbol{p}}$ of the unknown probability distribution $\boldsymbol{p}$ must be as similar as possible to an appropriate *a priori* distribution $\boldsymbol{q}$, constrained by the observed data. Specifically, the Cross-Entropy (CE) procedure estimates $\widetilde{\boldsymbol{p}}$ by minimizing the Kullback-Leibler divergence $D(\boldsymbol{p}\|\boldsymbol{q})$ (Kullback, 1959):

$$\underset{\boldsymbol{p}}{\operatorname{Min}}\,D(\boldsymbol{p}\|\boldsymbol{q})=\sum_{m=1}^{M}p_m\,ln\left(\frac{p_m}{q_m}\right) \tag{6}$$

The divergence $D(\boldsymbol{p}\|\boldsymbol{q})$ measures the dissimilarity of the distributions $\boldsymbol{p}$ and $\boldsymbol{q}$. This measure reaches its minimum (zero) when $\boldsymbol{p}$ and $\boldsymbol{q}$ are identical and its minimum is reached when no constrains are imposed. If some information (for example, observations on the variable) is available, each piece of information will lead to an update of the *a priori* distribution $\boldsymbol{q}$. When $\boldsymbol{q}$ is set as uniform (a situation without a priori information to favour some of the results), minimizing () is equivalent to maximizing the Shannon's entropy:

---

[3] Note that in this specification the parameters are allowed to vary across the T areas and K classes.

$$\underset{\boldsymbol{p}}{\text{Max}}\, H(\boldsymbol{p}) = -\sum_{m=1}^{M} p_m\, ln(p_m) \quad (7)$$

And the CE procedure is turned into a Maximum-Entropy (ME) problem.

The same underlying idea can be applied for estimating the parameters of general linear models, which leads us to the so-called Generalized Cross Entropy (GCE). The point of departure consists in assuming the parameters to be estimated ($\boldsymbol{\beta}_{ij}$ and $\boldsymbol{\gamma}_i$) as discrete random variables that can take values considered in some supporting vectors with $M \geq 2$ possible values (namely, $\boldsymbol{b}^{\boldsymbol{\beta}}$ and $\boldsymbol{b}^{\boldsymbol{\gamma}}$) with respective unknown probabilities $\boldsymbol{p}^{\boldsymbol{\beta}}$ and $\boldsymbol{p}^{\boldsymbol{\gamma}}$. The ε_{ij} errors are treated in terms of a discrete random variable with unknown probability distribution as well. The uncertainty about the realizations of these errors is introduced in the problem by considering each element ε_{ij} as a discrete random variable with $L \geq 2$ possible outcomes, contained in a supporting vector $\boldsymbol{v}' = \{v_1, \dots, 0, \dots, v_L\}$. The unknown probability distribution for the support vectors will be denoted as $\boldsymbol{w}$.

In general, the support spaces for parameters and errors are constructed as discrete, bounded entities, but it is possible to construct unbounded and continuous supports within the same framework (Golan, Judge and Miller, 1996). The support points in vectors $\boldsymbol{b}^{\boldsymbol{\beta}}$ and $\boldsymbol{b}^{\boldsymbol{\gamma}}$ for the parameters are chosen on the basis of some *a priori* information[4]. 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. The set of possible values for the ε_{ij} errors in vector $\boldsymbol{v}'$ are usually assumed to be symmetric ($-v_1 = v_L$) and centered on zero. With regard to the bounds in this support vector for the errors, the "three-sigma rule" can be used (Golan, 1996). This rule implies to set as upper and lower bounds $\pm$ three times the standard deviation of the dependent variable in a regression model, which in this case is the observable indicator $y_{i\cdot}$. The *a priori* distribution for the parameters (namely, $\boldsymbol{q}^{\boldsymbol{\beta}}$ and $\boldsymbol{q}^{\boldsymbol{\gamma}}$) and the error ($\boldsymbol{w}^{\boldsymbol{0}}$), without any additional prior information, can be naturally set as uniform and the GCE solution reduces to the Generalized Maximum Entropy (GME) one.

Under a GCE framework, the full distribution of each parameter and each error (within their support spaces) is simultaneously estimated under minimal distributional assumptions, by means of the following program:

$$\underset{p^{\beta},p^{\gamma},W}{Min}\; D(\boldsymbol{p}^{\boldsymbol{\beta}}, \boldsymbol{p}^{\boldsymbol{\gamma}}, \boldsymbol{w} \| \boldsymbol{q}^{\boldsymbol{\beta}}, \boldsymbol{q}^{\boldsymbol{\gamma}}, \boldsymbol{w}^{\boldsymbol{0}}) = \sum_{h=1}^{H}\sum_{i=1}^{T}\sum_{j=1}^{K}\sum_{m=1}^{M} p^{\beta}_{ijhm} ln\left(\frac{p^{\beta}_{ijhm}}{q^{\beta}_{ijhm}}\right) + \sum_{r=1}^{R}\sum_{i=1}^{T}\sum_{j=1}^{K}\sum_{m=1}^{M} p^{\gamma}_{irm} ln\left(\frac{p^{\gamma}_{irm}}{q^{\gamma}_{irm}}\right) + \sum_{i=1}^{T}\sum_{j=1}^{K}\sum_{l=1}^{L} w_{ijl}\, ln\left(\frac{w_{ijl}}{w^{0}_{ijl}}\right) \quad (8)$$

Subject to:

$$\sum_{m=1}^{M} p^{\beta}_{ijhm} = \sum_{m=1}^{M} p^{\gamma}_{irm} = \sum_{l=1}^{L} w_{ijl} = 1; \quad (9)$$
$$\forall i = 1, \dots, T;\ \forall j = 1, \dots, K;\ \forall h = 1, \dots, H;\ \forall r = 1, \dots, R$$

[4] The choice of M, and the choice of continuous support spaces and different priors, is discussed in Golan, Judge and Miller, (1996).

$$y_{i\cdot} = \sum_{j=1}^{K} \left[\sum_{h=1}^{H} \sum_{m=1}^{M} p_{ijhm}^{\beta} b_{ijhm}^{\beta} x_{ij,h} + \sum_{r=1}^{R} \sum_{m=1}^{M} p_{irm}^{\gamma} b_{irm}^{\gamma} z_{i,r} + \sum_{l=1}^{L} w_{ijl} v_l \right] \theta_{ij}; \quad \forall i = 1, \ldots, T \qquad (10)$$

In the case that only the aggregate indicators considered in (2) are available, the sample information is contained in (10). If, additionally, aggregate information across the T areas for each one of the K classes is available as in (3), the following additional constrain can be included in the GCE program:

$$y_{\cdot j} = \sum_{i=1}^{T} \left[\sum_{h=1}^{H} \sum_{m=1}^{M} p_{ijhm}^{\beta} b_{ijhm}^{\beta} x_{ij,h} + \sum_{r=1}^{R} \sum_{m=1}^{M} p_{irm}^{\gamma} b_{irm}^{\gamma} z_{i,r} + \sum_{l=1}^{L} w_{ijl} v_l \right] \eta_{ij}; \quad \forall j = 1, \ldots, K \qquad (11)$$

Once estimated the coefficients in (5), the estimates of the indicators for each class j in each area i will be given by:

$$\hat{y}_{ij} = \sum_{ij,h=1}^{H} \hat{\beta}_{ij,h} x_{ij,h} + \sum_{i,r=1}^{R} \hat{\gamma}_{i,r} z_{i,r} + \hat{\varepsilon}_{ij} \qquad (12)$$

The optimal solutions depend on the prior out-of-sample information (the *a priori* distributions and supporting vectors), the data in (10) and (11) and the normalization constrains in (9), which should be found by means of numerical optimization techniques.

4. A numerical simulation

In this section we try to find some empirical evidences, by means of some numerical simulations, on the comparative performance of the DWR estimator to recover a set of $(T \times K)$ disaggregated latent indicators. The point of departure of the experiment is the unknown elements of a target matrix, which are drawn from a log-normal distribution with mean 10 and standard deviation 2. The choice of a log-normal to simulate the target variable is motivated because economic variables like income or productivity often follow this distribution. Once these values are generated, they are divided by the (observable) corresponding population totals N_{ij} to obtain the y_{ij} indicators.

As usual in DWR estimation, regressors to explain the y_{ij} indicators should be available. In order to keep our simulation as simple as possible, we first consider that only one disaggregated regressor x_{ij} is assumed observable. The disaggregated regressor x_{ij} contains some imperfect information on the target indicators. To reflect this idea, in the experiment the elements of the $(T \times K)$ matrix $\boldsymbol{X}$ have been generated in the following way:

$$x_{ij} = y_{ij} + u_{ij}; \; \forall i = 1, \ldots, T; \; \forall j = 1, \ldots, K \qquad (13)$$

Where $\boldsymbol{u} \sim N(0, \sigma y_{ij})$ and σ is a scalar that adjusts the variability of this noise making it proportional to the respective element y_{ij} and it has been set to 0.1. In this context, the latent indicators will be modeled by means of a simple linear regression like:

$$y_{ij} = \alpha_{ij} + \beta_{ij} x_{ij} + \varepsilon_{ij} \qquad (14)$$

Being α_{ij} an area and class-specific intercept. Additionally, an aggregate regressor z_i is incorporated into the model, being the values of this aggregate indicator generated in a similar way, where:

$$z_i = y_{i\cdot} + \epsilon_i;\ \forall i = 1, \dots, T \quad (15)$$

Where $\boldsymbol{\epsilon} \sim N(0, \sigma y_{i\cdot})$ and σ is the same scalar previously defined. In such a case, the DWR model to estimate is:

$$y_{ij} = \alpha_{ij} + \beta_{ij} x_{ij} + \gamma_i z_i + \varepsilon_{ij} \quad (16)$$

The parameters in (14) and (16) will be estimated by the GCE program described in equations (8) to (11) with equal supporting vectors for all them (-100 ,0, 100) with $M = 3$. For the error terms, again the support with $L = 3$ values has been chosen applying the three-sigma rule with uniform a priori weights. The a priori probability distributions taken for the coefficients are uniform as well, so the CGE estimation is equivalent to a GME program.

Two different scenarios with various levels of available aggregate information will be assumed: i) the usual situation where only aggregates for each one the i areas $y_{i\cdot} = \sum_{j=1}^{K} y_{ij}\theta_{ij}$ are known, and ii) an alternative scenario where, additionally, aggregates for each j class $y_{\cdot j} = \sum_{i=1}^{T} y_{ij}\eta_{ij}$ are observable as well.

In the experiment we evaluate the performance of the DWR modeling with and without the additional constrain considered in (11) under different scenarios. Six different dimensions of the target matrix have been considered in the experiment. The six types of matrices reflect several situations with different number of regions (T) and classes (K). For example, in matrices 1, 2 and 3, the number of geographical areas is small ($T = 20$), whereas in cases 4, 5 and 6 more geographical areas are considered ($T = 200$). In each one of these simulated scenarios several possibilities for the number of classes (K) have been considered: namely 2, 4 and 8. In each one of the twelve resulting scenarios 1,000 trials have been carried out and the average of two measures of error have been computed: the root of the mean squared error (RMSE), and the weighted absolute percentage error (WAPE). Table 1 summarizes the results.

		Matrix 1 (20 × 2)		**Matrix 2 (20 × 4)**		**Matrix 3 (20 × 8)**		**Matrix 4 (200 × 2)**		**Matrix 5 (200 × 4)**		**Matrix 6 (200 × 8)**	
		WAPE	**RMSE**	**WAPE**	**RMSE**	**WAPE**	**RMSE**	**WAPE**	**RMSE**	**WAPE**	**RMSE**	**WAPE**	**RMSE**
One regressor	**DWR**	10.46	120.24	14.36	123.85	23.77	209.69	8.06	75.29	11.81	116.52	21.93	209.65
	DWR with additional information	6.22	55.35	9.83	75.85	16.59	123.68	7.74	75.02	11.32	103.01	21.26	203.32
Two regressors	**DWR**	5.18	47.93	12.64	117.13	22.74	205.67	6.73	66.65	9.84	96.11	21.38	207.32
	DWR with additional information	5.12	44.28	8.68	74.39	15.75	120.33	6.73	66.62	9.19	83.76	20.73	200.86

Table 1: *Results of the numerical experiments (1,000 replications)*

5. Empirical application

Complementarily to the numerical simulation carried out in the previous section, the two approaches are tested in an empirical application using a data set for Spain. Spain is administratively divided into 50 provinces for which official data on gross value added by industry (classified into 5 different sectors) are regularly published in the Regional Accounts by the National Statistical Institute (INE). However, the provincial and sectoral aggregates are available much sooner than the disaggregated information, whereas the disaggregated data by industry and province are made public with a time lag of several years. In this context, it would be interesting the application of an estimation procedure that produce disaggregated values quicker than the official ones. The empirical application will be conducted taking as reference year 2005. The target variable is the distribution of GVA per unit of labor by province and industry (now we aim at a latent indicator-target instead of a level-target). The industry aggregates are assumed as observable, as well as additional information required to define the weights θ_{ij}. Specifically, for weights we use that data on labor units (thousands of workers) by industry and province in year 2005. This is a realistic situation, given that the Spanish Labor Force Survey (EPA) publishes estimates of labor by industry and province with quarterly and annual frequency. With all this information the DWR equation has been estimated by means of GME. Tables 2 and 3 compare the results with the actual values by obtaining the absolute percentage error. Table 2 reports the average absolute deviations in percentage over the aggregate GVA by province, whereas Table 3 shows the same average deviation measures in relative terms to the industry aggregates.[5] As a first indicator of the accuracy of the DWR technique, the average absolute error in the estimation of value added per worker is approximately 16%. The general trend that can be observed is that the errors obtained are concentrated in the agriculture and energy activities, and they diminish for the industries with a major share in the economic structure in Spain (manufacturing and services). In terms of variability in the error of the DWR approach across sectors, the biggest provinces in terms of population presented deviation that are above the average error (with the exception of Madrid and Valencia).

Province	*GCE*	*DWR*	*Province*	*GCE*	*DWR*
Albacete	9.33	14.26	Jaén	5.47	9.19
Alicante	2.59	11.90	León	2.74	16.65
Almería	11.90	12.06	Lerida	4.18	12.43
Álava	6.44	8.26	Lugo	11.70	13.45
Asturias	0.78	13.30	Madrid	1.08	12.60
Ávila	21.68	13.93	Málaga	11.11	12.44
Badajoz	12.91	12.42	Murcia	1.59	15.51
Bal. Islands	10.00	14.02	Navarra	7.46	10.75
Barcelona	2.20	12.18	Orense	8.82	10.44
Vizcaya	2.74	11.99	Palencia	11.21	12.06
Burgos	9.84	7.90	Las Palmas	10.34	12.65
Cáceres	19.37	10.74	Pontevedra	1.87	11.93
Cádiz	2.59	14.30	Rioja	6.21	17.31
Cantabria	2.42	10.40	Salamanca	8.10	10.90
Castellón	3.60	11.34	Tenerife	9.53	11.05
Ciudad Real	12.27	15.47	Segovia	15.59	15.71
Córdoba	3.60	8.37	Sevilla	2.62	13.59
Coruña	1.66	11.48	Soria	27.56	11.27
Cuenca	17.00	13.02	Tarragona	5.38	10.05
Gipuzcoa	10.95	7.95	Teruel	19.71	9.08
Gerona	4.48	12.00	Toledo	6.79	15.05

[5] The results are weighted averages where each province and industry is weighted by their number of workers.

Granada	12.19	14.07	Valencia	0.76	13.50
Guadalajara	14.85	9.74	Valladolid	5.79	11.67
Huelva	4.43	17.66	Zamora	19.97	14.54
Huesca	8.52	12.25	Zaragoza	3.78	12.59
Average percentage abs. error DWR = 12.35					

Table 2: *Absolute percentage errors by province (real GDP vs. Estimates)*

Industry	DWR
Energy and manufacturing	8.17
Construction	17.14
Commerce, trade and transport and communic. services	6.61
Financial, insurance and real estate services	15.12
Non-market services	14.71
Average percentage abs. error DWR = 12.35	

Table 3: *Absolute percentage errors by industry (real GVA vs. Estimates)*

6. Final remarks

In this paper a distributionally weighted regressions (DWR) Entropy-based approach to Ecological Inference is formulated in presence of spatial heterogeneity throughout simulation experiments and a real data application. If compared to the traditional EI problem formulations, our approach is remarkably different and presents two distinctive differences in terms of model formulation: (i) spatial heterogeneity of parameters and (ii) data contraints for available aggregate information are introduced. The performance of the proposed approach is tested by means of numerical simulations under several scenarios. We studied the effect of the informative contribution contained in the “non target” variable to predict the sub-area values (or sub-area indicators) for target variables. We also evaluated the performance of the formulation in presence of different number of classes. The results observed in the simulation showed the goods results of the DWR estimator expecially in small samples cases and when additional information at aggregate level is included. The application of the proposed approach is illustrated by means of a real-world example with data of Spain, where the target is the estimation of GVA per unit of labor by province and industry in 2005. The average deviations are similar to those obtained in the numerical simulation and, the DWR equation considered reduces the variability of the deviations across provinces and industries. In this study we have considered the case of continuous target variables. Further work should be done to explore the performance of the competing methods: (i) within a panel data framework; (ii) by improving model specification when the covariates alone do not succeed in accounting for the spatial heterogeneity; (iii) by exploring new IT- based composite estimators. Future research is also needed to evaluate the predictive accuracy of the proposed approaches by the use of discrete target variables and count data.

7. Bibliography

1. Bernardini Papalia R., 2010. Incorporating spatial structures in ecological inference: an information theoretic approach, Entropy, 12, 10, 2171-2185.
2. Bernardini Papalia R., 2013. An Information Theoretic Approach to Ecological Inference in Presence of Spatial Dependence. Defining the Spatial Scale in Modern Regional Analysis: New Challenges from Data at Local Level. Fernández Vázquez E, Rubiera Morollón F. (EDS), pp. 157-172.
3. Bernardini Papalia R, C. Bruch, T. Enderle, S. Falorsi, A. Fasulo, E. Fernandez-Vazquez, M. Ferrante, J. P. Kolb, R. M□unnich, S. Pacei, R. Priam, P. Righi, T. Schmid, N. Shlomo, F. Volk, and T. Zimmermann. 2013. Best practice recommendations on variance estimation and small area estimation in business surveys. Technical report, BLUE-ETS, deliverable D6.2, 2013.
4. Bidani B. and Ravallion M., 1997. Decomposing social indicators using distributional data. Journal of Econometrics 77: 125–139.
5. Duncan, O. D. and Davis B., 1953. An Alternative to Ecological Correlation, American Sociological Review, 18, pp. 665–666.
6. Fernandez-Vazquez E. FR Morellon, EA Jaramillo, 2013. Estimating Spatially Disaggregate Data by Entropy Econometrics: an Excercise of Ecological Inference for Income in Spain. Research in Applied Economics 5(4), pp. 80-96.
7. Golan, A., 2006. Information and Entropy Econometrics. A review and synthesis, Foundations and Trends in Econometrics, 2, pp. 1-145.
8. Golan, A., Judge, G. and Miller, D., 1996. Maximum Entropy Econometrics: Robust Estimation with Limited Data, New York, John Wiley & Sons.
9. Goodman, L., 1953. Ecological Regressions and the Behavior of Individuals, American Sociological Review, 18, pp. 663–666.
10. Judge, G., Miller, D. J. and Cho W. T. K., 2004. An Information Theoretic Approach to Ecological Estimation and Inference, in King, G., Rosen, O. and M. A. Tanner (Eds. Ecological Inference: New Methodological Strategies, Cambridge University Press, pp. 162-187.
11. King, G., Rosen, O. and Tanner M. A., 2004. Ecological Inference: New Methodological Strategies, Cambridge University Press. Cambridge, UK.
12. King, G., 1997. A solution to the Ecological Inference Problem: Reconstructing individual behavior from aggregate data. Princeton, Princeton University Press.
13. Peeters, L. and Chasco, C., 2006. Ecological inference and spatial heterogeneity: an entropy-based distributionally weighted regression approach, Papers in Regional Science, 85(2), pp. 257-276, 06.
14. Robinson W.S., 1950. Ecological correlations and the behavior of individuals. American Sociological Review,15, pp. 351–357.

ACKNOWLEDGEMENTS: *This work 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).*