RAFAEL GARDUÑO-RIVERA AND ESTEBAN FERNÁNDEZ-VÁZQUEZ

Ecological Inference with Entropy Econometrics: Using the Mexican Census as a Benchmark

CIDE

Abstract

....

Most regional empirical analyses are limited by the lack of data. Researchers have to use information that is structured in administrative or political regions which are not always economically meaningful. The non-availability of geographically disaggregated information prevents to obtain empirical evidence in order to answer some relevant questions in the field of urban and regional economics. The objective of this paper is to suggest an estimation procedure, based on entropy econometrics, which allows for inferring disaggregated information from more aggregated data.

In addition to a description of the main characteristics of the proposed technique, the paper illustrates how the procedure works taking as an empirical application the estimation of income for different classes of Mexican municipalities. It would be desirable to apply the suggested technique to a study case where some observable data are available and confront the estimates with the actual observations. For this purpose, we have taken the information contained in the Mexican census as a benchmark for our estimation technique. Assuming that the only available data are the income aggregates per type of municipality and State, we make an exercise of ecological inference and disaggregate these margins to recover individual data.

Keywords: entropy econometrics, ecological inference, Mexico.

Resumen

....

La mayoría de los análisis empíricos regionales se ven limitados por la falta de datos. Los investigadores tienen que utilizar la información que se estructura en las regiones administrativas o políticas que no siempre son económicamente significativos. La falta de disponibilidad de información desagregada geográficamente impide obtener evidencia empírica con el fin de responder a algunas preguntas relevantes en el campo de la economía urbana y regional. El objetivo de este trabajo es sugerir un procedimiento de estimación, basado en la entropía econométrica, lo que permite inferir la información desagregada de más datos agregados.

Además de una descripción de las principales características de la técnica propuesta, el documento ilustra cómo funciona el procedimiento tomando como una aplicación empírica la estimación de ingresos para las diferentes clases de municipios mexicanos. Sería deseable aplicar la técnica sugerida en un estudio de casos donde algunos datos observables están disponibles y hacer frente a las estimaciones con las observaciones reales. Para ello, hemos tomado la información contenida en el censo mexicano

como un punto de referencia para nuestra técnica de estimación. Suponiendo que los únicos datos disponibles son los agregados de la renta por tipo de municipio y estado, hacemos un ejercicio de inferencia ecológica y desglosamos estos márgenes para recuperar los datos individuales.

Palabras clave: entropía econométrica, inferencia ecológica, México.

Ecological inference with Entropy Econometrics: using the Mexican Census as a benchmark

Esteban Fernandez-Vazquez
University of Oviedo, Spain

Rafael Garduño-Rivera
Centro de Investigación y Docencia Económicas, A.C. (CIDE), Mexico.

Abstract:

Most regional empirical analyses are limited by the lack of data. Researchers have to use information that is structured in administrative or political regions which are not always economically meaningful. The non-availability of geographically disaggregated information prevents to obtain empirical evidence in order to answer some relevant questions in the field of urban and regional economics. The objective of this paper is to suggest an estimation procedure, based on entropy econometrics, which allows for inferring disaggregated information from more aggregated data.

In addition to a description of the main characteristics of the proposed technique, the paper illustrates how the procedure works taking as an empirical application the estimation of income for different classes of Mexican municipalities. It would be desirable to apply the suggested technique to a study case where some observable data are available and confront the estimates with the actual observations. For this purpose, we have taken the information contained in the Mexican census as a benchmark for our estimation technique. Assuming that the only available data are the income aggregates per type of municipality and State, we make an exercise of ecological inference and disaggregate these margins to recover individual data.

Keywords: entropy econometrics, ecological inference; Mexico.

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. To overcome this problem, a process of Ecological Inference (EI) is required in order to recover the information at the required spatial scale.

Generally speaking, EI is the process of estimating disaggregated information from data reported at aggregate level. Research in this area has grown enormously in recent years, given its usefulness in many academic disciplines of social science as well as in policy analysis. 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. An extensive survey of recent contributions to the field can be found in King, Rosen and Tanner (2004).

The objective of this paper is to explore the potential of an estimation procedure based on entropy econometrics to recover disaggregated information from more aggregated data. The paper is divided into four further sections. Section 2 provides a picture of the availability of information in Mexico at small scale (municipalities). Section 3 presents the entropy econometrics solution to the estimation problem. In section 4, the performance of the technique is evaluated by applying to a real-world problem where the disaggregated data are observable. Specifically we take the Mexican 2009 *Censos Económicos* as a benchmark. This database is

published every five years and contains information about gross value added, wages and salaries and labor by industry disaggregated at the level of municipalities. The data contained in this census will be confronted with estimates obtained assuming that only aggregated information is observable. The main conclusions and possible further lines of research complete the paper.

2. Ecological Inference as a matrix balancing problem

The problem of estimating spatially disaggregated data can be described in the same terms as on the matrix-balancing problem depicted in Golan (2006, page 105), where the goal is to fill the (unknown) cells of a matrix using the information that is contained in the aggregate data of the row and column sums.

Graphically, the point of departure of our problem is a matrix where the cells z_{ij} are the unknown elements we would like to estimate and we define the following sums $\sum_{j=1}^{T} z_{ij} = z_{i\cdot}$, $\sum_{i=1}^{K} z_{ij} = z_{\cdot j}$, and $\sum_{i=1}^{K} \sum_{j=1}^{T} z_{ij} = z$. The z_{ij} elements can be expressed as a bi-dimensional probability distribution simply dividing the quantities of the matrix by the sum $\sum_{i=1}^{K} \sum_{j=1}^{T} z_{ij} = z$. In such a case, the previous matrix can be rewritten in terms of a new matrix $\boldsymbol{P}$ where the $p_{ij}'s$ are defined as the proportions $\frac{z_{ij}}{z}$, with new row and column margins defined as $R_i = \frac{z_{i\cdot}}{z}$ and $C_j = \frac{z_{\cdot j}}{z}$ respectively. Consequently, the followings equalities are fulfilled by the p_{ij} elements:

$$\sum_{j=1}^{T} p_{ij} = R_{i\cdot}\,; \ \forall i = 1, \dots, K \tag{1}$$

$$\sum_{i=1}^{K} p_{ij} = C_{.j} \; ; \; \forall j = 1, \dots, T \qquad (2)$$

These two sets of equations reflect all we know about the elements of matrix $\boldsymbol{P}$. Equation (1) shows the cross-relationship between the (unknown) $p_{ij}'s$ in the matrix and the (known) sums of each row and column. Additionally, equation (2) indicates that the $p_{ij}'s$ can be viewed as (column) probability distributions. In such a situation, the Cross Entropy (CE) principle can be applied to recover the unknown p_{ij} probabilities if we have available a prior distribution $\boldsymbol{Q}$ that reflects our initial assumptions about the target matrix $\boldsymbol{P}$. In other words, we want to transform an a priori probability matrix $\boldsymbol{X}$ into a posterior matrix $\boldsymbol{P}$ that is consistent with the vectors $\boldsymbol{R}$ and $\boldsymbol{C}$.

The solution to this type of problems is obtained by minimizing a divergence measure with the prior probability matrix $\boldsymbol{Q}$ subject to the set of constraints (1) and (2), which can be written in the following terms:

$$\underset{\boldsymbol{P}}{Min} \, D(\boldsymbol{P}\|\boldsymbol{Q}) = \sum_{i=1}^{K} \sum_{j=1}^{T} p_{ij} \, ln\left(\frac{p_{ij}}{q_{ij}}\right) \qquad (3)$$

The divergence measure $D(\boldsymbol{P}\|\boldsymbol{Q})$ is the Kullback-Liebler entropy divergence between the posterior and prior distributions. The entropy-based estimation techniques outlined above can be directly applied to the field of Ecological Inference (EI). Judge et al. (2004) suggested the use of information-based estimation techniques for EI problems, although in a different context (the estimation of individual voters' behavior from aggregate election data). In this paper we suggest an application of CE following the line of Judge et al. (2004) and posing the EI as a particular case of the more general matrix balancing problem.

Consider a geographical area that can be divided in T smaller spatial units (regions). Besides this first geographical partition, suppose that there is another dimension on which we would like to observe some variable. Consider that this second dimension is the classification into K different industries on which the economic activity can be divided. The objective of the estimation exercise would be to recover the values of the variable disaggregated by regions and industries from aggregate information at the industrial and regional scale. Graphically, this estimation problem can be represented by Figure 1.

Figure 1: The matrix balancing problem.

p_{11}	...	p_{1j}	...	p_{1T}	$R_{1\cdot}$
...		...		...	...
p_{i1}	...	p_{ij}	...	p_{iT}	$R_{i\cdot}$
...		...		...	...
p_{K1}	...	p_{Kj}	...	p_{KT}	$R_{K\cdot}$
$C_{\cdot 1}$	...	$C_{\cdot j}$	...	$C_{\cdot T}$	

Each one of the p_{ij} elements is defined as the proportion of the variable allocated in region i and industry j, forming a $(K \times T)$ matrix $\boldsymbol{P}$ of unknown values. The $(K \times 1)$ row vector $\boldsymbol{R}$ and the $(1 \times T)$ column vector $\boldsymbol{C}$ contain respectively the regional and sectoral shares of the variable across the country. If an a priori probability distribution $\boldsymbol{X}$ is also available, the Cross Entropy procedure outlined previously can be directly applied.

3. A flexible CE estimation with non-reliable margins

The above-sketched procedure assumes that we have perfectly reliable information of the margins $\boldsymbol{R}$ and $\boldsymbol{C}$, which can be considered as an unrealistic assumption. Suppose that we observe row and column margins as $\tilde{\boldsymbol{R}}$ and $\tilde{\boldsymbol{C}}$, where:

$$\tilde{R}_{i.} = R_{i.} + \varepsilon_i; \ \forall i \tag{4}$$

$$\tilde{C}_{.j} = C_{.j} + \epsilon_j; \ \forall j \tag{5}$$

Where ε_i and ϵ_j are random errors that make the observed margins diverge from the real margins of the target matrix. In this situation is still possible to adjust our prior $\boldsymbol{Q}$ with row and column margins not perfectly reliable by means of a Generalized Cross Entropy approach (GCE), following a similar approach to the ideas suggested in Golan and Vogel (2000) or Robinson et al. (2001).

The basic idea is to re-parameterize the errors ε_i and ϵ_j in terms of unknown probability distributions. The uncertainty about the realizations of these errors is introduced in the problem by considering each element ε_i and ϵ_j as discrete random variables with $L \geq 2$ possible outcomes (for the sake of simplicity L is assumed common for both). These values will be contained in two convex sets $\boldsymbol{v'} = \{v_1, \dots, 0, \dots, v_L\}$ and $\boldsymbol{u'} = \{u_1, \dots, 0, \dots, u_L\}$ respectively. We also assume that these possible realizations are symmetric $(-v_1 = v_L;$ $-u_1 = u_L)$ and centered on zero. The unknown probability distributions for the support vectors will be denoted as $\boldsymbol{w_\varepsilon}$ and $\boldsymbol{w_\epsilon}$ and, consequently, the random errors are defined as:

$$\varepsilon_i = \boldsymbol{v'} \boldsymbol{w_{\varepsilon i}} = \sum_{l=1}^{L} w_{\varepsilon il} v_l; \ \forall i$$

$$\epsilon_j = \boldsymbol{u'} \boldsymbol{w_{\epsilon j}} = \sum_{l=1}^{L} w_{\epsilon jl} u_l; \ \forall j$$

Consequently, the GCE problem can be written in the following terms:

$$\underset{X, w_{\varepsilon}, w_{\epsilon}}{\text{Min}} \, D\left(\boldsymbol{P}, \boldsymbol{w}_{\varepsilon}, \boldsymbol{w}_{\epsilon} \| \boldsymbol{Q}, \boldsymbol{w}_{\varepsilon}^{0}, \boldsymbol{w}_{\epsilon}^{0}\right) = \sum_{i=1}^{K} \sum_{j=1}^{T} p_{ij} \, ln\left(\frac{p_{ij}}{q_{ij}}\right)$$

$$+ \sum_{i=1}^{K} \sum_{l=1}^{L} w_{\varepsilon il} \, ln\left(\frac{w_{\varepsilon il}}{w_{\varepsilon il}^{0}}\right) \qquad (6a)$$

$$+ \sum_{j=1}^{T} \sum_{l=1}^{L} w_{\epsilon jl} \, ln\left(\frac{w_{\epsilon jl}}{w_{\epsilon jl}^{0}}\right)$$

Subject to:

$$\sum_{j=1}^{T} p_{ij} = R_{i\cdot} + \sum_{l=1}^{L} w_{\varepsilon il} v_{l} \, ; \, \forall i \qquad (6b)$$

$$\sum_{i=1}^{K} p_{ij} = C_{\cdot j} + \sum_{l=1}^{L} w_{\epsilon jl} u_{l} \, ; \, \forall j \qquad (6c)$$

$$\sum_{l=1}^{L} w_{\varepsilon il} = 1; \, \forall i \qquad (6d)$$

$$\sum_{l=1}^{L} w_{\epsilon jl} = 1; \, \forall j \qquad (6e)$$

Note that both the bounds specified in the support vectors as well as the a priori probability distributions ($\boldsymbol{w}_{\varepsilon}^{0}$ and $\boldsymbol{w}_{\epsilon}^{0}$) reflect our assumptions on the way the errors are affecting the observed margins. Larger bounds in $\boldsymbol{v}$ and $\boldsymbol{u}$ would allow, obviously, for larger errors. In the context of GCE problems, the values of the supporting vectors for the errors are usually fixed following the three-sigma rule (Pukelsheim, 1994), which in this case implies to take as upper and lower bound $\pm$ three times the standard deviation of $\widetilde{\boldsymbol{R}}$ and $\widetilde{\boldsymbol{C}}$ respectively; whereas the a priori distributions ($\boldsymbol{w}_{\varepsilon}^{0}$ and $\boldsymbol{w}_{\epsilon}^{0}$) are set as uniform.

4. The case of Mexico: available data for economic analysis at small-scale

Availability of economic data at small-scale varies very much depending on the country studied. In the specific case of Mexico both aggregated (at state

level) and disaggregated (at municipal level) economic information is available in several surveys published by the Mexican Statistical Agency INEGI.

If the researcher is interested in information observable at a municipal level, but not classified into industries, the most appropriate source of information is the *Sistema Estatal y Municipal de Bases de Datos* (SIMBAD). This database puts together information from several different surveys elaborated by INEGI and provides information at a municipal scale of different variables on a yearly basis. This information, however, does not allow for an industry classification and if we are interested in analysing the figures for some specific industry, the SIMBAD system only provides information at the State level (on a yearly basis as well).

Another inconvenient is that the estimates on GVA added are not available every year at a municipal level, but they are only published for the years when the *Censo Economico* is constructed. If the research we want to conduct implies the observation of data at a municipal level that refers to production, gross value added, etc., the reference is the *Censo Economico*. The National Economic Census in Mexico is the main source of economic data for the National System of Statistical and Geographical Information, and gives the basis for the development of many other economic measures. The information contained in the economic census refers to every economic unit in the country and it provides data for every level of geographic disaggregation, for each one of the more than 950 NAICS activities. Given the enormous amount of resources required to elaborate it, INEGI publishes one economic census every five years, corresponding the most recent one to 2009 (its cost was around 1.9 billion Pesos).

5. Estimates versus "real data": municipal GVA in the state of Aguascalientes

As commented in the previous section, state and industry aggregates are much easier to observe than the disaggregated data by industry and municipality contained in the Census. In this context, it would be interesting the application of an estimation procedure that produce disaggregated values quicker than the official ones. For practical purposes, we have focused on one specific Mexican state, Aguascalientes, although our analysis can be easily extended to other cases. This state is composed by eleven municipalities and we use a classification into five industries (we exclude agriculture). Table 1 presents the figures of gross value added by industry and municipality published in the Mexican Economic Census 2009 (the most recent one):

Table 1. Census data of gross value added by municipality and industry (thousands Mexican pesos)

	Energy	Manufact.	Commerce	Services to companies	Other services	Total (R)
Aguascalientes	3759828	19433189	5392094	4331919	1981483	**34898513**
Asientos	9180	3877	14126	18608	5014	**50805**
Calvillo	1935	199359	100619	11013	37053	**349979**
Cosío	3293	24473	7187	139	1506	**36598**
Jesús maría	17235	3519467	903192	177524	220660	**4838078**
Pabellón de arteaga	35516	185209	83707	18732	27129	**350293**
Rincón de romos	38647	653079	147031	10299	35608	**884664**
San José de Gracia	1713	11364	10085	456	3396	**27014**
Tepezalá	19821	208825	11244	87206	5902	**332998**
El llano	183	11494	15195	792	2718	**30382**
S. Francisco de los romos	10784	2344789	154763	272580	17613	**2800529**
Total (C)	**3898135**	**26595125**	**6839243**	**4929268**	**2338082**	

In our estimation problem we assume that not the individual cells, but only some aggregated information (the column and row totals in bold) is observable. More specifically, vector **R** with dimension (11×1), contains the municipal shares of gross value added across the state (row sums of Table 1). Moreover, vector C (1×5) is given by the proportion of gross value added by industry at the state level (column sums of Table 1). From these

aggregates, we will apply the entropy-based estimation strategy explained in previous sections to recover the distribution of gross value added by municipality and industry.

From this point of departure, the estimation problem can be written as a CE matrix adjustment as follows:

$$\underset{\boldsymbol{P}}{Min}\, D(\boldsymbol{P}\|\boldsymbol{X}) = \sum_{i=1}^{11}\sum_{j=1}^{5} p_{ij}\, ln\left(\frac{p_{ij}}{x_{ij}}\right) \tag{7}$$

Subject to:

$$R_{i\cdot} = \sum_{j=1}^{5} p_{ij}\,;\ \forall i = 1,\dots,11 \tag{8}$$

$$C_{\cdot j} = \sum_{i=1}^{11} p_{ij}\,;\ \forall j = 1,\dots,5 \tag{9}$$

Although several options are possible for the specification of an *a priori* distribution $\boldsymbol{X}$, we opted for taking as reference the values published in the previous Economic Census in 2004. This has the disadvantage that the estimates obtained for those cases (one specific industry in one municipality) where in 2004 there was no economic activity will be equal to zero. In other words, since the zeros in the $\boldsymbol{Q}$ matrix are kept, this means that the CE estimation cannot predict the "birth" of new industries in one municipality.[1]

The above-sketched adjustment method assumes that we have perfectly observable information on margins $\boldsymbol{C}$ and $\boldsymbol{R}$. Whereas assuming that vector $\boldsymbol{C}$ could be observable without error (SIMBAD provides official data yearly on GVA by industry at State level), it is difficult to assume that we have perfectly reliable information on $\boldsymbol{R}$, given that this information (GVA

---

[1] This problem can be solved, on the other hand, by specifying a prior different from the example used in our empirical analysis.

distribution by municipality) is only estimated in the Economic Census. Some proxy variable, however, could be used instead. For example, in the SIMBAD system we can find data on the distribution of public revenues by municipality, which can be supposed to be strongly correlated with the GVA level. In this context, the information of this variable is taken as vector $\widetilde{\boldsymbol{R}}$, being $\widetilde{\boldsymbol{R}} = \boldsymbol{R} + \boldsymbol{\varepsilon}$ and ε_i where each is defined as $\varepsilon_i = \boldsymbol{v}'\boldsymbol{w_{\varepsilon i}} = \sum_{l=1}^{L} w_{\varepsilon il} v_l$. In this context the following GCE problem can be solved:

$$\min_{X, w_\varepsilon} D(\boldsymbol{P}, \boldsymbol{w_\varepsilon} \| \boldsymbol{Q}, \boldsymbol{w_\varepsilon^0}) = \sum_{i=1}^{11} \sum_{j=1}^{5} p_{ij} \ln\left(\frac{p_{ij}}{q_{ij}}\right) + \sum_{i=1}^{11} \sum_{l=1}^{L} w_{\varepsilon il} \ln\left(\frac{w_{\varepsilon il}}{w_{\varepsilon il}^0}\right) \tag{10}$$

Subject to:

$$R_{i.} + \sum_{l=1}^{L} w_{\varepsilon il} v_l = \sum_{j=1}^{5} p_{ij} ; \; \forall i = 1, \dots, 11 \tag{11}$$

$$C_{.j} = \sum_{i=1}^{11} p_{ij} ; \; \forall j = 1, \dots, 5 \tag{12}$$

$$\sum_{l=1}^{L} w_{\varepsilon il} = 1; \; \forall i \tag{13}$$

Note that the last term in (6a) is removed in (10), and (12) is identical to (9), since we assume that $\boldsymbol{C}$ is observed without error. For the sake of simplicity, only three points ($L = 3$) are included in the support vectors of the errors in $\boldsymbol{R}$, which have been fixed using the three sigma rule and being always the central point equal to zero.

The solution to the CE program is presented in Table 2, whereas the result for the GCE adjustment with imperfect margins is presented in Table 3. The figures have been obtained as the respective estimate of p_{ij} multiplied by the total value added for the state of Aguascalientes as a whole (z).

Additional columns have been included, in order to have indicators of the accuracy of our estimates by municipality and industry. Specifically, we obtained the weighted absolute error percentage -WAPE-, given that is frequently used in the studies that evaluate the performance of matrix adjustment techniques. The municipal ($WAPE_i$), industry ($WAPE_j$) and total ($WAPE_T$) error indicators are defined as:

$$WAPE_i = \sum_{j=1}^{5} 100 \frac{|z_{ij} - \hat{z}_{ij}|}{\sum_{j=1}^{5} z_{ij}} \tag{12a}$$

$$WAPE_j = \sum_{i=1}^{11} 100 \frac{|z_{ij} - \hat{z}_{ij}|}{\sum_{i=1}^{11} z_{ij}} \tag{12b}$$

$$WAPE_T = \sum_{i=1}^{11} \sum_{j=1}^{5} 100 \frac{|z_{ij} - \hat{z}_{ij}|}{\sum_{i=1}^{11} \sum_{j=1}^{5} z_{ij}} \tag{12c}$$

where the $\hat{z}_{ij}$ elements denote the estimated gross values added.

Table 2. Deviations of the CE estimates (%)

	Energy	Manufact.	Commerce	Services to companies	Other services	$WAPE_i$
Aguascalientes	8.85	1.71	4.37	3.59	10.20	**3.63**
Asientos	27.61	25.68	10.44	32.20	63.92	**25.54**
Calvillo	98.76	118.12	54.27	43.24	105.02	**92.77**
Cosío	84.12	89.26	180.61	128.20	390.70	**95.33**
Jesús maría	12.62	11.65	55.77	16.41	366.67	**20.61**
Pabellón de arteaga	35.03	66.47	48.37	165.26	78.69	**55.70**
Rincón de romos	73.56	17.51	7.57	46.26	32.51	**26.31**
San José de Gracia	----	42.93	99.90	189.42	78.97	**63.23**
Tepezalá	----	35.12	29.85	5689.07	517.43	**67.90**
El llano	81.25	59.16	1160.50	----	4934.77	**114.83**
S. Francisco de los romos	71.21	0.63	30.73	67.95	0.69	**7.87**
$WAPE_j$	**16.88**	**4.65**	**11.56**	**8.66**	**19.27**	
$WAPE_T$	**7.99**					

Table 3. Deviations of the GCE estimates with noisy margins (%)

	Energy	Manufact.	Commerce	Services to companies	Other services	$WAPE_i$
Aguascalientes	19.94	8.17	0.37	1.86	8.19	**7.99**
Asientos	59.16	59.53	51.87	31.10	14.19	**3.22**
Calvillo	99.44	4.31	33.21	76.17	13.53	**22.37**
Cosío	96.54	60.22	41.78	54.08	0.82	**19.13**
Jesús maría	8.59	17.77	43.08	25.54	317.57	**39.12**
Pabellón de arteaga	49.77	24.24	61.97	89.47	28.21	**15.32**
Rincón de romos	77.19	2.16	11.61	57.17	6.08	**8.71**
San José de Gracia	----	28.69	146.52	246.18	115.01	**1.67**
Tepezalá	----	64.62	30.11	2921.75	223.73	**46.83**
El llano	98.36	96.56	4.93	----	308.31	**39.85**
S. Francisco de los romos	69.60	1.30	30.32	63.87	2.67	**16.58**
$WAPE_j$	**33.23**	**11.54**	**8.72**	**8.13**	**15.72**	
$WAPE_T$	**12.85**					

From the figures in Tables 2 and 3 we can see a huge variability in the accuracy of the estimates, but under the general rule that the largest cells produce comparatively smaller errors. At an industry dimension, the manufacturing activities (with a share of approximately 60% over the gross value added of the state) is the branch with the smallest error. Oppositely,

"Other services" with a GVA that weighted less than 6% in the 2009 state aggregate, was the industry with the largest error.

At a geographical dimension, this same pattern is observed as well, even to a greater extent. The errors range from more than 100% for the municipality of "El llano" (a very small municipality) to less than 4% for the capital city of Aguascalientes, which concentrated more than 78% of the state GVA according to the 2009 Economic Census. There are other medium-size towns like "Jesús María" or "San Francisco de los Romos", which weighted 11% and 6% respectively on the total GVA, where the errors are relatively low as well. Finally, is for the cases of small municipalities (with a weight of less than 2% over the total GVA according to the 2009 Economic Census) where the error figures increase dramatically. All in all, the average absolute error is around 8%, which seems a quite good indicator of the estimates obtained.

6. Conclusions and future research.

This paper applies a methodology based on Entropy Econometrics to estimate data at a municipal scale and classified into industries. The methodology proposed is evaluated with observable data from the Mexican Economic Census. This is an especially useful case because official data are available in a level of disaggregation not usual in other countries. Specifically, our objective was to estimate, by means of the CE procedure, municipal GVA data by industry published by INEGI in the Economic Census of 2009 for the state of Aguascalientes. In spite of some cases where the errors are exceptionally large, the general performance of the method seems to yield quite moderate errors suggesting that the CE approach can be taken as reasonably reliable.

References.

Duncan, O. D. and Davis B., 1953. An Alternative to Ecological Correlation, American Sociological Review, 18, pp. 665–666.

Golan, A., 2006. Information and Entropy Econometrics. A review and synthesis, Foundations and Trends in Econometrics, 2, pp. 1-145.

Golan, A., Judge, G. and Miller, D., 1996. Maximum Entropy Econometrics: Robust Estimation with Limited Data, New York, John Wiley & Sons.

Goodman, L., 1953. Ecological Regressions and the Behavior of Individuals, American Sociological Review, 18, pp. 663–666.

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).

Kapur, J. N. and Kesavan H. K., 1992. Entropy Optimization Principles with Applications. Academic Press. New York.

King, G., Rosen, O. and Tanner M. A., 2004. Ecological Inference: New Methodological Strategies, Cambridge University Press. Cambridge, UK.

King, G., 1997. A solution to the Ecological Inference Problem: Reconstructing individual behavior from aggregate data. Princeton, Princeton University Press.

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.

Robinson W.S. (1950) Ecological correlations and the behavior of individuals. American Sociological Review,15, pp. 351–357

Novedades

....

DIVISIÓN DE ADMINISTRACIÓN PÚBLICA

Ernesto Flores-Roux y Judith Mariscal, *The Development of Mobile Money Systems*, DTAP-256

David Arellano *et al.*, *Control de los conflictos de interés*, DTAP-255

David Arellano, Víctor Figueras y Walter Lepore, *Política de tránsito en el DF de México: una introducción a su estudio empírico*, DTAP-254

Sergio Cárdenas y Maximiliano Cárdenas, *La participación del poder legislativo en la definición de la política educativa en México*, DTAP-253

Sergio Cárdenas, *Administración centrada en la escuela,* DTAP-252

Joanna D. Lucio, Edgar Ramírez y Sergio Cárdenas, *¿Libertad para quién? El efecto de comunidades cerradas en el espacio urbano*, DTAP-251

Edgar E. Ramírez, *Land Development Permitting*, DTAP-250

Rodrigo Sandoval-Almazán, Luis F. Luna-Reyes y J. Ramón Gil-García, *Índice de Gobierno Electrónico Estatal: La medición 2009*, DTAP-249

J. Ramón Gil García y Armando Aldama, *Gobierno electrónico en Canadá: Antecedentes, objetivos, estrategias y resultados*, DTAP-248

J. Ramón Gil García y Luis F. Luna Reyes, *Teoría institucional y simulación dinámica para una mejor comprensión del gobierno electrónico*, DTAP-247

DIVISIÓN DE ECONOMÍA

David Mayer y Grodecz Ramírez, *Ciclo de vida humano y ciclo de vida urbano: Urbanización y desarrollo económico*, DTE-503

Kaniska Dam y Daniel Ruiz Pérez, *On the Existence of Sharecropping*, DTE-502

David Mayer, *Urbanization as a Fundamental Cause of Development*, DTE-501

Arturo Antón y Alan Villegas, *El papel de la tasa de interés real en el ciclo económico de México*, DTE-500

Víctor Carreón, *La arquitectura de mercado del sector eléctrico mexicano*, DTE-499

Sonia Di Giannatale *et al.*, *Confianza, redes sociales y hábitos financieros: un estudio empírico*, DTE-498

Antonio Jiménez, *Coordination Incentives for Information Acquisition with a Finite Set of Players*, DTE-497

Rodolfo Cermeño *et al.*, *Trade Flows and Volatility of their Fundamentals: Some Evidence from Mexico*, DTE-496

Kaniska Dam, *Principal-Agent Assignment*, DTE-495

Luciana Moscoso, *Who Runs Against the Incumbent? Candidate Entry Decisions*, DTE-494

DIVISIÓN DE ESTUDIOS INTERNACIONALES

Farid Kahhat, *Las industrias extractivas y sus implicaciones políticas y económicas*, DTEI-212

Mariana Magaldi de Sousa, *Trade Openness and the Channels of its Impact on Democracy*, DTEI-211

Mariana Magaldi de Sousa, *The Embedded-Agency Approach to Bank Regulation*, DTEI-210

Lorena Ruano, *The Europeanization of National Foreign Policies Towards Latin America...*, DTEI-209

Álvaro Morcillo, *Towards Europeanization?*, DTEI-208

Kimberly A. Nolan García, *Enforcement by Design: The Legalization of Labor Rights Mechanisms in US Trade Policy*, DTEI-207

Kimberly A. Nolan García, *Norms Socialization and NAFTA's Side Accord on Labor*, DTEI-206

Jorge Chabat, *Combatting Drugs in Mexico Under Calderon*, DTEI-205

David Crow, *(Can't Get No) Satisfaction: An Application of Dynamic Loglinear Models...*, DTEI-204

Ugo Pipitone, *Los daños del rey sabio: Mao y China*, DTEI-203

DIVISIÓN DE ESTUDIOS JURÍDICOS

María Solange Maqueo, *Mecanismos de tutela de los derechos de los beneficiarios*, DTEJ-53

Rodolfo Sarsfield, *The Mordida´s Game. How institutions incentive corruption*, DTEJ-52

Ángela Guerrero, Alejandro Madrazo, José Cruz y Tania Ramírez, *Identificación de las estrategias de la industria tabacalera en México*, DTEJ-51

Estefanía Vela, *Current Abortion Regulation in Mexico*, DTEJ-50

Adriana García and Alejandro Tello, *Salaries, Appelate Jurisdiction and Judges Performance*, DTEJ-49

Ana Elena Fierro and Adriana García, *Design Matters: The Case of Mexican Administrative Courts*, DTEJ-48

Gustavo Fondevila, *Estudio de percepción de magistrados del servicio de administración de justicia familiar en el Distrito Federal*, DTEJ-47

Jimena Moreno, Xiao Recio Blanco y Cynthia Michel, *La conservación del acuario del mundo*, DTEJ-46

Gustavo Fondevila, *"Madrinas" en el cine. Informantes y parapolicías en México*, DTEJ-45

María Mercedes Albornoz, *Utilidad y problemas actuales del crédito documentario*, DTEJ-44

DIVISIÓN DE ESTUDIOS POLÍTICOS

Francisco Javier Aparicio and Covadonga Meseguer, *Supply or Demand? Politics and the 3x1 Program for Migrants*, DTEP-228

Ana Carolina Garriga and Brian J. Phillips, *Foreign Aid and Investment in Post-Conflict Countries*, DTEP-227

Allyson Benton, *The Origins of Mexico's Municipal Usos y Costumbres Regimes*, DTEP-226

Ana Carolina Garriga, *Objetivos, instrumentos y resultados de política monetaria. México 1980-2010*, DTEP-225

Andreas Schedler, *The Limits to Bureaucratic Measurement. Observation and Judgment in Comparative Political Data Development*, DTEP-224

Andrea Pozas and Julio Ríos, *Constituted Powers in Constitution-Making Processes. Supreme Court Judges, Constitutional Reform and the Design of Judicial Councils*, DTEP-223

Andreas Schedler, *Transitions from Electoral Authoritarianism*, DTEP-222

María de la Luz Inclán, *A Preliminar Study on Pro and Counter Zapatista Protests*, DTEP-221

José Antonio Crespo, *México 2009: Abstención, voto nulo y triunfo del PRI*, DTEP-220

Andreas Schedler, *Concept Formation in Political Science*, DTEP-219

DIVISIÓN DE HISTORIA

Michael Sauter, *Human Space: The Rise of Euclidism and the Construction of an Early-Modern World, 1400-1800*, DTH-75

Michael Sauter, *Strangers to the World: Astronomy and the Birth of Anthropology in the Eighteenth Century*, DTH-74

Jean Meyer, *Una revista curial antisemita en el siglo XIX: Civiltá Cattolica*, DTH-73

Jean Meyer, *Dos siglos, dos naciones: México y Francia, 1810- 2010*, DTH-72

Adriana Luna, *La era legislativa en Nápoles: De soberanías y tradiciones*, DTH-71

Adriana Luna, *El surgimiento de la Escuela de Economía Política Napolitana*, DTH-70

Pablo Mijangos, *La historiografía jurídica mexicana durante los últimos veinte años*, DTH-69

Sergio Visacovsky, *"Hasta la próxima crisis". Historia cíclica, virtudes genealógicas y la identidad de clase media entre los afectados por la debacle financiera en la Argentina (2001-2002)*, DTH-68

Rafael Rojas, *El debate de la Independencia. Opinión pública y guerra civil en México (1808-1830)*, DTH-67

Michael Sauter, *The Liminality of Man: Astronomy and the Birth of Anthropology in the Eighteenth Century*, DTH-66

Ventas

....

El **CIDE** es una institución de educación superior especializada particularmente en las disciplinas de Economía, Administración Pública, Estudios Internacionales, Estudios Políticos, Historia y Estudios Jurídicos. El Centro publica, como producto del ejercicio intelectual de sus investigadores, **libros**, **documentos de trabajo**, y **cuatro revistas especializadas:** *Gestión y Política Pública*, *Política y Gobierno*, *Economía Mexicana Nueva Época* e *Istor*.

Para adquirir cualquiera de estas publicaciones, le ofrecemos las siguientes opciones:

VENTAS DIRECTAS:	VENTAS EN LÍNEA:
Tel. Directo: 5081-4003 Tel: 5727-9800 Ext. 6094 y 6091 Fax: 5727 9800 Ext. 6314 Av. Constituyentes 1046, 1er piso, Col. Lomas Altas, Del. Álvaro Obregón, 11950, México, D.F.	Librería virtual: **www.e-cide.com** Dudas y comentarios: **publicaciones@cide.edu**

¡¡Colecciones completas!!

Adquiere los **CDs de las colecciones completas** de los documentos de trabajo de todas las divisiones académicas del CIDE: **Economía**, **Administración Pública**, **Estudios Internacionales**, **Estudios Políticos**, **Historia** y **Estudios Jurídicos**.
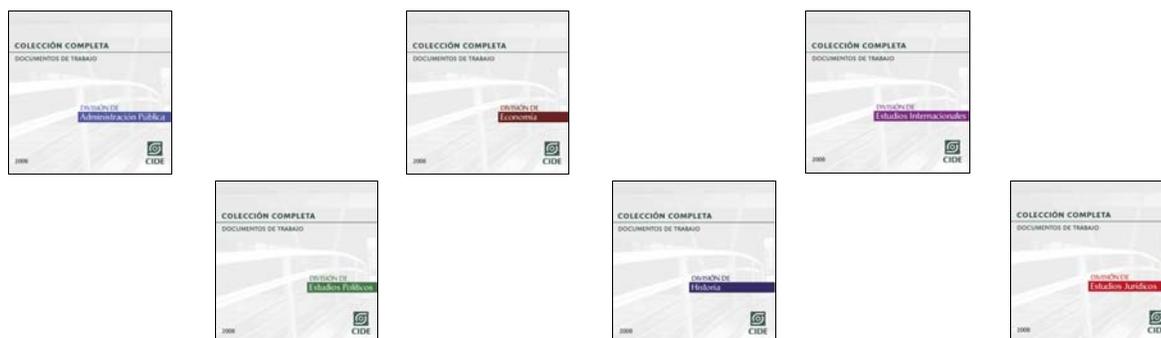
¡Nuevo! ¡¡Arma tu CD!!

Visita nuestra Librería Virtual www.e-cide.com y selecciona entre 10 y 20 documentos de trabajo. A partir de tu lista te enviaremos un CD con los documentos que elegiste.