Personal income distribution and inequality at the local level. An estimation for Spain using a tax microdata-based model.

Miriam Hortas-Rico[a,b]
Jorge Onrubia[a]
Daniele Pacifico[c]

January, 2013

Abstract

Local income data is a key element to analyze residents' standard of living and wellbeing as well as an important economic indicator, very used in a wide range of studies related to regional convergence, urban economics, fiscal federalism and spatial welfare analysis. Despite its importance, there is a lack of official data on local incomes and, most importantly, on local income distributions. In this paper we use official data on personal income tax returns and a reweighting procedure to derive a representative income sample at the local level. Unlike previous attempts in the literature to get local income estimates, the results obtained allow us to derive not only an average value of income but its local distribution, a valuable and informative tool for distributional and income inequality analysis. We apply this methodology to Spanish micro-data and illustrate its potential use in income inequality analysis by means of computed Gini and Atkinson coefficients for a set of municipalities.

Keywords: local income distribution, sample reweighting, income inequality.
JEL classification codes: C42, C61, D31, D63, O15

---

[a] Universidad Complutense de Madrid, Spain.
[b] Institut d'Economia de Barcelona (IEB), Spain.
[c] Department of the Treasury – Italian Ministry of Economy and Finance; Centre for North-South Economic Research, University of Cagliari, Italy
Corresponding author: miriamhortas@ccee.ucm.es

1. Introduction

Local income data is a key element to analyze residents' standard of living and wellbeing as well as an important economic indicator, very used in a wide range of studies related to regional convergence, urban economics, fiscal federalism and spatial welfare analysis. Despite its importance, local income data remains a key missing element within Spanish official statistics. This absence of official estimates of local income has led its measurement to direct or indirect methods by other research institutions, such as the Lawrence R. Klein Institute or the Research Department of La Caixa-Savings Bank. The direct method calculates the disposable income directly considering a production function and sectorial employment matrices with municipal data while the second approach relies on an econometric procedure where local income is estimated as a function of a set of socioeconomic indicators linked to the municipality. Nonetheless, both approaches present several limitations. On the one hand, they proxy personal income through territorialized macroeconomic magnitudes which could not adequately represent residents' ability to pay taxes or their share of disposable income allocated to consumption or saving, as these magnitudes include capital income under the criteria of where production activity is located instead of where their owners reside. On the other hand, these methods do not allow the researcher to obtain local income distributions as their result is a unique value without information about its dispersion.

Hence, in this paper we develop a model of sample reweighting intended to overcome these problems, particularly in the context of distributional and income inequality analysis. The reweighting procedure proposed here adapts the calibration approach proposed in Deville and Särndal (1992), Creedy (2003) and Creedy and Tuckwell (2004) for survey reweighting and allows us to derive local income distributions from micro-data of personal income tax returns. These local income distributions are a valuable and informative tool for distributional and income inequality analysis. They do not only summarize the information contained in thousands of observations (i.e. average taxable income) but provide us with useful inequality indexes drawn from these distributions (i.e. Gini and Atkinson indexes).

Thus, the objectives of the paper are twofold. On the one hand, we seek to provide a representative income sample at the local level based on official tax statistics. To that aim, we adapt a methodology for sample reweighting to the case of Spanish micro-data of personal income tax returns. On the other hand, we use this representative local income sample to derive local income distributions. Unlike previous attempts to get local income estimates, in this paper we obtain not only an average value of income for each Spanish municipality but its local distribution, allowing us to carry out income inequality analysis.

The article is organized as follows. In the next section we present the problem of estimating personal income at the local level and we review the related literature and data sources. The tax microdata-based model and the calibration approach implemented in the paper to get the new sample weights used to derive local income distributions is presented in the third section. Data used, main findings and the validation of estimates are presented in the fourth section. In the fifth section we report an illustration of income inequality analysis for the case of the Spain. Finally, in the last section, we conclude.

2. The problem of measuring the local income

In Spain, as in many other countries, there are not official statistics of personal income for territorial areas smaller than the provinces or the regions. However, the household income estimation is essential information to know the standard of living and wellbeing of the population of the municipalities. In our opinion, local personal income can be considered as one of the most important economic indicators, very used in a wide range of studies related to regional convergence, urban economics, fiscal federalism or spatial welfare analysis, among other topics, not forgetting the entrepreneurial, financial or commercial issues.

2.1. Background in personal disposable income estimation: the case of Spanish municipalities

The aforementioned absence of official estimates of local income has led its measurement to direct or indirect methods. The first proceeding calculates the disposable income directly, considering a production function and sectorial employment matrices with municipal data. This direct approach is based on the aggregation of the monetary flows of goods and services produced in each subsector of the municipal economy. Alternatively, it can be calculated by way of income, adding wages, interest, profits and other income earned by households in the municipality. From this first estimation, wage ratios and gross operating surplus are added to provide an estimate of the value for each element of the matrix (defined by the intersection of each activity with each municipality). It is a complex method that requires a large information database generally difficult to obtain and not always precise. Its main weakness is that it cannot reflect the underground economy of Spanish municipalities, even when estimating agricultural gross added value.

That is why the direct methodology has always needed to be complemented by indirect proceedings.

Thus, the estimation method most commonly used is the indirect one, given the complexity of the direct method discussed above. This methodology consists of an econometric estimation, in which municipal income for a given year is the dependent variable, using as explanatory variables selected socioeconomic indicators linked to the municipality, in addition to the aggregate size of the territorial level closer to municipality, usually the gross value added (GVA) of the province of reference[1].

Since 1992, the Lawrence R. Klein Institute of the Autónoma University of Madrid estimates personal disposable income for a sample of Spanish municipalities[2]. These data have been published, firstly in the "Atlas Comercial de España 1994" (Spanish Trade Atlas), and subsequently in the "Anuario Económico de España" (Spanish Economic Yearbook) sponsored and published by the Barcelona Pensions and Savings Bank, La Caixa, from 1999 to 2003[3]. This information about the personal income of each municipality was displayed on several levels, each of which is ranked in a set of thresholds[4]. This way of presenting information represents an important limitation for empirical work. Furthermore, the aggregate nature of the data prevents us from obtaining inequality measures. Other contributions from the academic field that have also estimated the municipal income usually with a regional scope should be mentioned: Arcarons *et al*. (1994) and Oliver *et al*. (1995) in Catalonia, Esteban and Pedreño (1992) in the Valencia Community, Fernández and Sierra (1992) in La Rioja, De las Heras (1992) and De las Heras and Murillo (1998) in Cantabria and Herrero (1998) in Castile and Leon. Some of them have introduced more complex estimation methods, such as multivariate factor and cluster analysis or econometric multiequational models. Likewise, Alañón (2002) offers an interesting study with estimates of gross value added for the Spanish municipalities using spatial econometric techniques.

The identification of the local income with some territorialized macroeconomic magnitude such as the GVA or even the gross domestic product (GDP) seriously limits the analysis of issues related to personal income distribution. Regardless of the problem of

---

[1] For a detailed description of the content of these models see Fernández-Jardón and Martínez-Cobas (2002).
[2] Only municipalities with more than 1,000 inhabitants are considered (i.e. around 3,200 out of 8,111 municipalities).
[3] http://www.anuarieco.lacaixa.comunicacions.com/java/X?cgi=caixa.le_RightMenuHemeroteca.pattern
[4] In particular, 10 income thresholds are defined: 1. From 0.01 € to 7,200 €; 2. From 7,200.01 € to 8,300 €; 3. From 8,300.01€ to 9,300 €; 4. From 9,300.01€ to 10,200 €; 5. From 10,200.01 € to 11,300 €; 6. From 11,300.01 € a 12,100 €; 7. From 12,100.01 € to 12,700 €; 8. From 12,700.01 € to 13,500 €; 9. From 13,500.01 € to 14,500 €; 10. From 14,500.01 onward. From 2003, this income variable was no longer available in the Spanish Economic Yearbook.

territorial assignation, these macroeconomic magnitudes do not adequately represent residents' ability to pay taxes or their share of disposable income allocated to consumption or saving, as they include capital income under the criteria of where production activity is located instead of where their owners reside. In fact, as the GAV is the value of output (goods and services) produced in an area less the value of intermediate consumption, it is a measure of the contribution to GDP made by an individual producer, industry or sector. In other words, GVA is the source from which the primary incomes of the National Accounts System (SNA) are generated and is therefore carried forward into the primary distribution of income account. Thereby, territorial GVA includes the return on capital income under the criteria of where production activity is located, instead of where their owners reside. For instance, we can think of a residential municipality with a high standard of living where owners of enterprises locate their activities in other municipalities, even in other regions or countries. Of course, there will also be municipalities whose residents do not have a high standard of living but where very profitable companies are located due to, for example, their lower wage costs.

A second limitation of using macro aggregates to estimate local income is related to the impossibility to obtain distributions of income for municipalities and, consequently, measures of inequality. Whatever the statistical or econometric method used to estimate the income of each municipality, the result is a unique value, which prevents having information about the dispersion of the magnitude.

2.2. Personal income estimation for municipalities using micro data

As argued above, the study of local income inequality, either directly or through its consideration in other analyzes in which it acts as an explanatory variable, is a matter of undoubted interest in economic research. Therefore, we believe it is necessary to try to find alternatives to overcome this shortfall. Furthermore, these alternatives should try to overcome the above limitations related to the need of approximating a more precise notion of personal income.

In this regard, the recent availability of data on personal income tax returns turns out to be a feasible approach to address this problem. Nevertheless, the utilization of micro-data from personal income tax returns to estimate local income tax involves using a tax definition of income. Given that our interest is in the estimation of personal income at the local level, we believe this is not a matter for concern. Usually, the taxable income for this kind of taxes includes all incomes obtained by residents in a territory regardless of its source (labor income, capital income -both financial and real estate incomes-, and income from personal business activities). Available information comes from the tax forms according to the rules of taxation and, as such, the income reported includes all essential components of personal income in an economic sense, with the exception of certain exemptions of income that are not taxed. The main limitation arises from the measurement criteria of some kinds of the incomes taxed, as it is the case of income from business activities (largely estimated by means of objective methods) or real estate imputed rents for homeowners, and the capital gains. However, when we look at the measurement of aggregated household disposable income at the national level we observe that it often offers lower income levels than the tax data[5].

Another limitation is related to the unit of analysis used. According to legislation applicable for the Spanish Personal Income Tax (PIT hereinafter), the observation units are not the taxpayers of the tax but rather the income tax returns presented by them. These tax returns can be of two types: individual returns with a sole taxpayer and joint returns with more than one taxpayer. Joint returns can be of two types: those filed by both spouses in a married couple when they decide to file jointly and those filed by a single parent, widow/widower or divorcees together with their under-age or disabled children. In accordance with tax cost of the family unit, married couples who opt for joint taxation have only one income earner. Therefore, in any case, the information contained in the tax returns limits the unit of analysis for the taxpayer unit, not being possible to integrate the tax information at the economic household level, as it is usual in the surveys conducted by statistical agencies (for instance, the household budget survey or the living conditions survey).

After comparing the advantages and limitations of using micro-data from personal income tax returns, this option turns out to be the most suitable one, to the extent that the use of micro-data is indispensable for conducting distributional analysis. The representativeness of these tax micro-data is appropriate for small territorial estimates, as in the case of municipalities. Perhaps the most controversial issue is the use of the individual as the unit of analysis. However, it has been widely used in the related literature on income inequality and redistribution and, therefore, we believe it is a reasonable choice.

[5] For instance, see the analysis made by Picos (2006) for the Spanish case.

3. Tax microdata-based model

3.1. The model

Let $F(Y)$ be the personal income distribution (measured by the variable taxable income) for a given year t (the year 2007 in our case) corresponding to the reference population N. In turn, $F(y)$ is the distribution function of the same variable for the sample obtained from population administrative census of tax returns.

For each of n tax units, micro-data sample contains information on this income variable and other variables of territorial identification, such as provincial and municipal codes. Insofar as the sample has been obtained using minimum variance stratification under Neyman allocation, a sample weight w_i^h was assigned to each observation i extracted. The strata used in the sampling are: a) the province[6]; b) 12 income levels; c) the type of return (separate or joint filing). So, this "original weight" was calculated as the ratio between the size of the population of the strata h and its sample size, $w^h = N^h/n^h$.

Let y_i be the taxable income corresponding to sample tax unit i. The estimated population total of taxable income $(\hat{Y})$ can be obtained using the original weights provided in the sample, such that[7]:

$$\hat{Y} = \sum_{i=1}^{n} y_i w_i \qquad [1]$$

In so far as the spatial stratification variable was fixed at the provincial level, both the population estimates for the provinces and for the whole national population keep the stated confidence level in the sample design. However, to obtain estimates at the municipal level is necessary to calculate new population weights, to the extent that our estimates would face now smaller spatial areas used as a strata sample extraction.

We define this "new weight" as $z_{i,j}$, such that the total population income estimated for the municipality j can be obtained as follows:

$$\hat{Y}_{j|z} = \sum_{i=1}^{n_j} y_{i,j} z_{i,j} \qquad [2]$$

Following Creedy and Tuckwell (2004), we use the distance criterion to assess the closeness between $z_{i,j}$ and $w_{i,j}$ in each of j spatial areas. In general terms, let denote this distance through the function, $\phi(w_{i,j}, z_{i,j})$, what must verified in aggregate terms that:

[6] 48 Spanish provinces with common fiscal regime, plus the Autonomous Cities of Ceuta and Melilla, as well as an additional group of Spanish non-resident taxpayers that paid taxes in Spain (article 10 of the Law 35/2006).
[7] For the sake of simplicity, the superscript h corresponding to each stratum is discarded from now on.

$$D = \phi\left(w_{i,j}, z_{i,j}\right), \; D \in \mathbb{R}_+ \quad [3]$$

Therefore, the method to obtain the new weights that allow estimates of income at the municipal level using micro-data sample consists in solving the following optimization program: to minimize distance function [3] subject to municipality restriction [2]. To carry out this reweighting we need information on true population totals for the taxable income variable, for each j municipality, so that the estimated value $\hat{Y}_{j|z}$ can be replaced in [2]. This information is taken from administrative census of personal income tax[8].

3.2. Computational settlement: the calibration approach

In this section we overview the method we use to adjust the original micro-data sample weights provided by the Spanish Tax Administration (AEAT hereinafter) so as to make them representative with respect to both the average income and the aggregate number of taxpayers in each Spanish municipality. The methodology closely follows Creedy (2003), Creedy and Tuckwell (2004) and Deville and Särndal (1992) and it was coded in Stata 12.

Following Creedy (2003), let us consider a sample of n taxpayers and K individual-level variables, both monetary (as taxable income or tax liability) and non-monetary (as age, sex, province and municipality of residence). We collect these variables for the generic taxpayer i in the following vector: $x_i = [x_{i1}, x_{i2}, \dots, x_{ik}]'$. If we define the original sample weight with the vector $w = [w_1, w_2, \dots, w_i, \dots, w_n]$, the estimated population values of each K individual-level variable is given by:

$$\hat{X}_{k|w} = \sum_{i=1}^{n} w_i x_{ik} \quad [4]$$

The AEAT provided us with the true population totals for some of these K variables (X_k). Specifically, we managed to obtain from the AEAT the aggregate income and the total number of taxpayers in each j Spanish municipality. With this information in hands it is possible to compute a new vector of sample weights for each municipality, $z_j = \left[z_{1j}, \dots, z_{n_j j}\right]$, where $\sum_{j=1}^{J} n_j = n$, that is as close as possible to the original sample weights, while satisfying the set of K calibration equations:

$$X_k^j = \sum_{i=1}^{n_j} z_j x_{ik} \quad [5]$$

[8] This population data have been provided by Spanish Tax Administration Agency (hereafter AEAT).

where X_k^j is the true population value of each K individual-level variable in each j municipality. Indeed, if we denote the distance between the original and the new sample weights with the function $\phi(w_{i,j}, z_{i,j})$, the new sample weights can be obtained by minimizing the following Lagrangian function with respect to $\boldsymbol{z}$:

$$L = \sum_{i=1}^{n} \phi(w_{i,j}, z_{i,j}) + \sum_{k=1}^{K} \lambda_k \left[X_k^j - \sum_{i=1}^{n_j} z_j x_{ik}\right] \quad [6]$$

where $\lambda = [\lambda_1, \lambda_2, ..., \lambda_K]'$ are the Lagrange multipliers.

Clearly, the solution of the minimization problem strongly depends on the property of the distance function $\phi(w_{i,j}, z_{i,j})$ and, in what follows, we require the function $\phi(w_{i,j}, z_{i,j})$ to respect two fundamental properties:

- The first derivative of $\phi(w_{i,j}, z_{i,j})$ with respect to $z_{i,j}$ must be expressed as a function of the ratio between the new and the original weights:

$$\frac{\partial \phi(w_{i,j}, z_{i,j})}{\partial z_{i,j}} = \phi'\left(\frac{z_{i,j}}{w_{i,j}}\right) \quad [7]$$

- The inverse of the first derivative of $\phi(w_{i,j}, z_{i,j})$ must be invertible explicitly.

If these properties hold, then the n first order conditions for the problem in [6] are:

$$\phi'\left(\frac{z_{i,j}}{w_{i,j}}\right) - x_i'\lambda = 0 \quad i = 1,2, \dots, n \quad [8]$$

The new weights can be obtained as:

$$z_{i,j} = w_{i,j}\phi'^{-1}(x_i'\lambda) \quad i = 1,2, \dots, n \quad [9]$$

and given a solution for the Lagrange multipliers, which can be obtained through an iterative procedure (Newton's method) after some algebraic manipulations of equations [9], [5]and [4]. Specifically, if we substitute equation [9] into equation [5] and then subtract from both sides equation [1] after some rearrangements we obtain:

$$\left(X_k - \hat{X}_{k|w}\right) - \sum_{i=1}^{n} w_i \left[\phi'^{-1}(x_i'\lambda) - 1\right] x_i = 0 \quad [10]$$

The root of this function can be computed by means of the following iterative recursion:

$$\lambda^{(I+1)} = \lambda^{(I)} - \left[\frac{\partial f(\lambda)}{\partial \lambda}\right]^{-1} f(\lambda) \quad [11]$$

where $f(\lambda)$ is given by the left hand side of equation [10] and, at each iteration *I+1*th, is evaluated using the value of the Lagrange multipliers in the previous Ith iteration, $\lambda^{[I]}$. Hence, given a set of initial values for λ, equation [11] can be repeatedly evaluated until convergence is reached, where possible.

The four distance functions used in this paper are presented in Table 1. The first function, the Chi-squared distance function, is probably one of the most popular choices in the applied literature because the constrained minimisation problem in equation [6] has an explicit solution and the new weights can be obtained immediately. However, this function place no constraints on the size of the adjustment to each weight and, therefore, it could happen that some of the new weights take negative values.

[Insert Table 1 around here]

To avoid this problem, the other three distance functions in Table 1 incorporate a non-negative constraint on the size of the adjustment. Nevertheless, for these functions a closed-form solution to the constrained minimisation problem is no longer available and the iterative procedure explained above has to be used. This implies that problems of no-convergence may rise, which could depend on the combination of a specific distance function with the original weights or on the starting values that enter the first iteration of the recursion.

Functions 2 and 3 force the new weight to be positive but they do not place an upper bound to the adjustment. Hence, implausible large weights with respect to the original ones could result after the calibration process. This issue is considered by the fourth distance function, Deville and Särndal (1992), because it constraints the new weights within a user-defined range. In particular, the ratio of the new to the original weight is bounded as follows:

$$l < \frac{z}{w} < u \qquad [12]$$

where l and u are known parameters that enter the distance function before the calibration process[9].

[9] The initial values for these parameters are 0.2 and 3, respectively. If convergence is not achieved after 100 iterations with different starting values the new bounds for these two parameters are drawn from two uniform distributions with supports: 0.1-1 and 1-6.

4. Empirical results

4.1. The data

4.1.1. Micro-data (PIT, 2007)

To carry out the estimation of the personal income of Spanish municipalities / the estimation of local income distributions we use micro-data contained in the Spanish PIT annual sample. In particular, in this paper we use the sample for the year 2007, which includes 1,351,802 records extracted from a population of 18,702,875 personal income tax returns (Picos *et al*., 2011). This database has been developed by the Spanish Institute for Fiscal Studies (Instituto de Estudios Fiscales, IEF hereinafter), in collaboration with the AEAT, entity in charge of extracting annual samples from its administrative registers of the Spanish personal income tax[10].

For the construction of this annual sample the minimum variance stratification under Neyman allocation has been used. Thereby population income may be estimated in a highly precise manner with a reasonable sample size. Three stratification variables had been used in the sampling process: a) the province, as territorial stratum (48 provinces with common fiscal regime, plus the Autonomous Cities of Ceuta and Melilla[11]); b) the income level of the tax filers (to that end, income sample places in 12 levels[12]); c) the type of tax return (separate or joint filing). So, the number of strata in the sample is h=1,152 (48x2x12). The sample income was calculated as the sum of net incomes, imputed income and capital gains and losses.

To select the sample, the tax returns were classified in each one of the 1,152 strata. Previously, the size of the total sample n was calculated for a specific relative sampling error (e < 0.011) with a confidence level of 3 per 1000. Next, the population for each stratum (N_h) was determined using the population quasi-variance of the sample income of each one of them (S^2_h). Finally, using the values N_h and S^2_h, the number of observations that had to be extracted randomly for each stratum (n_h) was determined, so that $\sum_h n_h = n$. Table 2 shows the final sample sizes and their distribution by Provinces.

[10] To date, micro-data samples are available to researchers and analysts, free of charge, on application to the IEF (http://www.ief.es) for the years 2002-2009.

[11] This territorial stratum also includes an additional group of Spanish non-resident taxpayers that paid taxes by article 10 of the law 35/2006.

[12] A. Negatives and zero; B. From 0.01 € to 6,000 €; C. From 6,000.01 € to 12,000 €; D. From 12,000.01€ to 18,000 €; E. From 18,000.01€ to 24,000 €; F. From 24,000.01 € to 30,000 €; G. From 30,000.01 € a 36,000 €; H. From 36,000.01 € to 42,000 €; I. From 42,000.01 € to 48,000 €; J. From 48,000.01 € to 54,000 €; K. From 54,000.01 € to 60,000 €; L. From 60,000.01 onward.

The original records provided by the AEAT are incorporated in one bi-dimensional file that contains the PIT returns extracted through a sampling process (one per row). For each observation the file offers a series of variables for which the source of information is, directly or indirectly, the return form for the corresponding year. According to the nature of the variables included in the file, these can split into two groups: non-monetary variables, which contain the main qualitative and personal characteristics of each return; and monetary variables, that contain information from the boxes of the annual PIT return form.

Regarding the non-monetary information the variables included provide a series of personal, family and territorial data: the taxpayer's year of birth and, if applicable, of the spouse, sex, marital status, number of descendants, ascendants or disabilities, autonomous community (region), province and municipality zip code. Besides, as of 2004, the annual samples give qualitative information on self-employment activities (based in the code of activity from Spanish economic activities registration tax, IAE) and as of 2005 they give information on real estate (housing and rent imputations, among others). In total 76 non-monetary variables are provided in 2007, while the monetary variables are 295.

[Insert Table 2 around here]

Regarding territorial representation, the annual sample of micro-data includes tax returns for 5,346 out of 7,024 Spanish municipalities, all of them belonging to the 15 autonomous communities with common tax system (database does not include observations for the Basque Country and Navarra, which have their own tax systems (so-called "foral tax systems").

4.1.2. Population data (PIT 2007)

Statistics with population data for the Spanish PIT are collected by the AEAT. To carry out this paper, the Department of Information Technology has provided us with a database containing information on the municipal income tax for the year 2007. This PIT database includes the following aggregate information for each of the 7,024 municipalities included in the common tax regime: the number of income tax returns filed in the municipality, the average taxable income and the average tax liability. For identification purposes, the database includes a specific municipal code established by the AEAT, and the name of the municipality[13].

4.2. Description of the Spanish municipal map

Spain is a decentralized country composed of three different levels of government: the central government, 17 regional governments named Autonomous Communities (created by mandate of the Spanish Constitution in 1978) and about 8,110 local governments. As it is shown in Table 3, the latter are characterized by their high degree of fragmentation. About 60% of existing municipalities have fewer than 1,000 inhabitants and represent just 3.37% of the total population, which implies a structure of many independent units of government with very small populations.

The aforementioned levels of governments coexist with a historically administrative division of the Spanish territory, the Province. The present division of the country into 50 Provinces has remained essentially unchanged since its design in 1833. Each province consists of a group of municipalities and one or more Provinces yield to an Autonomous Community. Central and Local Governments are formed according to direct election by universal suffrage and subject to a proportional representation criterion, whereas governmental institutions at the Province level respond to representativeness of political parties in each Province's municipalities. That is to say, members of the Provincial government are elected by the municipal councilors among themselves.

[Insert Table 3 around here]

There exists a high degree of heterogeneity among Spanish Provinces, in terms of both number of municipalities and population size (see Appendix 1). The most populated Provinces are Madrid, Barcelona, Valencia, Seville, Alicante and Malaga. Burgos, Salamanca, Barcelona, Zaragoza, Guadalajara, Navarra and Valencia are the Provinces with a greater number of municipalities. With the exception of Barcelona and Valencia, Provinces with the highest number of municipalities (above 200) are among the less populated, since their proportion of overall population rank between 0.32 to 2.02% of total Spanish population. With regard to the municipalities by population size, the greater dispersion is found in the lower population thresholds where, for instance, Provinces can have either 0 or 345

[13] There is an important previous task of linking tax codes (population data) to postal codes (sample data) and then to the 5-digit codes given by the Spanish National Statistics Institute to identify each municipality.

municipalities with less than 1,000 inhabitants. In contrast, municipalities with population above 20,000 inhabitants show much less dispersion.

4.3. Main findings and validation of estimates

As aforementioned, the AEAT provided us with a sample of 5,346 out of 7,024 Spanish municipalities, i.e. those with common fiscal regime (1,337,957 records out of 18,702,875 personal income tax returns). We discarded 18 municipalities that had only one observation in the sample, as for them it was not possible to apply any of the reweighting methods presented in Section 3[14]. The AEAT provided us with two population totals, that is, the number of tax-payers and the aggregate income of each municipality. Hence, the set of calibration equations in our exercise is defined from this data.

Table 4 shows the percentage of the 5,328 municipalities for which convergence has been achieved when the recursive algorithm was used. The table also reports the percentage of municipalities for which non-negative weights were observed after the calibration with the Chi-squared distance function.

[Insert Table 4 around here]

For 250 municipalities (1,953 personal income tax returns) none of the functions listed above produced a new vector of weights, either because of no-convergence issues or because the Chi-squared distance function produced negative weights[15]. However, from the Kernel density of the population size of these municipalities, it can be seen that they are quite small, with less than 1,000 inhabitants (see Figure 1). Accordingly, the total number of PIT taxpayers in these municipalities is also small (below 500 tax returns). As a result, from the Kernel density of the number of observations included in the AEAT sample it can be seen that the number is considerably small (below 30 tax returns included in the sample).

[Insert Figure 1 around here]

Table 5 shows the number of municipalities for which each distance function was chosen for the estimation of the new optimal vector of weights. For the selection among

[1]The estimation of the new weights requires at least two observations for each municipality.

[15] Note that whenever a new weight is not produced for a given observation of a given municipality, all observations of that municipality are dropped from the analysis.

different vectors of weights we follow Särndal (2007) and require the chosen vector for municipality *j*:

(i) not to take negative values:

$$z_{i,j} \geq 0 \quad \forall i \tag{13}$$

(ii) not to have too large values with respect to the original vector. In this regard, the goodness-of-fit criterion (minimizing the sum of the squared residuals) is used

$$\min \sum_{i=1}^{n} (w_{i,j} - z_{i,j})^2 \tag{14}$$

(iii) and to originate from a calibration exercise that converged as smooth as possible.

[Insert Table 5 around here]

As it can be seen, the Minimum Entropy distance is the function adopted in most of the cases, according to the selection criteria explained above. Then, it follows the Chi-squared and the DS distance function. However, as Deville and Särndal (1992) prove, all the above-listed functions generate asymptotically-equivalent calibration estimators. Hence, changes of the distance function will often have minor effects only on the variance of the calibration estimator, even if the sample size is rather small.

Figure 2 shows the distribution of the ratio of calibrated new sample weights with respect to the original sample weights. As it can be seen, the majority of these values are around one, meaning that the new weights are fairly close to the original sample weights. For the sake of clarity, the distribution of this ratio by percentiles is reported in Table 6. The results indicate that the values of the ratio between new and original sample weight ranges from 0.06 to 1.80. Besides, both the mean and the median are close to one, with a standard deviation of 0.98.

[Insert Figure 2 around here]

[Insert Table 6 around here]

Once the new sample weight is chosen and given the micro-data provided by the AEAT, we can derive representative personal income distributions for all Spanish municipalities included in the sample.

Figure 3 shows the income distribution for the entire sample, i.e. all municipalities included, before and after reweighting. As expected, the overall income distribution derived from the new sample weights replicates the overall income distribution when using the original sample weights. In general terms, differences are expected in local income distributions, as original weights were just representative at the provincial level while the new sample weights are now representative at the municipal level. In any case, the sample is always representative of the entire population, i.e. the weights are used for grossing up from the sample in order to obtain estimates of population values.

[Insert Figure 3 around here]

In Figures 4 and 5 we present some of the results obtained for the Spanish municipalities included in the sample. In particular, we display some local income distributions of poor and rich municipalities. In every graph, the local income distribution derived from the original sample weights is illustrated by the dash line, whereas the redefined local income distribution according to the new calibrated sample weights is given by the solid line.

[Insert Figure 4 around here]

[Insert Figure 5 around here]

5. Personal income inequality in Spanish municipalities

The estimated local income distributions obtained in the previous section are a valuable and informative tool for distributional and income inequality analysis. As an illustration, in this section we perform an analysis of local inequality for a sample of Spanish municipalities based on the computation of two of the most common measures of inequality, the Gini and the Atkinson indexes.

The Gini coefficient (Gini, 1912) is probably the standard in the income inequality literature. This index is defined as the area between the 45°(which indicates perfect equality) and the Lorenz curve

$$G(y) = 1 - 2\int_0^1 L(p;y)dp \qquad [15]$$

where the Lorenz curve of income $L(p;y)$ at such p-values of ranked relative cumulated-population (so that, $p \in (0,1)$) can be defined mathematically by the expression,

$$p = F(q) \Longrightarrow L(p; y) = \int_0^q yf(y)dy/\mu_y \qquad [16]$$

Accordingly, the Gini coefficient takes values between zero (perfect equality) and one (complete inequality).

As it is well known, we can infer income distribution $F(y)$ if we know mean income μ_y and the shape of the Lorenz curve of income $L(p; y)$[16]. Alternatively, the Gini coefficient can be calculated in terms of the covariance between income levels and their ranks, so that,

$$G(y) = \left(2/\mu_y\right) \operatorname{cov}\{y, F(y)\} \qquad [17]$$

An alternative formula for the Gini coefficient for the discrete approach of income distribution is,

$$G(y) = \sum_a \sum_b \frac{|y_a - y_b|}{2N^2\mu_y} \qquad [18]$$

where *a* and *b* are two generic observations of *N*-observations size income distribution *y*.

There are several plausible alternatives to calculate this expression when using micro-data. In particular, we use the Stata's do file provided by Haughton and Khandker (2009) and adapted for our stratified sample of micro-data.

The second income inequality measure used in our analysis is the Atkinson index (Atkinson, 1970). This index differs from the Gini index in its explicitly ethical foundation. In fact, the Atkinson index is based upon a social welfare function, including a weighting parameter ε which measures aversion to inequality, so that the index becomes more sensitive to changes at the lower end of the income distribution as approaches to 1, while if the level of inequality aversion falls (that is, as approaches to 0) the index becomes more sensitive to changes in the upper end of the income distribution. For $\varepsilon = 0$ the equally distributed equivalent income is simply the average level of income, while for $\varepsilon \rightarrow \infty$ the Rawlsian criterion is used (that is, social welfare function is close to the maximum concavity).

From a continuous approach of the income distribution, the Atkinson index is defined as,

$$A_\varepsilon(y) = 1 - \left(\int_0^\infty \left(\frac{y}{\mu_y}\right)^{1-\varepsilon} f(y)dy\right)^{\frac{1}{1-\varepsilon}} \qquad [19]$$

for $\varepsilon > 0$ and $\varepsilon \neq 1$, while if $\varepsilon = 1$, the expression is,

[16] For a demonstration of this lemma see Lambert (2001:32)

$$A_1(y) = 1 - \exp\left(\int_0^{\infty}\left(\frac{y}{\mu_y}\right) f(y)dy\right) \quad [20]$$

In both cases, the Atkinson index takes a value between 0 (if the income is distributed equally) and 1 (if the inequality is the highest).

We also use the Stata's do file included in Haughton and Khandker (2009) to calculate the Atkinson index with micro-data, according to the following expression in discrete terms (for N observations)[17]:

$$A_\varepsilon(y) = 1 - \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\mu_y}\right)^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}} \quad [21]$$

for $\varepsilon > 0$ and $\varepsilon \neq 1$. In particular, we have chosen for our calculations the value 0.5 for the ε inequality aversion parameter.

Confidence intervals via bootstrap re-sampling methods (Mills and Zandvakili, 1997) have been calculated for both inequality measures. In particular, two types of bootstrap confidence intervals are obtained, using respectively the alpha-percentile method and the normal-distribution method.

In the first case, the percentile confidence interval is defined as $\left[I^*_{\alpha/2}, I^*_{1-\alpha/2}\right]$, where I^* is the inequality coefficient (Gini or Atkinson index) estimated from a bootstrap sample and α is (100–confidence level)/100.

In the second case, the normal approximation confidence interval is defined as $\left[I^*_{\pi_1}, I^*_{\pi_2}\right]$, so that, $I_{\pi_1} = \mu_{I(y)} - \theta^{\frac{1-\alpha}{2}} \cdot \frac{\mu_{I(y)}}{\sqrt{B}}$ and $I_{\pi_2} = \mu_{I(y)} - \theta^{\frac{\alpha}{2}} \cdot \frac{\mu_{I(y)}}{\sqrt{B}}$, where $\mu_{I(y)}$ is the mean of inequality coefficients obtained from B bootstrap re-samples implemented, and θ represents the standard normal distribution values. Then for a 95% confidence interval based on the normal approximation is,

$$\left[I^*_{\pi_1}, I^*_{\pi_2}\right]_{(95\%)} = \mu_{I(y)} \pm 1.96\frac{\mu_{I(y)}}{\sqrt{B}} \quad [22]$$

Given the large size of the micro-data sample used in our analysis, the number of bootstrap replicates has been set at 100. Likewise, we have calculated the standard errors for both inequality indexes.

---

[17] For $\varepsilon = 1$ the above expression transforms into: $A_1(y) = 1 - \frac{\prod_{i=1}^{N} y_i^{1/N}}{\mu_y}$

Using the AEAT micro-data and the new sample weights, we calculate these two different income inequality measures at the municipality level. The results for both income inequality indexes are reported in Figures 6 and 7, respectively. Detailed results on these indexes as well as their bootstrapped standard errors and confidence intervals are presented in Appendix 2 and 3.

[Insert Figure 6 around here]

[Insert Figure 7 around here]

To the purpose of this empirical exercise we have selected a small sample of Spanish municipalities. In particular, only the results for the 56 municipalities with a population above 100,000 inhabitants are reported. For the sake of clarity and according to their size, this sample of municipalities has been divided into three groups. Two main finding arise from the results. On the one hand, the Gini coefficient has a wide range of variation, as it takes values from 0.38 to 0.52. On the other hand, there exists a clearly positive correlation between the Gini coefficient and the average taxable income of the municipality, with a correlation coefficient of 0.65. This result suggests that richer cities have more income inequality (more unequal income distributions) than the poorer ones. This result also holds for the Atkinson coefficients, whose results exhibit a very similar pattern of variation than those presented for the Gini coefficient.

6. Concluding remarks

Local income data is a key element to analyze residents' standard of living and wellbeing as well as an important economic indicator, very used in a wide range of studies related to regional convergence, urban economics, fiscal federalism and spatial welfare analysis. Despite its importance, there is a lack of official data on personal incomes for territorial areas smaller than the provinces or regions. This paper makes use of official data on personal income tax returns and a reweighting procedure to derive a representative income sample at the local level. The methodology implemented here relies on the calibration approach proposed in Deville and Särndal (1992), Creedy (2003) and Creedy and Tuckwell (2004) for survey reweighting. In doing so, we adjust the original micro-data sample weights so as to make them representative at the local level, given that our estimates would face now smaller spatial areas used as a strata sample extraction.

Unlike previous attempts in the literature to get local income estimates, the results obtained allow us to derive not only an average value of income but its local distribution, a valuable and informative tool for distributional and income inequality analysis. We apply this methodology to Spanish micro-data and illustrate its potential use in income inequality analysis. In particular, in this paper we use the micro-data contained in the Spanish Personal Income Tax annual sample for the year 2007, which includes 1,351,802 records extracted from a population of 18,702,875 personal income tax returns. As a result, we obtain local income distributions for 5,328 out of 7,024 Spanish municipalities with common fiscal regime. Next, we perform an analysis of local income inequality for a sample of those municipalities based on the computation of two of the most common measures of inequality, the Gini and the Atkinson indexes. Two main findings arise from the results. On the one hand, the Gini coefficient has a wide range of variation, as it takes values from 0.38 to 0.52. On the other hand, there exists a clearly positive correlation between the Gini coefficient and the average taxable income of the municipality, with a correlation coefficient of 0.65. This result suggests that richer cities have more unequal income distributions than the poorer ones. This result also holds for the Atkinson coefficients, whose results exhibit a very similar pattern of variation than those presented for the Gini coefficient.

Overall, the methodology presented here represents a starting point for income inequality analysis at the local level. A wide range of implementations arise from these results and should be addressed in future research. The illustration presented here could be extended to the whole set of municipalities, so as to get a picture of income inequality within municipalities in Spain. Besides, recent availability of PIT annual samples for several years would allow us to perform both cross-section and longitudinal income inequality analysis.

References

Arcaróns, J., Castells, A., García, G. and Parellada, M. (1992). Estimació de la renda familiar disponible a les comarques i municipis de Catalunya.1989. Generalitat de Catalunya, Departement d'Economia i Finances, Barcelona.

Alañón, A. (2002). "Estimación del valor añadido per cápita de los municipios españoles en 1991 mediante técnicas de econometría espacial", *Ekonomiaz*, 51: 172-194.

Atkinson, A. B. (1970). “On the measurement of inequality”, *Journal of Economic Theory*, 2: 244-263.

Creedy, J. (2003). Survey reweighting for tax microsimulation modelling. New Zealand Treasury Working Paper Series, 17/03.

Creedy, J. and Tuckwell, I. (2004). Reweighting household surveys for tax microsimulation modelling: An application to the New Zealand household economic survey. *Australian Journal of Labour Economics* 7 (1), 71–88.

De las Heras, A. (1992). “Un Modelo General de Estimación Indirecta de la Renta Familiar Disponible Municipal. Su Aplicación a la Comunidad Autónoma de Cantabria”. Tesis Doctoral. Universidad de Cantabria.

De las Heras, A. and C. Murillo (1998). “Información fiscal y estimación indirecta de la renta familiar disponible municipal en España”. I Congreso de Economía Aplicada. Barcelona.

Deville, J. and Särndal, C. (1992) Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87: 376–382.

Esteban, J. and A. Pedreño (1992). “La Articulación Territorial de la Economía Valenciana”. *Estructura Económica de la Comunidad Valenciana*. Madrid: Espasa Calpe, pp. 73-112.

Fernández, C. and Sierra, Y. (1992). “Estimación de la Renta Familiar Disponible a Nivel Municipal. Una Aplicación a La Rioja. Año 1985”. Actas de la VI Reunión Asepelt España. Granada.

Fernández-Jardón, C. M. and Martínez-Cobas, F. X. (2002). “Un método de estimación de la renta en unidades espaciales pequeñas”, *Revista Asturiana de Economía*, 23: 91-112

Gini, C. (1912) “Variabilità e mutabilità, contributo allo studio delle distribuzioni e relazioni statistiche”, *Studi Economico-Giuridici dell’ Universiti di Cagliari*, 3, (2): 1-158.

Herrero, L. C. (1998). “Perspectivas de Desarrollo Territorial: Renta Municipal y Desarrollo Económico en las Comarcas de Castilla y León”. Junta de Castilla y León, Consejería de Economía y Hacienda.

Haughton, J. H. and Khandker, S. R. (2009). *Handbook on poverty and inequality*. Washington DC: The World Bank.

Lambert, P. J. (2001). *The distribution and redistribution of Income*, 3rd edition. Manchester: Manchester University Press.

Mills J. A and Zandvakili, A. (1997). "Statistical inference via bootstrapping for measures of inequality", *Journal of Applied Econometrics*, 12: 133-150.

Oliver, J., Busóm, I. and Trullén, J. (1995). "Estimació de la renda familiar disponible per càpita de Barcelona, els seus districtes i els 27 municipis de la Corporació Metropolitana de Barcelona", Ajuntament de Barcelona, Barcelona.

Picos, F. (2006). "Microsimulación mediante fusión de Phogue y Panel de Declarantes para evaluar reformas fiscales". *Revista de Economía Aplicada*, 41: 33-60.

Picos, F., Pérez, C. and González, M. C. (2011). "La muestra de declarantes de IRPF de 2007: descripción general y principales magnitudes", *Documentos de Trabajo del Instituto de Estudios Fiscales*, 01/11.

Särndal, C. (2007): The calibration approach in survey theory and practice. *Survey Methodology* 33 (2), 99–119.

Tables and Figures

Table 1. Different distance functions

	$D(w,z)$
1. Chi-squared	$\frac{(z-w)^2}{2w}$
2. Minimum Entropy	$-sin\left(\frac{z}{w}\right)+z-w$
3. Modified Minimum Entropy	$z\ ln\left(\frac{z}{w}\right)-z-w$
4. Deville and Sändal (1992)	$\left(u-\frac{z}{w}\right)\ ln\left(\frac{u-\frac{z}{w}}{u-l}\right)+\left(\frac{z}{w}-l\right)\ ln\left(\frac{\frac{z}{w}-l}{1-l}\right)+\frac{u-l}{\alpha}w$

Note: u and l are known constants in the interval $l<1<u$; $\alpha=\frac{r_U-r_L}{(1-r_L)(r_U-1)}$.

Table 2. Final micro-data sample sizes and their distribution by Province

Province	**Province Code**	**Number of sample observations (used in estimations)**
Álava	1	-
Albacete	2	19,784
Alicante	3	44,072
Almería	4	24,353
Ávila	5	12,534
Badajoz	6	28,710
Balears (Illes)	7	32,885
Barcelona	8	86,880
Burgos	9	18,131
Cáceres	10	22,842
Cádiz	11	34,890
Castellón	12	25,682
Ciudad Real	13	21,542
Córdoba	14	33,076
Coruña (A)	15	37,749
Cuenca	16	14,172
Girona	17	24,974
Granada	18	33,254
Guadalajara	19	12,594
Guipúzcoa	20	-
Huelva	21	21,255
Huesca	22	14,167
Jaén	23	30,891
León	24	23,201
Lleida	25	20,342
Rioja (La)	26	16,820
Lugo	27	21,261
Madrid	28	110,208
Málaga	29	40,883
Murcia	30	38,140
Navarra	31	-
Ourense	32	19,439
Asturias	33	36,084
Palencia	34	12,065
Palmas (Las)	35	31,743

Table 2 (continued)

Province	Province Code	Number of sample observations (used in estimations)
Pontevedra	36	33,238
Salamanca	37	18,651
Santa Cruz de Tenerife	38	30,891
Cantabria	39	23,579
Segovia	40	11,297
Sevilla	41	44,700
Soria	42	8,624
Tarragona	43	27,661
Teruel	44	11,822
Toledo	45	24,773
Valencia	46	53,361
Valladolid	47	22,904
Vizcaya	48	-
Zamora	49	14,452
Zaragoza	50	36,454
Ceuta	51	5,244
Melilla	52	5,068
Non residents	99	615
Total of observations		**1,337,957**

Source: own elaboration using data drawn from the Spanish Personal Income Tax 2007 annual sample.

Table 3. Spanish municipalities according to population size, 2007.

Population threshold	Number of municipalities	% of Total Population
< 1,000 inhab.	4,877	3.37%
1,000 – 5,000 inhab.	1,968	10.06%
5,000 – 20,000 inhab.	895	19.37%
20,000 – 50,000 inhab.	235	15.50%
50,000 – 100,000 inhab.	77	12.05%
> 100,000 inhab.	59	39.66%

Source: Own elaboration using population counts from the Spanish National Statistics Institute.

Table 4. Percentage of municipalities for which a new non-negative vector of weights was obtained

Distance function	Percentage
Chi-squared	82.2%
Minimum Entropy	91.6%
Modified Minimum Entropy	94.8%
Deville and Sändal (1992)	73.3%

Source: Own elaboration

Table 5. Chosen distance function for each municipality

Distance function	Number of municipalities	%
Chi-squared	1,607	31.65%
Minimum Entropy	2,496	49.15%
Modified Minimum Entropy	473	9.31%
Deville and Särndal (1972)	502	9.89%
Total:	*5,078*	*100*

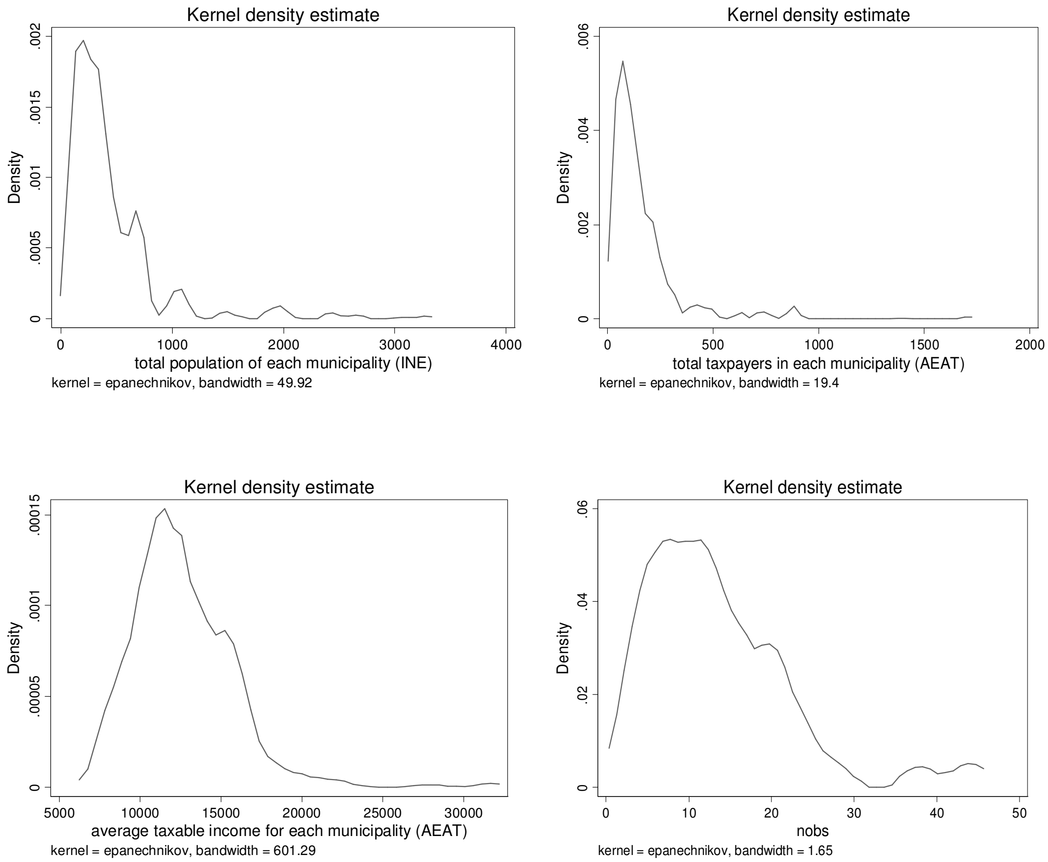
Source: Own elaboration

Table 6. Distribution of the ratio of new to original sample weights.

Percentiles	Ratio z/w
1%	0.06013
5%	0.31796
10%	0.62805
25%	0.91277
50%	0.99691
75%	1.04968
90%	1.14317
95%	1.24089
99%	1.80791
Mean	0.97445
St. Dev.	0.98183

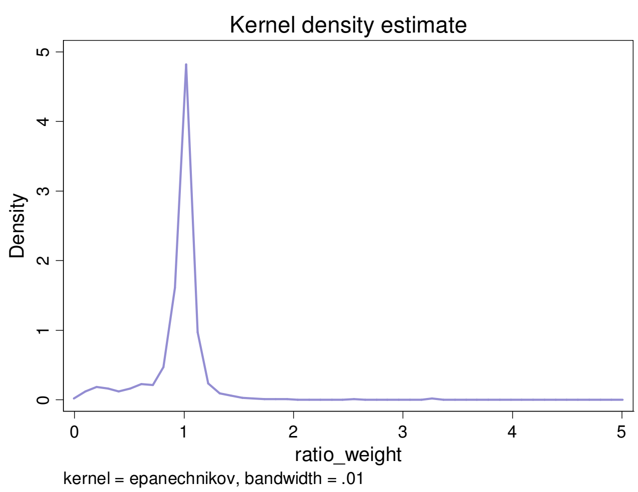
Source: Own elaboration

Figure 1. Kernel density of municipalities without a new vector of weights.



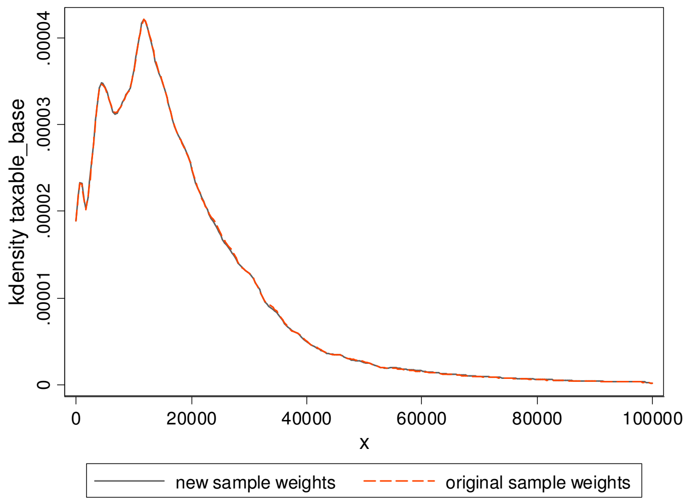
Source: Own elaboration

Figure 2. Ratio of new sample weights to original sample weights



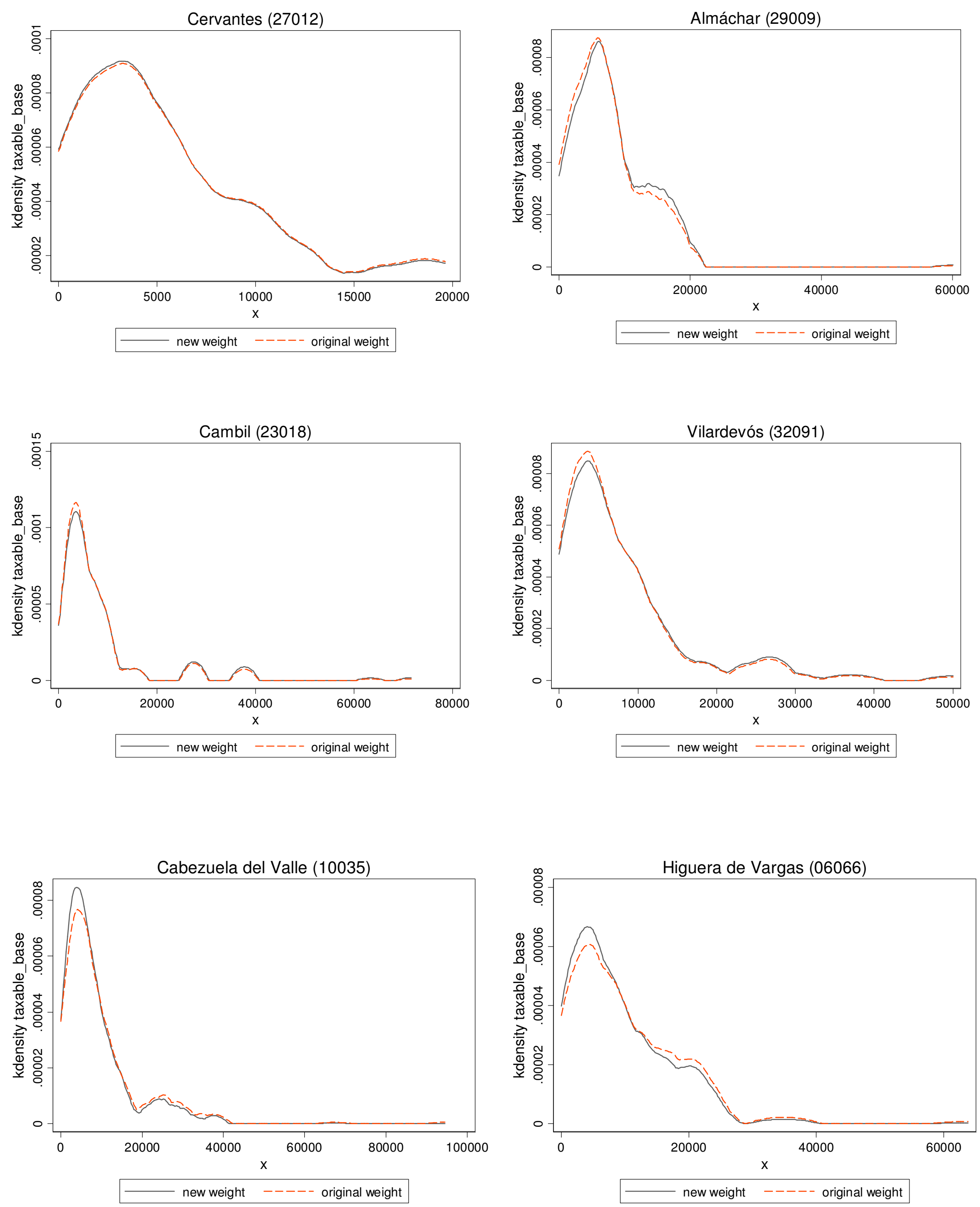
Source: Own elaboration

Figure 3. Overall income distribution



Own elaboration

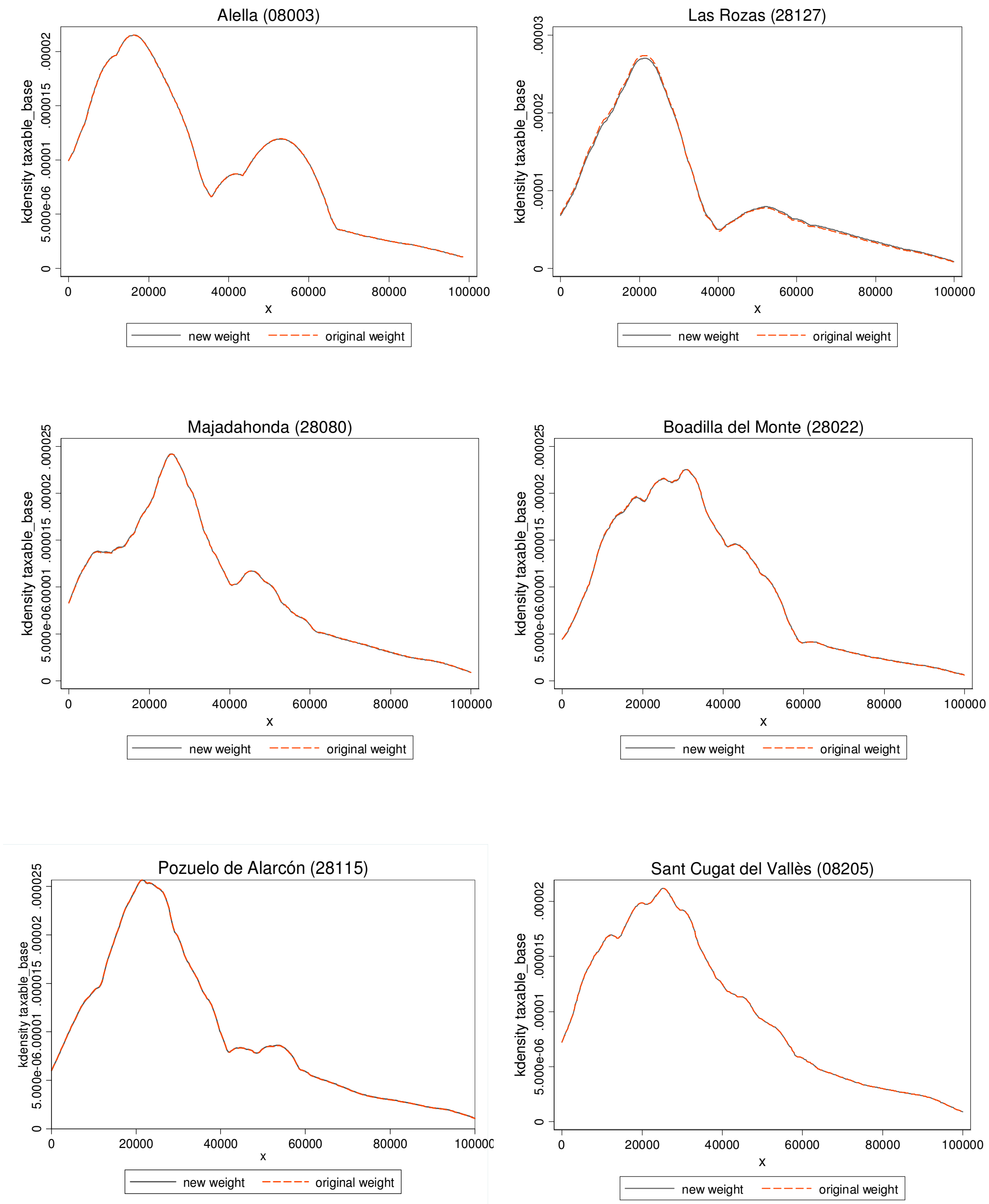
Figure 4. Local income distributions of poor Spanish municipalities*



Source: Own elaboration

* The poorest municipalities have been selected from the sample of municipalities with populations above 2,000 inhabitants.

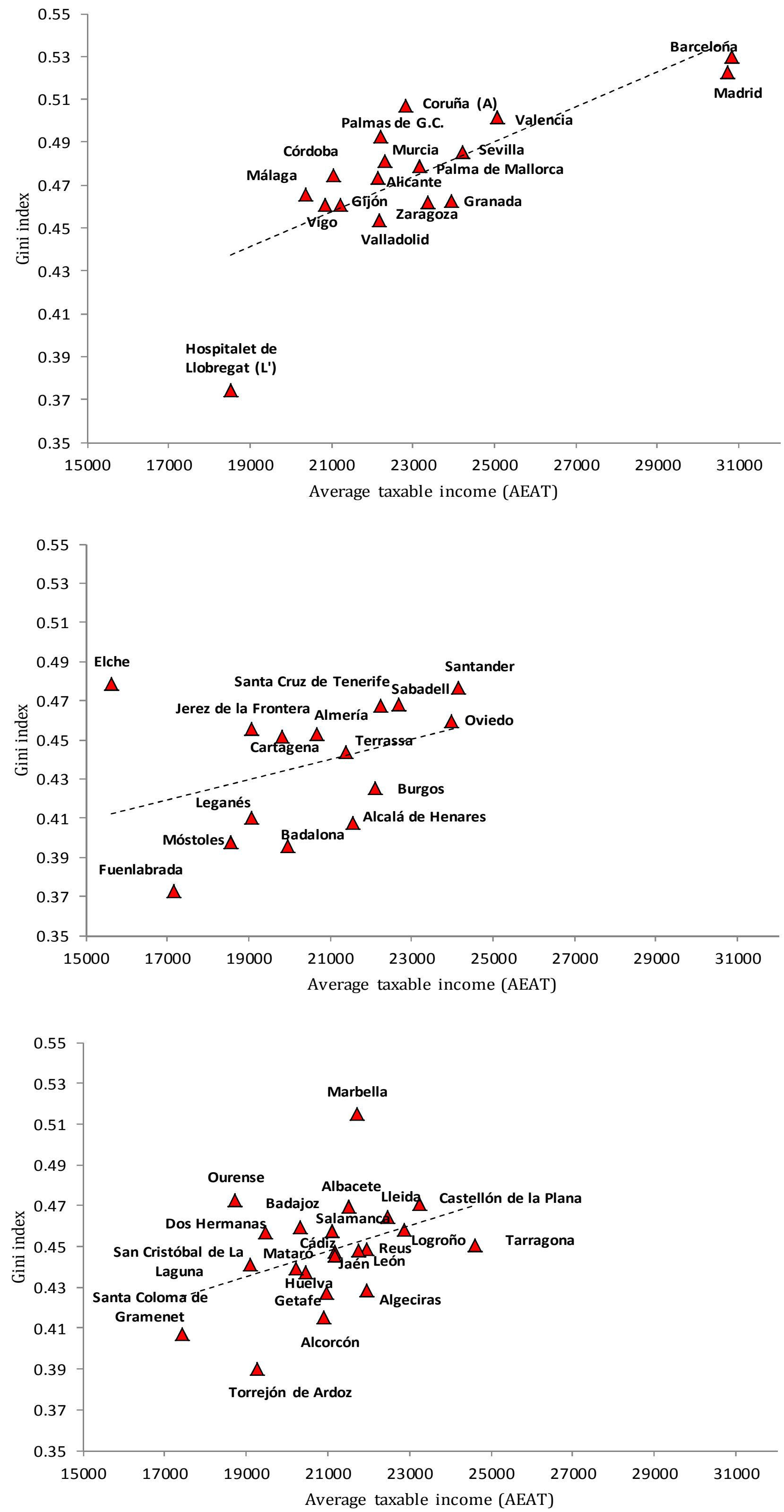
Figure 5. Local income distributions of the richest Spanish municipalities*



Source: Own elaboration

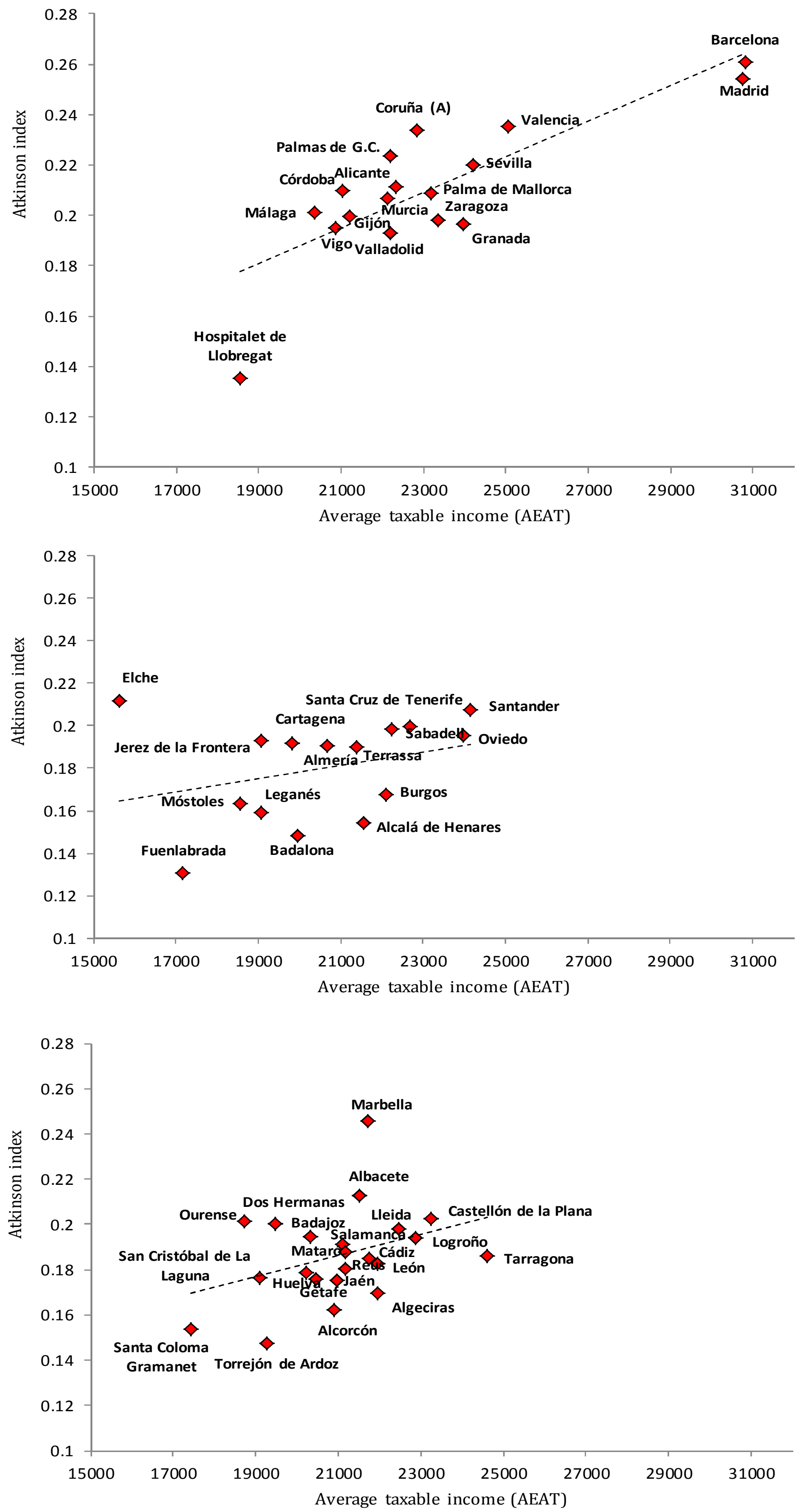
* For present purposes, the distributions are truncated at 100,000 euros.

Figure 6. Gini index of selected Spanish municipalities according to the new weights, 2007.



Source: Own elaboration.

Figure 7. Atkinson index of selected Spanish municipalities according to the new weights, 2007.



Source: Own elaboration.

Appendix 1. Spanish municipalities by province and population size, 2007.

	Population thresholds													
Province	*< 1.000*		*1.000 - 5.000*		*5.000 - 20.000*		*20.000 - 50.000*		*50.000 - 100.000*		*> 100.000*		*Total*	
	Number of municipalities	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*pop Province / total pop*
Álava	29	4,25%	19	11,36%	2	9,26%	0	0,00%	0	0,00%	1	75,13%	**51**	0,68%
Albacete	40	5,23%	39	21,86%	4	10,22%	3	20,67%	0	0,00%	1	42,02%	**87**	0,87%
Alicante	50	1,17%	35	4,88%	33	18,91%	16	25,49%	5	19,70%	2	29,86%	**141**	4,04%
Almería	50	3,44%	32	12,75%	14	21,05%	3	10,79%	2	23,10%	1	28,87%	**102**	1,43%
Asturias	16	0,86%	31	6,62%	24	23,96%	4	15,16%	1	7,75%	2	45,65%	**78**	2,38%
Ávila	226	31,16%	17	22,04%	4	14,90%	0	0,00%	1	31,90%	0	0,00%	**248**	0,37%
Badajoz	52	4,40%	84	26,72%	23	25,98%	3	13,40%	1	8,09%	1	21,41%	**164**	1,50%
Baleares	7	0,43%	23	6,09%	26	24,54%	10	31,77%	0	0,00%	1	37,17%	**67**	2,28%
Barcelona	98	0,73%	86	4,08%	83	15,55%	25	14,00%	12	14,96%	7	50,69%	**311**	11,80%
Burgos	345	18,78%	21	10,87%	2	3,55%	2	19,22%	0	0,00%	1	47,57%	**371**	0,81%
Cáceres	144	16,54%	63	29,34%	10	22,34%	1	9,72%	1	22,06%	0	0,00%	**219**	0,91%
Cádiz	3	0,17%	9	1,95%	17	13,71%	7	15,65%	5	31,64%	3	36,88%	**44**	2,67%
Cantabria	30	2,62%	54	20,71%	14	24,73%	2	10,53%	1	9,67%	1	31,74%	**102**	1,27%
Castellón	88	5,60%	28	10,43%	11	16,81%	7	37,05%	0	0,00%	1	30,11%	**135**	1,27%
Ceuta	0	0,00%	0	0,00%	0	0,00%	0	0,00%	1	100,00%	0	0,00%	**1**	0,17%
Ciudad Real	39	4,25%	39	15,78%	19	37,68%	3	18,41%	2	23,89%	0	0,00%	**102**	1,13%
Córdoba	11	0,90%	35	12,19%	21	23,40%	7	22,67%	0	0,00%	1	40,85%	**75**	1,75%
Coruña, A	0	0,00%	37	9,49%	46	34,02%	8	20,01%	2	14,91%	1	21,57%	**94**	2,51%
Cuenca	202	26,58%	29	25,90%	6	22,46%	0	0,00%	1	25,06%	0	0,00%	**238**	0,47%
Girona	131	7,44%	60	19,36%	22	29,41%	7	30,74%	1	13,05%	0	0,00%	**221**	1,56%
Granada	56	3,49%	76	18,54%	29	32,01%	5	12,62%	1	6,62%	1	26,72%	**168**	1,96%
Guadalajara	258	15,54%	24	23,06%	4	14,06%	1	12,56%	1	34,78%	0	0,00%	**288**	0,50%
Guipúzcoa	32	1,72%	25	8,60%	25	38,78%	4	15,85%	1	8,69%	1	26,35%	**88**	1,54%
Huelva	25	2,43%	34	18,44%	16	36,72%	3	13,04%	0	0,00%	1	29,37%	**79**	1,10%
Huesca	171	23,20%	24	19,04%	6	35,13%	1	22,63%	0	0,00%	0	0,00%	**202**	0,49%
Jaén	14	1,43%	53	20,52%	24	33,33%	4	17,99%	1	9,22%	1	17,51%	**97**	1,47%
León	138	13,28%	62	25,11%	8	15,21%	1	5,81%	1	13,44%	1	27,15%	**211**	1,10%
Lleida	163	15,19%	53	24,76%	14	29,30%	0	0,00%	0	0,00%	1	30,75%	**231**	0,92%
Lugo	4	0,75%	51	39,48%	11	33,34%	0	0,00%	1	26,42%	0	0,00%	**67**	0,79%
Madrid	54	0,38%	50	2,06%	45	6,86%	12	6,84%	9	10,57%	9	73,29%	**179**	13,45%

Appendix 1 (continued)

Province	*< 1.000*		*1.000 - 5.000*		*5.000 - 20.000*		*20.000 - 50.000*		*50.000 - 100.000*		*> 100.000*		*Total*	
	Number of municipalities	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*% of total provincial population*	*Number of municipalities*	*pop Province / total pop*
Málaga	25	0,86%	50	8,66%	10	6,71%	7	13,95%	6	24,50%	2	45,32%	**100**	3,36%
Melilla	0	0,00%	0	0,00%	0	0,00%	0	0,00%	1	100,00%	0	0,00%	**1**	0,15%
Murcia	2	0,12%	7	1,18%	21	18,74%	11	24,00%	2	10,70%	2	45,27%	**45**	3,08%
Navarra	187	9,08%	64	23,94%	18	25,80%	2	9,01%	0	0,00%	1	32,17%	**272**	1,34%
Ourense	10	2,12%	72	41,31%	9	24,76%	0	0,00%	0	0,00%	1	31,81%	**92**	0,75%
Palencia	170	20,46%	16	16,49%	4	15,56%	0	0,00%	1	47,49%	0	0,00%	**191**	0,38%
Palmas de G.C.	1	0,07%	3	0,73%	19	21,79%	7	20,73%	3	20,49%	1	36,20%	**34**	2,31%
Pontevedra	1	0,08%	23	8,42%	29	32,59%	7	19,33%	1	8,46%	1	31,11%	**62**	2,10%
Rioja, La	144	9,90%	21	14,92%	7	20,28%	1	7,69%	0	0,00%	1	47,21%	**174**	0,68%
Salamanca	334	25,76%	20	11,08%	7	18,78%	0	0,00%	0	0,00%	1	44,38%	**362**	0,78%
S.C. Tenerife	0	0,00%	17	4,99%	24	22,77%	9	27,76%	1	7,35%	2	37,13%	**53**	2,18%
Segovia	187	27,79%	18	22,16%	3	14,87%	0	0,00%	1	35,18%	0	0,00%	**209**	0,35%
Sevilla	5	0,18%	37	5,98%	49	27,02%	11	19,09%	1	3,57%	2	44,16%	**105**	4,09%
Soria	170	24,89%	10	22,67%	2	11,62%	1	40,82%	0	0,00%	0	0,00%	**183**	0,21%
Tarragona	98	5,58%	57	16,81%	20	24,11%	6	21,97%	0	0,00%	2	31,54%	**183**	1,68%
Teruel	214	32,38%	19	27,37%	2	16,48%	1	23,77%	0	0,00%	0	0,00%	**236**	0,32%
Toledo	89	6,96%	92	38,01%	21	29,36%	0	0,00%	2	25,67%	0	0,00%	**204**	1,42%
Valencia	84	1,65%	107	9,91%	48	20,63%	22	24,68%	4	11,06%	1	32,08%	**266**	5,50%
Valladolid	182	8,25%	33	13,27%	7	9,74%	2	8,06%	0	0,00%	1	60,68%	**225**	1,15%
Vizcaya	42	1,94%	37	7,77%	22	19,36%	8	24,38%	2	15,61%	1	30,94%	**112**	2,53%
Zamora	225	38,18%	20	13,85%	2	14,44%	0	0,00%	1	33,53%	0	0,00%	**248**	0,44%
Zaragoza	231	6,50%	52	12,46%	8	8,61%	1	2,26%	0	0,00%	1	70,18%	**293**	2,06%
Total:	**4877**	3,37%	**1968**	10,06%	**895**	19,37%	**235**	15,50%	**77**	12,05%	**59**	39,66%	**8111**	

Population thresholds

Own elaboration using population counts from the Spanish National Statistics Institute.

Appendix 2. Gini indexes for selected Spanish municipalities* (new weights used)

Municipality	Gini Index	α-percentile method at 95%		Normal method at 95%		Bootstrap (100 replications)	
		Lower limit	Upper limit	Lower limit	Upper limit	Mean of the estimator	Standard error
Madrid	0.52257	0.51587	0.52970	0.52162	0.52319	0.52241	0.00400
Barcelona	0.53002	0.51815	0.54292	0.52832	0.53073	0.52953	0.00615
Valencia	0.50176	0.48732	0.51690	0.49874	0.50158	0.50016	0.00722
Sevilla	0.48521	0.47665	0.49368	0.48377	0.48564	0.48470	0.00477
Zaragoza	0.46188	0.45531	0.46840	0.46112	0.46264	0.46188	0.00388
Málaga	0.46565	0.45852	0.47294	0.46455	0.46638	0.46547	0.00467
Murcia	0.48116	0.47327	0.48950	0.48010	0.48196	0.48103	0.00474
Palma de Mallorca	0.47876	0.47032	0.48634	0.47716	0.47904	0.47810	0.00480
Palmas G.C.	0.49246	0.47945	0.50908	0.49088	0.49424	0.49256	0.00857
Córdoba	0.47483	0.46449	0.48709	0.47337	0.47577	0.47457	0.00614
Alicante	0.47337	0.46262	0.48560	0.47246	0.47484	0.47365	0.00607
Valladolid	0.45337	0.44518	0.46234	0.45224	0.45412	0.45318	0.00481
Vigo	0.46074	0.45238	0.46750	0.45978	0.46140	0.46059	0.00414
Gijón	0.46088	0.45174	0.47114	0.45975	0.46176	0.46076	0.00512
Hospitalet Llobregat	0.37441	0.35401	0.39218	0.37091	0.37528	0.37309	0.01116
Coruña (A)	0.50698	0.49446	0.52088	0.50636	0.50941	0.50788	0.00777
Granada	0.46257	0.45335	0.47053	0.46172	0.46356	0.46264	0.00469
Elche	0.47830	0.46342	0.49019	0.47645	0.47956	0.47801	0.00792
Santa Cruz de Tenerife	0.46752	0.45935	0.47626	0.46676	0.46848	0.46762	0.00438
Oviedo	0.45946	0.44686	0.46998	0.45845	0.46093	0.45969	0.00634
Badalona	0.39552	0.37485	0.41643	0.39473	0.39916	0.39695	0.01130
Cartagena	0.45181	0.43910	0.46380	0.45000	0.45296	0.45148	0.00755
Móstoles	0.39790	0.34810	0.46170	0.38667	0.39975	0.39321	0.03339
Jerez de la Frontera	0.45574	0.44634	0.46740	0.45422	0.45642	0.45532	0.00562
Terrassa	0.44346	0.40923	0.47609	0.43578	0.44377	0.43977	0.02038
Sabadell	0.46838	0.44227	0.49242	0.46355	0.46905	0.46630	0.01402
Alcalá de Henares	0.40751	0.38397	0.43487	0.40802	0.41361	0.41081	0.01425
Fuenlabrada	0.37292	0.34778	0.39975	0.37048	0.37648	0.37348	0.01531
Almería	0.45299	0.44288	0.46273	0.45205	0.45394	0.45300	0.00482
Leganés	0.40992	0.37374	0.43635	0.40179	0.40862	0.40521	0.01741
Santander	0.47659	0.46458	0.48536	0.47498	0.47732	0.47615	0.00596
Burgos	0.42545	0.41916	0.43236	0.42547	0.42696	0.42621	0.00381
Castellón de la Plana	0.47092	0.46264	0.48400	0.47051	0.47250	0.47150	0.00509
Alcorcón	0.41525	0.38455	0.44873	0.41365	0.42109	0.41737	0.01897
Albacete	0.46965	0.44891	0.49192	0.46642	0.47145	0.46894	0.01282
Getafe	0.42737	0.39226	0.47757	0.41994	0.42826	0.42410	0.02122
Salamanca	0.45741	0.44467	0.46758	0.45601	0.45842	0.45722	0.00615
Huelva	0.43721	0.42823	0.44446	0.43542	0.43719	0.43630	0.00451
Logroño	0.45804	0.44844	0.46776	0.45700	0.45928	0.45814	0.00583
Badajoz	0.45968	0.44474	0.47278	0.45680	0.45960	0.45820	0.00715
San Cristóbal de La Laguna	0.44123	0.42915	0.45562	0.43962	0.44235	0.44099	0.00698

Appendix 2. (continued)

Municipality	Gini Index	α-percentile method at 95%		Normal method at 95%		Bootstrap (100 replications)	
		Lower limit	Upper limit	Lower limit	Upper limit	Mean of the estimator	Standard error
León	0.44852	0.44068	0.45604	0.44703	0.44884	0.44793	0.00464
Tarragona	0.45087	0.43762	0.46264	0.44877	0.45183	0.45030	0.00780
Cádiz	0.44729	0.43584	0.46181	0.44650	0.44954	0.44802	0.00774
Lleida	0.46426	0.45379	0.47694	0.46298	0.46560	0.46429	0.00669
Marbella	0.51489	0.49542	0.53378	0.51384	0.51782	0.51583	0.01014
Mataró	0.43900	0.40682	0.47274	0.43706	0.44389	0.44048	0.01741
Dos Hermanas	0.45715	0.43225	0.48305	0.45317	0.45907	0.45612	0.01504
Santa Coloma de Gramenet	0.40724	0.36949	0.43502	0.40030	0.40757	0.40393	0.01856
Jaén	0.44532	0.43571	0.45539	0.44412	0.44629	0.44520	0.00554
Algeciras	0.42852	0.41106	0.44498	0.42719	0.43047	0.42883	0.00836
Torrejón de Ardoz	0.39015	0.35569	0.42681	0.38661	0.39372	0.39017	0.01816
Ourense	0.47292	0.46388	0.48239	0.47239	0.47457	0.47348	0.00558
Alcobendas	0.63962	0.59375	0.68292	0.63307	0.64368	0.63838	0.02707
Reus	0.44789	0.42813	0.46723	0.44524	0.44932	0.44728	0.01040

Own elaboration

* Spanish municipalities with population size above 100,000 inhabitants.

Appendix 3. Atkinson indexes for selected Spanish municipalities* (new weights used)

Municipality	Atkinson Index (risk aversion 0.5)	α-percentile method at 95%		Normal method at 95%		Bootstrap (100 replications)	
		Lower limit	Upper limit	Lower limit	Upper limit	Mean of the estimator	Standard error
Madrid	0.25433	0.24676	0.26296	0.25332	0.25510	0.25421	0.00454
Barcelona	0.26068	0.24821	0.27294	0.25872	0.26129	0.26001	0.00658
Valencia	0.23535	0.21870	0.25485	0.23179	0.23549	0.23364	0.00944
Sevilla	0.22003	0.21044	0.23141	0.21855	0.22058	0.21957	0.00520
Zaragoza	0.19791	0.19106	0.20548	0.19721	0.19887	0.19804	0.00424
Málaga	0.20113	0.19356	0.20867	0.20008	0.20185	0.20097	0.00452
Murcia	0.21130	0.20264	0.21888	0.20998	0.21184	0.21091	0.00474
Palma de Mallorca	0.20897	0.20085	0.21754	0.20738	0.20921	0.20829	0.00469
Palmas de G.C.	0.22352	0.20792	0.24505	0.22166	0.22561	0.22363	0.01009
Córdoba	0.20989	0.19815	0.22441	0.20827	0.21102	0.20964	0.00701
Alicante	0.20666	0.19614	0.21989	0.20579	0.20815	0.20697	0.00603
Valladolid	0.19275	0.18450	0.20281	0.19165	0.19362	0.19263	0.00503
Vigo	0.19503	0.18812	0.20100	0.19436	0.19573	0.19505	0.00350
Gijón	0.19978	0.19219	0.20894	0.19887	0.20072	0.19980	0.00472
Hospitalet de Llobregat	0.13554	0.11949	0.14914	0.13368	0.13686	0.13527	0.00810
Coruña (A)	0.23380	0.22032	0.25179	0.23325	0.23671	0.23498	0.00885
Granada	0.19664	0.19044	0.20332	0.19609	0.19758	0.19683	0.00382
Elche	0.21167	0.19789	0.22326	0.21015	0.21299	0.21157	0.00726
Santa Cruz de Tenerife	0.19853	0.19223	0.20512	0.19793	0.19934	0.19864	0.00359
Oviedo	0.19538	0.18559	0.20562	0.19470	0.19686	0.19578	0.00549
Badalona	0.14829	0.13433	0.16373	0.14778	0.15134	0.14956	0.00907
Cartagena	0.19198	0.17959	0.20360	0.19022	0.19310	0.19166	0.00734
Móstoles	0.16316	0.11659	0.22770	0.15284	0.16608	0.15946	0.03378
Jerez de la Frontera	0.19317	0.18419	0.20390	0.19206	0.19408	0.19307	0.00517
Terrassa	0.18991	0.15886	0.22967	0.18353	0.19107	0.18730	0.01923
Sabadell	0.19984	0.17976	0.22228	0.19611	0.20105	0.19858	0.01260
Alcalá de Henares	0.15428	0.14122	0.17341	0.15507	0.15859	0.15683	0.00896
Fuenlabrada	0.13051	0.11700	0.14503	0.12992	0.13324	0.13158	0.00845
Almería	0.19054	0.18139	0.20137	0.18959	0.19148	0.19053	0.00481
Leganés	0.15910	0.13560	0.18868	0.15348	0.15907	0.15627	0.01426
Santander	0.20738	0.19686	0.21523	0.20622	0.20822	0.20722	0.00510
Burgos	0.16743	0.16192	0.17438	0.16732	0.16869	0.16800	0.00351
Castellón de la Plana	0.20238	0.19481	0.21275	0.20197	0.20383	0.20290	0.00474
Alcorcón	0.16207	0.13903	0.19147	0.16056	0.16665	0.16361	0.01555
Albacete	0.21249	0.18764	0.24197	0.20854	0.21464	0.21159	0.01556
Getafe	0.17527	0.14739	0.22481	0.16930	0.17733	0.17332	0.02048
Salamanca	0.19092	0.18216	0.19950	0.18993	0.19196	0.19095	0.00518
Huelva	0.17553	0.16770	0.18066	0.17392	0.17545	0.17469	0.00392
Logroño	0.19412	0.18432	0.20456	0.19312	0.19545	0.19429	0.00593
Badajoz	0.19456	0.18199	0.20826	0.19203	0.19467	0.19335	0.00673

Appendix 3. (continued)

Municipality	Atkinson Index (risk aversion 0.5)	α-percentile method at 95%		Normal method at 95%		Bootstrap (100 replications)	
		Lower limit	Upper limit	Lower limit	Upper limit	Mean of the estimator	Standard error
San Cristóbal de La Lagu	0.17644	0.16596	0.18881	0.17479	0.17733	0.17606	0.00648
León	0.18236	0.17576	0.18959	0.18126	0.18284	0.18205	0.00402
Tarragona	0.18595	0.17317	0.20232	0.18420	0.18738	0.18579	0.00810
Cádiz	0.18780	0.17836	0.20154	0.18746	0.18990	0.18868	0.00621
Lleida	0.19822	0.18769	0.21085	0.19709	0.19960	0.19834	0.00638
Marbella	0.24560	0.22362	0.26428	0.24445	0.24849	0.24647	0.01029
Mataró	0.17865	0.15747	0.20410	0.17882	0.18396	0.18139	0.01309
Dos Hermanas	0.20014	0.17631	0.22948	0.19627	0.20216	0.19922	0.01504
Santa Coloma Gramanet	0.15343	0.13244	0.17526	0.14963	0.15473	0.15218	0.01301
Jaén	0.18025	0.17312	0.18923	0.17926	0.18108	0.18017	0.00464
Algeciras	0.16941	0.15827	0.18310	0.16863	0.17136	0.16999	0.00697
Torrejón de Ardoz	0.14753	0.12319	0.17403	0.14526	0.15086	0.14806	0.01428
Ourense	0.20149	0.19238	0.21193	0.20102	0.20313	0.20207	0.00540
Alcobendas	0.37615	0.31233	0.43396	0.36849	0.38257	0.37553	0.03592
Reus	0.18472	0.17021	0.19876	0.18266	0.18591	0.18429	0.00831

Own elaboration

* Spanish municipalities with population size above 100,000 inhabitants.