

A municipal database from the 2011 Spanish census

Francisco J. Goerlich

University of Valencia, Valencia, Spain and

Instituto Valenciano de Investigaciones Económicas, Valencia, Spain

Received 9 April 2018
Revised 4 May 2019
Accepted 21 June 2019

Abstract

Purpose – The paper aims to describe the process to obtain a complete municipal database from the 2011 Spanish Census information. By complete, the authors mean variables for the full sample of the 8,116 municipalities as of the census reference date. In addition, the database should be consistent with the public census information released by the National Statistical Institute: microdata and customized tables.

Design/methodology/approach – The authors use mainly small area demographic and synthetic estimators that are reconciled using biproportional adjustment (iterative proportional fitting), when needed.

Findings – As a result, the authors obtain a complete and consistent municipal database composing 55 variables related to socio-demographic characteristics of persons.

Originality/value – The provision of a complete and consistent municipal database, available for download, which is absent in the original 2011 Spanish Census.

Keywords Municipalities, 2011 Spanish Census, Small area

Paper type Technical paper

1. Introduction

The 2011 Census marked a significant methodological turning point in the Spanish census tradition. It moved away from the classic census methodology, based on exhaustive fieldwork, toward a mixed system in which the population count and its most basic demographic characteristics are taken from administrative records –the *Padrón*, or Municipal Register– and the remaining population characteristics come from a large-scale survey of around 10 per cent of the population [Instituto Nacional de Estadística (INE), 2011].

© Francisco J. Goerlich. Published in *Applied Economic Analysis*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The author wishes to thank Jorge Luis Vega Valle, Carmen Teijeiro Breijo, Antonio Argüeso Jimenez and Ignacio Duque Rodriguez de Arellano from the Spanish Statistical Institute (INE) for their generous support in resolving innumerable methodological questions related to the census information, and is also grateful for feedback from members of the technical staff at the Instituto Valenciano de Investigaciones Económicas (Ivie), especially Irene Zaera and Carlos Albert, whose comments contributed to the iteration process in developing the disaggregation algorithms mentioned in the paper. The author is grateful for support from the FBBVA-Ivie research program, and from project ECO2015-70632-R. An extended version of this work (in Spanish) is available as a Working Paper, Goerlich (2016), at http://dx.medra.org/10.12842/MUNICIPIOS_CENSO_2011



Although this methodological change does not necessarily imply a loss in the quality of the resulting information (Goerlich *et al.*, 2015), a number of caveats must be mentioned, not only in light of the final information published by the INE through its various census breakdowns –Persons, Households, Dwellings and Buildings– but also in relation to the territorial areas referred to –National, Autonomous Communities, Provinces, Municipalities or Census Sections.

The 2011 Census provides scant information for smaller territorial areas, including municipalities, which are the basic administrative unit in the division of the territory, and for which censuses offer the only opportunity to gather homogenous and comparable data that goes beyond purely demographic information.

This study describes and applies a simple method to obtain estimations for the large majority of census variables, and for the full set of 8,116 municipalities included in the 2011 Census. These estimations are consistent with the published census information. The frame of reference for obtaining variables at the municipal level is the census microdata, which is the source of information for all the non-demographic population characteristics. The final aim is to create a complete and consistent municipal database for a wide set of variables.

The paper is structured as follows. In Section 2, the basic elements of the census methodology are described. This stage is necessary to understand the process followed to disaggregate the information at the municipal level, which is described in Section 3. The resulting database and how it can be accessed are presented in Section 4. The paper ends with some brief conclusions in Section 5.

2. Structure of the information in the 2011 Census

2.1 *Information on persons and households*

The information on persons and their characteristics for the 2011 Census is based on two fundamental sources: the Municipal Register, for purely demographic information; and a large survey, in principle designed to be representative at the municipal level, for all other population characteristics (Instituto Nacional de Estadística [INE], 2011). These two pillars provide different information and an understanding of how they interrelate is needed to understand the process followed to create the database.

First, one might naturally ask why the basic demographic characteristics of the population are taken from Municipal Register, even though they are not exactly the same as those from the Municipal Register for the census reference date, 1 November 2011. This is due to the very nature of the Municipal Register as a legally regulated administrative registry, which means that any alteration to it must have a legal basis; in other words, alterations cannot be made in the statistical adjustments. When the continuous Municipal Register was introduced in 1998, the population figures from the Municipal Register were disassociated from the census population figures, such that the population of the 2001 Census does not coincide with the population figures derived from the Municipal Register. It is well known that the way the Municipal Register is managed leads to an over estimation of the population, essentially associated with the register of foreign people, although problems have also been found in the upper and lower age distributions (Goerlich, 2007, 2012).

As a result, to find out the “population figure” of Spain and its territories, the Municipal Register – as the best statistical estimation of the resident population – had to be adjusted to give a more accurate reflection of the real situation. The INE therefore used the Municipal Register to build a pre-census file (PCF) that was adjusted as necessary to the increases and decreases in the natural population movement, and in which each registry entry had a count factor equal to 1 if the person could be proven to reside in Spain by crossing with other administrative records such as Social Security data, or was unknown if no conclusive proof

was available that the person was a resident in Spain. These registry entries were known as “doubtful”. Of the PCF registry entries, 97.2 per cent had a count factor equal to 1.

At the same time a large sampling survey was conducted with two objectives:

- (1) to determine the count factor in the PCF doubtful registry entries; and
- (2) to estimate the population characteristics.

The reference population for the sample is the population residing in main dwellings. The population living in institutional residences was therefore excluded from the sample and treated in a separate statistical operation: *Encuesta de Colectivos del Censo de Población y Viviendas 2011* (Instituto Nacional de Estadística [INE], 2013a).

2.2 The fit between the sample and the information from the pre-census file

The PCF and the sample are independent operations that must be reconciled. This reconciliation process, carried out by the INE, is based on two actions that are not wholly independent.

First, to determine the count factor of the doubtful entries both sets of information were partitioned into classes based on observable characteristics – age, nationality and place of residence – and a nominal crossing was made between the sample gathered from the fieldwork and the PCF, so the registry entries could be linked and those appearing in the PCF as doubtful could be identified if they were actually gathered in the sample.

From this identification, using the principle of analogy at the class level the count factors were estimated for the doubtful registry entries. The detailed procedure is described in Instituto Nacional de Estadística (INE) (2012) and Goerlich *et al.* (2015, chapter 1). What interests us here is that following this operation, each PCF entry has an assigned count factor. We therefore have a final weighted census file which determines the census population figure and its basic demographic characteristics. The resident population deriving from the census through this procedure was 46,815,916.

Second, the sample must be calibrated to the population to ensure consistency between the two in various dimensions referring to both population characteristics and territorial areas. However, the reference population of the survey is not derived from the final weighted census file, but the population in main family dwellings and excludes the population living in institutional residences. This population cannot be identified from the PCF.

The population living in institutional residences was estimated by the *Encuesta de Colectivos* as 444,101. However, not all the population living in institutional accommodation is officially registered as living there. According to this survey, only 241,187 people living in institutional establishments were officially registered as living there, whereas the remaining 202,914 were registered as living in main family dwellings, and are counted in the family dwellings for the effects of the sample, which is where they are officially registered. As a result, the population residing in main family dwellings is: $46,815,916 - 241,187 = 46,574,729$ persons. That is, the elevation factors of the survey must include this population. The calibration process uses the standard INE method: CALMAR (Deville and Sæmndal, 1992; Deville *et al.*, 1993), is carried out at the municipal level, and is a function of the municipality size [Instituto Nacional de Estadística (INE), 2014].

Having two reference population groups – the resident population and the population living in main dwellings – significantly complicates the process of disaggregating the microdata to create the municipal database, since the disaggregated variables must be adjusted to population marginals that cannot be taken directly from the PCF. The PCF provides information for the total resident population, whereas the municipal database,

constructed from the microdata, must be adjusted to the population living in main dwellings.

For this reason, we first had to estimate the population living in main dwellings at the municipal level by sex and in two age groups: under the age of 16, and aged 16 and over. The methods used for this purpose are described in Section 3.

2.3 Territorial structure in the microdata from the 2011 census

The microdata from the census only provide information at the municipal level for municipalities with more than 20,000 inhabitants. The remaining municipalities are grouped into four strata by size for each province as follows:

- (1) up to 2,000 inhabitants (Code 991);
- (2) between 2,001 and 5,000 inhabitants (Code 992);
- (3) between 5,001 and 10,000 inhabitants (Code 993); and
- (4) between 10,001 and 20,000 inhabitants (Code 994).

The distribution of the municipalities by province and strata are reported in [Table I](#).

The 394 municipalities with more than 20,000 inhabitants can be perfectly identified in the microdata. In addition, the eight cases in which there is only one municipality per stratum can also be identified, together with the smallest municipality in Spain in demographic terms, Illán de Vacas in the province of Toledo, which has just one inhabitant. We can therefore directly identify 403 municipalities in the microdata; for the remaining 7,713 municipalities, we can only know the aggregated values of the stratum to which they belong. The database in this study obtains information on certain variables for these municipalities.

2.4 The customized tables system in the published census information

In addition to the microdata file, the published findings from the 2011 Census include a *Customized Table* query system in which users can select the variables they are interested in from within a geographical area and domain.

The *Customized Tables* system is constructed from the sample and the reference population is therefore those living in main dwellings and as such is consistent with the microdata. However, for various reasons the system is fairly limited for obtaining complete generalized information for all the municipalities. On one hand, it is subject to a series of confidentiality norms that restricts the information provided, and which in no case covers all the municipalities. On the other hand, to ensure statistical secrecy all data is rounded to the closest multiple of five.

The information in the *Customized Tables* is, however, of unquestionable value since, following some experimentation, their incorporation was shown to notably improve the municipal estimations using the procedure described below. The information available in the *Customized Tables* was therefore incorporated as the starting point for the disaggregation process.

3. Methodology: from microdata to municipalities

The previous section describes the census information structure with regard to small areas – municipalities. The next question is how to combine all this information so we obtain estimations for all municipalities for a large set of variables. Whatever method is followed it must comply with a basic condition: the estimations must be consistent with the microdata. The reference population is therefore the population living in main dwellings.

Province	Up to 2,000 inhab.	2,001 to 5,000 inhab.	5,001 to 10,000 inhab.	10,001 to 20,000 inhab.	Over 20,000 inhab.	Total
01 Alava	42	6		2	1	51
02 Albacete	62	17	2	2	4	87
03 Alacant/Alicante	66	18	20	13	24	141
04 Almeria	62	19	9	6	6	102
05 Avila	233	10	4		1	248
06 Badajoz	97	41	17	4	5	164
07 Illes Balears	14	13	17	11	12	67
08 Barcelona	121	58	51	37	44	311
09 Burgos	360	6	2		3	371
10 Cáceres	188	21	7	3	2	221
11 Cádiz	6	6	10	7	15	44
12 Castellón/Castelló	104	11	9	3	8	135
13 Ciudad Real	62	16	11	8	5	102
14 Córdoba	23	24	14	6	8	75
15 A Coruña	12	29	31	11	11	94
16 Cuenca	222	9	5	1	1	238
17 Girona	159	29	14	11	8	221
18 Granada	95	34	18	14	7	168
19 Guadalajara	267	13	4	2	2	288
20 Guipúzcoa	45	10	13	14	6	88
21 Huelva	35	24	7	7	6	79
22 Huesca	189	6	1	5	1	202
23 Jaen	33	36	13	9	6	97
24 León	178	21	5	4	3	211
25 Lleida	193	23	10	4	1	231
26 La Rioja	153	12	5	2	2	174
27 Lugo	24	30	8	4	1	67
28 Madrid	69	31	31	15	33	179
29 Málaga	44	29	9	3	16	101
30 Murcia	5	4	6	13	17	45
31 Navarra	213	37	12	7	3	272
32 Ourense	61	21	4	5	1	92
33 Asturias	36	11	10	14	7	78
34 Palencia	180	6	4		1	191
35 Palmas de Gran Canaria (Las)	2	2	8	9	13	34
36 Pontevedra	4	21	12	16	9	62
37 Salamanca	349	3	6	3	1	362
38 Santa Cruz de Tenerife	6	16	12	8	12	54
39 Cantabria	55	27	9	6	5	102
40 Segovia	198	7	3		1	209
41 Sevilla	14	25	30	19	17	105
42 Soria	175	5	2		1	183
43 Tarragona	122	32	14	6	10	184
44 Teruel	225	8	1	1	1	236
45 Toledo	112	63	15	11	3	204
46 Valencia/València	132	55	28	20	31	266
47 Valladolid	201	13	7	1	3	225
48 Vizcaya	60	19	13	9	11	112
49 Zamora	244	1	1	1	1	248

Table I.
Geography by
municipality size in
2011 census
microdata

(continued)

Province	Up to 2,000 inhab.	2,001 to 5,000 inhab.	5,001 to 10,000 inhab.	10,001 to 20,000 inhab.	Over 20,000 inhab.	Total	Database from the 2011 Spanish census
50 Zaragoza	256	22	9	4	2	293	231
51 Ceuta					1	1	
52 Melilla					1	1	
<i>Spain</i>	<i>5,808</i>	<i>1,000</i>	<i>553</i>	<i>361</i>	<i>394</i>	<i>8,116</i>	

Source: Instituto Nacional de Estadística (INE) (2013a)

Table I.

Consistency with the microdata implies that: (i) for each municipality, values disaggregated by categories of one variable must coincide with the value for the same variable at the municipal level, and (ii) for each stratum of the microdata, the sum of the values disaggregated at the municipal level must coincide with the values for that stratum. The information (i) must be found externally, and the information (ii) comes from the microdata. In addition, the estimations for the 403 municipalities that can be identified in the microdata are taken directly from that source and are used to validate the method.

3.1 Disaggregation of the population living in main dwellings

As noted above, the reference population for creating the municipal database is the population living in main dwellings. In some cases, the corresponding group is the total of the population living in main dwellings (PRVP), but in other cases the group is limited to the classification by sex or age groups –below the age of sixteen, and sixteen years and over– and occasionally it is necessary to cross these variables or previous estimations of the microdata classification variables. These are the groups that act as marginals to which the estimations must be adjusted.

For this reason, the first stage was to disaggregate the PRVP according to the above-mentioned criteria. The procedure followed was very simple. For the 5,608 municipalities that do not have a population registered as living in institutional accommodation this information is available in the PCF and is taken from there. These municipalities are not estimated and form part of the validation set. For the rest we distinguish between two cases:

- (1) municipalities with more than 20,000 inhabitants; and
- (2) municipalities with up to 20,000 inhabitants.

The first group is also identified in the microdata and information for this group was taken directly from there. For the second group, following an initial estimation, an iterative proportional fitting (*ipf* – Deming and Stephan, 1940; Stephan, 1942) procedure was applied at the stratum level, more commonly known in economics as the RAS method (Bacharach, 1965)[1].

3.2 Disaggregation of variables of persons in the microdata

We start from the following general frame. Let us consider a categorical variable, X , for a municipality m , which takes J possible values. For example, the variable “Relation to economic activity”, $RELA$, takes 6 possible values, and is not applied when the person is below the age of 16 years. Therefore, when the population is restricted to the population aged 16 and over, in this example $J = 6$.

Given that each person in the municipality estimated must belong to one of the possible J categories, the population of that municipality, N^m , can be written as $N^m = \sum_{j=1}^J X_j^m$, where the superscript m indicates the corresponding municipality. The values of X_j^m are unknown for each j and m , and are the variables we are trying to estimate. We know the population of the municipality, N^m , from the final weighted census file, and also X_j for the stratum to which the municipality belongs, $X_j = \sum_{m \in S} X_j^m$ where S represents the stratum, taken from the microdata. In other words, seen in table format we know the marginal distributions, but not the whole distribution.

A mechanical application using an iterative proportional fitting process based on an initial uniform distribution yields very poor results, indicating that the key is to incorporate auxiliary information into the estimation of this joint distribution, in other words, to look for a reasonable initial estimation for each municipality that serves as an initial value in the iterative fitting process.

For the municipalities for which information is available in the *Customized Tables* system, this initial value can be taken from that source. Because this information is not available for the remaining municipalities we must find a reasonable alternative estimation. Let us suppose that we have another partition of the municipality's population into K exhaustive and mutually exclusive classes. We can also now write the population of the municipality as $N^m = \sum_{k=1}^K N_k^m$, where N_k^m is now known from the information in the final weighted census file.

Let us now consider the problem of estimating X_j^m . By definition:

$$X_j^m = \sum_{k=1}^K X_{k,j}^m = \sum_{k=1}^K N_k^m \frac{X_{k,j}^m}{N_k^m} \tag{1}$$

The estimator proposed for these municipalities estimates the rates that appear in (1), $\frac{X_{k,j}^m}{N_k^m}$, from the stratum to which the municipality belongs, S , with the information available in the microdata, and applies these rates to the partition of the population considered at the municipal level. That is:

$$\hat{X}_j^m = \sum_{k=1}^K N_k^m \frac{X_{k,j}^S}{N_k^S} \tag{2}$$

where $N_k^S = \sum_{m \in S} N_k^m$ and $X_{k,j}^S = \sum_{m \in S} X_{k,j}^m$. Consequently, (2) substitutes the real rates in (1), $\frac{X_{k,j}^m}{N_k^m}$, $\forall k$, with estimated rates at the level of the stratum to which the municipality belongs, $\frac{X_{k,j}^S}{N_k^S}$, $\forall k$, and applies these rates to all the municipalities in that stratum.

The method for obtaining \hat{X}_j^m from (2) is simple and falls within the so-called traditional demographic methods in the context of small area estimations (Rao, 2003, chapter 3), or synthetic estimators (Rao, 2003, chapter 4.2) and can be implemented in a generalized and automatic way for several different census microdata variables when the *Customized Tables* system provides no information for the municipality in question.

An estimator is known as synthetic if a reliable direct estimator for a large area covering several small areas is used to obtain an indirect estimator for these small areas, under the assumption that the small areas have the same characteristics as the large area. Clearly (2) falls within this definition, where the implicit assumption is that all the municipalities in

stratum S present the same rates, $\frac{X_{kj}^S}{N_k^S}$, $\forall k$, and the municipalities of this stratum are only differentiated by their demographic structure. This method is also known as the propensity method (Bell *et al.*, 1995), and is applied by the Instituto Nacional de Estadística (INE) (2013b) in a range of contexts.

An alternative way of looking at (2) is:

$$\hat{X}_j^m = \sum_{k=1}^K \frac{N_k^m}{N_k^S} X_{kj}^S \quad (3)$$

which highlights the way the value of X_j at the stratum level for each element in the partition, X_{kj}^S , is rescaled by the proportion that the population of the municipality represents in the stratum, $\frac{N_k^m}{N_k^S}$.

Colom *et al.* (2015) provide an explanation of the method within the framework of traditional sampling superpopulation models when it is not possible to identify the registers of the specific units within a broader domain. This is the case of the microdata structure in the 2011 Census. These authors show how in this context, (2) is an unbiased although inefficient estimator. Nonetheless, the estimated standard errors are very small and of a similar magnitude to that provided by the INE in many of its sample surveys. In addition, this procedure yields practically identical results to those obtained by modeling the variable to be disaggregated using discrete choice models.

Once we have \hat{X}_j^m for the J categories of the variable, and for all the municipalities in the stratum, either from the procedure described above or from the information provided by the *Customized Tables* system, these initial estimations are adjusted to the total known marginals, N^m and X_j , by means of an iterative bi-proportional fitting process (Deming and Stephan, 1940; Stephan, 1942). The estimation is therefore carried out at the stratum level and yields a final estimator \tilde{X}_j^m .

We use as a partition the municipal population by sex and simple ages up to 100 years and above since this partition is available from the final weighted census file, which generates a total of 202 cells, 101 for each sex, and therefore $K = 202$ in (2).

The application of (2) rests on the assumption that the municipality for which we perform the estimation has the same characteristics as the stratum to which it belongs, and that the differences between the municipalities in this stratum reside in their demographic structure. This implies that the closer the variable in question is related to the demography, and the more homogenous the municipalities within the stratum, the lower the estimation errors will be.

Because the method we describe above can be applied to municipalities that are clearly identified in the microdata, these data constitute the validation set against which to measure the aggregate estimation error. It should be noted, however, that these are mostly municipalities with more than 20,000 inhabitants, which undoubtedly means it is a biased validation set.

For these municipalities, X_j^m is known, so we can calculate the absolute error (AE): $|\tilde{X}_j^m - X_j^m|$. From this discrepancy we calculate standard error means, the mean of the absolute relative errors (*MARE*), as a percentage:

$$MARE = \frac{100}{M \times J} \times \sum_{m=1}^M \sum_{j=1}^J \frac{|\tilde{X}_j^m - X_j^m|}{X_j^m} \quad (4)$$

and an overall error mean, as the total absolute relative error (*TARE*), as a percentage:

$$TARE = 100 \times \frac{\sum_{m=1}^M \sum_{j=1}^J |\tilde{X}_j^m - X_j^m|}{2 \times N} \quad (5)$$

ranging between 0 and 1, since the sum of the *AE*, $\sum_{m=1}^M \sum_{j=1}^J |\tilde{X}_j^m - X_j^m|$, ranges between 0 when no error is made, $\tilde{X}_j^m = X_j^m, \forall m, j$, and twice the reference population, $N = \sum_{m=1}^M \sum_{j=1}^J X_j^m$, when the error is the maximum possible in each case, and can be interpreted as the percentage of the population erroneously distributed in the set[2]. An analysis of errors showed negligible errors for the validation municipalities in all cases.

4. Database: content and access

The procedure described above allowed us to disaggregate the 55 variables reported in Table II, together with the variables related to the population living in main dwellings according to certain classification criteria, and that are not generally available at the municipal level from the 2011 Census.

The advantages of this database derive from the availability of data for all municipalities without exception, unlike the information available from the census, yet at the same time it is wholly consistent with the published census information. It can therefore be used in research whose territorial scope is the municipality or certain arbitrary aggregations of municipalities such as, for example, districts or rural areas (Reig *et al.*, 2016), and morphological (Goerlich and Cantarino, 2013) or functional urban areas (Goerlich *et al.*, 2019).

The database is available in an Access file at this link (<https://nuvol.uv.es/owncloud/index.php/s/aWLV2KzUbodR5bQ>). It should be used in conjunction with the design of the census microdata register, and it is structured as follows. For each variable included in Table II, a table is provided in which the rows represent the municipalities, identified by a code, and include as many columns as there are values for the corresponding variable. The columns are named according to the following criterion: given the variable in question, the name of which appears in the last column of Table II, and the values it takes, each column is identified with the name of the variable to which its code is added. The final column indicates the marginal to which the variable in question is added.

For example, the variable “*Relation to economic activity*”, *RELA*, takes 6 possible values: 1 – *Employed*, 2 – *Unemployed with previous work experience*, 3 – *Unemployed in search of first job*, 4 – *Person with permanent work disability*, 5 – *Retired, early retiree, pensioner or rentier* and 6 – *Other situation*; and is defined for the population living in main dwellings aged 16 years or over, *PRVP16M*. Thus, the first column in the table “*20_RELA*” in the Access file has the code for the municipality, *codmun*, followed by 6 columns, *RELA#*, # = 1 to 6, and a final column, *PRVP16M*, such that *RELA1* gives the number of people in employment in each municipality, and *RELA5* the retired, early retirees, pensioners or rentiers.

A final table contains only the codes and names of the municipalities as they appear in the census.

5. Conclusions

This study describes the process followed to create a municipal database for a large set of variables based on the 2011 Census. This information is not available in a general form for all municipalities. The methods for creating the database are simple, although time-consuming, but have the advantage that they are compatible with the published census information, and allow the incorporation of external information derived from the INE’s *Customized Tables* system, which is essential to improve the accuracy of the estimations.

Variables acting as marginals in the disaggregation process

1	Population living in main dwellings by age group Under the age of 16 16 years old and above	PRPVM16 PRVP16M
2	Population living in main dwellings by sex Male Female	PRVPVAR PRVPMUJ
3	Population living in main dwellings by sex and age Males under the age of 16 Males aged 16 years old and above Females under the age of 16 Females aged 16 years old and above	PRVPVARM16 PRVPVAR16M PRVPMUJM16 PRVPMUJ16M

Microdata classification variables

4	Current municipality of residence and Previous municipality of residence	RES_ANTERIOR
5	Current municipality of residence and Municipality of residence 1 year ago	RES_UNANO
6	Current municipality of residence and Municipality of residence 10 years ago	RES_DANO
7	Spending more than 14 nights in second municipality	SEG_VIV
8	Having a dwelling in second municipality	SEG_DISP
9	Marital status	ECIVIL
10	Attending school	ESCOLAR
11	Level of completed studies (qualifications)	GRADOS
12	Level of completed studies (details)	ESREAL
13	Type of studies undertaken	TESTUD
14	Caring for a child under the age of 15	TAREA1
15	Caring for a person with health problems	TAREA2
16	Charitable work or social volunteering	TAREA3
17	Responsible for most of the domestic tasks in the home	TAREA4
18	Indicator of whether the woman has had children	HIJOS
19	Principal relation with economic activity (employed/unemployed)	ACTIVO
20	Principal relation with economic activity (detail)	RELA
21	Type of working day Occupation code	JORNADA OCUPACION
22	to 1 digit	OCUPACION
23	to 2 digits Economic activity code to 2 digits	CNO
24	Branch	RAMA
25	Letter	LETRA
26	to 2 digits	CNAE
27	Professional situation	SITU
28	Socioeconomic status	CSE
29	Students (ESCUR1): Yes/No Current studies: Type of Studies	ESTUDIANTE
30	01 – Compulsory secondary education (ESO), Adult secondary education	ESCUR01
31	02 – Initial Professional Qualification Programs	ESCUR02
32	03 – High school (baccalaureate)	ESCUR03
33	04 – Middle Grade Vocational Training, Plastic Arts and Design, and Sports Education or equivalent	ESCUR04
34	05 – Official Language School Education	ESCUR05
35	06 – Professional Music and Dance Education	ESCUR06
36	07 – Higher Grade Vocational Training, Plastic Arts and Design, and Sports Education or equivalent	ESCUR07
37	08 – University diploma, Technical architecture, Technical engineering or equivalent	ESCUR08

(continued)

Table II.
Variables of persons
disaggregated by the
methods described in
the paper

38	09 – University first degree studies, Artistic studies or equivalent	ESCUR09
39	10 – Bachelor’s degree, Architecture, Engineering or equivalent	ESCUR10
40	11 – Official university Master’s degree, Specialities (medicine) or similar	ESCUR11
41	12 – Post graduate studies	ESCUR12
42	13 – Other official educational courses (Initial adult education programs, . . .)	ESCUR13
43	14 – Public Employment Service training courses	ESCUR14
44	15 – Other non-regulated training courses	ESCUR15
45	Students (Yes/No) according to relation to economic activity (3 categories): 6 categories	ESTURELA
46	Population in work or studying: Yes/No	TRABAEST
47	Place of work or study	LTRABA
48	Number of daily journeys	NVIAJE
	Means of travel	
49	01 – Car or van (driver)	MDESP01
50	02 – Car or van (passenger)	MDESP02
51	03 – Bus, coach, minibus	MDESP03
52	04 – Subway/underground	MDESP04
53	05 – Motorbike	MDESP05
54	06 – On foot	MDESP06
55	07 – Train	MDESP07
56	08 – Bicycle	MDESP08
57	09 – Other means	MDESP09
58	Journey time	TDESP

Table II. Source: Instituto Nacional de Estadística (INE) (2013a, 2013b) – 2011 Census

The procedures used must overcome numerous small inconsistencies between the two main pillars of the 2011 Census –the final weighted census file and the survey– which provide all the population characteristics beyond simple demographic data. Apart from these small inconsistencies the estimations generated are wholly consistent at the municipal level and at the level of the strata to which the municipalities in the microdata belong. Although all the disaggregated variables in the database are at the individual person level, identical methods can be used for household variables. Similar methods could also be used for the dwellings and buildings variables.

Finally, a few words of caution. The results must be interpreted for what they are – estimations based on a census sample– with the aim of providing statistics for all municipalities, and they should be used with that caution in mind. The information derived from the *Customized Tables* system has been exploited to the full, but in some cases it is limited or partial and in no case is it available in a general sense for all municipalities.

Notes

1. There are two exceptions to the above rules due to the lack of consistency between the PCF and the calibration of the microdata. In both cases, to maintain consistency with the final database we prioritized the use of the microdata. The details of the process followed in these cases are described in [Goerlich \(2016\)](#).
2. That is, assigned to a cell to which it does not correspond.

References

- Bell, M., Cooper, J. and Les, M. (1995), *Household and Family Forecasting Models. A Review*, Department of Housing and Regional Development, Canberra, p. 68.
- Colom, M.C., Goerlich, F.J., Molés, M.C. and Murgui, S. (2015), "Estimación de proporciones a partir de diseños no aleatorios: aplicación al censo de población de 2011", *trabajo presentado en XXIX Congreso Internacional de Economía Aplicada. Métodos Cuantitativos para la Economía y la Empresa. ASEPELT 2015, Cuenca*, 24-27 de junio de.
- Bacharach, M. (1965), "Estimating nonnegative matrices from marginal data", *International Economic Review*, Vol. 6 No. 3, pp. 294-310.
- Deming, W.E. and Stephan, F.F. (1940), "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", *The Annals of Mathematical Statistics*, Vol. 11 No. 4, pp. 427-444.
- Deville, J.-C. and Särndal, C.-E. (1992), "Calibration estimators in survey sampling", *Journal of the American Statistical Association*, Vol. 87 No. 418, pp. 376-382.
- Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993), "Generalized raking procedure in survey sampling", *Journal of the American Statistical Association*, Vol. 88 No. 423, pp. 1013-1020.
- Goerlich, F.J. (2007), "Cuántos somos? Una excursión por las estadísticas demográficas del instituto nacional de estadística (INE)", *Boletín de la Asociación de Geógrafos Españoles*, Vol. 45, pp. 123-156.
- Goerlich, F.J. (2012), "Estimaciones de la población actual (ePOBa) a nivel municipal. Discrepancias Censo- Padrón a pequeña escala", *Boletín de la Asociación de Geógrafos Españoles*, Vol. 58, pp. 83-104.
- Goerlich, F.J. (2016), "Es posible construir una base de datos municipal completa y consistente a partir del censo de 2011?", *Ivie 2016-03. Valencia, España. Documentación en línea*, available at: www.ivie.es/es/informes/2016-3-es-posible-construir-una-base-de-datos-municipal-completa-y-consistente-a-partir-del-censo-de-2011.php (accessed 1 April 2019).
- Goerlich, F.J. and Cantarino, I. (2013), "A population density grid for Spain", *International Journal of Geographical Information Science*, Vol. 27 No. 12, pp. 2247-2263, doi: [10.1080/13658816.2013.799283](https://doi.org/10.1080/13658816.2013.799283).
- Goerlich, F.J., Reig, E., Albert, C. and Robledo, J.C. (2019), *Las Áreas Urbanas Funcionales en España: Economía y Calidad de Vida*, Fundación BBVA, Bilbao.
- Goerlich, F.J., Ruiz, F., Chorén, P. and Albert, C. (2015), "Cambios en la estructura y localización de la población. Una visión de largo plazo (1842-2011)", Fundación BBVA. 2015, *Bilbao*. p. 354.
- Instituto Nacional de Estadística (INE) (2011), "Proyecto de los censos demográficos 2011: Subdirección general de estadísticas de la población", (Febrero), INE, Madrid.
- Instituto Nacional de Estadística (INE) (2012), "Metodología de cálculo de las cifras de población censal", available at: www.ine.es/censos2011/censos2011_meto_calculo.pdf (accessed 20 September 2013).
- Instituto Nacional de Estadística (INE) (2013a), "Población residente en establecimientos colectivos (encuesta de colectivos del censo de población y viviendas 2011", Metodología, available at: www.ine.es/censos2011/censos2011_meto_pobla_colectivos.pdf (accessed 20 May 2016).
- Instituto Nacional de Estadística (INE) (2013b), "La producción de información demográfica en el INE a partir del censo de 2011", Curso de la Escuela de Estadística de las Administraciones Públicas (EEAP), INE, Madrid, 14-15 de marzo de.
- Instituto Nacional de Estadística (INE) (2014), "Censo 2011. Productos Para consultar esta información", Curso de la Escuela de Estadística de las Administraciones Públicas (EEAP), INE, Madrid, 3 de marzo de.
- Rao, J.N.K. (2003), *Small Area Estimation*, Wiley Series in Survey Methodology. John Wiley and Sons. Hoboken, NJ.

-
- Reig, E.; Goerlich, F.J. and Cantarino, I. (2016), "Delimitación de áreas rurales y urbanas a nivel local", *FBVA - Informe Técnico*, pp. 1-138.
- Stephan, F.F. (1942), "Iterative method of adjusting frequency tables when expected margins are known", *The Annals of Mathematical Statistics*, Vol. 13 No. 2, pp. 166-178.

Further reading

- Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003), "Micro-level estimation of poverty and inequality", *Econometrica*, Vol. 71 No. 1, pp. 355-364.
- Goerlich, F.J. and Cantarino, I. (2016), "Zonas de morfología urbana. Coberturas del suelo y demografía", *FBVA - Informe Técnico*, pp. 1-125.
- Goerlich, F.J. and Cantarino, I. (2017), "Grid poblacional 2011 Para España. Evaluación metodológica de diversas posibilidades de elaboración", *Estudios Geográficos*, Vol. 78 No. 282, pp. 135-163, available at: <http://doi.org/10.3989/estgeogr.201705> (accessed 1 April 2019).
- Instituto Nacional de Estadística (INE) (2019), "Qué tipos de cifras de población publica el INE?", available at: www.ine.es/daco/daco43/epoba/cifras.pdf (accessed 20 May 2016).

Corresponding author

Francisco J. Goerlich can be contacted at: Francisco.J.Goerlich@uv.es