

NOW-CAST POPULATION ESTIMATES AT MUNICIPAL LEVEL DISCREPANCIES BETWEEN THE CENSUS AND THE MUNICIPAL REGISTER AT SMALL SCALE

Francisco J. Goerlich Gisbert

Universidad de Valencia e Instituto Valenciano de Investigaciones Económicas

The sources of information for the study of demographic phenomena are a recurrent research topic in social sciences (Melón 1951, Arango 1981, Reher y Valero 1995, Puyol 1997, Reher 1997, Goerlich, Mas, Azagra y Chorén 2006). This paper aims to contribute to this literature, focusing on the current population statistics in Spain and their discrepancies at a municipal level.

At present, the Spanish National Statistical Institute (NSI, *Instituto Nacional de Estadística –INE*) publishes two types of population figures:

(i) On the one hand, it publishes population estimates (inter-census or now-cast figures) and projections (for the short, medium and long term). For the most part, these are published at a provincial level of disaggregation (NUTS 3), and they are anchored at the level given by the censuses (which take place every 10 years). They use all the available information on demographic fluxes (births, deaths and migrations), measured or estimated. From the point of view of the present work, what matters is that the now-cast population estimates (*ePOBa* in *INE*'s terminology) take as a base the population level from the last available census (currently the 2001 census). Another minor point worth taking into consideration here is that now-cast population estimates are the ones that *INE* supplies to international institutions as the actual Spanish population.

(ii) On the other hand, official population figures (in a legal sense) are based on a municipal administrative registry called *Padrón Municipal*. Each municipal council is responsible for the management of the registry, including the updating procedures. However, the National Statistical Institute (*INE*) performs the necessary supervision and coordination of the system: that is, crossing records among municipalities so as to ensure that the system as a whole is consistent. Because of the current management system of the municipal registry (the so-called continuous registry —*Padrón continuo*—), *INE* obtains legal population figures with 1st January of each year as a reference date since 1998, when this system was implemented for the first time. The legal character of these population figures derive from the fact that they are published each year in a Royal Decree. This figure is, therefore, the official

population of each municipality and is the one taken into consideration in negotiations between different levels of administrations when public funds have to be distributed between different levels of government. From the point of view of the present work, what matters here is that the legal population figures at municipal level come from an administrative registry, subject to legal requirements, and alterations have to be made at the level of the individual record with a legal basis.

It is true, however, that from the municipal registry the *INE* derives other statistical products from the municipal registry. Thus, with the same periodicity and reference date, we have infra-municipal population figures either at the level of census tracts or at the level of settlements that appear in the nomenclator. Although they are not official in the legal sense, they come from the municipal population registry.

It is well known that not only do both types of population figures not coincide, but they differ by a significant amount. Now-cast population figures (*ePOBa*) and aggregate population figures derived from the municipal population registry do not agree in their levels (García Coll y Sánchez Aguilera 2001; Goerlich 2007). This discrepancy can be traced back to the fact that both population figures were not forced to coincide when the last census took place, as was the case prior to the implementation of the current management system of the municipal registry. Before the 2001 census, the recounting of people in the municipal registry (which took place every 5 years in the old system) was made coincident at the time of a new census. This was the case with the 1981 and 1991 censuses, when both population figures were made coincident. The new system implemented in 1998, however, is completely based on administrative records that are updated on a continuous basis.

Graph 1 clearly shows the discrepancy between both sources of population figures. We can observe that the divergence increases at the end of the 20th century and shows a relatively stable pattern in the first decade of this century. In the last few years the change in both population figures is quite similar with the discrepancy remaining stable. Fortunately, we have only a level and not a rate of change discrepancy.

Graph 2 illustrates the same profile from a different perspective: the difference between both sources of population figures, i.e. the height between the blue and pink points in graph 1. From the last census, the difference between both population sources remains relatively stable at around one million people, showing a stable pattern. The last legislative change concerning foreign people without a permanent resident permit, implemented on 1st January 2006, has been unable to close this discrepancy in any significant amount.

Summing up, both population sources demonstrate the same tendency in growth rates, and disagree in the levels. This discrepancy is relatively stable from the last census onwards and has been shown to be extremely difficult to close in practice.

Graph 2 shows that the absolute discrepancy between the now-cast population estimates and the municipal registry of 1st January 2002 is slightly less than one million inhabitants (874 thousand people), representing a relative discrepancy of 2.13% (the municipal registry gives a higher population estimate). The paper shows that this divergence between information sources is highly heterogeneous across different geographical units with observable and unobservable characteristics. Moreover, these discrepancies increase as we move to smaller geographical units. That is, we observe a greater variability at provincial level (NUTS 3)

than at regional level (the so-called Autonomous Regions that correspond to NUTS 2 in the European nomenclature).

Some figures are highly illustrative. At national level the aggregate discrepancy between population sources is 2.13%, as was mentioned above. At NUTS 2 level the range of discrepancies oscillates between 1.05% (Castile and León) and 5.06% (Balearic Islands), with a standard deviation of 1.3%. (This excludes the tiny autonomous cities of Ceuta and Melilla). At NUTS 3 level the range of discrepancies increases significantly, going from a slightly negative value, -0.35% (Granada) to more than 3 times the discrepancy at national level, 6.54% (Alicante), with a standard deviation of 1.5%.

The first contribution of the paper is to show that these discrepancies are huge at municipal level. Taking as a case study the province of Alicante, graph 3 shows that discrepancies go from more than 50% (in three cases exceeding 40%) to less than -20%. The standard deviation of discrepancies at municipal level for this province is 10.1%, 6 times higher than the one observed at national level. In this context, map 2 provides the geography of the discrepancies, clearly showing that neighborhoods matter because they tend to cluster together. In fact, a correlation analysis (table 5) shows that not only does neighborhood matter for these discrepancies, but also demography and physical geography given that they correlate with altitude, coast proximity, mobility, and foreign people in the municipalities, in addition to first order neighborhood discrepancies.

The second contribution of the paper is to show how we can obtain municipal population figures that are consistent in their levels with the now cast-population estimates. That is, municipal population figures anchored in the levels of the 2001 census and which, in turn, aggregate to the *ePOBa* provincial figure given by *INE*. These can be called now-cast municipal population estimates, despite deriving from a very different method than the *ePOBa* figures at regional level provided by *INE*. Again, we use Alicante as a case study.

It is fairly obvious that redistribution methods typically used in these cases (such as imposing the municipal structure derived from the municipal population registry on the aggregate figure at provincial level in the now-cast population estimate) are likely to give very poor results due to the heterogeneity of the discrepancies. In fact, this implies significant abrupt changes in municipal figures between the census data and the next year estimates, especially in big cities.

Given the bad results with statistical modeling, and the relative stability of the discrepancies in the levels between the now-cast population figures and the municipal registry observed at the aggregate level (graph 1), we use a method of components approach to reconstruct municipal population figures from the last census onwards. We use the dynamic equation of population determination, and therefore add births, subtract deaths and adjust for migrations at municipal level. When aggregated at provincial level, the resulting figure from this yearly updating process is surprisingly close to the provincial *ePOBa* figure offered by *INE*, and so eventually this small discrepancy is distributed proportionally using municipal foreign people as a proxy. This simple method works extremely well in our case study, is very easy to apply, and can be used more generally in other provinces. Moreover, it can be extended to other dimensions, such as sexes and/or ages, and discrepancies resolved by mechanical bi-proportional methods such as the RAS method. Contrary to *INE*'s methodology to obtain regional *ePOBa* estimates, which is top-down, we adopt here a bottom-up approach.

Of course, how good the methodology is, and the accuracy of our methods can only be judged from the municipal data derived from the next 2011 census, whose methodology will be slightly different from the one used so far. As this is true at any geographical level of analysis, the next census will provide us with benchmark data against which to compare now-cast population estimates, not only at municipal level as described in the present paper, but also regional and aggregate data as offered by *INE*.