

HOW MANY ARE WE?

The demographic statistics of the Spanish National Statistical Institute (INE)

Francisco José Goerlich Gisbert
University of Valencia and IVIE

I. INTRODUCTION

This paper looks at the internal and external consistency of the demographic statistics produced by the Spanish National Statistical Institute (INE). In particular we focus on two things; on the one hand we look carefully at the different sources for population *stocks*, on the other we look at the consistency between *stocks* and flows of births, deaths and net migration to see if the basic demographic accounting is verified.

Because demographic statistics are produced at a regional level we perform our analysis for regions, *Comunidades Autónomas* (NUTS 2). This offers us the possibility to see that discrepancies are not uniform across space. Eventually, a simple exercise is done for the age distribution of population at a national level.

Two comments are important before proceeding. Firstly, we think demographic statistics are important enough to take the discrepancies we show in this investigation seriously. If we don't know how many we are, it is difficult to know how much we produce. Moreover, population figures are essential information in the distribution of public funds, either at regional or international level, so it is important to have accurate information.

Secondly, this paper should not be taken as a critique of the behaviour of INE as an organization which produces and disseminates statistical information. In fact our work has been made possible because INE has decided to offer to the public more (and better) statistics. The calculations we have done can be reproduced by any researcher interested in the topic and this work is offered with the aim to improve our demographic statistics.

II. POPULATION STOCKS: CENSUS VERSUS MUNICIPAL REGISTER

There are two basic sources of population *stocks*: the census carried out approximately every ten years, (the last one in 2001 November 1st), and the municipal register, a continuous

administrative register that, given the present legislation, delivers the official (legal) population figures every year dated January 1st.

The starting point of the paper is the discrepancy between both figures of the population stock for the year they coincide with, despite the fact that both sources are using the same definition of population (resident or *de jure* population). In absolute terms this discrepancy is close to 1 million people (873.650), and in relative terms this discrepancy is around 2,13%. These are big figures. Moreover, for some regions relative discrepancies are greater than 5%.

Given this, we do some exercises that show the importance of the different sources of population in computing basic demographic indicators, such as birth, fertility and mortality rates or life expectancy. The differences here are not big, but not inconsiderable either.

Eventually we review the framework of the law which applies to the municipal register, since this has changed quite a lot since the mid nineties. In fact, we find that some legal changes are partly responsible for the discrepancies in population figures shown previously, since the municipal registers find it difficult to track immigrants once registered, if afterwards they go abroad or even move within the country. As a result there is a presumption that the municipal registered is overvalued, as it shows a higher number of population than really exists.

Some recent changes in the legislation, which were in effect for the first time in 2006, are expected to partly correct this problem, but the final result on population statistics remains to be seen.

III. POPULATION FLOWS: BIRTHS, DEATHS AND MIGRATIONS

After reviewing the different sources of population figures, we turn to the consistency between *stocks* and flows: births, deaths and migrations, either external or internal.

Our starting point is the basic accounting relation in demography. Given a territorial unit the stock of population can be described by the following equation

$$P^{(t+1)} = P^{(t)} + N^t - D^t + I^t - O^t \quad (1)$$

where $P^{(t)}$ and $P^{(t+1)}$ are population *stocks* at the beginning of the years t and $t + 1$, respectively, D^t and N^t are deaths and births, I^t is the flow of immigrants and O^t is the flow of emigrants during the year t .

The INE produces statistics separately for each of the elements in equation (1), for the nation as a whole and for the different regions. The data for D^t and N^t come from the administrative civil registers, whose data also go to the municipal register, and the data from I^t and O^t come from the statistical exploitation of the municipal register, which nowadays is centralized by the INE. Hence, in principle at least, all the population statistics come from a few, consistent sources, so we expect *a priori* that equation (1) should be satisfied or, at a minimum, no big discrepancies should be observed.

This turns out not to be the case when we put all the statistics together as a lot of discrepancies come from the net migration term, $I^t - O^t$, if it is calculated as a residual from equation (1), $I^t - O^t = P^{(t+1)} - P^{(t)} - (N^t - D^t)$, or directly from the Residential Variations Statistics (*Estadística de Variaciones Residenciales*). The INE, aware of this problem, offers

some explanations for this fact on its *web* page (<http://www.uv.es>), but in essence the main problem seems to be a combination of the legal framework of the municipal register, which favours the registration of an immigrant even he is illegally staying in the country, with an inadequate tracking of migration, especially external migration. This makes it very difficult to know if an immigrant has eventually returned to their country, since there is no incentive to de-register, but rather the incentive is to try to maintain the registration, even if he is not in the country anymore.

This is not, however, the only problem. A final exercise looking at the age distribution of population at the national level shows that the information from the civil register does not flow with the required fluency from this register to the municipal register; so some additional inconsistencies are found at the extremes of the age distribution.

IV. FINAL REMARKS

This paper has presented some inconsistencies between *stocks* and flows in the demographic statistics published by the Spanish National Statistical Institute (INE). Even if it is true that this fact is common in other countries; the magnitudes of some discrepancies are quite puzzling and some of them can be traced back to the legal framework and the lack of communication between administrative registers.

Demographic *stocks* and flows obey an accounting relationship. In fact this relationship is used by demographic projection methods to forecast population figures. What seems puzzling is that when projected statistics are substituted by real statistics then the accounting relationship is lost.

We argue in favour of developing accounting methods for demographic statistics similar to the accounting methods developed a long time ago in National Accounting Statistics (Stone and Stone, 1966). In fact some eminent demographers, like Frans Willekens or Phil Rees, among many others, have argued in favour of this, since accounting has a great advantage, it should balance.

Finally, the paper offers an annex with the statistical information that was publicly available at the INE's *web* as in October 2006.

