An Annual Estimate of Spatially Disaggregated Populations: Spain, 1900-2011

Rafael González-Val (corresponding author)

Universidad de Zaragoza

Facultad de Economía y Empresa (Campus Paraíso). c/ Gran Vía 2, 50005, Zaragoza, Spain

Institut d'Economia de Barcelona (IEB)

Facultat d'Economia i Empresa, Universitat de Barcelona, c/ John M. Keynes, 1-11, 08034 Barcelona, Spain

Email: rafaelg@unizar.es

Javier Silvestre

Universidad de Zaragoza, Instituto Agroalimentario de Aragón (IA2) and Grupo de Estudios Población y Sociedad (GEPS)

Facultad de Economía y Empresa (Campus Paraíso). c/ Gran Vía 2, 50005, Zaragoza, Spain

Email: javisil@unizar.es

Acknowledgements

We are especially grateful to the journal's reviewers, and David S. Reher, who helped us to define the estimation procedure, as well as Jordi Martí-Henneberg and Mateu Morillas-Torné, who shared data with us. We also received helpful bibliographical suggestions from Leandro Prados de la Escosura, comments from Vicente Pinilla, and assistance in obtaining data from Carlos Serrano.

Funding

Financial support was provided by the Spanish Ministry of Economy, Industry and Competitiveness (research projects ECO2015-65582, ECO2017-82246-P, ECO2016-75941-R and PGC2018-095529-B-I00); and the Government of Aragon–European Fund for Economic and Regional Development (2014-2020) (research projects S27 and ADETRE).

An Annual Estimate of Spatially Disaggregated Populations: Spain, 1900-2011

Abstract

Long-term population estimates are usually based on decennial (or roughly decennial), census-type data, especially those that go beyond the aggregate or national level. The lack of annual population datasets, however, restricts the range of analysis and therefore our knowledge of populations' behaviour over time. This paper contributes to this scant literature by providing two new long-term, annually and spatially disaggregated population estimates for Spain, a key case study. We use a set of official sources and propose a straightforward method to estimate Spanish annual populations at two levels of spatial disaggregation, the provinces (NUTS III) and their capital-cities (a selection of LAU II/NUTS V), between 1900 and 2011. For the cities, our estimates account for changes in boundaries. Our proposed method is especially suitable for countries with a lack of long-term, systematic, historical migration statistics, a common issue. To test the robustness of our estimates, some comparisons are made with previous more aggregated-level or decennial census-based estimates, and a preliminary analysis of the evolution of disaggregated Spanish population is presented.

Keywords

Population estimation method; annual data; spatial disaggregation; long-term; Spain

JEL Classifications

C82 – J11 – N33 – N34 – R12

1 **Introduction**

The long-term estimation of population is a cornerstone of research in historical demography, historical geography, economic history and regional science. It is not only the importance of knowing populations by themselves, but also their strong relationships to economic growth and any other societal change. In Europe, national population estimates are a principal part of collections of historical statistics comparing nations (e.g. Mitchell 1992). Within countries, a favourite unit of analysis has been the city (e.g. De Vries 1984; Bairoch et al. 1988).

However, long-term estimates of populations tend to be based on population censuses, which are generally conducted every ten years or so. As recently recalled by Ronsse and Standaert (2017), while this approach is helpful for some analyses, it limits the scope of research. A common solution is the use of interpolation, which admits several variations. One simple method consists of using only consecutive censuses to estimate the annual in-between observations. An alternative uses additional information in order to obtain these unobserved values, although there is no single criterion to link different sources. In this regard, the recent application of state-space models helps the researchers to combine census data with related demographic variables, such as births, deaths and migration (Ronsse and Standaert 2017).

As far as we know, there are only three long-term estimates of annually and spatially disaggregated populations. Sharma (2003) used a sample of 100 Indian cities for the 1901-1991 period; Bosker et al. (2008) constructed a dataset of 62 West-German cities for the 1925-1999 period; and Ronsse and Standaert (2017) constructed a dataset of 2,681 Belgian municipalities for the 1880-1970 period. Sharma (2003) and Bosker et al. (2008) consider only the largest cities, while Ronsse and Standaert's (2017) sample includes all Belgian municipalities.

In this paper, we provide an annual estimate of the Spanish population for the 1900-2011 period based on the decennial censuses and two further, related variables: the annual numbers of births and deaths. We, therefore, go beyond the use of a simple interpolation method. Instead, we apply the so-called inter-census balance method to estimate the difference between census (total) population growth and the natural increase (births minus deaths). The resulting residual provides us with an estimate of inter-census net migration, which acts as the third related variable. The three variables, births, deaths and estimated migration, are then combined with population figures at census dates. Principally, this straightforward approach obviates the need to use 'direct' annually and spatially disaggregated measures of internal and international migration, which are not always available. In fact, in the case of Spain, they are inaccurate, incomplete or non-existent for a significant part of the period covered here. Our approach also assumes that the population data and the number of births and deaths were reliably collected—an issue to which we will turn below.

To establish a time-series model, we propose two complementary spatial units of analysis, for which the data necessary to apply our approach are available: the province (NUTS III) and its capital city (representing a selection of the main LAU II/NUTS V units). The use of these boundaries is a compromise between the whole country, or regions, and municipalities, which will allow scholars to establish links with different, available indicators. For the cities, we account for changes in boundaries.

We extend the scant literature, providing annually and spatially disaggregated populations, to Spain, an important case study. The demographic evolution of Spain associated with modern economic growth has been extensively studied and is a topic deeply rooted in the debate on the country's relative backwardness over a significant part of its recent history (e.g. Prados de la Escosura 2017). However, almost all research on the long-term evolution of Spanish population has only used census data, thus missing all the inter-census population dynamics.

In this regard, recent studies include those of Martí-Henneberg (2005), Ayuda et al. (2010a, 2010b) and Gregory et al. (2010), at the province level; and Reher (1994), Lanaspa et al. (2003), De Cos and Reques (2005), Goerlich et al. (2006, 2015), Tafunell (2005), Espinha da Silveira et al. (2013), and González-Val et al. (2014, 2017), at the municipal level, using different population size thresholds.[1] For the most recent years, Bayona-i-Carrasco and Pujadas (2019) and Pujadas and Bayona-i-Carrasco (2019) complete the census series with the annual revision of the municipal register. Also at census dates, Beltrán-Tapia et al. (2018) use the *partido judicial* (district), which is composed of several municipalities and disaggregated at a lower level than the province. As a key aspect of long-term comparisons, sometimes research has paid special attention to geographic changes in municipal boundaries. For example, Goerlich et al. (2006, 2015), Franch-Auladell et al. (2014), Martí-Henneberg et al. (2016) and Chatel et al. (2017), based on geographical information system (GIS) tools, use the 2001 or 2011 census as the base reference to homogenize the population series.

A major strand of the literature covers the most recent decades, and usually draws on annual data to estimate the contribution of internal migration to population growth or decline, and the spatial distribution of population at different levels of disaggregation (e.g. Cabré et al. 1985; Recaño 2004; Recaño and Domingo 2006; Franch-Auladell et al. 2009; Franch-Auladell and Recaño 2011; López-Gay 2014; Bayona-i-Carrasco et al. 2018). While, the estimation by Alcaide-Inchausti et al. (2007) is an intermediate step between long-term approaches that only use decennial censuses and those that apply interpolation to obtain

[1] De cos and Reques (2005) and Cardesín and Mirás (2017) provide comprehensive reviews.

annual data: by using related demographic variables to construct five-year estimates of provincial populations for the twentieth century.

Few studies for Spain provide long-term annual estimates at any geographical level. At the national level, Maddison (2001) and Nicolau (2005) essentially apply linear interpolation to census years; and Maluquer de Motes (2008) and Prados de la Escosura (2017) use methods based on related variables. Also based on related variables, De la Fuente (2016) has constructed an annually and spatially disaggregated estimate of the Spanish population, but he uses Autonomous Communities (NUTS II), instead of provinces (NUTS III), as the unit of analysis, and the period covered, 1950-2015, is comparatively short. Finally, Almarcha et al. (1975) provides annual values at the provincial level for the 1900-1970 period, but they are simply calculated by interpolation—and data for 22 scattered years are missing. Almarcha's et al. (1975) estimates tend to coincide with those of the Spanish Statistical Office (*Instituto Nacional de Estadística*, INE)—which also provides an annual estimate for capital cities.

This paper focuses on the construction of two new, long-term, annually and spatially disaggregated estimates of the Spanish population. The rest of the paper is organised as follows. In Section 2, we present the data sources and the methodology. Section 3 compares our estimates with those from previous authors. Section 4 provides an analysis of the evolution of province and capital-city size distributions, as a first exploration of the basic characteristics of our dataset. Section 5 concludes.

2 Data sources and estimation method

2.1 Data sources

We use two main data sources established by the Spanish Statistical Office (*Instituto Nacional de Estadística*, INE), available at http://www.ine.es/: the decennial censuses and the yearly *Movimiento natural de la población* (natural growth of population). The censuses provide population figures, and were conducted in years ending in zero, between 1900 and 1970, and one, from 1981 onwards. The *Movimiento natural de la población* provides the difference between the number of births and deaths.

Scholars have considered Spanish censuses—including the census carried out in 1940, right after the Spanish Civil War—, to be reliable and consistent over time (e.g. Reher and Valero-Lobo 1995, pp. 29, 51, 54; Goerlich et al. 2006, pp. 29, 53; Maluquer de Motes 2008). Up to 1991, the censuses provided both 'de facto' and 'resident' populations. Only resident population is provided thereafter. De facto population refers to all persons present in the area of reference at the census date. Whereas resident or 'de jure' population refers to all persons with legal residence, present or not, in the area of reference at the census date. Resident population, therefore, includes international and internal out-migrants. For the 1900-1991 period, we follow the most common choice in long-term population estimations, including the Spanish Statistical Office's own estimations, that is, the use of de facto population (e.g. Nicolau 2005; Maluquer de Motes 2008; Prados de la Escosura 2017; see also Sánchez-Alonso 2010). From 1991 onwards, we follow the unavoidable procedure of splicing the (so far) de facto population series with the resident population series, as done by Maluquer de Motes (2008) and Prados de la Escosura (2017). Certainly, this criterion breaks the coherence of the series. However, the comparison of both population definitions yields a negligible difference.[2]

As for the *Movimiento natural de la población*, a systematic yearly estimate starts in 1900, and that is the reason for the period covered here. In spite of some problems associated with data collection at some points in time, the *Movimiento* has been portrayed as very

[2] Prados de la Escosura (2017, pp. 190-191). See also Goerlich et al. (2006, p. 4) and De la Fuente (2006, p. 4).

reliable (e.g. Reher and Valero-Lobo 1995, pp. 85, 90-91; Alcaide-Inchausti et al. 2007, p. 12; Maluquer de Motes 2008). We used the definitive revisions of preliminary estimates as provided by the Spanish Statistical Office.[3]

A further official source at the capital-city level available from 1996 onwards is the annual revision of the municipal register (*Padrón continuo*), which we used for the inter-census years between 1996 and 2001, and 2001 and 2011. The municipal register includes those individuals who reside regularly in each municipality, and it is updated with births and deaths. Registration is compulsory.

The units of analysis of our two estimates are the province (NUTS III) and its capital city. Except for one case up to 1927 (see Table 1), the number of inhabitants of capital cities is above 10,000, which is a standard urban threshold in Spain (Tafunell 2005, p. 457). The number of provinces, and therefore the capitals, 49, has remained unchanged. The two provinces corresponding to the Canary Islands are joined together (due to the census reporting criteria during the first decades of the twentieth century), and the Spanish enclaves of Ceuta and Melilla in northern Africa have been excluded. As discussed above, several studies use decade-by-decade observations at the, more disaggregated, municipal level. However, except for capital cities, our proposed method cannot be applied to all municipalities, since the necessary data to estimate annual net migration, that is, annual records of births and deaths, are not consistently available for such a long period: at best, from 1975 onwards, and recording criteria may vary (Reher and Valero Lobo 1995, p. 90; Egea 2000; Nicolau 2005). Furthermore, the bulk of economic and social historical information in Spain that is susceptible to be matched with demographic series is mostly available at our chosen geographical levels (especially the province).

[3] The list of all specific sources, including their date of publication, is available from the authors upon request.

2.2 Estimation method

To obtain annual estimates of population, our method is comprised of two main procedures, in relation to boundary changes in the spatial units considered, and, particularly, the estimation of values in-between censuses.

A fundamental feature of long-term population estimations is the homogenization of spatial units, since boundaries may have changed over time. There have been practically no relevant administrative changes in the case of the provinces (Goerlich et al. 2006, p. 76; 2015, p. 114). However, 31 out of 49 capital cities, of different size, have changed their boundaries at some point in time over the 1900-2011 period, as a result of annexations and segregations of (usually neighbouring) municipalities. Previous estimates of homogeneous long-term populations in Spain include those of García-Fernández (1985) and Goerlich et al. (2006, 2015) for the twentieth century. More recently, using GIS Martí-Henneberg et al. (2016) and Chatel et al. (2017) have extended Goerlich's et al. (2006) series of resident, municipal populations, estimating backward (until 1870) and forward (until 2011).[4]

Our estimate, which, as explained above, is based on two linked populations, de facto and resident, follows the homogenization criteria proposed by the Spanish Statistical Office ('Alterations to the municipalities in the Population Censuses since 1842').[5] Using the 2011 census as reference, for each—in our case—capital city these criteria account for 13 types of change since 1900, regarding the, predominantly, annexation and, often more recently, segregation of municipalities.[6] Below, in section 3, we will offer a comparison between our population estimate and that of Goerlich et al. (2015), who apply a similar, official sources-

[4] The areal interpolation method (e.g. Gregory 2002; Gregory and Ell 2006) is used as reference.

[5] Available at: https://www.ine.es/intercensal/?L=1.

[6] The list of annexations and segregations is available from the authors upon request.

based, methodology. Table 1 summarizes the descriptive statistics of provinces and capital cities at census dates.

[Insert Table 1 around here]

In order to estimate the annual population, both for provinces and capital cities, we use population figures at census dates, and the natural increase—births minus deaths—for each inter-census year. Then, we applied the inter-census balance method to obtain the difference between census population growth and the natural increase over the period between two consecutive censuses. The resulting residual provides us with an estimate of inter-census net migration—out-migrants minus in-migrants, which is thereafter annualized, that is, divided by the number of years between censuses. It is therefore assumed that net migration is homogeneous throughout all inter-census years. The final annual estimate of population for each inter-census year is the result of adding the annual natural increase and the annualized net migration to the first census figure, and then operating recursively. For the capital cities, we applied this method to inter-census periods up to 1991 and the 1991-1996 period. From 1996s onwards, we used the annual revision of the municipal register instead, as explained above.

The approach described in the previous paragraph is similar to those used in annual estimations at the national level (Maluquer de Motes 2008; Prados de la Escosura 2017), annual estimations at the regional level (De la Fuente 2016), and five-year estimations at the provincial level (Alcaide-Inchausti et al. 2007). The main difference between previous estimations and ours is that the former, for some periods, use additional, specific data on international emigration and immigration as a result of which the *estimated* residual obtained via the inter-census balance method is somewhat reduced and assumed to correspond to internal, but not international, migration.

Our contribution is to estimate annual populations at more disaggregated spatial units, and for a longer period. Therefore, we favour the use of the estimated inter-census net migration instead of any specific migration data, for two main reasons. First, there are shortcomings related to the use of international migration statistics at different points of time (Losada and Viso 1997; Ortega and Silvestre 2006; Alcaide-Inchausti et al. 2007, p. 24; Maluquer de Motes 2007, 2008; Sánchez-Alonso 2010; Prados de la Escosura 2017, pp. 189-190). In fact, for some rather long periods, scholars simply rely on the inter-census balance method (Maluquer de Motes 2008, pp. 146-147; De la Fuente 2016, p. 5). Not to mention the lack of international migration statistics at the level of capital cities. Second, provincial statistics may not capture internal-to-international stepwise migration adequately. Thus, some rural migrants may have spent a period in a city or elsewhere within the country before undertaking a subsequent, international movement. The extent of this phenomenon remains a topic of debate (Silvestre 2005; Sánchez-Alonso 2010).

3 Estimates compared

In this section, we compare our estimate of the Spanish population at the national, provincial and capital-city levels with key previous estimates. As most of the existing ones only consider the national level, we begin with the total Spanish population. We aggregate our estimates of annual populations for all 49 provinces, comprising the entire territory of the country, in order to compare the total with the result of annual estimations of the Spanish population by Nicolau (2005), Maluquer de Motes (2008), De la Fuente (2016), Prados de la Escosura (2017), and the Spanish Statistical Office. Time spans may differ. Figure 1, panel *a*, brings together the estimates in logarithmic scale.[7] The estimates tend to be very much alike, and no extraordinary deviations can be observed. The main differences are observed over the 1930s and 1940s, in particular during the years of the Spanish Civil War (1936-1939) and its aftermath. Figure 1, panel *b*, zeroes in on that period. Estimates only based on interpolation, that is, those of Nicolau (2005) and the Spanish Statistical Office, seem to underestimate the total population for the war period, while Maluquer de Motes (2008) and Prados de la Escosura (2017) provide the highest population values. Our estimate hovers in the mid-range and displays a smooth evolution during the war years.

[Insert Figure 1 around here]

We also compare our database with others at more disaggregated levels. To do that, we estimate the empirical density functions using adaptive kernels for the whole 1900-2011 period in order to assess whether our dataset under- or overestimates the population of large, medium or small provinces and capital-cities. Figure 2, panel *a*, shows the empirical density functions for provinces. We compare our annual estimates with those of Alcaide-Inchausti et al. (2007), which—based on interpolation—provide five-year values. Therefore, we interpolate these values to fill in the gaps in order to obtain an annual series. Apparently, the two series display a similar size distribution, with no remarkable deviations, but there seem to be small deviations for medium-sized provinces, for which Alcaide-Inchausti et al. (2007) slightly overestimates the density. We run the Kolmogorov-Smirnov (KS) test to check whether both distributions are significantly different. The KS test's null hypothesis is that the two samples come from the same distribution, and this hypothesis is rejected at the 5% significance level (p-value = 0.011).

[7] Unreported results without applying the log scale are similar.

[Insert Figure 2 around here]

Figure 2, panel *b*, shows the empirical density functions for capital cities. Here our annual estimate is compared with that of Goerlich et al. (2015). Like ours, this estimation method accounts for changes in city boundaries. The authors provide decennial values coinciding with census dates, so again we interpolate these values to fill in the inter-census gaps. The results show slight deviations in density for medium-sized cities, and the KS test points to the existence of non significant deviations between both distributions (p-value = 0.465).

Overall, as shown in Figures 1 and 2, we do not find important deviations between our population estimates and previous estimates at the national level, with the main exception of the civil war years; but we find significant differences at the provincial level. As for the capital-city level, our estimate as a whole is consistent with the homogeneous estimate by Goerlich et al. (2006, 2015), which would indicate that we have successfully controlled for changes in boundaries. Therefore, it would be fair to say that our new annual datasets are reasonable and within the limits established using previous estimation methods and, at the same time, provide new evidence at spatially disaggregated levels, especially the province.

4 The evolution of province and capital-city size distributions

A standard analysis of the evolution of population over time involves fitting a Pareto distribution to the data in order to estimate the shape of its size distribution (e.g. Cheshire 1999; Gabaix and Ioannides 2004). Aimed at assessing changes in the Spanish population system, previous research has applied this method to different spatial units over different periods (e.g. Reher 1994; Lanaspa et al. 2003; Goerlich and Mas 2010; González-Val et al. 2017). For instance, regularities and changes in city sizes at different points have been found, related to major economic transformations, such as industrialisation and domestic market integration (Gónzalez-Val et al. 2017). However, all these studies are based on decennial (census) data. So, the main advantage of our population estimates is the use of annual data.

Let us denote S as the size or population of a unit—that is, a province or capital-city—and R as its corresponding rank (1 for the largest, 2 for the second largest and so on). Within a long-term temporal framework, it is necessary to use a relative measure of size, since, for example, 10,000 inhabitants represented a larger population size in 1900 than in 2011. Thus, we define the relative size of a unit as the quotient between its population and the contemporary average,

$$s = \frac{S}{\overline{S}} = \frac{S}{\frac{1}{N}\sum_{1}^{N} S},$$

where N is the sample size. A power law (Pareto distribution) links the province or capital-city relative size and rank as follows: $R(s) = As^{-a}$. This expression is usually specified and estimated in its logarithmic version:

$$ln(R) = b - a\,ln(s) + \xi, \qquad (1)$$

where ξ is the error term, and b and a are the parameters that characterize the distribution. The latter is known as the Pareto exponent. When $\hat{a} = 1$, Zipf's law is observed, which points to a constant relationship between rank and size. In other words, when ordered from largest to smallest, the size of the second unit is half that of the first one, the size of the third unit is a third of the first one and so on. Thus, the greater the coefficient, the more homogeneous are the province or capital-city sizes; in addition, an increase in the coefficient over time would imply a process of convergence in province or capital-city sizes. Correspondingly, the smaller the coefficient, the less homogeneous are the province or capital-city sizes; and a decrease in the coefficient over time would imply a process of divergence in sizes. As proposed by Gabaix and Ibragimov (2011), we specify equation (1) by subtracting $1/2$ from the rank to obtain an unbiased estimation of a:

$$ln\left(R-\frac{1}{2}\right)=b-a\,ln(s)+\varepsilon \,. \qquad (2)$$

We use our annual dataset based on a consistent set of provinces and capital-cities. Figure 3 reports the results of an OLS estimation of Equation (2). For the whole 1900-2011 period, panels *a* and *b* refer to provinces and capital-cities, respectively. Regarding provinces (panel a), the results show that the estimated Pareto coefficient decreases at a relatively constant rate over time, which indicates a continuous process of divergence in size. Estimated values at the end of the period are close to one, and from the 1940s onwards we cannot reject that $\hat{a}=1$, which means that Zipf's law holds from the middle of the twentieth century to 2011. Results for capital-cities (panel b) show that the Pareto exponent tends to stay at the same level throughout the period covered, with the estimated coefficients close to one in all possible sub-periods, thereby finding evidence in favour of the fulfilment of Zipf's law.

[Insert Figure 3 around here]

Therefore, these results point to different behaviours in the dynamics of population depending on the spatial unit considered. At the level of provinces, the estimated coefficient (as reported in Figure 3, panel *a*) decreases over time. That is, a process of divergence between provincial populations and, as a result, an increase in the spatial concentration of the *national* population is shown—since the sum of provincial populations accounts for the whole Spanish population. This result is consistent with that of Ayuda et al. (2010a, 2010b), but here we provide year-by-year, instead of decennial, evidence of the evolution of provincial population growth over time. On the other hand, the distribution of population across capital-cities was quite stable throughout the period covered (see Figure 3, panel *b*). This is a different result from those produced by previous research that uses decennial data and different sample sizes and city definitions, which shows a decreasing trend in the Pareto exponent (Lanaspa et al. 2003; Goerlich and Mas, 2010; González-Val et al. 2017). Although the change in the trend observed from the early 1970s is in line with the decentralization process of population from the largest cities shown by Garcia-López et al. (2015) and Martí-Hennenberg et al. (2016).

One final confirmation of the evolution in the degree of evenness or unevenness of the population distribution may be provided by the Gini coefficient—which, as is well-known, does not impose a specific size distribution (Pareto for rank-size coefficients). Figure 4 shows the annual results at both the provincial and capital-city levels, again considering relative sizes. The different evolution in the coefficients between provinces and cities is notable. The coefficient clearly increases over time in the case of the provinces, which indicates a divergent behaviour of relative populations; while in the case of the capital-cities the coefficient tends to remain constant until around 1970, when it decreases slightly and points to a soft convergence in relative sizes.

[Insert Figure 4 around here]

The possibility of estimating further analysis by year will also allow us to delve deeper into the dynamics of size distributions for particular periods. Thus, the short- and long-term effects of key demographic events such as the flu pandemic (1918-1919) and, particularly, the Spanish Civil War (1936-1939) may be approached from a different angle. For example, even though the flu pandemic and the civil war disrupted national population growth (e.g. Maluquer de Motes 2008; Prados de la Escosura 2017), the preliminary analysis presented here (in Figures 3 and 4) would suggest that none of these events seems to have had a significant effect on the size distribution of provinces and capital cities. Another possibly meaningful effect that would require further research is that of the baby boom in the 1950s and 1960s. Although small when compared with other countries, as shown by Van Babel and Reher (2013), our results, particularly for the provinces (Figures 3 and 4), nevertheless point to rising unevenness in the population distribution over those decades. This is indicated by the acceleration in the fall of the Pareto exponent (Figure 3) and the rise in the Gini coefficient (Figure 4). A contemporary rise in emigration and internal migration may also be considered here. However, Spanish emigration in the 1960s has recently been depicted as modest, again as compared to other countries (Sánchez-Alonso 2010); internal migration may have been of more importance (Paluzie et al. 2009). In any case, a further, more specific time-series approach is needed in order to pinpoint the precise extent of these and other impacts in both the short and the long term.

5 Conclusions

Long-term estimates of population, beyond those only built on decennial (or roughly decennial), census-type data, have to deal with the lack of annual information. That may be the reason why, to the best of our knowledge, there are only three annually and spatially disaggregated population estimates, for Belgium, Germany and India (Ronsse and Standaert 2017; Bosker et al. 2008; Sharma 2003). By using information from official Spanish sources from 1900 to 2011, we have proposed an estimation method based on the decennial censuses, the annual numbers of births and deaths, and an estimate of the inter-census net migration. Our approach to migration numbers is elementary. However, because it avoids the use of ‘direct’ measures of historical internal and international migration, it is eligible to be applied to other countries for which these measures are imprecise, recorded under different criteria or simply not available.

This paper adds to this scant literature by providing information on the particular case of Spain, a latecomer to modern European economic growth regarding which there is a rich debate on the role played by population in this process. However, the estimation of long-term, annually and spatially disaggregated populations had not previously been addressed. We have applied this method to provinces (NUTS III) and their capital-cities (a selection of LAU II/NUTS V), which, given available data on births and deaths, are two suitable levels of spatial disaggregation for the case under study. For the cities, our estimates account for changes in boundaries. Future research may also match these new data to other available demographic, economic and social indicators.

We have made robust comparisons between ours and previous population estimates. If we aggregate our provincial estimates, national values are plausible and consistent with those of previous literature estimating the national, annual, population. However, the main contribution of this paper is to provide a new and distinctive long-term, annual dataset at two sub-levels. Here, our data provide new evidence of the spatial distribution of populations. Thus, an initial examination of the evolution of the population estimates at the two disaggregated geographical levels corroborates some previous findings obtained only using decennial (census) data: basically, the long-term spatial concentration of the total population of the country, and the recent suburbanization of the largest cities. However, the comparison of density functions reveals deviations in, particularly, medium-sized provinces and capitals; and the fitting of a Pareto distribution suggests a different evolution of the capital-size distribution.

This paper presents a new dataset, but it only touches upon its uses. To go a step further, in ongoing research we try to expand on the exploitation of the *fully dynamic* nature of annual population datasets (Bosker et al. 2008, 343). For instance, by investigating the persistence of different types of historical population shocks.

References

Alcaide-Inchausti, J., P. Alcaide-Guindo, P. Alcaide Guindo, and P. García-Fernández. 2007. *Evolución de la población española en el siglo XX, por provincias y comunidades autónomas*. Bilbao: Fundación BBVA.

Almarcha, A., J. Martín, J. Martínez, J. de Miguel, B. Oltra, J. Salcedo, and I. Sebastian. 1975. *Estadísticas básicas de España 1900-1970*. Madrid: CECA.

Ayuda, M.I., F. Collantes, and V. Pinilla. 2010a. From locational fundamentals to increasing returns: the spatial concentration of population in Spain, 1787-2000, *Journal of Geographical Systems* 12(1): 25-50.

Ayuda, M.I., F. Collantes, and V. Pinilla. 2010b. Long-run regional population disparities in Europe during modern economic growth: a case study of Spain, *The Annals of Regional Science* 44(2): 273-295.

Bairoch, P., J. Batou, and P. Chèvre. 1988. *La Population des Villes Européenes*. Geneva: Droz.

Bayona-i-Carrasco, J., F. Gil-Alonso, M. Rubiales, and I. Pujadas. 2018. New spatial mobility patterns in large Spanish cities: from the economic boom to the great recession. *Applied Spatial Analysis and Policy* 11(2): 287-312.

Bayona-i-Carrasco, J., and I. Pujadas. 2019. Las grandes áreas metropolitanas en España: del crecimiento y la expansión residencial al estancamiento poblacional. *Documents d'Anàlisi Geogràfica* (forthcoming).

Beltrán-Tapia F., A. Díez-Minguela, and J. Martínez-Galarraga. 2018. Tracing the evolution of agglomeration economies: Spain, 1860-1991. *The Journal of Economic History* 78(1): 81-117.

Bosker, E. M., S. Brakman, H. Garretsen, and M. Schramm. 2008. A century of shocks: the evolution of the German city size distribution 1925-1999. *Regional Science and Urban Economics* 38(4): 330-347.

Cabré, A., A. Moreno, and I. Pujadas (1985): Cambio migratorio y 'reconversión territorial' en España. *Revista de Investigaciones Sociológicas* 32: 43-65.

Cardesín, J.M., and J. Mirás. 2017. Historic urbanization process in Spain (1746-2013): from the fall of the American empire to the real estate bubble. *Journal of Urban History* 43(1) 33-52.

Chatel, C., M. Morillas-Torné, A. Esteve, and J. Martí-Henneberg. 2017. Patterns of population and urban growth in southwest Europe. *Journal of Urban History* 43(6): 1021-1040.

Cheshire, P. 1999. Trends in sizes and structure of urban areas. In *Handbook of Regional and Urban Economics, Vol. 3*, ed. P. Cheshire, and E.S. Mills, 1339-1373. Amsterdam: Elsevier.

De Cos, O., and P. Reques. 2005. Los cambios en los patrones territoriales de la población española (1900-2001). *Papeles de Economía Española* 104: 167-192.

De la Fuente, A. 2016. Series largas de algunos agregados demográficos regionales, 1950-2015. Estudios sobre la Economía Española 2010/14, FEDEA, Madrid.

De Vries, J. 1984. *European urbanization 1500-1800*. Cambridge, MA: Harvard University Press.

Egea, C. 2000. Fuentes alternativas al movimiento natural de la población (INE) en el estudio de la natalidad: el registro civil y el archivo parroquial. *Estudios Geográficos* 61(3): 529-535.

Espinha da Silveira, L., D. Alves, M. Painho, A.C. Costa, and A. Alcântara. The evolution of population distribution on the Iberian Peninsula: a transnational approach

(1877- 2001). *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 46(3): 157-174.

Franch-Auladell, X., A. Esteve, and J. Recaño. 2009. Los modelos de población en España, 1986-2006. *In Envejecimiento, despoblación y territorio. Un análisis sobre la población española*, ed. L. López and A. Abellán, 731-745. León: Publicaciones de la Universidad de León.

Franch-Auladell, X., and J. Recaño. 2011. La evolución de las migraciones internas en España: un estudio de la distancia de las migraciones intermunicipales en las dos últimas décadas. In *Población y espacios urbanos*, ed. I. Pujadas et al., 109-126. Barcelona: Departament de Geografía Humana de la UB y Grupo de población de la AGE.

Franch-Auladell, X., M. Morillas-Torné, and J. Martí-Henneberg. 2014. The railway network and the process of population concentration in Spain, 1900-2001. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History* 32(3): 351-379.

Gabaix, X., and Y.M. Ioannides. 2004. The evolution of city size distributions. In *Handbook of urban and regional economics, Vol. 4*, ed. J.V. Henderson, and J.F. Thisse, 2341-2378. Amsterdam: Elsevier.

Gabaix, X., and R. Ibragimov. 2011. Rank-1/2: a simple way to improve the OLS estimation of tail exponents. *Journal of Business Economics and Statistics* 29(1): 24-39.

García-Fernández, P. 1985. *Población de los actuales términos municipales, 1900-1981. Poblaciones de hecho según los censos*. Madrid: Instituto Nacional de Estadística.

Garcia-López, M.A., A. Holl, and E. Viladecans-Marsal. 2015 Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics* 85: 52-67.

Goerlich, F.J., J. Azagra, M. Mas, and P. Chorén. 2006. *La localización de la población española sobre el territorio un siglo de cambios. Un estudio basado en series homogéneas (1900-2001)*. Bilbao: Fundación BBVA.

Goerlich, F.J., and M. Mas. 2010. La distribución empírica del tamaño de las ciudades en España, 1900-2001 ¿Quién verifica la ley de Zipf?. *Revista de Economía Aplicada* 54(18): 133-159.

Goerlich, F.J., F. Ruiz, P. Chorén, and C. Albert. 2015. *Cambio en la estructura y localización de la población. Una visión de largo plazo (1842-2011)*. Bilbao: Fundación BBVA.

González-Val, R., L. Lanaspa, and F. Sanz-Gracia. 2014. New evidence on Gibrat's law for cities. *Urban Studies* 51(1): 93-115.

González-Val, R., D.A. Tirado-Fabregat, and E. Viladecans-Marsal. 2017. Market potential and city growth: Spain 1860-1960. *Cliometrica* 11(1): 31-61.

Gregory, I.N. 2002. The accuracy of areal interpolation techniques: standardising 19th and 20th century census data to allow long-term comparisons. *Computers, Environment and Urban Systems* 26(4): 293-314.

Gregory, I.N., and P.S. Ell. 2006. Error-sensitive historical GIS: Identifying areal interpolation errors in time-series data. *International Journal of Geographical Information Science* 20(2): 135-152.

Gregory, I.N., J. Martí-Henneberg, and F.J. Tapiador. 2010. Modelling long-term pan-European population change from 1870 to 2000 by using geographical information systems. *Journal of the Royal Statistical Society* 173(1): 31-50.

Lanaspa, L., F. Pueyo, and F. Sanz. 2003. The evolution of Spanish urban structure during the twentieth century. *Urban Studies* 40(3): 567-580.

López-Gay, A. 2014. Population growth and re-urbanization in Spanish inner cities: The role of internal migration and residential mobility. *Revue Quetelet/Quetelet Journal* 2(1): 67-92.

Losada, A., and P. Viso. 1997. Fontes estatísticas oficiais españolas de movementos migratorios exteriores: unha revisión crítica. *Estudios Migratorios* 4: 83-108.

Maddison, A. 2001. *The world economy. A millennial perspective*. Paris: OECD.

Maluquer de Motes, J. 2007. La incidencia de la Gran Depresión y de la Guerra Civil en la población de España (1931-1940). Una nueva interpretación. *Revista de Demografía Histórica* 25(2): 131-166.

Maluquer de Motes, J. 2008. El crecimiento moderno de la población de España de 1850 a 2001: una serie homogénea anual. *Investigaciones de Historia Económica* 4: 129-162.

Martí-Henneberg, J. 2005. Empirical evidence of regional population concentration in Europe, 1870-2000. *Population, Space and Place* 11(4): 269-281.

Martí-Henneberg, J., X. Franch-Auladell, and J. Solanas-Jiménez. 2016. The use of digital tools for spatial analysis in population geography. *Frontiers in Digital Humanities* 3: article 9.

Mitchell, B.R. 1992. *International historical statistics: Europe, 1750-1988*. New York: Palgrave.

Nicolau, R. 2005. La población española, siglos XIX-XX. In *Estadísticas históricas de España. Siglos XIX y XX*, ed. A. Carreras, and X. Tafunell, 51-90. Madrid: Fundación BBVA.

Ortega, J.A., and J. Silvestre. 2006. Las consecuencias demográficas. In *La economía de la guerra civil*, ed. P. Martín-Aceña, and E. Martínez-Ruíz, 53-105. Madrid: Marcial Pons.

Paluzie, E., J. Pons, J. Silvestre, and D. Tirado. 2009. Migrants and market potential in Spain over the twentieth century: a test of the new economic geography. *Spanish Economic Review* 11(4): 243-265.

Prados de la Escosura, L. 2017. *Spanish economic growth, 1850-2015*. Hampshire: Palgrave.

Pujadas, I., and Bayona-i-Carrasco. 2019. Municipios pequeños en España. Evolución y características de sus residentes. *Índice. Revista de Estadística y Sociedad* 74: 28-31.

Recaño, J. 2004. Migraciones internas y distribución espacial de la población española. In *Informe sobre la situación demográfica en España*, ed. J. Leal, 187-230. Madrid: Fundación Fernando Abril Martorell.

Recaño, J., and A. Domingo. 2006. Evolución de la distribución territorial y movilidad geográfica de la población extranjera en España. In *Veinte años de inmigración en España. Perspectivas jurídica y sociológica (1985-2004)*, ed. E. Aja, and J. Arango, 303-338. Barcelona: Fundación CIDOB.

Reher, D.S. 1994. Ciudades, procesos de urbanización y sistemas urbanos en la Península Ibérica, 1550-1991. In *Atlas histórico de ciudades europeas*, ed. M. Guàrdia, F.J. Monclús, and J.L. Oyón, 1-29. Barcelona: Salvat-Centre de Cultura Contemporània.

Reher, D.S., and A. Valero-Lobo. 1995. *Fuentes de información demográfica en España*. Madrid: CIS.

Ronsse, S., and S. Standaert. 2017. Combining growth and level data: An estimation of the population of Belgian municipalities between 1880 and 1970. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 50(4): 218-226.

Sánchez-Alonso, B. 2010. El final del ciclo emigratorio español: desiderátum. *Historia y Política* 23(1): 135-162.

Sharma, S. 2003. Persistence and stability in city growth. *Journal of Urban Economics* 53(2): 300-320.

Silvestre, J. 2005. Internal migrations in Spain, 1877-1930. *European Review of Economic History* 9(2): 233-265.

Tafunell, X. 2005. Urbanización y vivienda. In *Estadísticas históricas de España. Siglos XIX y XX*, ed. A. Carreras, and X. Tafunell, 455-499. Madrid: Fundación BBVA.

Van Babel, J., and D.S. Reher. 2013. The baby boom and its causes: what we know and what we need to know. *Population and Development Review* 39(2): 257-288.

Tables

Table 1. Selected characteristics of provinces and capital cities at census dates

a. Provinces

	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2011
Population	18,594,405	19,926,910	21,303,102	23,563,867	25,877,971	27,976,755	30,430,698	33,823,918	37,616,495	39,201,748	40,709,455	46,651,078
Mean	379,478	406,672	434,757	480,895	528,122	570,954	621,035	690,284	767,684	800,036	830,805	952,063
Median	337,964	349,923	358,149	365,512	407,497	422,089	451,474	464,867	489,636	520,401	536,731	640,129
Coef. var.	0.48	0.50	0.55	0.64	0.66	0.71	0.85	1.08	1.18	1.18	1.21	1.24
Min.	96,385	97,181	98,668	104,176	112,876	118,012	138,934	114,956	98,803	94,130	90,717	94,610
Max.	1,054,541	1,141,733	1,349,282	1,800,638	1,931,875	2,232,119	2,877,966	3,929,194	4,726,986	4,935,642	5,423,384	6,421,874

b. Capital cities

	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2011
Population	3,266,440	3,637,291	4,279,178	5,366,342	6,619,822	7,774,377	9,427,088	12,123,727	13,930,234	14,435,417	13,933,130	15,128,440
Mean	65,329	72,746	85,584	107,327	132,396	155,488	188,542	242,475	278,605	288,708	278,663	302,569
Median	29,395	32,657	34,447	42,183	58,808	72,108	80,826	109,873	136,040	152,490	149,646	165,491
Coef. var.	1.68	1.69	1.79	1.94	1.79	1.85	2.00	2.04	1.80	1.68	1.64	1.64
Min.	7,736	8,144	8,167	10,588	13,441	17,297	19,589	23,030	28,225	31,068	31,158	35,660
Max.	576,538	659,775	848,383	1,137,943	1,326,674	1,645,215	2,259,931	3,146,071	3,188,297	3,084,673	2,938,723	3,198,645
No. <10,000	1	1	1	0	0	0	0	0	0	0	0	0
% of provincial population	17.57	18.25	20.09	22.77	25.58	27.79	30.98	35.84	37.03	36.82	34.23	32.43

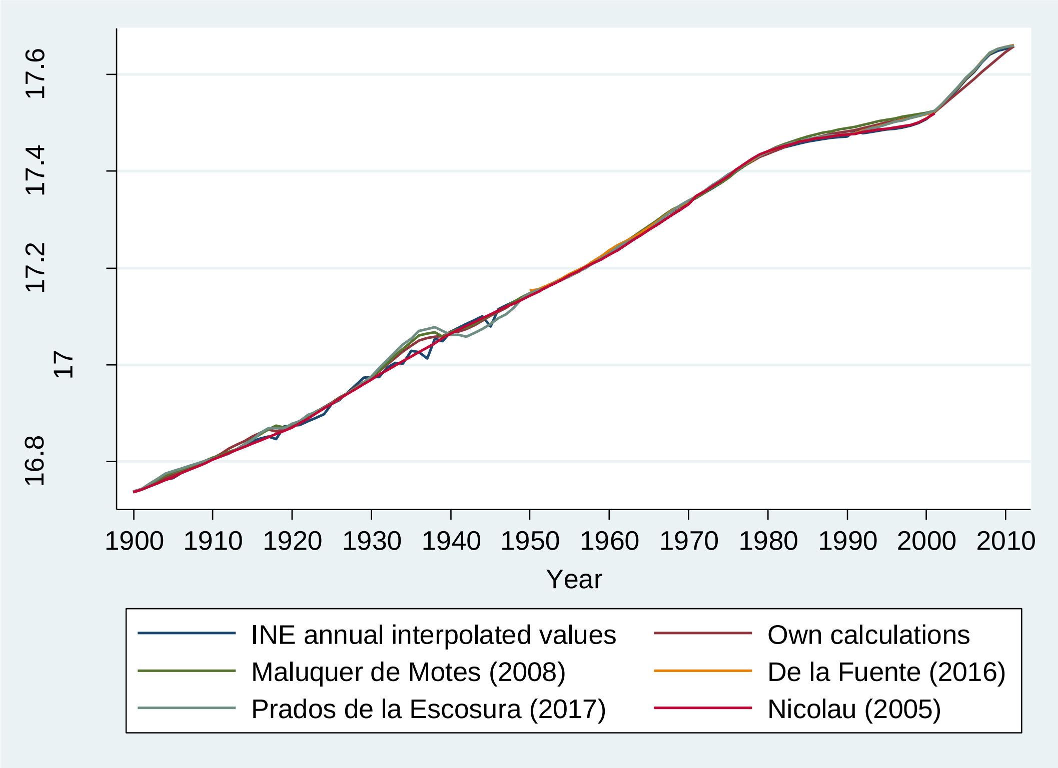
Notes: The only capital city below 10,000 inhabitants up to 1927 is Soria. City definitions in Panel *b* account for changes in boundaries as a result of annexations and segregations of municipalities.

Sources: See section 2.

Figures

Figure 1. Annual estimates of the Spanish population

a. All years (1900-2011)



b. Zoom: Pre- and post-Spanish Civil War (1936-1939) years

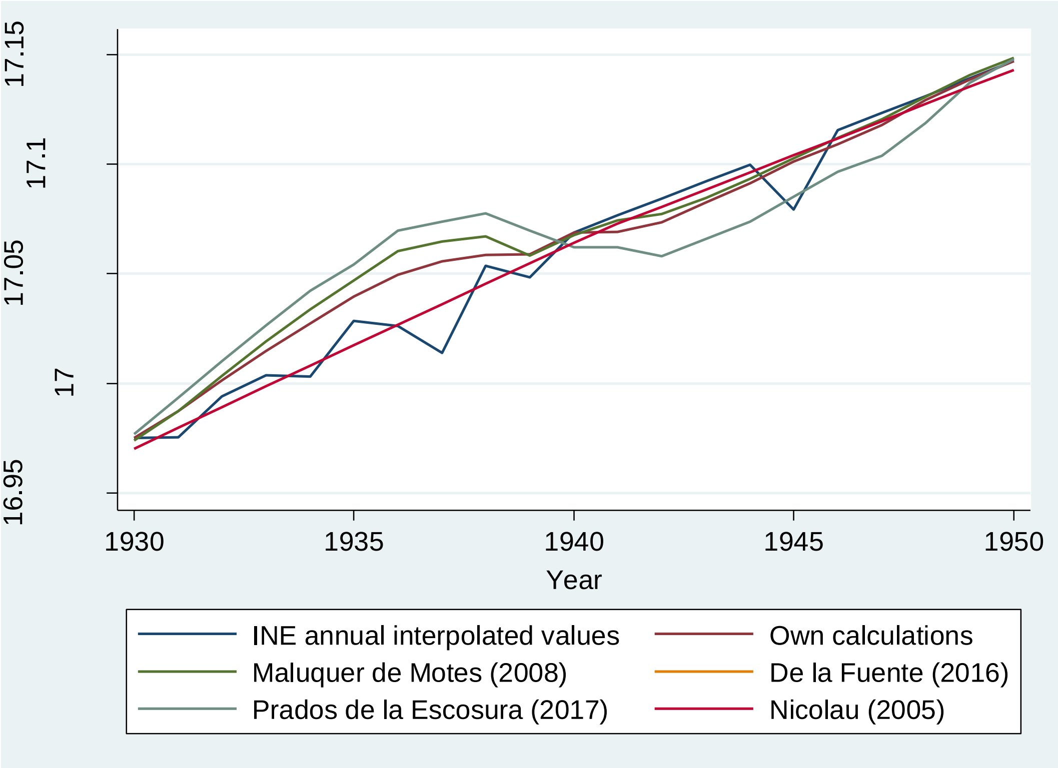
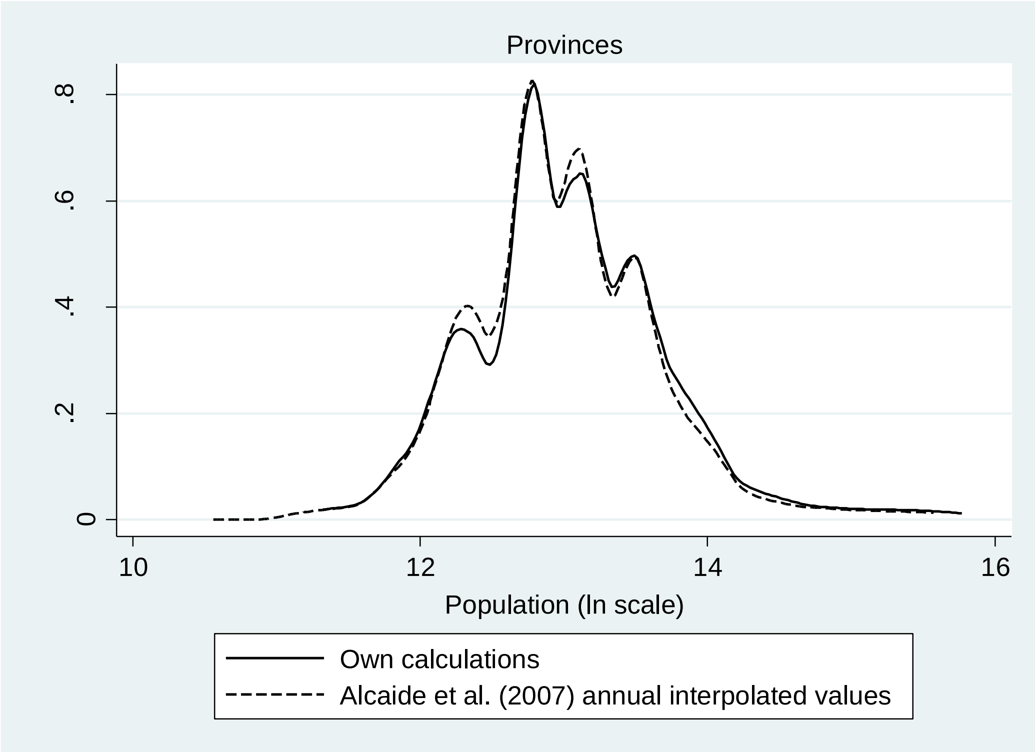


Figure 2. Empirical density functions, pool 1900-2011

a. Provinces



b. Capital cities

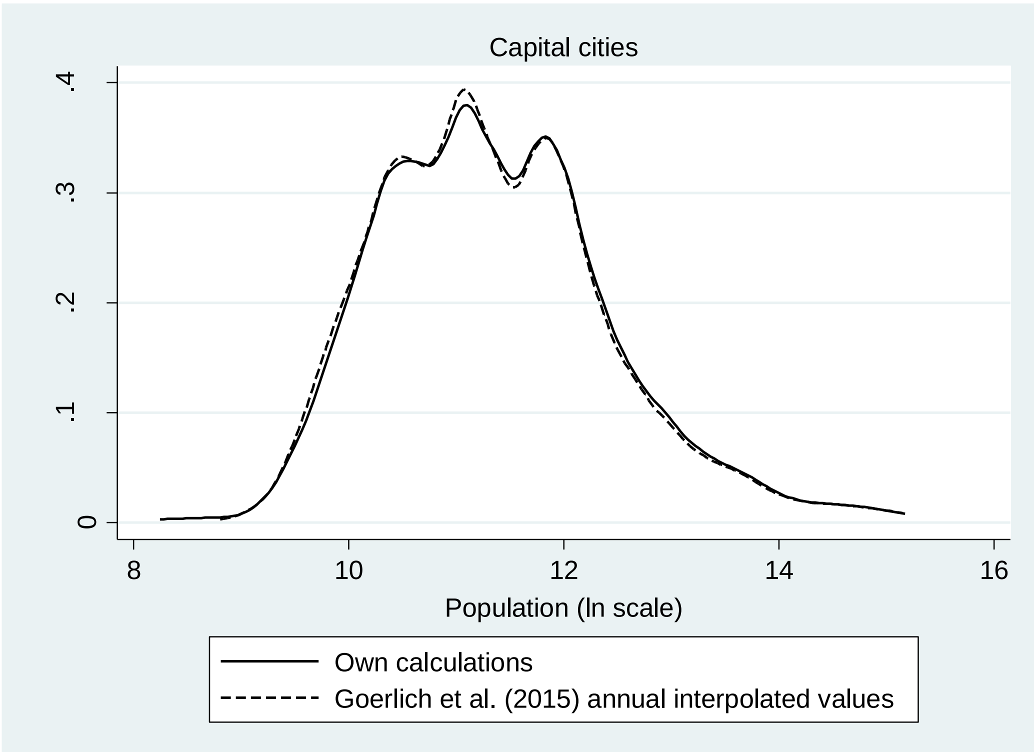
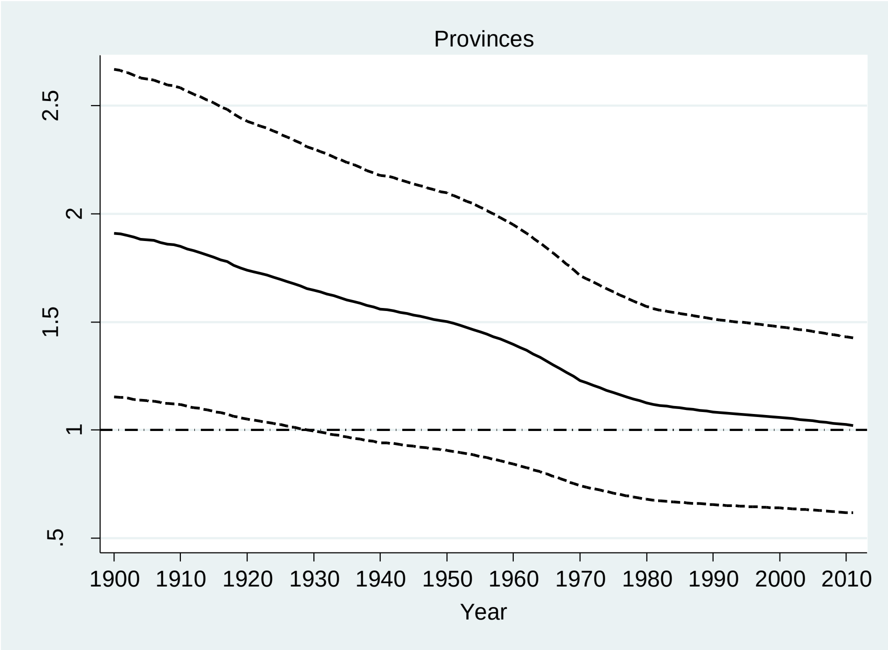
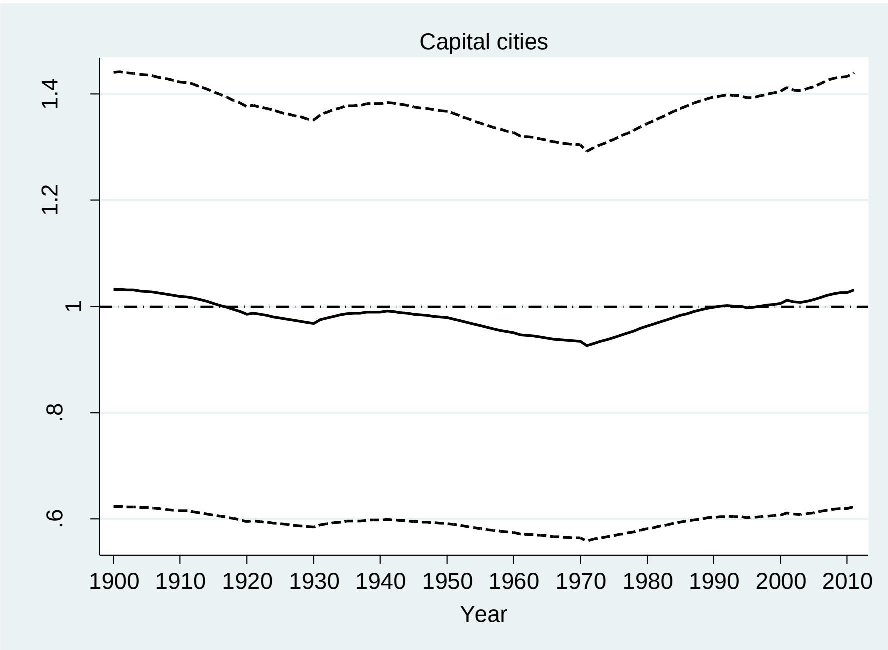


Figure 3. Evolution of the estimated Pareto exponents

a. Provinces



b. Capital cities



Notes: The Pareto exponent is estimated using Gabaix and Ibragimov's Rank-$1/2$ estimator. Dashed lines represent the standard errors calculated applying Gabaix and Ioannides's (2004) corrected standard errors: $\text{GI s.e.} = \hat{a} \cdot (2/N)^{1/2}$, where N is the sample size.

Figure 4. Gini coefficients by year