DRIVERS OF AGGLOMERATION: GEOGRAPHY VS. HISTORY*

Francisco J. Goerlich

(Corresponding author) Universidad de Valencia, Departamento de Análisis Económico, Campus de Tarongers, Avda. de Tarongers s/n, 46022-Valencia (Spain) Phone: 963 82 82 46 / Fax: 963 82 82 49 E-mail: Francisco.J.Goerlich@uv.es

and Instituto Valenciano de Investigaciones Económicas, C/ Guardia Civil, 22, Esc. 2, 1º, 46020 Valencia-Spain

Matilde Mas

Universidad de Valencia, Departamento de Análisis Económico, Campus de Tarongers, Avda. de Tarongers s/n, 46022-Valencia (Spain)

and Instituto Valenciano de Investigaciones Económicas, C/ Guardia Civil, 22, Esc. 2, 1º, 46020 Valencia-Spain. Phone: 96-319 00 50 / Fax: 96 319 00 55. E-mail: matilde.mas@ivie.es

ABSTRACT

This paper focuses on the influence of two classical drivers of population agglomeration: geography and history. Geography is identified by two co-ordinates: coastal position and altitude. The prominence of history is also captured by two characteristics: the initial size of the municipalities, and their status as the administrative centre of the area. Our reference is census population data for Spanish municipalities for the period 1900-2001. The eleven censuses have been homogenised according to the municipal structure of the 2001 Census.

Key Words: Population, Municipalities, Census, Agglomeration.

JEL Classification: J10, J11

---

* Acknowledgements: The authors wish to express their gratitude to Pilar Chorén for her help with information processing and for the financial support through the Spanish Ministry of Science and Technology/FEDER, SEC2005-02776 project and the BBVA Foundation-Ivie research programme. A statistical appendix to this study and the results mentioned but not included in the paper can be requested at Francisco.J.Goerlich@uv.es.

This paper discusses the importance of two classical drivers of population agglomeration: geographical determinants versus historical importance. The references used in this study are the population data for Spanish municipalities gathered over the 20th century. The two geographical conditioning factors used in the analysis refer to: 1. coastal or inland location; and 2. height above sea level, in other words, whether a municipality is situated in a mountainous region or on the plains. Historical importance is also examined through two variables: 1. size –measured by the number of inhabitants- at the start of the period, that is to say the initial size of the municipality; and 2. whether it has provincial capital status, and thus represents the political-administrative centre of the area.

Throughout the 20th century, the Spanish population became increasingly concentrated (Vinuesa 1997; Zoido and Arroyo 2004; De Cos and Reques 2005; Goerlich et al. 2006). The country's uneven population distribution was already evident in 1900 and this imbalance was acutely intensified by the development and industrialisation of Spanish society. Economic development during the 20th century did not create its own urban system in a vacuum, but rather it operated within a network of existing cities, formed in the 18th and 19th centuries (or perhaps much earlier). A brief look at the *Atlas de la Industrialización de España, 1750-2000* by Jordi Nadal (2003) shows that, with some relevant exceptions –many of which are linked to the mining industry, by its very nature based essentially on immobile resources- the population has remained in the same locations for centuries.

The Spanish experience is similar to those of other large European cities (De Vries 1984), although with a certain time lag, and our calculations corroborate those made at a provincial level by Ayuda et al. (2004, 2005, 2007) with a longer time span, although their use of a larger geographical unit of analysis moderates the process of physical population agglomeration to a large degree. Martí-Henneberg (2005) obtains similar results at a regional level in Europe.

The process of population location at a municipal level during the 20th century is thoroughly described in Goerlich et al. (2006) and in Goerlich and Mas (2007). In these studies, we detail the varied pace of gradual depopulation in small towns and villages (the rural environment), as compared to the growth of medium-sized cities and the burgeoning large cities (the metropolitan areas), all of which followed a marked spatial pattern. While the inland areas became increasingly depopulated, the coastal strip grew more densely populated. Madrid, the country's capital, is the most notable exception in this process of population dispersion towards the coast, although this is not in any way surprising since national capitals have always had their own demographic dynamic (Ades and Glaeser 1995).

In this paper, we set out to explore these general patterns of population agglomeration in greater depth. We aim to uncover the location patterns and the timing of these patterns from the eleven censuses conducted in the 20th century, although we are aware we have no general explanatory model for the origins of population agglomerations in certain places, and their subsequent dynamics. In contrast to the work of other scholars (Ayuda et al. 2004, 2005, 2007; Esteve and Devolver 2004 for the Spanish case; and Wheaton and Shishido 1981; De Vries 1984; Suarez-Villa 1988; Van der Woude, De Vries and Hayami 1990; Glaeser et al. 1995; Eaton and Eckstein 1997; Gabaix 1999; Overman and Ioannides 2001; Ioannides and Overman 2003, 2004, for other countries), our interest does not lie solely in urban agglomerations or large cities. Rather, our analysis in this paper includes the smaller municipalities, of limited importance in terms of population figures but significant in number and land surface area.

The paper is structured as follows. The next section reviews the information sources used and the procedures followed in creating the homogeneous series. Section 2 introduces some methodological issues. Section 3 describes two geographical characteristics of Spanish municipalities. Section 4 presents two potentially determining historical features of the current population agglomeration. Section 5 contrasts geographical and historical factors. Finally, Section 6 provides a synthesis of the main conclusions.

1. Statistical sources

The primary information source for the research is the resident (*de jure*) municipal population recorded in the eleven Spanish censuses conducted between 1900 and 2001 (the latest available census). Of all the Spanish administrative divisions, municipalities are the smallest administrative units with assigned precise boundaries and are the base for gathering information on demographic effects at different moments in time.[1] Furthermore, this information has a long historical tradition. The first census to cover all the municipalities in Spain was the 1842 *Censo de la Matrícula Catastral* (property register census). This census was conducted using a imputation procedure and, as a result, the figures it provides lack rigour and reliability. The 1857 census is

[1] There is also a further administrative unit below that of municipality, namely the *Local territorial entity smaller than a municipality* (smaller local entities), defined as a unit for the management, decentralised administration and political representation within a municipality (Law 7/1985, of 2 April, regulating the bases of local government). However, no systematic demographic statistics exist for these entities, and unlike the municipality, they do not have a delimited physical surface area.

therefore considered to be the first modern census. However, other censuses of great historical value go back as far as the 16th century.[2]

The municipal unit is clearly inadequate to provide a full picture of how the population is distributed across the territory. Nevertheless, there is a subdivision of Spanish municipalities that, although not official, is traditionally highly relevant. These subdivisions are the collective and individual population entities and their corresponding nuclei and outlying properties. These units represent the true population settlements. However, information on these units, historically compiled in local records, is neither consistent over time nor adequately systemised. Moreover, these units have no precise boundaries on which to calculate, for instance, population densities.

The Spanish municipal structure witnessed major changes during the 20th century. The number of municipalities fell considerably from 9,267 in 1900 to 8,108 in the 2001 census. Numerous modifications also occurred in the municipal structure, due to mergers, divisions and other types of alterations made to existing municipalities in periods between censuses. This is a latent problem in many of the studies on population location conducted from a municipal perspective (Zoido and Arroyo 2004; De Cos and Reques 2005), but the complexity of adjustments has meant that only one author[3], García Fernández (1985), aware of the problem, approached the task of homogenisation by taking as a reference the municipal structure of the 1981 census, and using the *de facto* population as his study variable. Unfortunately, the 2001 census did not include this variable in its analysis and centred only on the registered or usual resident population; moreover, the number of municipalities grew between 1981 and 2001 as a result of a certain locally based independentist tendency. These two reasons provide sufficient grounds for undertaking the work of García Fernández (1985) afresh, based on the municipal structure from the most recent census, 2001, and taking the registered population as the study variable.

As a result, Goerlich et al. (2006) created homogenised municipal populations starting from *two basic principles*:

1. populations are defined on the basis of a territorial criterion, the municipal boundaries, and

---

[2] For a historical view of Spanish censuses (particularly the earliest), see the excellent work of García España (1991). On the censuses used in this study, see Goerlich, Mas, Azagra and Chorén (2006, chapter 1), and the references cited therein.

[3] The *Ministerio de Fomento* (Ministry of Development) *Atlas estadístico de las áreas urbanas en España* (2000) carried out some homogenisation of municipalities for most recent years with the 1996 *Padrón* (Register) as its reference date.

2. the criterion that determines these territories is the existing municipalities recorded in the 2001 census.

Hence, this study uses information on the homogenised registered municipal populations from the censuses conducted between 1900 and 2001, where this homogeneity is based on the municipal boundaries in existence in the 2001 census, with the registered populations of the 8,108 municipalities in the 2001 census reconstructed and backdated to 1900. Goerlich et al. (2006) provide a detailed description of the homogenisation process and the resulting series. Data on municipal land area and height above sea level of the municipal capital are taken from the *Instituto Geográfico Nacional* (IGN) (National Geographical Institute) municipal database and provincial land area data come from the aggregation of the municipal land area.

2. Methodological considerations

Throughout the paper, we use two concentration indicators commonly found in the inequality literature: the Gini indices and the mean logarithmic deviation or (second) Theil index. Both indeces are described briefly below, together with the *decomposability property* of the latter, since it will be widely applied in the following sections.

If y_i is the population of municipality *i*, we can define the *Gini index*, G, as ½ the relative mean difference,

$$G = \frac{1}{2}\frac{1}{\mu n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right| \qquad (1)$$

where μ is the mean of the distribution, $\mu = \frac{1}{n}\sum_{i=1}^{n} y_i$, and *n* the number of municipalities studied. Thus, we measure the distance, in terms of population, of each municipality from each of the others, and *G* takes the average of all the distances. The Gini index is bounded between zero, if all the municipalities were of the same size, and one, in the case of maximum concentration.[4]

[4] For discrete distributions, the maximum value of *G* is given by $G = \frac{n-1}{n}$, which tends towards 1 as $n \to \infty$.

We also use another common index, with a property of particular interest, namely the (second) Theil index (1967) or mean logarithmic deviation, T^*, which can be written as

$$T^* = \frac{1}{n}\sum_{i=1}^{n}\log\frac{\mu}{y_i} = \log\frac{\mu}{\tilde{\mu}} \tag{2}$$

where $\tilde{\mu}$ is the geometric mean of the distribution, $\log\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n}\log y_i$. The mean logarithmic deviation also takes a value of zero if all the municipalities were the same size, but in contrast to G, it is not bounded above, so that a higher concentration is shown as a higher index value without it tending towards a specific value.

Note that both G and T^* are relative indices; in other words, if population growth had been proportional in all municipalities, the dispersion, measured by G or T^*, would have remained constant. If the observed concentration increases, it is precisely because population growth has not occurred proportionally; some municipalities have grown more than others, or (as in our case) while some grow, others become smaller.

The Theil index T^* presents the *additive decomposability property* explained below. Let us assume that we consider the total set of Spanish municipalities to contain the combination of H different groups, all exhaustive and mutually exclusive, denoted by the index $h = 1, 2, 3, \ldots, H$. We designate the number of municipalities from group h by n_h, and its vector of populations by $y^h = (y_1^h, y_2^h, \ldots, y_{n_h}^h)$, so that y_i^h is the population of municipality i from group h. Let $\mu = (\mu_1, \mu_2, \ldots, \mu_H)$ be the vector of the means of each group, where μ_h is the mean municipal size of group h. This notation enables us to write the overall mean, μ, as a weighted sum of the means of the different groups, where the weighting is given by the importance –measured by the number of municipalities- of each group,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{h=1}^{H}\sum_{i=1}^{n_h} y_i^h = \frac{1}{n}\sum_{h=1}^{H} n_h \mu_h = \sum_{h=1}^{H}\frac{n_h}{n}\mu_h \tag{3}$$

Now we can express the overall dispersion, measured by T^*, as the sum of two components,

(i) the existing dispersion *within* each one of the groups, or *intra-group* dispersion and

(ii) the existing dispersion *among* the different groups, *inter-group* dispersion

Moreover, the dispersion *within* the groups is obtained as a weighted average of the dispersion indices applied to each one of the groups, where the weights add up to unity and reflect the relative weight (in terms of the number of municipalities) of these groups. On the other hand, the dispersion *among* groups is simply the application of the T^* index to the mean municipality size of each group (thus the dispersion within each of the groups is not considered in this calculation).

Specifically,

$$
\begin{aligned}
T^* &= \frac{1}{n}\sum_{h=1}^{H}\sum_{i=1}^{n_h}\log\frac{\mu}{y_i^h} = \sum_{h=1}^{H}\sum_{i=1}^{n_h}\frac{1}{n}\log\frac{\mu_h}{y_i^h}.\frac{\mu}{\mu_h} \\
&= \sum_{h=1}^{H}\sum_{i=1}^{n_h}\frac{1}{n}\left(\log\frac{\mu_h}{y_i^h}+\log\frac{\mu}{\mu_h}\right) = \sum_{h=1}^{H}\frac{n_h}{n}\left[\frac{1}{n_h}\sum_{i=1}^{n_h}\log\frac{\mu_h}{y_i^h}+\log\frac{\mu}{\mu_h}\right] \\
&= \sum_{h=1}^{H}\frac{n_h}{n}\underbrace{\frac{1}{n_h}\sum_{i=1}^{n_h}\log\frac{\mu_h}{y_i^h}}_{T^{*h}}+\sum_{h=1}^{H}\frac{n_h}{n}\log\frac{\mu}{\mu_h} \qquad (4) \\
&= \underbrace{\sum_{h=1}^{H}\frac{n_h}{n}T^{*h}}_{\text{Intra-group component}}+\underbrace{\sum_{h=1}^{H}\frac{n_h}{n}\log\frac{\mu}{\mu_h}}_{\text{Inter-group component}}
\end{aligned}
$$

3. The importance of geographical location: from the inland areas to the coast and from the mountains to the plains

Spain is clearly a coastal country. Of the 47 peninsular provinces, 19 have direct sea access and 13 of their capitals are located on the coast.[5] The total length of the Spanish coastline (including the islands, Ceuta and Melilla) is around 8,000 kilometres. Despite this extension, only 460 of Spain's present 8,108 municipalities have direct sea access, a scant 5.7% representing only 7.0% of the land surface area. Additional information is provided in table 1.

At the same time, compared to its European neighbours Spain is a very mountainous country. Not only is the extent of its mountain chains considerable, but they are also relatively high. According to IGN data, 39.3% of Spain's land area lies between 600 and 1,000 metres above sea level, and 18.5% is above that height. Since

[5] The six exceptions are: Girona, Granada, Lugo, Murcia, Oviedo and Bilbao; however, note that the coastal city of Gijón in Asturias is equally or even more important than the capital Oviedo in terms of population, and that Bilbao, although not on the coast, is located on a navigable estuary. Seville is a similar case, located on the river Guadalquivir, although the province does not have its own coastline.

the population is not distributed evenly across the country, but in population nuclei, we can, for practical purposes, take the height of the municipal capital (main nucleus) as the altitude of population settlement. Table 2 shows that 3,080 municipalities are located at an altitude of between 600 and 1,000 metres above sea level (42.9%) and 1,022 at over 1,000 metres (12.6%). The table also shows that in general, the altitude of the capital is lower than the average altitude of the province.

If we take a (simple) average of municipal capital altitude as the average altitude of Spain (in terms of population settlement) the average altitude is 615 metres.[6] But, as can be seen in table 2, there are huge differences between provinces, from an average altitude of 113 metres in Vizcaya, to an average of over 1,000 in Ávila or Soria.

It is particularly interesting to contrast population concentration in terms of these two parameters, proximity to the coast and altitude. The geographical factor clearly has an impact on population agglomeration. We focus on aggregated aspects since the diversity across provinces is such that greater detail would entail an excessively long study.

[6] If, instead of the simple average, we consider the average weighted by the number of inhabitants in the municipalities, the altitude would be lower, and moreover, it would have fallen from 424.6 metres in 1900 to 304.8 metres in 2001. Consequently, the average altitude of where the population resides fell by more than 100 metres in 100 years.

TABLE 1. **Length of coastline. Coastal municipalities and their surface area**

	Province	Length of coast		Isles	Coastal municipalities			
					Number	%	Land area	%
01	Álava	-	-	-	-	-	-	-
02	Albacete	-	-	-	-	-	-	-
03	Alicante/Alacant	244	3.1%	7	19	13.5%	1,625	27.9%
04	Almería	249	3.1%	2	13	12.7%	2,148	24.5%
05	Ávila	-	-	-	-	-	-	-
06	Badajoz	-	-	-	-	-	-	-
07	Balears (Illes)	1,428	18.1%	-	37	55.2%	3,806	76.2%
08	Barcelona	161	2.0%	-	28	9.0%	480	6.2%
09	Burgos	-	-	-	-	-	-	-
10	Cáceres	-	-	-	-	-	-	-
11	Cádiz	285	3.6%	-	16	36.4%	2,389	32.1%
12	Castellón/Castelló	139	1.8%	7	16	11.9%	919	13.9%
13	Ciudad Real	-	-	-	-	-	-	-
14	Córdoba	-	-	-	-	-	-	-
15	Coruña (A)	956	12.1%	47	41	43.6%	2,726	34.3%
16	Cuenca	-	-	-	-	-	-	-
17	Girona	260	3.3%	7	22	10.0%	663	11.2%
18	Granada	81	1.0%	-	9	5.4%	448	3.5%
19	Guadalajara	-	-	-	-	-	-	-
20	Guipúzcoa	92	1.2%	2	10	11.4%	280	14.7%
21	Huelva	122	1.5%	1	9	11.4%	1,846	18.2%
22	Huesca	-	-	-	-	-	-	-
23	Jaén	-	-	-	-	-	-	-
24	León	-	-	-	-	-	-	-
25	Lleida	-	-	-	-	-	-	-
26	Rioja (La)	-	-	-	-	-	-	-
27	Lugo	144	1.8%	5	8	11.9%	642	6.5%
28	Madrid	-	-	-	-	-	-	-
29	Málaga	208	2.6%	-	14	14.0%	1,385	18.9%
30	Murcia	274	3.5%	16	8	17.8%	2,946	26.0%
31	Navarra	-	-	-	-	-	-	-
32	Ourense	-	-	-	-	-	-	-
33	Asturias	401	5.1%	2	19	24.4%	2,053	19.4%
34	Palencia	-	-	-	-	-	-	-
35	Palmas (Las)	815	10.3%	-	27	79.4%	3,798	93.4%
36	Pontevedra	398	5.0%	109	22	35.5%	928	20.6%
37	Salamanca	-	-	-	-	-	-	-
38	Sta. Cruz de Tenerife	768	9.7%	-	49	92.5%	3,139	92.8%
39	Cantabria	284	3.6%	7	26	25.5%	875	16.7%
40	Segovia	-	-	-	-	-	-	-
41	Sevilla	-	-	-	-	-	-	-
42	Soria	-	-	-	-	-	-	-
43	Tarragona	278	3.5%	-	21	11.5%	1,018	16.1%
44	Teruel	-	-	-	-	-	-	-
45	Toledo	-	-	-	-	-	-	-
46	Valencia/València	135	1.7%	-	23	8.7%	702	6.5%
47	Valladolid	-	-	-	-	-	-	-
48	Vizcaya	154	1.9%	4	21	18.9%	271	12.2%
49	Zamora	-	-	-	-	-	-	-
50	Zaragoza	-	-	-	-	-	-	-
51	Ceuta	20	0.3%	-	1	100.0%	19	100.0%
52	Melilla	9	0.1%	-	1	100.0%	13	100.0%
	España	**7,905**	**100.0%**	**216**	**460**	**5.7%**	**35,119**	**7.0%**

Note: The coastline and isles are measured in Kms. The coastline percentage is the vertical percentage of the national total. Municipal land area in Km^2. The percentage of coastal municipalities and their land area is the percentage of the provincial coastline; in the case of Spain, the percentage is of the national total**.**

'*Source*: INE, IGN and authors' own calculations**.**

TABLE 2. **Statistics on height above sea level**

Province	Average altitude Spain		Municipalities according to altitude zones: Km.²				Municipalities according to altitude zones (%)				Altitude of the provincial capital		
	Meters	100	Up to 200 m.	From 201 to 600 m	From 601 to 1,000 m	From 1,001 to 2,000 m.	Up to 200 m.	From 201 to 600 m	From 601 to 1,000 m	From 1,001 to 2,000 m.	Meters	INE code	Name
01 Álava	532	86.5	1	37	13	-	2.0%	72.5%	25.5%	-	540	01059	Vitoria-Gasteiz
02 Albacete	796	129.4	-	7	69	11	-	8.0%	79.3%	12.6%	686	02003	Albacete
03 Alicante/Alacant	299	48.6	62	59	20	-	44.0%	41.8%	14.2%	-	8	03014	Alicante/Alacant
04 Almería	561	91.2	20	34	37	11	19.6%	33.3%	36.3%	10.8%	16	04013	Almeria
05 Ávila	1,030	167.5	-	5	103	140	-	2.0%	41.5%	56.5%	**1,131**	05019	Avila
06 Badajoz	422	68.7	7	138	19	-	4.3%	84.1%	11.6%	-	186	06015	Badajoz
07 Balears (Illes)	122	19.9	58	9		-	86.6%	13.4%		-	15	07040	Palma
08 Barcelona	376	61.1	111	131	60	9	35.7%	42.1%	19.3%	2.9%	12	08019	Barcelona
09 Burgos	858	139.5	-	16	310	45	-	4.3%	83.6%	12.1%	929	09059	Burgos
10 Cáceres	467	76.0	-	185	32	2	-	84.5%	14.6%	0.9%	459	10037	Cáceres
11 Cádiz	246	40.0	25	14	5	-	56.8%	31.8%	11.4%	-	69	11012	Cádiz
12 Castellón/Castelló	478	77.8	27	62	33	13	20.0%	45.9%	24.4%	9.6%	27	12040	Castellón de la Plana/Castelló
13 Ciudad Real	690	112.2	-	17	85	-	-	16.7%	83.3%	-	628	13034	Ciudad Real
14 Córdoba	444	72.2	12	48	15	-	16.0%	64.0%	20.0%	-	106	14021	Córdoba
15 Coruña (A)	168	27.3	59	35	-	-	62.8%	37.2%	-	-	26	15030	A Coruña
16 Cuenca	925	150.3	-	-	180	58	-	-	75.6%	24.4%	999	16078	Cuenca
17 Girona	276	44.9	152	34	16	19	68.8%	15.4%	7.2%	8.6%	70	17079	Girona
18 Granada	831	135.1	5	20	98	45	3.0%	11.9%	58.3%	26.8%	683	18087	Granada
19 Guadalajara	987	160.6	-	-	157	131	-	-	54.5%	45.5%	685	19130	Guadalajara
20 Guipúzcoa	188	30.6	49	39	-	-	55.7%	44.3%	-	-	8	20069	Donostia-San Sebastián
21 Huelva	318	51.7	35	30	14	-	44.3%	38.0%	17.7%	-	30	21041	Huelva
22 Huesca	599	97.4	10	106	61	25	5.0%	52.5%	30.2%	12.4%	488	22125	Huesca
23 Jaén	651	105.9	-	39	53	5	-	40.2%	54.6%	5.2%	568	23050	Jaen
24 León	848	137.9	-	18	158	35	-	8.5%	74.9%	16.6%	838	24089	León
25 Lleida	533	86.7	14	140	54	23	6.1%	60.6%	23.4%	10.0%	182	25120	Lleida
26 Rioja (La)	680	110.5	-	76	81	17	-	43.7%	46.6%	9.8%	385	26089	Logroño
27 Lugo	402	65.3	14	42	10	1	20.9%	62.7%	14.9%	1.5%	454	27028	Lugo
28 Madrid	810	131.7	-	24	115	40	-	13.4%	64.2%	22.3%	655	28079	Madrid
29 Málaga	444	72.2	19	52	29	-	19.0%	52.0%	29.0%	-	11	29067	Málaga
30 Murcia	218	35.4	29	11	5	-	64.4%	24.4%	11.1%	-	39	30030	Murcia

TABLE 2 (continued). **Statistics on height above sea level**

	Province	Average altitude Spain		Municipalities according to altitude zones: Km.^{2}				Municipalities according to altitude zones (%)				Altitude of the provincial capital		
		Meters	100	Up to 200 m.	From 201 to 600 m	Meters	Meters	Meters	From 201 to 600 m	From 601 to 1,000 m	From 1,001 to 2,000 m.	Meters	INE code	Name
31	Navarra	503	81.8	13	192	66	1	4.8%	70.6%	24.3%	0.4%	490	31201	Pamplona/Iruña
32	Ourense	519	84.4	11	46	35	-	12.0%	50.0%	38.0%	-	139	32054	Ourense
33	Asturias	243	39.4	36	35	7	-	46.2%	44.9%	9.0%	-	232	33044	Oviedo
34	Palencia	854	138.9	-	-	173	18	-	-	90.6%	9.4%	734	34120	Palencia
35	Palmas (Las)	376	61.1	12	16	4	2	35.3%	47.1%	11.8%	5.9%	13	35016	Palmas de Gran Canaria (Las)
36	Pontevedra	170	27.6	44	15	3	-	71.0%	24.2%	4.8%	-	27	36038	Pontevedra
37	Salamanca	825	134.1	-	4	336	22	-	1.1%	92.8%	6.1%	800	37274	Salamanca
38	Sta. Cruz de Tenerife	396	64.4	13	30	8	2	24.5%	56.6%	15.1%	3.8%	*5*	38038	Santa Cruz de Tenerife
39	Cantabria	236	38.4	66	20	16	-	64.7%	19.6%	15.7%	-	11	39075	Santander
40	Segovia	964	156.7	-	-	132	77	-	-	63.2%	36.8%	1,002	40194	Segovia
41	Sevilla	195	31.8	68	34	3	-	64.8%	32.4%	2.9%	-	11	41091	Sevilla
42	Soria	**1,045**	169.9	-	-	55	128	-	-	30.1%	69.9%	1,063	42173	Soria
43	Tarragona	274	44.5	81	84	18	-	44.3%	45.9%	9.8%	-	69	43148	Tarragona
44	Teruel	991	161.1	-	32	83	121	-	13.6%	35.2%	51.3%	912	44216	Teruel
45	Toledo	583	94.9	-	114	90	-	-	55.9%	44.1%	-	529	45168	Toledo
46	Valencia/València	214	34.7	168	69	26	2	63.4%	26.0%	9.8%	0.8%	13	46250	Valencia
47	Valladolid	766	124.6	-	-	225	-	-	-	100.0%	-	698	47186	Valladolid
48	Vizcaya	*113*	18.3	97	14	-	-	87.4%	12.6%	-	-	6	48020	Bilbao
49	Zamora	759	123.4	-	-	241	7	-	-	97.2%	2.8%	649	49275	Zamora
50	Zaragoza	578	93.9	24	129	127	12	8.2%	44.2%	43.5%	4.1%	199	50297	Zaragoza
51	Ceuta	40	6.5	1	-	-	-	100.0%	-	-	-	40	51001	Ceuta
52	Melilla	15	2.4	1	-	-	-	100.0%	-	-	-	15	52001	Melilla
	España	**615**	**100.0**	**1.344**	**2.262**	**3.480**	**1.022**	**16.6%**	**27.9%**	**42.9%**	**12.6%**	**655**	**28079**	**Madrid**

Note: The mean altitude is obtained as the simple average of the altitudes of each municipal capital.
Municipality distribution according to altitude zones is based on the altitude of the municipal capital.
The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.
For Spain, we take data for the capital, Madrid, as Provincial Capital Altitude data.

Source: INE, IGN and authors' own calculations.

From the inland areas to the coast: coastal Spain

Our very restricted definition of coastal considers only those municipalities with direct access to the sea. Table 3 presents some statistics to illustrate the gradual concentration of the population on this very narrow strip of land. This definition of coastal ensures that our results will not be biased towards a higher concentration.

Hence, while the population multiplied by a factor of 2.2 during the 20th century, the population living "right on the coast" multiplied by a factor of 3.3 and the inland population by 1.9. The proportion of the population living in coastal municipalities rose by over 10 percentage points throughout the century. However, note that the level of concentration on the coastline was already quite high in 1900, although at that time, 460 municipalities represented only 7.0% of the entire Spanish land area (Balearics and Canary Islands).

In the case of the coast, the concentration indices show an increasing trend until the beginning of the eighties, when a slight trend towards dispersion began. Since these indices refer only to coastal municipalities, what they indicate is a certain tendency towards dispersion within the coastal strip itself. Thus, in the last quarter of the 20th century, residential destinations on the coast appear to diversify (all of which occurred within a context of high saturation). In fact, there were many coastal municipalities in the first decades of the century, fishing villages which at that time had no economic future and lost significant numbers of residents, but that in the second half of the century became tourist nuclei of the first order, with large demographic gains (Goerlich et al. 2006, chapter 4). What can be clearly seen is that during the second half of the 20th century, the distribution of municipal sizes is less concentrated on the coast than in Spain as a whole, although the average size is substantially higher (some five times higher). In the case of inland municipalities, dispersion has always been lower than that of the country as a whole.

Table 4 presents the decomposition of the Theil index for the coastal – inland division. It is interesting to note how both components, *inter*- and *intra*-groups, grow continuously over the whole period. Thus, on average, the contrast between the coast and the inland areas gradually becomes sharper. These two groups of municipalities show little homogeneity and marked internal differences.

TABLE 3. **Coastal concentration of the population. Spain. 1900-2001**

Spain	Zone	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
Population	Coast	*3,954,429*	4,372,354	4,828,658	5,473,142	6,324,963	6,991,715	7,953,848	9,640,136	11,441,430	12,109,295	**12,934,862**
	Inland	*14,876,220*	15,987,952	17,184,005	18,553,429	20,061,891	21,180,553	22,823,087	24,401,346	26,240,925	26,762,973	**27,912,509**
	Spain	*18,830,649*	20,360,306	22,012,663	24,026,571	26,386,854	28,172,268	30,776,935	34,041,482	37,682,355	38,872,268	**40,847,371**
% of total population	Coast	*21.0%*	21.5%	21.9%	22.8%	24.0%	24.8%	25.8%	28.3%	30.4%	31.2%	**31,7%**
	Inland	**79.0%**	78.5%	78.1%	77.2%	76.0%	75.2%	74.2%	71.7%	69.6%	68.8%	*68,3%*
Mean municipal size	Coast	*8,597*	9,505	10,497	11,898	13,750	15,199	17,291	20,957	24,873	26,325	**28,119**
	Inland	*1,945*	2,090	2,247	2,426	2,623	2,769	2,984	3,191	3,431	3,499	**3,650**
	Spain	*2,322*	2,511	2,715	2,963	3,254	3,475	3,796	4,199	4,648	4,794	**5,038**
Spain = 100	Coast	*370.1*	378.5	386.6	401.5	422.5	437.4	455.5	499.1	535.2	549.1	**558.2**
	Inland	**83.8**	83.2	82.8	81.9	80.6	79.7	78.6	76.0	73.8	73.0	*72.4*
Gini index	Coast	*0.653*	0.654	0.666	0.680	0.701	0.720	0.730	0.747	**0.755**	0.744	0.722
	Inland	*0.599*	0.604	0.623	0.640	0.662	0.680	0.715	0.781	0.826	0.840	**0.847**
	Spain	*0.637*	0.643	0.660	0.678	0.701	0.719	0.750	0.808	0.846	0.857	**0.862**
Theil index	Coast	*0.814*	0.816	0.848	0.888	0.958	1.026	1.072	1.148	**1.205**	1.165	1.088
	Inland	*0.652*	0.665	0.714	0.763	0.829	0.886	1.012	1.316	1.608	1.724	**1.813**
	Spain	*0.754*	0.771	0.823	0.880	0.958	1.024	1.156	1.474	1.777	1.893	**1.979**

Note: The coast is represented by municipalities with direct access to the sea, a total of 460, representing 7.0% of the total land area.
The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.

Source: INE and authors' own calculations.

TABLE 4. **Decomposition of the Theil Index (Mean logarithmic deviation). Coastal-inland classification. 1900 – 2001**

Component	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
***Inter*-groups (External)**	*0.093*	0.097	0.102	0.110	0.122	0.130	0.141	0.168	0.191	0.200	**0.207**
%	12.3%	12.6%	12.4%	12.5%	12.7%	**12.7%**	12.2%	11.4%	10.8%	10.6%	*10.4%*
***Intra*-groups (Internal)**	*0.661*	0.673	0.721	0.770	0.837	0.894	1.015	1.307	1.585	1.693	**1.772**
%	87.7%	87.4%	87.6%	87.5%	87.3%	*87.3%*	87.8%	88.6%	89.2%	89.4%	**89.6%**
Total	*0.754*	0.771	0.823	0.880	0.958	1.024	1.156	1.474	1.777	1.893	**1.979**
%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note: The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.

Source: INE and authors' own calculations.

The process of population concentration on the coast has tended to generate a more homogeneous coastal area in a certain sense,[7] as opposed to a heterogeneous inland area with a few large nuclei (essentially Madrid, its surroundings and provincial capitals) and many less consequential municipalities scattered across the rest of inland Spain.

Thus, while Spain was already a coastal country in 1900, it is now much more so at the beginning of the 21st century. In addition, enormous internal changes have taken place in the size structure along the coast. The Spanish case is not unique however; the US also shows similar levels of coastal population concentration (Rappaport and Sachs 2003), and although the historical processes that have led to this situation are very different, the results appear to be quite similar.

- **From the mountains to the plains: mountainous Spain**

We define 4 altitude zones: up to 200 metres (the *plains*, which includes much of the coastal strip, but also the "second line" of coastal development and the shores of many important rivers such as the Ebro or the Guadalquivir); from 200 to 600 metres; from 600 to 1,000 metres; and above 1,000 metres (the *mountains*). Table 5 illustrates the gradual movement of the population from the mountains to the plains.

Population distribution tends to be polarised between the two extremes. On one hand, the zone covering territories up to 200 metres above sea level accommodates a growing percentage of the population, exceeding 50% from 1981 onwards (despite covering a limited land area of 16.6%); on the other, mountain settlements (above 1,000 metres) start off with a very scant population in 1900 (5.0%, representing somewhat less than one million inhabitants), but after experiencing a sharp decline beginning in 1950 (Collantes 2004, 2005) they fall to a current minimum both in relative (1.5% of the population) and absolute terms (below 600 thousand inhabitants), despite the fact that three provincial capitals, Ávila, Segovia and Soria, are located over 1,000 metres above sea level.

The two intermediate zones, covering 200 to 1,000 metres, begin the period with very similar population figures and, although they gain numbers in absolute terms, they lose to the plains in relative terms.

[7] We might call this "homogenously concentrated". In 2001, 224 of the 460 coastal municipalities had over 10,000 inhabitants and were home to 91.9% of all the population residing on the coast. In 1900, the situation was of 67 municipalities accommodating 63.4% of the coastal population.

TABLE 5. **Population distribution according to altitude.** **Spain. 1900 – 2001**

Spain	Altitude zone	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
Population	Below 200 m.	*6,640,844*	7,292,811	8,107,283	9,109,359	10,391,336	11,481,940	13,335,042	16,477,811	19,471,384	20,417,731	**21,566,916**
	Between 200 and 600 m	*5,655,262*	6,101,241	6,477,095	6,860,610	7,289,642	7,545,305	7,827,727	7,675,204	7,969,112	8,133,878	**8,568,007**
	Between 600 and 1,000 m	*5,588,569*	5,981,234	6,428,681	7,034,791	7,682,780	8,108,491	8,625,042	9,116,379	9,603,782	9,720,458	**10,118,650**
	Over 1,000 m	945,974	985,020	999,604	1,021,811	1,023,096	**1,036,532**	989,124	772,088	638,077	600,201	*593,798*
% of total population	Below 200 m.	*35.3%*	35.8%	36.8%	37.9%	39.4%	40.8%	43.3%	48.4%	51.7%	52.5%	**52.8%**
	Between 200 and 600 m	**30.0%**	30.0%	29.4%	28.6%	27.6%	26.8%	25.4%	22.5%	21.1%	*20.9%*	21.0%
	Between 600 and 1,000 m	**29.7%**	29.4%	29.2%	29.3%	29.1%	28.8%	28.0%	26.8%	25.5%	25.0%	*24.8%*
	Over 1,000 m	**5.0%**	4.8%	4.5%	4.3%	3.9%	3.7%	3.2%	2.3%	1.7%	1.5%	*1.5%*
Mean municipal size	Below 200 m.	*4,941*	5,426	6,032	6,778	7,732	8,543	9,922	12,260	14,488	15,192	**16,047**
	Between 200 and 600 m	*2,500*	2,697	2,863	3,033	3,223	3,336	3,461	3,393	3,523	3,596	**3,788**
	Between 600 and 1,000 m	*1,606*	1,719	1,847	2,021	2,208	2,330	2,478	2,620	2,760	2,793	**2,908**
	Over 1,000 m	926	964	978	1,000	1,001	**1,014**	968	755	624	587	*581*
	Spain	2,322	2,511	2,715	2,963	3,254	3,475	*396*	4,199	4,648	4,794	**5,038**
Spain = 100	Below 200 m.	*212.8*	216.1	222.2	228.7	237.6	245.9	261.4	292.0	311.7	316.9	**318.5**
	Between 200 and 600 m	**107.6**	107.4	105.5	102.4	99.0	96.0	91.2	80.8	75.8	*75.0*	75.2
	Between 600 and 1,000 m	**69.1**	68.4	68.0	68.2	67.8	67.1	65.3	62.4	59.4	58.3	*57.7*
	Over 1,000 m	**39.9**	38.4	36.0	33.7	30.8	29.2	25.5	18.0	13.4	12.2	*11.5*
Gini index	Below 200 m.	*0.671*	0.672	0.685	0.698	0.720	0.739	0.753	0.776	**0.791**	0.788	0.777
	Between 200 and 600 m	*0.556*	0.559	0.566	0.576	0.595	0.608	0.633	0.676	0.722	0.740	**0.753**
	Between 600 and 1,000 m	*0.595*	0.601	0.626	0.649	0.673	0.690	0.730	0.808	0.856	0.872	**0.884**
	Over 1,000 m	*0.486*	0.487	0.499	0.513	0.526	0.544	0.570	0.641	0.710	0.739	**0.765**
	Spain	*0.637*	0.643	0.660	0.678	0.701	0.719	0.750	0.808	0.846	0.857	0.862
Theil index	Below 200 m.	*0.879*	0.889	0.930	0.978	1.058	1.136	1.208	1.338	1.452	**1.456**	1.414
	Between 200 and 600 m	*0.567*	0.574	0.590	0.617	0.664	0.700	0.772	0.913	1.091	1.170	**1.230**
	Between 600 and 1,000 m	*0.633*	0.646	0.710	0.774	0.847	0.901	1.041	1.402	1.725	1.866	**1.988**
	Over 1,000 m	*0.396*	0.399	0.420	0.446	0.475	0.513	0.569	0.761	0.995	1.101	**1.208**
	Spain	*0.754*	0.771	0.823	0.880	0.958	1.024	1.156	1.474	1.777	1.893	1.979

Note: The four altitude zones are defined by the altitude of the corresponding municipal capital.
The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.

Source: INE and authors' own calculations.

On average, these differences tend to become accentuated, as shown by the mean municipality sizes. In fact, from 1940 onwards, the only municipalities with mean sizes above the national average were those located in the plains. In addition, a remarkable uniformity can be observed: the higher above sea level the municipality, the lower its mean size, a tendency that remains constant across all the periods analysed.

The last section of table 5 presents the inequality indices for each of the four altitude zones. There is a clear tendency towards concentration within each zone; symptoms of stability only emerge in the last decade for the plains' municipalities, although much less perceivably than the picture given in table 3 for the coast. At the beginning of the 20^{th} century, the population concentration appears to be lower than the national average in all zones except the plains, below 200 metres. Over time, this situation changes such that by the end of the century, the plains show a lower concentration than the national average. This result is similar to what we have observed on the coast, the process of displacement towards the plains has tended to generate a more homogeneous altitude zone. The opposite process can be observed in the 600 to 1,000 metre zone, which appears to be where population concentration is most acute.

The decomposition of the Theil index is presented in table 6. In addition to the generalised growth of both components, the *inter*-group component emerges as having a greater relative importance, and also shows a slight tendency to increase. The message is therefore that classification of municipalities by altitude zones shows a lower degree of contrast than the coastal-inland classification, all, as before, within the context of a high degree of saturation in the lowest altitude zone.

4. The relevance of history

The importance of history as a conditioning factor in future evolution has been highlighted by numerous authors. For instance, Krugman (1991, chapter 2) puts forward some very compelling examples. In the present paper, we identify two potentially conditioning factors in agglomeration processes: 1. the selection, at a certain moment in time, of a municipality as the seat of political/administrative power by designating it a territorial capital, and 2. the municipality's capacity for agglomeration in the past, for reasons that are not generally explained.

TABLE 6. **Decomposition of the Theil index (Mean logarithmic deviation). Classification by altitude zones. 1900 – 2001**

Component	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
Inter-groups (External)	*0.129*	0,136	0,147	0,157	0,174	0,189	0,222	0,300	0,366	0,386	**0,396**
%	*17.1%*	17,6%	17,8%	17,9%	18,2%	18,5%	19,2%	20,4%	**20,6%**	20,4%	20,0%
Intra-groups (Internal)	*0.625*	0,635	0,676	0,723	0,784	0,835	0,934	1,174	1,411	1,507	**1,583**
%	**82.9%**	82,4%	82,2%	82,1%	81,8%	81,5%	80,8%	79,6%	*79,4%*	79,6%	80,0%
Total	*0.754*	0,771	0,823	0,880	0,958	1,024	1,156	1,474	1,777	1,893	**1,979**
%	100.0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Note: The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.

Source: INE and authors' own calculations.

The Spanish provinces were created by the Royal Decree of 30 November 1833. This project, led by Javier de Burgos, created a decentralised state divided into 49 provinces. The provinces were known by the name of their capital city (with the exceptions of the provinces of Navarra, Álava, Guipúzcoa and Vizcaya whose capitals are in Pamplona, Vitoria, San Sebastián and Bilbao, respectively). This project was practically the same as that of 1822, formulated following the Riego coup during the *Trienio Liberal* or three years of Liberal rule (1820-1823). The most substantial changes were the abrogation of the provinces of Calatayud, Villafranca and Játiva, and name changes to others, following changes to their capitals. Some provinces appear for the first time in 1833, such as Almería (separated from the Kingdom of Granada), Huelva (from the Kingdom of Seville), or Logroño, and others appear with new names such as Murcia or the Basque provinces.

The provincial division proposed by Javier de Burgos was consolidated and continues today, with only a few exceptions of interest. The most noteworthy is the division in 1927 of the province of Santa Cruz de Tenerife into the two provinces it has today, Las Palmas and Santa Cruz. The provincial capitals were immediately endowed with basic government institutions and political heads were created at the same time. Consequently, the present provincial capitals go back to at least the first third of the 19th century, and were selected as such at that time because they were the municipalities with the highest number of inhabitants in the province. In only seven provinces has the capital not been the largest municipality during censuses carried out in the 20th century. The most notable is Pontevedra, whose capital, Pontevedra, has always fallen behind the municipality of Vigo in terms of population size. The other cases are: Cádiz, whose largest municipality has been Jerez de la Frontera since 1950; Ciudad Real, where the largest municipality was Valdepeñas between 1900 and 1930, and Puertollano between 1950 and 1981; Jaén, whose largest municipality was Linares between 1900 and 1930; Asturias, where Gijón was the largest municipality in various years (1910, 1930, 1940, 1950, 1970, 1981, 1991 and 2001); Tarragona, where Reus was the largest municipality in 1910 and 1920; and finally Toledo, where the capital lost ground to Talavera de la Reina between 1970 and 2001.

Table 7 provides the same information as above, but for the division between provincial capitals and non-capitals. The population in the capitals more than quadrupled during the period analysed, which in turn has led the percentage of the population residing in provincial capitals to double, from 17.3% in 1900 to 34.1% in 2001. The concentration indices reveal an interesting pattern. In relative terms, the concentration in the capitals sub-set is fairly stable. A slight tendency towards concentration persists until 1970, but then indices fall to levels slightly below those seen at the beginning of the 20^{th} century. In contrast, the concentration in the non-capital municipalities group increases throughout the whole period. In both cases, the concentration in the two groups is always lower than the overall concentration, which is a consequence of the enormous and increasing discrepancies between the mean sizes of the municipalities in the two groups.

Table 8 shows the decomposition of the Theil index. The *inter*-group component shows an increasing trend until the seventies, followed by certain stability. Since this component is the index applied to the mean values of the two groups, its evolution is due to the growth of the large non-capital cities. However the *intra*-group component reveals a continued increasing trend throughout the entire period, practically in line with the evolution of the overall index[8].

- **<u>The importance of initial conditions</u>**

The story of Catherine Evans, told by Krugman (1991), perfectly describes what for him and many other authors illustrates the importance of initial conditions. In 1895 Catherine Evans was an adolescent living in the small city of Dalton in the state of Georgia. When Catherine made a rug as a wedding present, this apparently trivial occurrence became the embryo of one of the most important carpet and rug manufacturing centres in the United States after the Second World War. This story, together with others he relates, leads Krugman to conclude “that when one tries to understand the reasons for that localization, one finds that it can be traced back to some seemingly trivial historical accident” (Krugman 1991, p. 35).

Only through the study of each individual case can we attempt to identify this “seemingly trivial historical accident”. To gain a more aggregate picture, the importance of history to the subsequent evolution of an activity and, hence, the settlement of the population in a certain location, can be approached from various perspectives. In this paper, we focus on two approaches. The first is the calculation of a simple correlation coefficient between the situation in 1900 and that in 2001, either in absolute population figures or in rankings. Table 9 shows that for all the municipalities considered, this correlation is extremely high, 0.93 and 0.80 in the case of levels and rankings respectively, even in this case which spans a time interval of over 100 years. From the aggregate point of view, persistence is therefore extremely marked.

[8] It should be remembered that the *intra*-groups component in (4) is a weighted mean of the inequality indices of the different groups, and consequently, it is dominated by the non-capital group index.

TABLE 7. **Population concentration in provincial capitals. Spain. 1900 – 2001**

Spain	Zone	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
Population	Capitals	*3,256,794*	3,597,921	4,313,125	5,219,615	6,492,167	7,627,904	9,294,128	12,009,442	13,740,930	**13,940,513**	13,920,609
	Non-capitals	*15,573,855*	16,762,385	17,699,538	18,806,956	19,894,687	20,544,364	21,482,807	22,032,040	23,941,425	24,931,755	**26,926,762**
% of total population	Capitals	*17.3%*	17.7%	19.6%	21.7%	24.6%	27.1%	30.2%	35.3%	**36.5%**	35.9%	34.1%
	Non-capitals	**82.7%**	82.3%	80.4%	78.3%	75.4%	72.9%	69.8%	64.7%	*63.5%*	64.1%	65.9%
Mean municipal size	Capitals	*62,631*	69,191	82,945	100,377	124,849	146,690	178,733	230,951	264,249	**268,087**	267,704
	Non-capitals	*1,933*	2,081	2,197	2,335	2,470	2,550	2,667	2,735	2,972	3,095	**3,342**
	Spain	*2,322*	2,511	2,715	2,963	3,254	3,475	3,796	4,199	4,648	4,794	**5,038**
Spain = 100	Capitals	*2,696.7*	2,755.4	3,055.1	3,387.3	3,836.3	4,221.8	4,708.6	5,500.8	**5,685.8**	5,591.8	5,313.8
	Non-capitals	**83.2**	82.9	80.9	78.8	75.9	73.4	70.3	65.1	*63.9*	64.6	66.3
Gini index	Capitals	0.581	0.565	0.582	0.599	0.601	0.598	0.613	**0.623**	0.594	0.573	*0.558*
	Non-capitals	*0.572*	0.577	0.588	0.599	0.615	0.626	0.654	0.715	0.771	0.789	**0.802**
	Spain	*0.637*	0.643	0.660	0.678	0.701	0.719	0.750	0.808	0.846	0.857	**0.862**
Theil index	Capitals	0.584	0.551	0.592	0.630	0.636	0.630	0.668	**0.697**	0.626	0.577	*0.541*
	Non-capitals	*0.593*	0.606	0.635	0.666	0.708	0.742	0.831	1.077	1.363	1.490	**1.604**
	Spain	*0.754*	0.771	0.823	0.880	0.958	1.024	1.156	1.474	1.777	1.893	**1.979**

Note: Capitals are the provincial capitals including Ceuta and Melilla, a total of 52 municipalities representing 3.1% of the total national land surface area.
The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.

Source: INE and authors' own calculations.

TABLE 8. **Decomposition of the Theil Index (Mean logarithmic deviation). Classification by capitals-non-capitals. 1900 – 2001**

Component	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
Inter-groups (External)	*0.161*	0.166	0.188	0.214	0.251	0.283	0.326	0.400	**0.418**	0.409	0.382
%	21.4%	21.5%	22.9%	24.4%	26.2%	27.7%	**28.2%**	27.1%	23.5%	21.6%	*19.3%*
Intra-groups (Internal)	*0.593*	0.605	0.635	0.666	0.707	0.741	0.830	1.074	1.358	1.484	**1.597**
%	78.6%	78.5%	77.1%	75.6%	73.8%	72.3%	*71.8%*	72.9%	76.5%	78.4%	**80.7%**
Total	*0.754*	0.771	0.823	0.880	0.958	1.024	1.156	1.474	1.777	1.893	**1.979**
%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note:.The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.

Source: INE and authors' own calculations.

TABLE 9. **Correlations between the municipal population in 1900 and 2001**

	Province	Levels	Rankings
01	Álava	0.987	0.775
02	Albacete	0.800	0.803
03	Alicante/Alacant	0.872	0.845
04	Almería	0.786	0.755
05	Ávila	0.883	0.764
06	Badajoz	0.805	0.809
07	Balears (Illes)	0.960	0.805
08	Barcelona	0.970	0.723
09	Burgos	0.896	0.799
10	Cáceres	0.678	0.681
11	Cádiz	0.905	0.856
12	Castellón/Castelló	0.902	0.807
13	Ciudad Real	0.779	0.858
14	Córdoba	0.893	**0.887**
15	Coruña (A)	0.932	0.822
16	Cuenca	0.822	0.842
17	Girona	0.885	0.771
18	Granada	0.951	0.605
19	Guadalajara	0.654	0.698
20	Guipúzcoa	0.961	0.852
21	Huelva	0.758	0.799
22	Huesca	0.826	0.765
23	Jaén	0.835	0.887
24	León	0.815	0.684
25	Lleida	0.901	0.686
26	Rioja (La)	0.844	0.859
27	Lugo	0.738	0.678
28	Madrid	**0.990**	0.547
29	Málaga	0.961	0.872
30	Murcia	0.907	0.785
31	Navarra	0.890	0.735
32	Ourense	0.765	0.746
33	Asturias	0.789	0.866
34	Palencia	0.807	0.825
35	Palmas (Las)	0.973	0.555
36	Pontevedra	0.835	0.693
37	Salamanca	0.897	0.653
38	Sta. Cruz de Tenerife	0.917	0.530
39	Cantabria	0.959	0.652
40	Segovia	0.910	0.723
41	Sevilla	0.970	*0.468*
42	Soria	*0.591*	0.804
43	Tarragona	0.905	0.710
44	Teruel	0.831	0.807
45	Toledo	0.800	0.744
46	Valencia/València	0.986	0.786
47	Valladolid	0.986	0.794
48	Vizcaya	0.951	0.678
49	Zamora	0.833	0.758
50	Zaragoza	0.986	0.792
51	Ceuta	-	-
52	Melilla	-	-
	España	0.931	0.804

Note: The *minimum* value of each province is shown in *italics*.
The **maximum** value of each province is shown in **bold**.

Source: INE and authors' own calculations.

The correlations at a provincial level reveal that persistence is generalised. In terms of levels, correlation coefficients below 0.7 only appear in three provinces, Cáceres, Guadalajara and Soria. In terms of rankings, only four provinces present correlation coefficients below 0.6, Madrid, Las Palmas, Santa Cruz de Tenerife and Seville, with a minimum coefficient of 0.47.[9] Note that, with the exception of the two provinces in the Canary archipelago, none of these provinces is coastal.

An alternative way of examining these results is by means of an equation that relates the initial population with the subsequent growth rate. This is the β-convergence (unconditional) equation from the economy of growth literature (Barro and Sala-i-Martín 1992, 1995). A negative relation between initial size and subsequent growth indicates convergence in municipality sizes, in that the smallest municipalities tend to grow more than the largest municipalities. In contrast, a positive relation indicates divergence; the municipalities that started out large tend to grow more, on average, than the smaller ones, and consequently, we can observe a tendency towards population concentration in a limited number of localities, those that, broadly speaking, had larger populations at the beginning of the period.

Using logarithms for the entire period, we obtain,

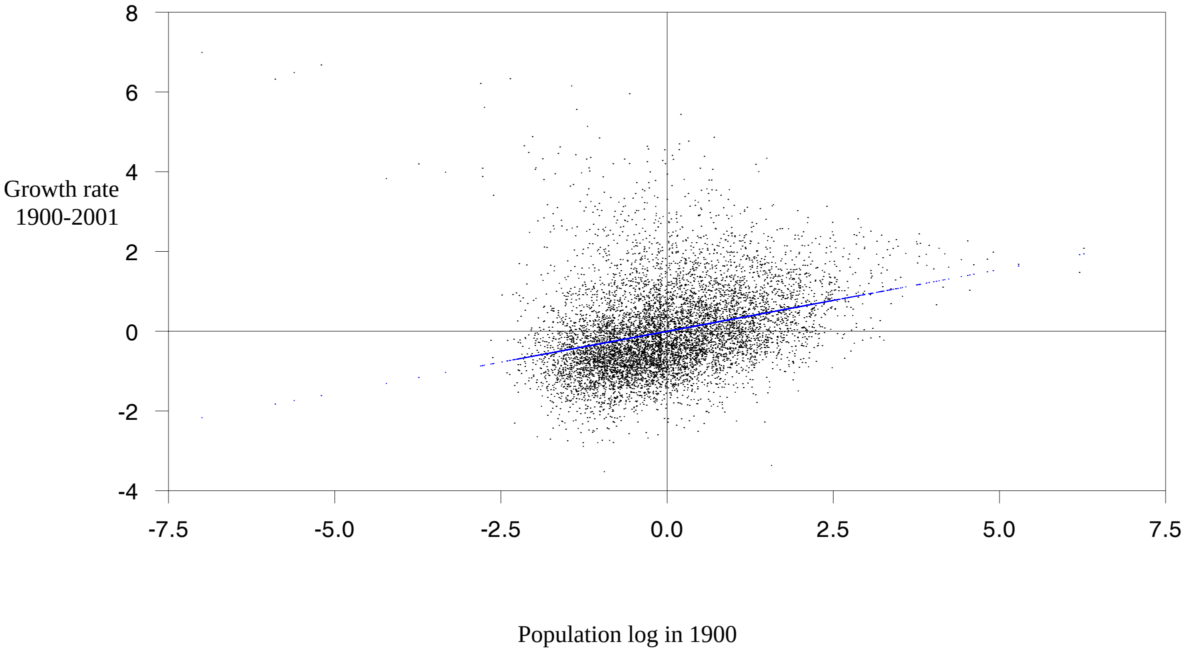
$$\log(Pob_{2001}) - \log(Pob_{1900}) = \hat{\alpha} + \underset{(0,0159)}{0,3098}\log(Pob_{1900}) + \hat{u} \qquad n = 8.108 \quad R^2 = 0,090 \qquad (5)$$

where $\log(Pob_{2001}) - \log(Pob_{1900})$ represents the average growth over the entire century. The equation is estimated by ordinary least squares and the heteroskedasticity-robust standard error (White 1980) is given in parenthesis. Figure 1 illustrates regression (5) and shows the coefficient of the initial population to be positive and highly significant (*t*-ratio 19.47). This result confirms, from an alternative perspective, the trend towards population concentration in the same places that were already important at the beginning of the century, and supports the notion of history as an important factor in the

[9] From a statistical perspective, all these coefficients are, without exception, highly significant under the null hypothesis of independence between initial and final distribution. Hence, history is important, and would seem to be very much so.

way the population settles across the territory. The previous result is robust to various types of weighted least squares to correct the heteroskedasticity present in the data.[10]

FIGURE 1: **Growth 1900-2001 *versus* (log) population in 1900**



Source: INE and authors' own calculations.

5. Geography versus History

In the previous sections, we have reviewed the importance of geographical and political-historical factors in population agglomeration across a territory. As a synthesis, we now present two exercises that illustrate the importance of these factors. The first is an analysis of variance and the second, the estimation of a conditional convergence equation. The analysis of variance considers the two geographical factors: coast and altitude, and provincial capital status. The following equation is estimated for each census year

[10] From the time series point of view, equation (5) represents an unstable AR(1) process; in this case the usual estimators do not have the appropriate properties to perform standard inference. However, the estimation of (5) only rests on the cross-section dimension of our data and is perfectly valid to perform the inference presented in the text. Work in progress (tentatively) shows that the same qualitative results are obtained when we use more complex dynamic panel techniques. In general terms, a tendency towards dispersion is observed.

$$\log(Pob_i) = \sum_{j=1}^{52} \alpha_j P_{ji} + \beta L_i + \sum_{j=1}^{3} \gamma_j A_{ji} + \delta C_i + u_i \quad (6)$$

where L_i is a dummy variable that takes the value of one if municipality i has direct sea access and zero otherwise; A_{ji} are dummy variables that take the value of one if municipal capital i has an altitude of 200 metres or below for $j = 1$, between 200 and 600 metres, for $j = 2$, between 600 and 1,000 metres for $j = 3$, and zero otherwise; C_i is a dummy variable that takes the value of one if municipality i is a provincial capital and zero otherwise; and finally P_{ji} are dummy variables that take the value of one if municipality i belongs to province $j = 1,2,\ldots,52$, and zero otherwise, and is introduced to capture heterogeneous behaviours in the different provinces.

Thus, the reference category in equation (6) is, for a given province, an inland municipality, mountainous (the capital of which lies over 1,000 metres above sea level) and is not a provincial capital. The importance of population movements from the inland areas to the coast, from the mountains to the plains and from the rural areas to the cities should be expressed in positive, statistically significant estimations for the parameters β, γ_j and δ. Furthermore, an increasing tendency in the estimations denotes the increasing importance of these attributes in the demographic movements. This should be seen as an average, and does not exclude specific cases of particular relevance.[11] Specifically, the cases of the cities of Madrid and Barcelona should be analysed with caution. On one hand, the physical boundaries of these municipalities may be conditioning certain results, and furthermore, these cities already appear as exceptional cases well before the 20th century (de Vries 1984, chapter 6).

The results of estimating equation (6), by ordinary least squares, are shown in table 12. The estimations could not be more conclusive. As only dummy variables are used, the R^2 is moderately high (between 42.3% and 55.2%), but what is more relevant is that it shows a clearly increasing tendency, and therefore the coast, low altitude and status of provincial capital are factors of increasing importance in explaining the size of Spanish municipalities.

---

[11] In fact, this type of regression yields a large number of what statisticians call atypical observations or outliers. However, there is nothing atypical in this case, as they are simply municipalities that, because of their own particular circumstances, deviate widely from the average behaviour. These particular cases are worth studying in their own right, but they are not cases that must be statistically “corrected” to improve the fit of the equation in question. The results of this type of regression should be taken as descriptive of average behaviour.

TABLE 10. **Analysis of Variance. Spain. 1900-2001**

Spain	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001
Coast	*0.51*	0.52	0.52	0.54	0.55	0.56	0.58	0.68	0.79	0.88	**0.97**
Below 200 m.	*0.39*	0.45	0.50	0.60	0.70	0.78	0.96	1.37	1.73	1.88	**1.99**
Between 200 and 600 m.	*0.36*	0.40	0.43	0.47	0.53	0.56	0.62	0.85	1.08	1.16	**1.21**
Between 600 and 1,000 m.	*0.16*	0.17	0.18	0.21	0.25	0.27	0.28	0.39	0.51	0.54	**0.57**
Capital	*3.07*	3.09	3.18	3.26	3.43	3.58	3.73	4.07	4.37	4.45	**4.46**
R^2	*42.3%*	43.6%	45.5%	46.9%	48.5%	48.9%	50.1%	51.3%	52.5%	53.8%	**55.2%**

Note: All the coefficients are statistically significant at levels well below 1%.
The *minimum* value of each province is shown in *italics*.
The **maximum** valued of each province is shown in **bold**.

Source: INE and authors' own calculations.

The results of the estimated coefficients are also extremely revealing. All of them are positive and highly significant,[12] and also show the correct magnitude. The highest estimated coefficient in all cases is for the *capital status* factor, and the *altitude* factor coefficients decrease evenly with increased height above sea level towards the mountains. Moreover, note that the magnitude of the coefficients increases continuously over time, and all them reach maximum values in 2001. This result simply indicates that while the three identifying factors –altitude, coastal location and capital status- were important at the beginning of the century, they are even more so today.

The second exercise is the estimation of a convergence equation similar to (5) but conditional on the three components considered in the analysis of variance (6). The result of this estimation, once again for the total period, is as follows,

$$\begin{aligned}\log(Pob_{2001}) - \log(Pob_{1900}) = \Sigma_{j=1}^{52}\hat{\alpha}_j P_{ji} + &\underset{(0,0207)}{0,0715}\log(Pob_{1900}) + \\ &\underset{(0,0569)}{0,4305}L_i + \\ &\underset{(0,0505)}{1,5698}A_{1i} + \underset{(0,0388)}{0,8256}A_{2i} + \underset{(0,0312)}{0,3997}A_{3i} + \\ &\underset{(0,1499)}{1,1708}C_i + \hat{u}\end{aligned} \qquad (7)$$

$$R^2 = 0,494 \qquad n = 8.108$$

Thus, although the introduction of the three previous components as conditioning factors in the convergence equation does not eliminate the trend towards divergence –in other words, towards population concentration- since the coefficient of the initial population is still positive and significant, its order of magnitude is substantially lower, from 0.30 in the non-conditional convergence equation to 0.07 in equation (7).[13] Consequently, the population concentration would not have been as acute as our findings show if we had been able to live in a *virtual world* and remove ourselves from the geographical and historical characteristics represented by the conditioning variables considered.

---

[12] The significance, not shown here, is obtained from the heteroskedasticity-robust standard errors (White 1980). This significance increases over time and the lowest *t*-ratio values are obtained in 1900. The lowest *t*-ratio is 5.33.

[13] Significance is also much lower since the *t*-ratio is 3.45. In addition, if the dummy variables indicating the altitude zones are substituted in (7) by the quantitative variable indicating the altitude of each municipality, the coefficient of the initial condition is even further reduced (to a value of 0.0565), but remains positive and significant.

Two further characteristics are important in equation (7). First the capacity of the dummy variables to explain growth is substantial, the R^2 increases sharply as compared to equation (5). Second, all the dummy variables are significant and present the expected sign (positive) and magnitude. The greatest effects on growth are, as before, for capital status and lowest altitude zone (which includes part of the coastal effect). Furthermore, population growth diminishes with altitude, and coastal location contributes an additional growth factor that is added to capital status or low altitude. Equation (7) therefore strengthens our previous conclusions.

6. Conclusions

In contrast to most studies on population agglomerations, the present paper does not focus solely on cities, however they may be defined, but on all municipalities. Spain has a very high number of small municipalities; around half the existing municipalities have, or have had 1,000 inhabitants or fewer and therefore represent a major section of Spanish rural geography. These municipalities have also suffered more intense depopulation and have fed population growth in a relatively limited number of cities.

In 1900, Spain's population was settled predominantly on the coast, in the plains and the provincial capitals. The results presented in this paper provide evidence, and through various techniques, quantify the importance of geographical and historical factors in population agglomeration. The coast and the plains attract growing numbers of inhabitants, both in absolute and relative terms, in a process that shows no signs of remission. The provincial capitals, representative of the urban system, have acted as magnets in the "local population markets", despite the fact that many of them lie in the inland areas, and some even in the mountains. It is precisely the characteristic of capital status that has emerged as the most important factor of agglomeration. However, it is highly possible that at least some of this influence may lie in the *initial conditions* since, in the end, the capitals were selected for this role on the grounds that they had the largest populations in each of the provinces as established by Royal Decree in 1833.

Despite the fact that behaviour of some large cities, particularly Madrid and Barcelona, would seem to indicate that the *capital status* factor has reached a point of inflection in recent decades, our results indicate that when the complete set of provincial capitals are considered, the population shift towards the provincial capitals does not appear to be coming to an end. In all events, the experiences of these two cities indicate that the rigidity of the municipal boundaries may be conditioning some of the results. The analysis therefore needs to be extended beyond the provincial capitals.

The above warnings should not obscure the main messages that derive from the exercises presented in this paper. The location of a municipality on the coast or in the plains, or its provincial capital status are highly relevant in explaining the capacity Spanish municipalities have for attracting inhabitants. Moreover, and surely still more relevant, the influence of these factors does not appear to be waning; indeed, it has become stronger with the passing of time. If these factors were already important in 1900, they are even more so today.

References

Ades, A., Glaeser, E., 1995. Trade and circuses: Explaining Urban Giants, Quarterly Journal of Economics, 110, 195-228.

Ayuda, M. A., Collantes, F., Pinilla, V., 2004. From locational fundamentals to increasing returns: The spatial concentration of population in Spain, 1787 - 2000. Mimeo. Universidad de Zaragoza, Zaragoza (Spain).

Ayuda, M. A.., Collantes, F., Pinilla, V., 2005. Explicando la concentración a largo plazo de la población española, 1860-2000. Working Paper nr. 2005-3. Centro de Estudios sobre la Despoblación y Desarrollo de Áreas Rurales, Zaragoza (Spain).

Ayuda, M. A., Collantes, F., Pinilla, V., 2007. Long-run regional population divergence and modern economic growth in Europe: A case study of Spain. Working Paper nr. 310-2007, Fundación de las Cajas de Ahorro (FUNCAS), Madrid.

Barro, R. J., Sala-i-Martín, X., 1992. Convergence. Journal of Political Economy, 100, 2, 223-251.

Barro, R. J., Sala-i-Martín, X., 1995. Economic Growth, McGraw Hill, New York.

Collantes Gutiérrez, F., 2004. Convergencia económica 'por defecto' en el medio rural español: El caso de las zonas de montaña, 1970-2000. Revista Asturiana de Economía, 29, 135-155.

Collantes Gutiérrez, F., 2005. Declive demográfico y cambio económico en las áreas de montaña españolas, 1860-2000. Revista de Historia Económica – Journal of Iberian and Latin American Economic History, 23, 3, 515-540.

Collantes Gutiérrez, F., Pinilla Navarro, V., 2003. La evolución a largo plazo de la población española, 1860-2000: Tipología provincial y análisis del caso aragonés, in: Políticas Demográficas y de Población, Gobierno de Aragón, Zaragoza (Spain), pp. 43-70.

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

De Vries, J., 1984. European Urbanization 1500 – 1800, Methuen and Co. Ltd, London.

Eaton, J., Eckstein, Z., 1997. City and growth: Theory and evidence from France and Japan. Regional Science and Urban Economics, 27, 443-474.

Esteve, A., Devolder, D., 2004. De la ley rango-tamaño (rank-size) a la ley log-normal: Los procesos aleatorios en el crecimiento demográfico de los agregados de población. VII Congreso Asociación de Demografía Histórica, Granada 1-3 April 2004. Session 11. Dinámicas espaciales de la población en el largo plazo (siglos XIX y XX).

Gabaix, X., 1999. Zipf´s law and the growth of cities. American Economic Review, Papers and Proceedings, 89, 129-132.

García España, E., 1991. Censos de población españoles. Estadística Española, 33, 128, (September-December). [http://www.ine.es]

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

Glaeser, E., Scheinkman, J., Shleifer, A., 1995. Economic growth in a cross-section of cities. Journal of Monetary Economics, 36, 117-143.

Goerlich, F. J., Mas, M., 2001. La Evolución Económica de las Provincias Españolas (1955 – 1998). Volume I. Capitalización y Crecimiento. Fundación BBVA. Bilbao.

Goerlich, F. J., Mas, M., 2007. Pautas de localización de la población a lo largo del siglo XX", Investigaciones Regionales, forthcoming.

Goerlich, F. J., Mas, M., Azagra, J., Chorén, P., 2006. La Localización de la Población sobre el Territorio. Un Siglo de Cambios. Un Estudio Basado en Series Homogéneas 1900-2001. Fundación BBVA. Bilbao.

Ioannides, Y. M., Overman, H. G., 2003. Zipf´s law for cities: an empirical examination. Regional Science and Urban Economics, 33, 127-137.

Ioannides, Y. M., Overman, H. G., 2004. Spatial evolution of the US urban system. Journal of Economic Geography, 4, 131-156.

Krugman, P., 1991. Geography and Trade, Leuven University Press, Leuven, Belgium and The MIT Press, Cambridge (Massachusetts).

Kuznets, S., 1973. Modern Economic Growth: Findings and Reflections. American Economic Review, 63, 3 (June), 247-258.

Lanaspa, L.; Perdiguero, A. M., Sanz, F., 2004. La distribución del tamaño de las ciudades en España, 1900-1999. Revista de Economía Aplicada, XII, 34, 5-16.

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

Ministerio de Fomento, 2000. Atlas estadístico de las áreas urbanas en España, Ministerio de Fomento (Spanish Ministry of Public Works), Madrid.

Nadal, J., (Director) 2003. Atlas de la Industrialización de España, 1975-2000. Fundación BBVA-Editorial Crítica, Bilbao.

Overman, H. G., Ioannides, Y. M., 2001. Cross-sectional evolution of the U.S. city size distribution. Journal of Urban Economics, 49, 543-566.

Rappaport, J., Sachs, J. D., 2003. The United States as a Coastal Nation. Journal of Economic Growth, 8, 5-46.

Reher, D. S., 1990. Urbanization and demographic behaviour in Spain, 1860-1930, in: Van der Woude, A., De Vries, J., Hayami, A. (Eds.): Urbanization in History. A Process of Dynamic Interaction, Clarendon Press, Oxford, pp. 282-299.

Suarez-Villa, L., 1988. Metropolitan evolution, sectoral economic change, and the city size distribution". Urban Studies, 25, 1-20.

Van der Woude, A., De Vries, J., Hayami, A., 1990. Introduction: The hierarchies, provisioning, and demographic patterns of cities in: Van der Woude, A., De Vries, J., Hayami, A. (Eds.): Urbanization in History. A Process of Dynamic Interaction, Clarendon Press, Oxford, 1-19.

Vinuesa Angulo, J., 1997. El crecimiento de la población y los desequilibrios en la distribución espacial, in: Puyol, R. (Ed.): Dinámica de la población en España: cambios demográficos en el último cuarto del siglo XX, Ed. Síntesis, Madrid, 265-311.

Wheaton, W., Shishido, H., 1981. Urban concentration, agglomeration economies and the level of economic development. Economic Development and Cultural Change, 30, 17-30.

White, H. A., 1980. A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity, Econometrica, 48, 4, (May), 721-746.

Zoido, F., Arroyo, A., 2004. La población de España, in Arroyo, A. (Coord.): Tendencias demográficas durante el siglo XX en España, available at: http://www.ine.es/prodyser/pubweb/tend_demo_s20/tend_demo_s20.htm, Instituto Nacional de Estadística, Madrid, pp. 22.