UNIVERSITAT
ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

CENTRE DE RECERCA EN ECONOMIA
INDUSTRIAL I ECONOMIA PÚBLICA

WORKING PAPERS

Col·lecció "DOCUMENTS DE TREBALL DEL
DEPARTAMENT D'ECONOMIA - CREIP"

Spatial Exploration of Age Distribution in Catalan Municipalities

Daniel Liviano
Josep Maria Arauzo

Document de treball n.24- 2013

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

UNIVERSITAT ROVIRA I VIRGILI
DEPARTAMENT D'ECONOMIA

CREIP
CENTRE DE RECERCA EN ECONOMIA INDUSTRIAL I ECONOMIA PÚBLICA

Edita:

Departament d'Economia
www.fcee.urv.es/departaments/economia/public_html/index.html
Universitat Rovira i Virgili
Facultat d'Economia i Empresa
Avgda. de la Universitat, 1
43204 Reus
Tel.: +34 977 759 811
Fax: +34 977 300 661
Email: sde@urv.cat

CREIP
www.urv.cat/creip
Universitat Rovira i Virgili
Departament d'Economia
Avgda. de la Universitat, 1
43204 Reus
Tel.: +34 977 558 936
Email: creip@urv.cat

Adreçar comentaris al Departament d'Economia / CREIP

Dipòsit Legal: T - 1279 - 2013

ISSN edició en paper: 1576 - 3382
ISSN edició electrònica: 1988 - 0820

DEPARTAMENT D'ECONOMIA – CREIP
Facultat d'Economia i Empresa

Spatial Exploration of Age Distribution in Catalan Municipalities

Daniel Liviano ♣

Josep-Maria Arauzo-Carod ♠ ♦

Abstract

This paper takes the shelf and digs into the complex population's age structure of Catalan municipalities for the year 2009. Catalonia is a very heterogeneous territory, and age pyramids vary considerably across different areas of the territory, existing geographical factors shaping municipalities' age distributions. By means of spatial statistics methodologies, this piece of research tries to assess which spatial factors determine the location, scale and shape of local distributions. The results show that there exist different distributional patterns across the geography according to specific local determinants.

Keywords: Spatial Models.

JEL Classification: C21.

This research was partially funded by CICYT's SEJ2010-19088/ECON and SEJ2010-16934/ECON), the "Xarxa de Referència d'R+D+I en Economia i Polítiques Públiques" and the SGR Program (2009-SGR-322) of the Catalan Government. The usual disclaimer applies.

(♣) Universitat Oberta de Catalunya (UOC); Av. Tibidabo 39-43; 08035 Barcelona (Catalonia); Phone: +34 932542169.

(♠) Universitat Rovira i Virgili, Departament d'Economia (QURE-CREIP); Av. Universitat 1; 43204 Reus (Catalonia); Phone: +34 977758902; Fax: +34 977759810.

(♦) Institut d'Economia de Barcelona (IEB); Av. Diagonal 690; 08034 Barcelona (Catalonia).

1 Motivation

Population Economics is a topic that has received considerable attention by scholars and leading journals in recent years. Inside this area, many contributions rely on issues related with population growth, which is analysed from a broad type of perspectives like, among others, relationship between population and jobs' growth, internal and international migration, locational preferences of individuals, commuting and counterurbanization, etc. However, there is a key issue that deserves to be taken into account: the structure of population pyramids, which show the distribution of various age groups in a population. When we talk about age distribution we refer to the distribution of population in age categories, namely, individuals between 5 and 10 years old, 11 and 15 and so on. This distribution varies considerably among countries, as it depends on specific characteristics as gender distribution, income levels, fertility rates and, generally speaking, on variables related with the overall development level of a community. Although there is plenty of data about age cohorts (and their trends and evolution) at national level, unfortunately less is known about whether they also vary at more spatial disaggregated levels inside a single country, as municipalities, for instance.

Previous well established national differences in terms of age distribution lead to different population pyramids, which is the most usual way to graphically represent age distribution. In this sense, growing populations show expanding pyramids (triangle-shaped with the base at the bottom), stable populations show stationary pyramids (square-shaped) and elder populations show contracting pyramids (triangle-shaped with the base at the top). According to population projections by United Nations Population Division (World Population Prospects, the 2010 Revision) the share of individuals aged 65 years and over is expected to grow from 19.40% in 2010 to 32.03% in 2050[1], as well as those between 15 and 64 are expected to shrink from 64.77% in 2010 to 52.75% in 2050[2]. These changes in age structure will obviously affect future societies in terms of, for instance, higher demands for health care and nursing (Uhlenberg, 2006), so age structure really matters and needs to be placed on

[1] For a general approach about implications of population ageing see the Journal of Population Ageing, specialised in these issues.

[2] In any case, it is important to notice that population forecasting can be biased for long term projections (Keilman, 1997).

the agendas of both researchers and policy makers.

As we have previously introduced, age distribution issues have been approached mainly at a national level in terms of intercountry comparisons, ignoring the potential role played by spatial issues over these distributions. This shortcoming, regrettably, has been usual for demographic analysis in general. Although spatial effects in population analysis have been largely theorized, many demographic studies lack a spatial perspective (Tiefelsdorf, 2000). This implies, when it comes to empirical research, that the formal modelling of spatial effects is still a matter pending. This omission entails the assumption that population dynamics (e.g., migration, aging, etc.) in a specific area has no relationship with what happens in neighbouring areas, which is clearly unrealistic and contradictory with, for instance, Waldo Tobler's first law of geography (1970, p. 7), which states that *"everything is related to everything else, but near things are more related than distant things"*.

Recently, however, scholars (see, among others, Chi and Marcouiller, 2012) have introduced into the analysis the spatial dimension of population by studying how spatial dependence holds for these issues, demonstrating that this is an important dimension that needs to be taken into account. Otherwise, researchers will end up with biased and misleading results. Fortunately, there have been considerable research advances regarding such spatial issues, so researchers in this area can largely benefit from what has been theorized and applied in similar disciplines. In terms of empirical tools, we can highlight the role played by Geographical Information Systems (GIS), which allow not only to collect and represent geographically referenced data (White and Lindstrom, 2006), but also to infer relationships among variables.

Given the importance of previously discussed age distribution implications, our main goal in this paper is to analyse age distribution at a smaller geographical units, this is, municipalities. Specifically, our aim is fourfold: i) to obtain detailed and accurate estimations of these distributions in order to correctly identify specific local patterns, ii) to explain the heterogeneity of these distributions by means of local factors and variables, iii) to identify spatial dependence processes linked to age distributions, and iv) to discuss the policy implications of the results derived from our research.

This paper is structured as follows. In Section 2 contributions on population economics and geography are discussed and reviewed. Many of these contributions focus on population growth, which is related to the research undertaken in this paper. In Section 3 the dataset and the constructed variables used in this paper are presented and justified. Section 4 is devoted to the empirical analysis. This section has two parts: a descriptive analysis at the univariate level giving information about the variables' characteristics, including spatial patters, and a regression analysis that tries to explain the relationship between local age distributions and local determinants. Finally, Section 5 summarises and discusses the main conclusions and gives some indications for future research.

2 Literature Review

Age distribution deserves further attention because it has plenty of implications for public and private welfare. Surely one of the most evident issues is health care, since the demand for health care services strongly depends on the age structure of the population (both for younger and older individuals). In this sense, as a consequence of population aging in most developed countries, caused by general improvements on health care and medicine, there have been important increases in health care demand by elder people as well as resource transfer via social security systems (Lee and Edwards, 2001). This growing demand of health care modifies the labour market composition, as it requires more workers to be allocated at health care industries from non-health care industries, which turns to lower aggregate per capita income growth rates (Hashimoto and Tabata, 2010). Obviously, changes in age distribution, typically by population aging as we have explained previously, reduce population growth (Chi and Marcouiller, 2012; Rickman and Rickman, 2011; Partridge et al. 2009), as they imply lower fertility rates (Hashimoto and Tabata, 2010; Waldorf and Byun, 2005), influence migration patterns (Stillwell and García Coll, 2000; Nivalainen, 2004; Greenwood and Hunt, 1989; Greenwood, 1975) and is an important determinant of residential preferences (Fuguitt and Brown, 1990), given that locational preferences (which areas are preferred) vary with age. A typical example of this heterogeneity could be presented in terms of housing. As Greenwood and Stock (1990, p. 274) point out, *"(...) since the demand for different housing is often conditioned by the birth and aging of children, these important societal changes could well have had implications* *for central city versus suburban settlement patterns"*. Hence, urban structure also depends on age distribution, as younger families with children prefer to live in suburban environments while senior families need to be closer to health services provided at the city centre.

As preferences of individuals vary according to their age, different age compositions will also determine different composite preferences in terms of consumption (housing, commodities, leisure activities, etc.), public services (education, health care, transportation, etc.) and migration patterns (Douglas, 1997). The effect of age composition over consumption has been widely analysed by several scholars (Hock and Weil, 2012; Elmendorf and Sheiner, 2000; Cutler et al., 1990) being that the typical argument about this relationship considers that (Hock and Weil, 2012, p. 1021): *"(...) rising old-age dependency reduces the disposable income of the working population, resulting in lower fertility and further population aging"*. Apart from the distributional issues of age cohorts, it is also remarkable that individuals from different ages will have different lifestyles and will tend to consume different types of goods (Lee et al. 2008, Cutler et al., 1990), so population aging (as any other change in age cohorts) modifies consumption patterns favouring certain amenities preferred by seniors (Rickman and Rickman, 2011).

Accordingly to previous empirical evidence, it seems reasonable to argue that policy makers should care about age distribution trends in order to properly account for all the consequences that derive from it. Nevertheless, most of the efforts have been put to analysis at a national level, trying to demonstrate how aggregate changes in age cohorts can affect welfare levels for the whole country. In this sense, there is plenty of empirical evidence showing that there are important differences at a national level in terms of age distribution, but very little is known about to what extent population's age structure changes occur at a much more disaggregated level, and thus whether they are caused by local determinants, among which spatial effects are prominent. Therefore, there are still unanswered questions regarding whether age distribution depends on characteristics (institutional, economic, etc.) that vary at a national level or, on the contrary, it depends on characteristics (amenities, climate, labour markets, etc.) that are mainly local.

If we assume that age distribution is mainly a *national* phenomenon, then we should rely on some institutional and macro characteristics that have nothing to do with smaller geographical areas, but if this is not the case and it is assumed that this is mainly a *local* phenomenon, then we must take into account some spatial issues that can strongly influence this distribution. These spatial effects determine whether this phenomenon is geographically isolated (e.g., age distribution of a given municipality is independent from that of neighbour municipalities) or, on the contrary, neighbour areas have similarities in their age cohorts.

Specifically, previous spatial effects can be categorized into two groups. On the one hand, spatial dependence[3] can be defined as a similarity (and also dissimilarity) measure between spatial variables located nearby. Spatial dependence is composed of large-scale spatial irregularities and local-scale spatial interaction effects. On the other hand, spatial heterogeneity is a much broader concept that refers to differences in variable distributional parameters across space. Spatial heterogeneity often affects mean, variance and covariance structures of the data.

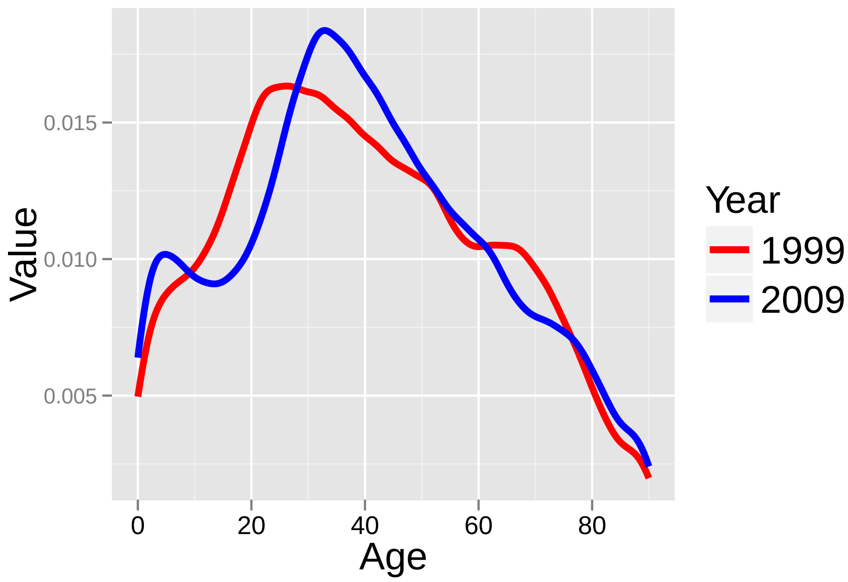
Therefore, although overall changes in age distribution (at the aggregate level) are expected to be important in incoming years (e.g., continued population aging in developed countries), it is also necessary to analyse whether age structure differs from a spatial point of view in a cross-section approach. In this sense, differences in age distribution across spatial units (municipalities, counties, provinces, regions, etc.) have plenty of implications in terms of public policies affecting health, transport demand, education and social services. Concretely, typologies of public services demanded by individuals are strongly dependent on their age, so spatial units with different age distributions will differ in terms of the weight of public services demanded by them, especially if their demographic structures are quite different than those at upper levels, like the whole country. Accordingly, it is of key importance to accurately explain how and why these age distributions vary across spatial units and, to the same extent, whether is it possible to predict future trends.

[3] Spatial dependence is also known as spatial autocorrelation and spatial interaction, although some geographers and demographers understand them differently. See Chi and Zhu (2008) for a discussion.

3 Data and Variables

Before getting into the data analysis, it is useful to contextualize the phenomenon under analysis. Figure 1 portraits the estimated density of the aggregated population's age for years 1999 and 2009 in Catalonia, and shows the overall evolution of the land's age structure over 10 years. With the naked eye we can see that the centre of the distribution has moved to the right side, which indicates an underlying overall ageing process. However, the shape of both distributions reveal more information than that. For instance, there is a bump between age values 0 and 15 for the year 2009, showing a birth rate boost in the period 1999-2009. Likewise, the population percentage in the age interval 60-80 has decreased throughout this period. However, the distributions depicted in Table 1 do not reveal to what extent population's age structure changes occur at a much more disaggregated level. This is the contribution that our paper wants to make: to split the distribution for the year 2009 into smaller local distributions and study them thoroughly.

Fig. 1: Empirical density of Catalan population's age (1999 and 2009)



The data used in this article refer to local units (municipalities) in Catalonia for the year 2009. On the one hand, data on the population classified into age intervals come from the Spanish Statistical Institute (INE), and on the other hand data referring to geographical characteristics of municipalities are taken from the Catalan Statistical Institute (IDESCAT). Moreover, the polygon map and related shape files

Tab. 1: Original Structure of Demographic Data

Category (k)	Interval	Mid-point (x_k)	Category (k)	Interval	Mid-point (x_k)
1	[0,4]	2	10	[45,49]	47
2	[5,9]	7	11	[50,54]	52
3	[10,14]	12	12	[55,59]	57
4	[15,19]	17	13	[60,64]	62
5	[20,24]	22	14	[65,69]	67
6	[25,29]	27	15	[70,74]	72
7	[30,34]	32	16	[75,79]	77
8	[35,39]	37	17	[80,84]	82
9	[40,44]	42	18	[+85]	88

have been obtained from the Catalan Cartographic Institute (ICC).

Formally, the data is divided into $N = 941$ municipalities[4]. Municipalities are denoted by the subscript i, so that P_i is the total population in municipality i. $P = \sum_{i=1}^{N} P_i$ is the total population of Catalonia. Population data for each municipality and year is available in a frequency table with $n = 18$ age intervals. This structure is shown in Table 1.

Certainly, the available data is limited to age intervals, so that the exact age distribution has to be inferred and estimated. We propose an estimation procedure: for each municipality i[5], let p_k be the number of people whose age belongs to the k-th age category ($k \in 1,\ldots,n$). The first step is to create the vector $\vec{v}$, whose length is the total number of inhabitants, this is, $P = \sum_{k=1}^{n} p_k$. The elements of $\vec{v}$ are the mid-points x_k repeated p_k times, respectively. More formally, vector $\vec{v}$ can be expressed as the concatenation of n vectors $(v_1,\ldots,v_n)$, where each v_k is computed by means of the scalar product of mid-points by a ones vector $\vec{v}_k = x_k \cdot \mathbf{1}_{(1\times p_k)}$. Once $\vec{v}$ has been computed, the next step is to estimate its density function, and this is done non-parametrically by means of a Kernel density estimation, which is a method for data smoothing based on a finite data sample. Analytically, the aim is to obtain an estimate of the density function $f(\cdot)$ that has generated the values in $\vec{v} = (v_1,...,v_P)$. The shape of $f(\cdot)$ is approximated by:

[4] Data for five new municipalities (Gimenells i el Pla de la Font, Riu de Cerdanya, Sant Julià de Cerdanyola, Badia del Vallès and La Palma de Cervelló) have been left out due to lack of data.

[5] For the sake of simplicity, individual subscripts (i) are omitted in this exposition.

$$\hat{f}(v) = \frac{1}{h \cdot P} \sum_{r=1}^{P} K\left(\frac{v - v_r}{h}\right). \tag{1}$$

In this function, $K(\cdot)$ is chosen to be a Gaussian density estimation, and h is the bandwidth or smoothing parameter, whose value has been selected so as to guarantee that $\hat{f}(\cdot)$ integrates to one, this is, $\int \hat{f}(v) = 1$. For each municipality, $\hat{f}(\cdot)$ is evaluated over a grid of $1,001$ equally spaced nodes within the range $[0,90]$[6].

Once we have estimated the N local age distributions, the next step is to obtain certain indicators that contain information about its shape characteristics. We have chosen four parameters to fulfill this aim, what we call *distributional parameters*. The smoothness of the estimated age distributions, and the fact that the area under each of these integrates to one, allow the computation of these distributional parameters more efficiently than the analysis of a simple discrete histogram.

Next, we introduce and explain these distributional parameters, and also introduce the local characteristics assumed to influence local age distributions. These local characteristics have been classified into three categories: i) human variables, which include basically the population density and the share of immigrants, ii) economic structure, which refers to employment and sector structure, and iii) geographical variables, including altitude and coastal location, among others. These four grupus of variables are detailed below and summarised in Table 2.

A) Distributional parameters

In order to capture the whole shape of municipal age distributions, four different indicators have been selected and computed. These are described next, and are graphically depicted in Figure 2.

1. **Median** ($\hat{\eta}$): this measure captures the central tendency of the sample space, so that the distributional mass on both sides is 50%. This is a simple measure

[6] The upper limit of this range has been chosen rather arbitrarily, due to the fact that the data are right-censored because the last age interval is not closed (population above 85 years). Therefore, we are assuming that the end of the distribution is 90 years. Due to the low percentage of population above this age, this restriction is expected to have a negligible effect on the results.

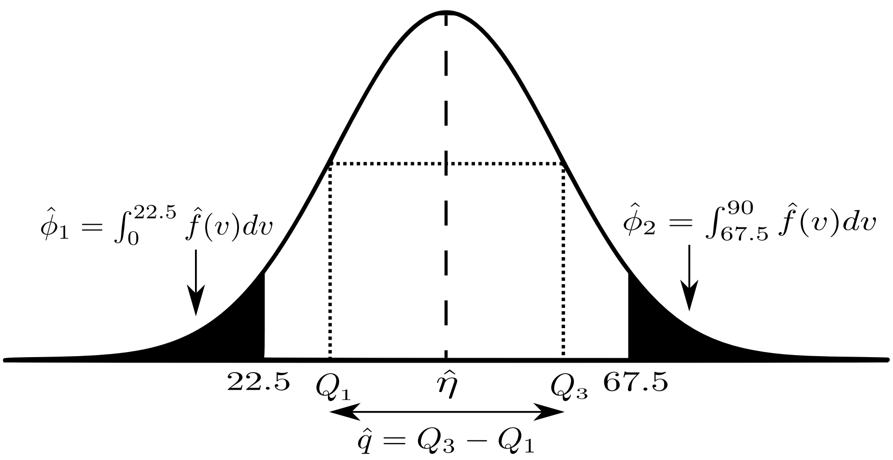


Fig. 2: Description of Distributional Parameters

and indicates to what extent a municipality is, on average, young, middle-aged or old.

2. **Interquartile Range** ($\hat{q}$): also called midspread, it measures dispersion as the difference between the upper and lower quartiles, so that $\hat{q} = Q_3 - Q_1$. This way, it measures the difference between the ages including 50% of the distribution mass. A high value of $\hat{q}$ indicates a highly dispersed population, whilst a low value points out a population where a large part of its distribution mass gathers around the mean.

3. **Distribution Tails ($\hat{\phi}$):** these empirical densities capture the thickness and magnitude of the distribution tails, this is, the population mass falling in the age interval [0,22.5] and, at the other side of the distribution, the population share above 67.5 years. For each municipality, these two parameters are obtained by integrating the corresponding estimated density function $\hat{f}(v)$ over these age intervals:

$$\hat{\phi}_1 = \int_0^{22.5} \hat{f}(v)\,dv \tag{2}$$

$$\hat{\phi}_2 = \int_{67.5}^{90} \hat{f}(v)\,dv. \tag{3}$$

Considering that $\hat{f}(v)$ has been computed assuming the closed range [0,90], the first interval includes the first 25% of this range, whilst the second interval encompasses the last 25% of this range.

B) Human Variables

This group of covariates has two variables. DEN is the log of population density, computed as population divided by the urban area, and MIG stands for the share of immigrants. There are few sites where MIG takes the value zero, which means that there are no immigrants among the native population. The opposite situation is represented by some municipalities, where nearly 50% of inhabitants were born outside Spain.

C) Economic Structure

The variables included in this group reflect the economic structure of each municipality, in the sense of employment rate and sector diversity. The variable EMP is the employment rate, which is the ratio of number of workers over the municipality's whole population. The variables AGR, MAN, CON and SER stand for the proportion of workers in the sectors agriculture, manufactures, construction and services, respectively. These variables have been computed so that, for each municipality, they add up to 100%.

D) Geographical Position

These are two variables controlling for the geographical position of each municipality. ALT is the average municipality's altitude with respect to sea level, which controls for accessibility. TMC is the transport time to the main cities, which captures to what extent a municipality is isolated from the main urban areas. CC is a dummy variable taking a value one if the municipality is the capital of a county, and CL is a dummy variable with a value one if the municipality is coastal, and zero otherwise.

4 Empirical Analysis

The goal of the empirical analysis conducted in this paper is twofold. The first subsection consists in a collection of descriptive and spatial statistics, whereby the main characteristics of the variables under study are summarised and studied. The second subsection tries to explain the spatial heterogeneity of local age distributions by

Tab. 2: Description of Variables

Description	Variable
A) Distributional Parameters	
Median	$\hat{\eta}$
Interquartile Range	$\hat{q} = Q_3 - Q_1$
Left Tail $[0, 22.5]$	$\hat{\phi}_1 = \int_0^{22.5} \hat{f}(v) dv$
Right Tail $[67.5, 90]$	$\hat{\phi}_2 = \int_{67.5}^{90} \hat{f}(v) dv$
B) Human Variables	
Population Density	DEN
Share of Immigrants	MIG
C) Economic Structure	
Employment Rate	EMP
Agriculture (%)	AGR
Manufacture (%)	MAN
Construction (%)	CON
Services (%)	SER
D) Geographical Position	
Altitude	ALT
Transport time	TMC
County Capital	CC
Coast Location	CL

Source: Own elaboration and Catalan Statistical Institute (IDESCAT).

relating them to the three groups of local variables, i.e. human variables, economic structure and geographical position. In practice, this is carried out by estimating five regression models, where the distributional parameters are the dependent variables and the local variables are the covariates. Because of the existence of spatial effects in the residuals, spatial econometric techniques have been used to account for these effects.

4.1 Descriptive Statistics

The first exploratory univariate analysis presented in this section is shown in Table 3. It consists of a correlation matrix containing all four distributional parameters, and it shows how these parameters relate to each other. Municipalities with a higher median parameter ($\hat{\eta}$) show higher dispersion around the mean ($\hat{q}$) and a higher percentage of the population above 67.5 years ($\hat{\phi}_2$). These results confirm that there are several patterns describing different types of age distributions: the areas where the center of the distribution ($\hat{\eta}$) is lower are also characterised by a lower dispersion, a higher percentage of young people ($\hat{\phi}_1$) and a lower mass of older people ($\hat{\phi}_2$).

Tab. 3: Correlation of Distributional Parameters

	Median	IQR	Left T.	Right T.
Median	1.00			
IQR	0.58	1.00		
Left T.	−0.89	−0.33	1.00	
Right T.	0.88	0.75	−0.79	1.00

Note: all coefficients are significant at a 1% significance level.

Table 4 shows a collection of statistics, divided into three groups. The first group consists of the mean and standard deviation of the variables, capturing the centre and dispersion. These measures are not weighted, i.e. they give the same weight to municipalities with different populations. For this reason, they shall not be interpreted as aggregated statistics. The second group consists of these two statistics computed weighting by population, so that they can be interpreted as aggregate statistics. Comparing these two statistics one-by-one is interesting: the parameter $\hat{\eta}$ is higher when not weighted, indicating that less populated municipalities have a higher median age. Similarly, we can infer that these less populated municipalities have a thinner left tail and a fatter right tail (a sign of ageing population). Besides, these areas show a lower density, a lower share of immigrants and a lower employment rate, which is consistent with the fact that these municipalities tend to be in rural areas with low levels of urbanization. Lastly, the statistics for the coast location (CL) dummy variable indicates that, although only 7% of the municipalities lie on the coast, 43% of the population lives in these coastal areas[7].

The third column displays the Moran's I statistic and its P-value. They capture the spatial autocorrelation in each variable. The concept of spatial correlation is more complex than the simple linear correlation, since the former is bi-dimensional (the map used in this paper is an euclidean plane with two dimensions) and multi-directional. A zero value indicates a random spatial pattern, whereas a positive value indicates positive spatial correlation, i.e. high (low) values of the variable tend to be surrounded by high (low) values of the same variable in neighbouring spatial units. In order to compute this test, we have used a standardized contiguity weights matrix, whereby two municipalities are considered neighbours only if they share a border. The results indicate that the four distributional parameters yield a spatial correlation statistic's value around 0.5, which is a rather high value and shows a strong spatial pattern, stronger than the same statistic for the remaining variables. This result has deep implications, this is, the shape of local age distributions is not randomly distributed, and shows a strong spatial pattern.

The spatial correlation present if the four distributional parameters can be also expressed by means of maps and scatterplots, as shown in Figure 3. The first column shows a map of Catalonia divided into the $N = 941$ municipalities, where the value of the distributional parameters is displayed in different colours according to the chromatic scale on the right side of the map. The second column shows a scatterplot where the x-axis corresponds to the values of each distributional parameter, and the y-axis is the weighted mean of each municipality's neighbours. The positive slope in all four graphs is the indication of positive spatial autocorrelation, i.e. higher values of the parameter are positively correlated with this parameter's values for the neighbours. All in all, these graphs indicate that the broad area closer to the sea is more populated

[7] The fact that two of the main cities are coastal (Barcelona and Tarragona) helps explaining this fact.

Tab. 4: Descriptive Statistics of Variables

	Non weighted stats.		Weighted stats.		Spatial autocorrelation	
	Mean	Std. Dev.	Mean	Std. Dev.	Moran's I	P-value
A) Distributional Parameters						
Median ($\hat{\eta}$)	42.85	4.91	39.41	2.42	0.55	0.00
Interquartile Range ($\hat{q}$)	33.32	3.19	32.05	2.00	0.43	0.00
Left Tail ($\hat{\phi}_1$)	0.22	0.05	0.24	0.03	0.51	0.00
Right Tail ($\hat{\phi}_2$)	0.13	0.05	0.10	0.03	0.58	0.00
B) Human Variables						
Population Density (DEN)	5.81	16.45	12.76	10.21	0.01	0.29
Share of Immigrants (MIG)	0.10	0.07	0.16	0.07	0.32	0.00
C) Economic Structure						
Employment Rate (EMP)	0.21	0.22	0.33	0.18	0.21	0.00
Agriculture (AGR)	0.04	0.09	0.00	0.02	0.21	0.00
Manufacture (MAN)	0.26	0.23	0.19	0.15	0.19	0.00
Construction (CON)	0.14	0.13	0.09	0.06	0.07	0.00
Services (SER)	0.57	0.23	0.71	0.17	0.14	0.00
D) Geographical Position						
Altitude (ALT)	373.56	322.24	119.63	173.54	0.86	0.00
Transport time (TMC)	87.41	23.30	67.56	16.21	0.96	0.00
County Capital (CC)	0.04	0.20	0.43	0.51	-0.03	0.15
Coast Location (CL)	0.07	0.26	0.43	0.51	0.52	0.00

and younger, while rural areas (bottom-left and medium-left) and mountain areas (top) show smaller municipalities with ageing population.

4.2 Spatial Models

This section's goal is to perform a regression analysis in order to explain the heterogeneity and variations in the four distributional parameters. This is, the estimation of an equation for each distributional parameter is intended to shed light upon the effects of local covariates upon these parameter's variations in 2009. Formally, the models to be estimated take the following form:

$$y_i = \alpha + x_i^{'}\beta + \epsilon_i, \qquad i = 1, \ldots, N. \tag{4}$$

In this model, the dependent variables are all four distributional parameter, i.e. $y = \{\hat{\eta}, \hat{q}, \hat{\phi}_1, \hat{\phi}_2\}$. The set of covariates x_i includes the human, economic and geographical variables introduced in the previous section[8]. An obvious concern regarding the estimation of these models is the presence of spatial effects. These effects are of different nature, but they have in common that they render classical inference unreliable[9]. A possible spatial effect is that the variance of the model is not distributed randomly across the space, and thus follows some pattern. Besides, the hypothesis of independence among sample observations is often violated, and thus causality relationships jump over each observation's borders and affect neighbouring observations.

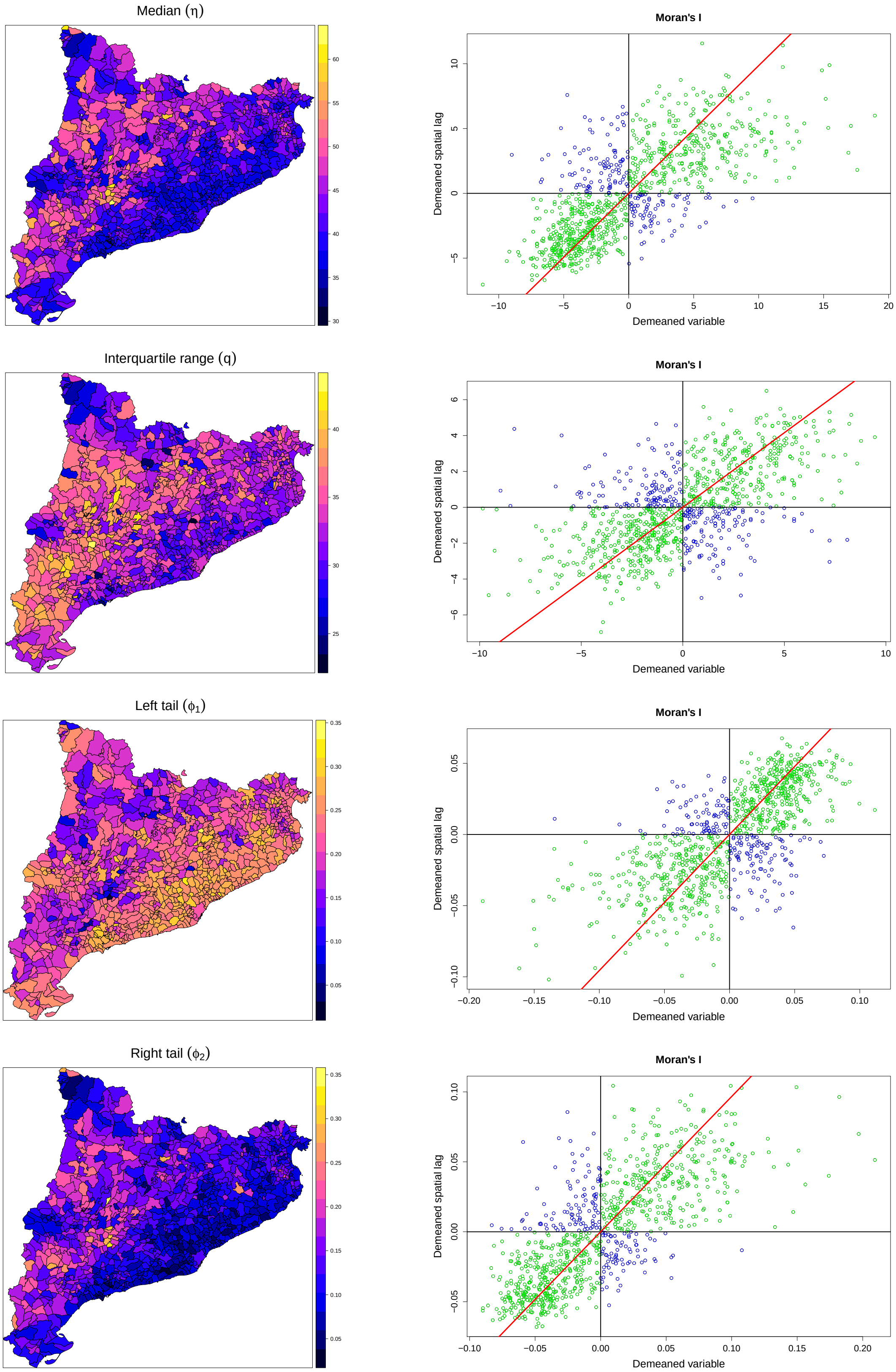
In a regression framework, a common procedure to detect spatial effects is the analysis of estimated residuals' distribution. To do so, the first step is to estimate the four equations and then run the Moran's I test on the residuals. The results are shown in Table 5, where we can see that there is strong evidence of spatial effects in the residuals.

In order to explain these effects away, they have to de explicitly modelled in a spatial model. In the spatial econometrics literature, the two classic models are the spatial lag model, includes a spatial lag of the dependent variable, and the spatial

[8] A multicollinearity test (*Variance Inflation Factor*) suggested us to drop the variables CON and SER.

[9] For seminal works on this matter, see Anselin (1988) and Cressie (1993).

Fig. 3: Spatial Distribution of Parameters



Tab. 5: Spatial Autocorrelation in OLS Models' Residuals

	Moran's I	P-value
Median ($\hat{\eta}$)	0.43	0.00
IQ Range ($\hat{q}$)	0.34	0.00
Left Tail ($\hat{\phi}_1$)	0.35	0.00
Right Tail ($\hat{\phi}_2$)	0.45	0.00

error model, in which the disturbances exhibit spatial dependence[10]. After running both models for our four equations, a comparison considering the *Akaike Information Criterion* (AIC) have pointed out that the spatial error model is slightly preferable. This model consists of the following equation:

$$y_i = \alpha + x_i^{'}\beta + u_i$$
$$u_i = \rho W u_i + +\epsilon_i$$

Where ϵ_i is assumed to be a classical error term. This model assumes that the spatial influence comes only through the error terms, being W the spatial weights matrix and ρ a parameter indicating the strength of the spatial effects. Rearranging the previous equation the following form is obtained:

$$y_i = \alpha + x_i^{'}\beta + (I - \rho W)^{-1}\epsilon_i$$

The Maximum Likelihood (ML) estimation of this equation for all parameters is shown in Table 6. The results show that the share of immigration (MIG) appears to be a key variable shaping the local age distributions, this is, municipalities with a higher value of MIG have a lower median parameter, less dispersion and a thicker left tail, since immigrants tend to boost natality. The coefficients of the employment rate (EMP) point in the same direction, which is consistent with the fact that immigrants tend to locate where there is demand for workers. Another interesting result comes from the share of agricultural workers (AGR), whose coefficient clearly indicates that in rural areas with a high predominance of agricultural activities the ageing process

[10] For a detailed explanation of these models, see Anselin (1988), and for a description of diagnostics tests for spatial dependence, see Anselin et al. (1996).

is more intense, i.e. the median of the distribution is moving rightwards, with values approaching 45 years and higher, and the right tail is thicker. Lastly, geographical factors also matter: those areas with a higher transport time to urban areas (TMC) are more isolated and, therefore, show an ageing population. On the contrary, the cities which are county capitals (CC) show an opposite effect, which is derived from the fact that younger populations are found in larger and more populated areas.

Tab. 6: ML Estimation of Spatial Error Models

	Median ($\hat{\eta}$)		IQ Range ($\hat{q}$)		Left Tail ($\hat{\phi}_1$)		Right Tail ($\hat{\phi}_2$)	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Intercept	38.61	0.00	34.18	0.00	27.61	0.00	8.71	0.00
Human Variables								
log(DENS)	0.13	0.34	-0.02	0.85	-0.13	0.31	0.15	0.26
MIG	-22.90	0.00	-10.05	0.00	13.58	0.00	-17.31	0.00
Economic Structure								
EMP	-1.06	0.05	-1.44	0.00	0.92	0.09	-1.54	0.01
AGR	3.35	0.01	1.97	0.05	-2.87	0.02	1.68	0.19
MAN	-0.94	0.08	0.34	0.43	0.98	0.07	-0.32	0.57
Geographical Position								
ALT	0.16	0.02	-0.20	0.00	-0.24	0.00	-0.08	0.26
TMC	0.06	0.00	0.02	0.04	-0.06	0.00	0.06	0.00
CC	-0.89	0.08	0.08	0.86	1.39	0.01	-0.07	0.89
CL	0.38	0.47	-0.56	0.17	-0.54	0.30	-0.94	0.09
ρ		0.11^{***}		0.10^{***}		0.10^{***}		0.11^{***}
AIC		5283		4737		5182		5410

5 Summary and Conclusions

In this paper spatial trends and implications of age structure in local areas have been addressed, which is a topic of population economics that has not received yet sufficient attention according to its importance. The structure of local population pyramids has been approached by the computation of four distributional parameters, i.e. the median and the interquartile range (standing for centre and dispersion), plus both left and right tails of the distribution. The distribution of these parameters over space has been found to be not random at all. Rather, strong evidence of spatial heterogeneity has been found. Besides, the distribution of these parameters has been related to human, economic and geographical variables for each municipality. A spatial regression analysis has proven that local age distributions are strongly influenced by these local determinants, being of importance the spatial location of the municipality, the share of immigration and the economic structure. Overall, this analysis roughly supports the prior expectations of such inequalities made in this paper, which fits perfectly into previous empirical evidence (both from the area analysed in this paper as well as from other similar areas) regarding spatial heterogeneity in terms of population distribution. The policy implications of these results are that locally focused actions fostering natality and immigration attraction are potentially more effective than those policies affecting the whole geography. Besides, if no immediate action is taken, a large area of Catalonia with small and old urban municipalities faces the risk of becoming a wasteland.

Further extensions of this research should point into specific determinants of age distribution, as well as better identify causality relationship among different economic activities and whether there is still room for policy measures affecting the shape of the distribution in the long term.

References

Anselin, L. (1988). *Spatial econometrics: methods and models*. London: Kluwer.

Anselin, L., Bera, A., Florax, R., and Yoon, M. (1996). Simple diagnostic tests for spatial dependence. *Regional science and urban economics*, 26(1):77–104.

Arauzo-Carod, J. and Teruel-Carrizosa, M. (2005). An urban approach to firm entry: The effect of urban size. *Growth and Change*, 36(4):508–528.

Chi, G. and Marcouiller, D. (2011). Isolating the effect of natural amenities on population change at the local level. *Regional Studies*, 45(4):491–505.

Chi, G. and Ventura, S. (2011). An integrated framework of population change: Influential factors, spatial dynamics, and temporal variation. *Growth and Change*, 42(4):549–570.

Chi, G. and Voss, P. (2011). Small-area population forecasting: borrowing strength across space and time. *Population, Space and Place*, 17(5):505–520.

Chi, G. and Zhu, J. (2008). Spatial regression models for demographic analysis. *Population Research and Policy Review*, 27(1):17–42.

Cressie, N. (1993). *Statistics for spatial data*. Wiley (New York).

Cutler, D. M., Poterba, J. M., Sheiner, L. M., Summers, L. H., and Akerlof, G. A. (1990). An aging society: opportunity or challenge? *Brookings papers on economic activity*, 1990(1):1–73.

Douglas, S. (1997). Estimating relative standard of living in the united states using cross-migration data. *Journal of Regional Science*, 37(3):411–436.

Elmendorf, D. W. and Sheiner, L. M. (2000). Should america save for its old age? fiscal policy, population aging, and national saving. *The Journal of Economic Perspectives*, 14(3):57–74.

Fossett, M. (2005). Urban and spatial demography. *Handbook of population*, pages 479–524.

Fritsch, M., Brixy, U., and Falck, O. (2006). The effect of industry, region, and time on new business survival–a multi-dimensional analysis. *Review of Industrial Organization*, 28(3):285–306.

Fuguitt, G. and Brown, D. (1990). Residential preferences and population redistribution: 1972–1988. *Demography*, 27(4):589–600.

Gómez, R. and Hérnandez de Cos, P. (2008). The importance of being mature: the effect of demographic maturation on global per capita gdp. *Journal of population economics*, 21(3):589–608.

Greenwood, M. (1975). Research on internal migration in the united states: a survey. *Journal of Economic Literature*, 13(2):397–433.

Greenwood, M. and Hunt, G. (1989). Jobs versus amenities in the analysis of metropolitan migration. *Journal of Urban Economics*, 25(1):1–16.

Hashimoto, K. and Tabata, K. (2010). Population aging, health care, and growth. *Journal of Population Economics*, 23(2):571–593.

Hock, H. and Weil, D. N. (2012). On the dynamics of the age structure, dependency, and consumption. *Journal of Population Economics*, 25(3):1019–1043.

Keilman, N. (1997). Ex-post errors in official population forecasts in industrialized countries. *Journal of official statistics*, 13:245–278.

Kelejian, H. and Prucha, I. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2):509–533.

Lee, R. and Mason, A. (2008). Charting the economic life cycle. *Population aging, human capital accumulation, and productivity growth*, page 208.

Lee, R. D. and Edwards, R. D. (2001). The fiscal impact of population change. *Seismic shifts: the economic impact of demographic change*, 46:189.

Leichenko, R. (2001). Growth and change in us cities and suburbs. *Growth and Change*, 32(3):326–354.

Nivalainen, S. (2004). Determinants of family migration: short moves vs. long moves. *Journal of Population Economics*, 17(1):157–175.

Nyström, K. (2007). An industry disaggregated analysis of the determinants of regional entry and exit. *The Annals of Regional Science*, 41(4):877–896.

Parr, J. (2002). Missing elements in the analysis of agglomeration economies. *International Regional Science Review*, 25(2):151–168.

Partridge, M., Rickman, D., Ali, K., and Olfert, M. (2009). Do new economic geography agglomeration shadows underlie current population dynamics across the urban hierarchy?*. *Papers in Regional Science*, 88(2):445–466.

Rickman, D. and Rickman, S. (2011). Population growth in high amenity nonmetropolitan areas: what's the prognosis? *Journal of Regional Science*, 51(5):863–879.

Stillwell, J. and Coll, A. (2000). Inter-provincial migration of the spanish workforce in 1988 and 1994. *Regional Studies*, 34(7):693–711.

Tiefelsdorf, M. (2000). *Modelling spatial processes*. Springer Berlin/Heidelberg.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography*, 46:234–240.

Uhlenberg, P. (2005). Demography of aging. *Handbook of Population*, pages 143–167.

Voss, P. (2007). Demography as a spatial social science. *Population Research and Policy Review*, 26(5):457–476.

Waldorf, B. and Byun, P. (2005). Meta-analysis of the impact of age structure on fertility. *Journal of Population Economics*, 18(1):15–40.

White, M. and Lindstrom, D. (2006). Internal migration. *Handbook of population*, page 311.