

Spatial analysis of urban growth in Spain, 1900–2001

Julie Le Gallo · Coro Chasco

Received: 15 July 2006 / Accepted: 15 January 2007 / Published online: 27 July 2007
© Springer-Verlag 2007

Abstract The purpose of this paper is to improve the knowledge of the Spanish urban system. We study the evolution of population growth among the group of 722 municipalities included in the Spanish urban areas over the period 1900–2001. A spatial SUR model is estimated for Zipf's law and shows the existence of two main phases: divergence (1900–1980) and convergence (1980–2001). Then, the cross-sectional distribution of urban population is characterized by means of nonparametric estimations of density functions and the growth process is modeled as a first-order stationary Markov chain. Spatial effects are finally introduced within the Markov chain framework using regional conditioning. This analysis shows a low interclass mobility, i.e., a high-persistence of urban municipalities to stay in their own class from one decade to another over the whole period, and the influence of the geographical environment on urban population dynamism.

Previous versions of this paper were presented at the 45th Congress of the European Regional Science Association (Vrije Universiteit, Amsterdam, Netherlands, August 23–27, 2005) and at the International Workshop on Spatial Econometrics and Statistics (Rome, Italy, May 25–27, 2006). We would like to thank two anonymous referees, M. Bosker, P. Cheshire, A. Carrington, B. Fingleton, R. Guillain, E. Lopez-Bazo, J. Paelinck and the other participants of these meetings for their valuable comments. Coro Chasco acknowledges financial support from the Spanish Ministry of Education and Science SEJ2006-14277-C04-01. The usual disclaimers apply.

J. Le Gallo (✉)

CRESE, Université de Franche-Comté, 45D, avenue de l'Observatoire, 25030 Besancon Cedex, France
e-mail: jlegallo@univ-fcomte.fr

C. Chasco

Departamento de Economía Aplicada, Universidad Autónoma de Madrid, 28049 Madrid, Spain
e-mail: coro.chasco@uam.es

Keywords Convergence · Urban growth · Spatial autocorrelation · Spatial SUR models · Markov chains

JEL Classifications C14 · C21 · O18

1 Introduction

Economic development is associated with the movement of population from the countryside to cities. This observation raises the question of how cities of different sizes grow during the process of development. The size distribution of cities may become more even over time if smaller cities catch up with larger ones. At the other extreme, urbanization may take the form of the expansion of the largest cities. In this case, the size distribution would become more unequal.

In this paper, we consider the particular experience of Spain between 1900 and 2001. Our aim is thus to improve the knowledge of the Spanish urban system and answer the following questions. How has the size distribution evolved over the last century? Has it become more even or more unequal? Is there a lot of mobility of cities within this size distribution? These questions are particularly relevant since the Spanish urbanization process has mainly taken place during the twentieth century producing significant processes of industrialization and economic growth. Specifically in Spain, this process has not been uniform and different results may be found depending on the definition of an “urban area”. In fact, there is no official definition of an “urban area” in Spain and it is not easy to obtain statistical data at the level of municipalities. Hence, analyses of the Spanish urban system are still scarce. Nevertheless, some authors have considered the group of “main cities” -above 50,000 inhabitants—as urban units (e.g. Lanaspa et al. 2003, 2004, Mella and Chasco 2006).

In this paper, we propose to work with the set of municipalities that form the Spanish “urban areas”, as defined by the [Ministerio de Fomento \(2000\)](#). It is a heterogeneous group of municipalities that not only includes the main cities but also all the satellite towns that make up the complete metropolitan area. We study the evolution of population growth among this set of 722 municipalities included in the present Spanish urban areas over the period 1900–2001.

In order to examine urban evolution and answer the preceding questions, we first examine the city size distribution by centering on the question of whether Zipf’s law or its deterministic equivalent, the rank-size rule, holds for Spanish cities. Zipf’s law has been applied in numerous studies (see [Gabaix and Ioannides 2004](#), for a recent review) but none of them considers the possibility of spatial effects. However, due to the geographical nature of the empirical data used, we emphasize in this paper the need to pay attention to the appropriate econometric methodology needed for a reliable statistical inference. Therefore, we formally test for spatial autocorrelation and spatial heteroskedasticity in the spatial SUR framework suggested by [Anselin \(1988\)](#).

This empirical work on the rank-size rule is essentially involved with one particular characteristic of the distribution of city sizes: the shape of that distribution. However, some papers have also paid attention to the intra-distribution dynamics ([Eaton and Eckstein 1997](#); [Black and Henderson 1999, 2003](#); [Lanaspa et al. 2004](#)). Indeed, as

Quah (1996) has forcefully argued, typical cross-section or panel data econometric techniques do not allow inference about patterns in the intertemporal evolution of the entire cross-sectional distribution. Making such inferences requires estimating directly the full dynamics of the entire distribution of cities. We therefore follow this strand of literature by focusing on how cities develop relatively to the rest of the urban system, both in terms of rankings and relative sizes. For that purpose, the cross-sectional distribution of urban population is analyzed by means of nonparametric estimations of density functions and the growth process is modeled as a first-order stationary Markov chain. The evolution of the shape of the population cross-sectional distribution and the changes in the municipalities' relative positions within this distribution is then able to uncover the existence of alternate divergence/convergence trends. Moreover, as in the analysis of Zipf's law, we also adopt an explicit spatial approach by measuring the extent to which the geographical environment influences the urban municipalities' relative position within the population cross-sectional distribution.

Hence, we extend previous studies focusing on the Spanish case (mainly Lanaspá et al. 2003, 2004) in two ways. First, not only do we consider a broader set of cities, but we also analyze the *complete set* of municipalities that actually belong to the Spanish urban areas: bigger cities and metropolitan towns. This feature allows for a better knowledge of the evolution of the complete urban system in Spain, which is not exactly the same as the one experienced by other groups of cities in the same country. Second, we explicitly introduce spatial dependence specifications and tests in both Zipf's law and Markov Chains analyses to capture the influence of space on convergence and transition probabilities and we perform a more complete analysis of movement speed and form of convergence in the city size distribution.

The paper is organized as follows. In the first section, the evolution of the disparities between the Spanish urban municipalities is characterized by examining the population cross-sectional distribution over the period from 1900 to 2001. We also test the validity of Zipf's law over this period. In the second section, we estimate a first-order Markov chain and analyze its ergodic properties. The article concludes with a summary of key findings.

2 The evolution of the Spanish urban system 1900–2001

This section examines growth in the Spanish urban system and changes in the relative size distribution of urban municipalities over a 100-year period.

2.1 Data

In order to explore these issues, we need a data set with urban areas defined consistently over the century. For that purpose, we have considered the classification proposed by the Spanish Ministry of Urbanism and Public Works (Ministerio de Fomento 2000). It divides the Spanish territory into urban areas, which include a set of 722 municipalities: (1) a set of 495 towns included in the 65 "Large Urban Areas" (areas above 50000 inhabitants each); (2) the group of 227 municipalities considered as "Small Urban Areas" (towns above 10000 inhabitants not included in the Large Urban Areas, with

minor corrections). This is rather different from the approach in Lanaspá et al. (2003), who operated with a sample of 100 largest cities¹ as proxies of the Spanish urban system.

Moreover, these 722 urban settlements are located across the whole Spanish territory: Andalusia (137), Aragón (10), Asturias (25), Balearics (13), Canary Islands (29), Cantabria (5), Castille and León (23), Castille-La Mancha (17), Catalonia (192), Valencian Community (96), Extremadura (11), Galicia (32), Madrid (33), Murcia (19), Navarre (18), Bask Country (58), Rioja (2), Ceuta and Melilla (2).

The evolution of population distribution is analyzed using the Census data over the period from 1900 to 2001. There are 11 decades under consideration: 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1991 and 2001. The data on population are extracted from the Spanish Office for Statistics (INE) databank.²

2.2 The evolution of the shape of urban population distribution

Using this dataset, we first analyze the evolution of the shape of urban size. For that purpose, we estimate non-parametric kernel density estimates of the urban population distribution for each decade and we analyze its unimodality or multimodality characteristics. More precisely, we examine relative urban municipality size distribution in 1900 and the way this distribution has changed over time in 1950, 1970 until 2001. Relative size distributions are considered, where size for each decade is normalized by dividing by the average municipal area size. Figure 1 shows the relative log urban municipality size distributions in 1900, 1950, 1970 and 2001. This density plot may be interpreted as the continuous equivalent of a histogram in which the number of intervals has been set to infinity and then to the continuum. From the definition of the data, 1 on the horizontal axis indicates Spanish average city size, 2 indicates twice this average, and so on.

Figure 1 plots an interesting graph where the distribution is bimodal in 1900 but becomes a unimodal distribution in 2001 (minor mode around 70–80%). This may reflect the existence, in 1900, of a group of urban municipalities with sizes below the average, converging toward a lower population level than the rest of the towns. Compared with 1900, more urban municipalities reported in 2001 population about the Spanish average. The distributions in 1900 and 1950 are quite similar, while the central mass significantly increased in 1970 to reach the highest point in the 2001 distribution. This progressive concentration of probability mass around 100% can be interpreted as evidence for slight convergence. This result is similar to others in the literature (Lanaspá et al. 2003, 2004, for Spanish largest cities; Anderson and Ge 2005, for Chinese cities), though differs from Black and Henderson (2003) results for US metropolitan areas.

¹ These authors chose a relatively arbitrary number of “largest cities” after finding that the results were qualitatively robust to different sample sizes. In a posterior paper, Lanaspá et al. (2004) chose different subsets with the 100, 300, 500 and 700 most-populated municipalities.

² This data are available in the INE webpage: <http://www.ine.es>.

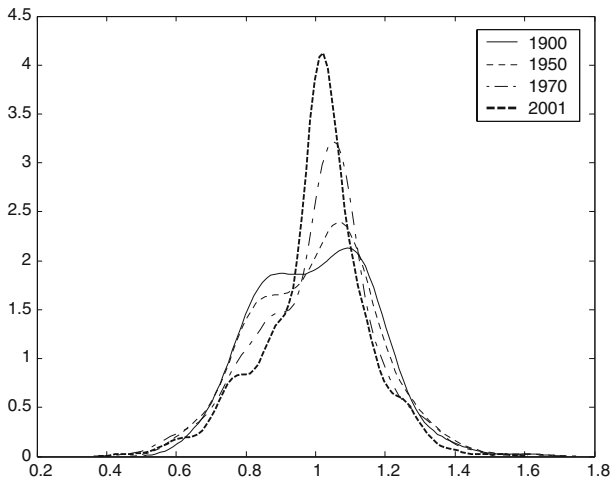


Fig. 1 Densities of log relative urban municipality size

2.3 Zipf's law, or the rank-size rule

We continue our exploration of the evolution of the Spanish urban municipality size distribution by using Zipf's law, or the rank-size rule. Zipf (1949) claimed that the size distribution of cities follows a Pareto law (Pareto 1897) when it holds that:

$$R = a \cdot S^{-b} \quad (1)$$

where R is the city rank order of the population distribution; S is the population of the cities; and a and b are parameters, with the latter being the Pareto exponent, always positive by construction.

The rank size rule, which emerged from regularly observed features of the data lacking any economic theoretic foundation, has recently been analyzed, among others, by Krugman (1996), Eaton and Eckstein (1997), Overman and Ioannides (2001), Dobkins and Ioannides (2000), Davis and Weinstein (2002), Ioannides and Overman (2003), Cordoba (2003), Rossi-Hansberg and Wright (2004), Gabaix and Ioannides (2004), Soo (2005), Gabaix and Ibragimov (2006). Gabaix (1999a,b) has derived a statistical explanation of Zipf's law for cities. He shows that if different cities grow randomly with the same expected growth rate and the same variance, the limit distribution of city size will converge so as to obey Zipf's law. Duranton (2006) provides some economic foundations for Zipf's law: he embeds the endogenous growth model suggested by Grossman and Helpman (1991) into an urban framework and views investments in R&D as the main driver of city growth. This model can then generate Zipf's law under some particular condition detailed in the paper. Finally, Nitsch (2005) carries out a meta-analysis combining 515 estimates from 29 studies and finds that cities are on average more evenly distributed than suggested by Zipf's law.

Formally, in this framework, the size distribution of cities is more or less even, depending on the value of the Pareto exponent (b). At the limit, if b tends to infinity, then

all the cities will be of an equal size. When b is equal to one, we obtain the well-known rank-size rule or Zipf's law. According to this rule, city populations among any group of cities at any time are proportional to the inverse of the ranking of their populations in that group. The Pareto exponent can therefore be interpreted as a convergence indicator. Indeed, values that fall over time indicate relatively more important roles (increasing weights) for the largest cities. More precisely, as b decreases, a 1% increase in city size produces a smaller fall (in %) in rank and the city size distribution becomes more spread out. Therefore, this will cause a divergence trend inside the group of urban municipalities or greater metropolitan concentration. Likewise, a 1% increase in city size produces a larger fall (in %) in rank as b increases. Therefore, increasing values of the Pareto exponent represent convergence dynamics, or in other words, greater dispersion of the population outside the large metropolitan areas and a more balanced population distribution between urban centers of different sizes.

Empirically, departing from Eq. (1), we take logarithms on both sides and estimate the resulting linear expression for the set of 722 urban municipalities ($i = 1, \dots, 722$) for each of the eleven decades ($t = 1, \dots, 11$) under consideration:

$$\ln R_{it} = \ln a_t - b_t \cdot \ln S_{it} + \varepsilon_{it} \quad (2)$$

Gabaix and Ioannides (2004) have shown by Monte-Carlo simulations that OLS estimation of this equation presents several pitfalls in small samples.³ Moreover, OLS may be affected by the omission of spatial autocorrelation. More precisely, if spatially autocorrelated residuals represent the effects of an unmodelled spatially autoregressive error process, then the parameter estimates remain unbiased but become inefficient. Statistical inference is biased in this case. Conversely, if they are due to the omission of spatial autocorrelated variables, then the parameter estimates are biased.

Therefore, and since we have allowed for different intercepts and slopes in each period, we have followed the strategy suggested by Anselin (1988, pp. 203) for the specification of spatial SUR models. In a first stage, we have estimated Eq. (2) by Ordinary Least Squares (OLS) for the 11 spatial equations individually considered (i.e., one equation for each decade under consideration). For each model, we have tested for the presence of spatial effects. As shown in Table 1, the OLS residuals of the 11 equations are non-normal and exhibit both heteroskedasticity and spatial autocorrelation (as pointed out by Jarque-Bera, Koenker-Basset and Kelejian-Robinson tests, respectively). Therefore, we can conclude that both spatial effects (spatial autocorrelation and spatial heterogeneity characterized by heteroskedasticity) are present in the 11 models. We will focus in the remaining discussion on the former problem, leaving an explicit treatment of heteroskedasticity problems for further research. Regarding spatial autocorrelation, the non-normality of the error terms implies that the

³ Gabaix and Ibragimov (2006) also point out that the OLS estimator of b in equation (2) is strongly biased in small samples. To overcome this problem, they provide a simple practical remedy and show, by Monte-Carlo simulations, that this bias is considerably reduced when using the Rank $-1/2$, and run $\log(\text{Rank} - 1/2) = \alpha - b \log(\text{Size})$. Although we do not have a small sample and we use estimation methods other than OLS, we have also estimated equation (2) introducing this change. As expected, the results obtained are robust to the modification suggested by Gabaix and Ibragimov, and they are available upon request from the authors.

Table 1 Rank-size regressions, Spanish urban municipalities 1900–2001

	OLS Basic model				Spatial SUR model (ML)				Spatial SUR spatial-error model (ML)				Spatial SUR spatial lag model (ML)							
	$\hat{\alpha}$	\hat{b}	JB	KB	KR	$\hat{\alpha}$	\hat{b}	JB	KB	KR	$\hat{\alpha}$	\hat{b}	$\hat{\lambda}$	JB	KB	KR	$\hat{\alpha}$	\hat{b}	$\hat{\rho}$	
1900	11.04	-0.66	728	116	188	10.66	-0.62	728	116	188	10.68	-0.62	0.16	728	116	188	10.43	-0.62	0.04	
1910	11.06	-0.66	748	111	249	10.65	-0.61	748	111	249	10.67	-0.61	0.19	748	111	249	10.29	-0.60	0.05	
1920	11.04	-0.65	794	100	287	10.59	-0.60	794	100	287	10.61	-0.60	0.19	794	100	287	10.18	-0.59	0.06	
1930	11.05	-0.64	668	89	313	10.55	-0.58	668	89	313	10.56	-0.58	0.20	668	89	313	9.99	-0.57	0.08	
1940	10.91	-0.62	757	85	418	10.40	-0.56	757	85	418	10.41	-0.56	0.22	757	85	418	9.85	-0.55	0.08	
1950	10.82	-0.60	682	82	493	10.30	-0.54	682	82	493	10.32	-0.54	0.22	682	82	493	9.69	-0.53	0.09	
1960	10.81	-0.59	616	62	540	10.30	-0.53	616	62	540	10.32	-0.53	0.23	616	62	540	9.66	-0.52	0.09	
1970	10.78	-0.57	614	37	580	10.31	-0.52	614	37	580	10.32	-0.52	0.26	614	37	580	9.48	-0.50	0.12	
1981	10.67	-0.54	721	31	509	10.21	-0.50	721	31	509	10.23	-0.50	0.27	721	31	509	9.25	-0.48	0.14	
1991	10.90	-0.56	744	25	369	10.39	-0.51	744	25	369	10.41	-0.51	0.25	744	25	369	9.51	-0.49	0.13	
2001	11.39	-0.60	837	13	237	10.83	-0.54	837	13	237	10.87	-0.55	0.30	837	13	237	9.81	-0.52	0.14	
Diagonality tests																				
LM test:	30.274																			
LR test:	22.531																			
Wald homogeneity test	2.928																			
b parameter:	687																			
spatial parameter:	616																			
Spatial dependence	21																			
LM-spatial-error:	217																			
LM-spatial lag:	81																			
Goodness of fit																				
LIK	6.833																			
AIC	26.34																			
$\alpha = \ln \alpha$	6.778																			
	48.36																			

OLS Ordinary least squares estimation, ML Maximum likelihood estimation. All coefficients are significant at the 1% level, JB The Jarque–Bera non-normality test on the residuals, KB The Koenker–Basset test for heteroskedasticity, KR The Kelejian–Robinson test for spatial autocorrelation in the error term, LIK The log-likelihood ratio test, AIC The Akaike Information Criterion. All statistics lead to the rejection of the corresponding null hypothesis at 1%

Lagrange Multiplier (LM) tests are less reliable. The Kelejian–Robinson statistics—though highly significant—cannot orient towards a spatial lag or spatial error formulation.

In addition, we can also test for the existence of temporal correlation between the 11 equations under the form:

$$E[\varepsilon_t \varepsilon_s'] = \sigma_{ts} I_N \text{ with } s, t = 1, \dots, 11 \quad (3)$$

where ε_t is an $(N, 1)$ vector containing the N error terms for time period t and $N = 722$. This assumption of dependence between equations can be tested for by means of a LM test or a likelihood ratio (LR) test of the diagonality of the error covariance matrix. Note that this specification differs from the most familiar SUR design (Zellner 1962) with N fixed and $T \rightarrow \infty$, where the regression coefficients are assumed to vary by cross-sections (but are constant over time) and where the error terms are contemporaneously correlated. When the cross-sectional units pertain to spatial units, this latter assumption allows estimating nonparametrically cross-sectional dependence, interpreted as spatial autocorrelation, which is left unspecified as a general covariance (see Hordijk and Nijkamp 1977; White and Hewings 1982 for applications). In our case, $N > T$, so that the standard SUR is not appropriate and spatial autocorrelation should instead be expressed as a parameterized function.

The SUR model can be estimated using FGLS or maximum likelihood (ML). The latter corresponds to iterated FGLS, yielding consistent and asymptotically normal estimates under the assumption of normality of errors. Spatial autocorrelation can be incorporated either in the form of a spatial lag or in the form of a spatial error term. In the first case, the model can be written as follows:

$$\ln R_{it} = \rho_t \cdot \sum_{j=1}^N w_{ij} \ln R_{jt} + \ln a_t - b_t \cdot \ln S_{it} + \varepsilon_{it} \quad (4)$$

where the error terms are as in (3) and where w_{ij} is an element of a spatial weights matrix W . It is equal to 1 if urban municipality i is, at most, 160 km away from urban municipality j . The role of the spatial weights matrix is to introduce the notion of a neighborhood set for each of the urban municipalities.⁴ This model can also be estimated with ML. Note that for the spatial lag spatial SUR model, the three stage least squares estimation method has also been suggested when the assumption of normality is untenable and/or to avoid the computational problems associated with the Jacobian term in the ML estimation. However, in this case, appropriate instruments must be found and the estimation can yield explosive estimation of the spatial parameter, whereas it remains bounded with ML (see Anselin 1988; Anselin et al. 2007 for further technical details and Fingleton 2001 for an empirical application). Conversely,

⁴ We have also used a contiguity spatial weights matrix using a Thiessen polygonalization of the Spanish territory for the 722 urban municipalities. The results are similar and can be obtained upon request from the authors.

the spatial error SUR model is as follows:

$$\begin{cases} \ln R_{it} = \ln a_t - b_t \cdot \ln S_{it} + \varepsilon_{it} \\ \varepsilon_{it} = \lambda_t \cdot \sum_{j=1}^N w_{ij} \varepsilon_{jt} + u_{it} \end{cases} \quad (5)$$

where u_{it} are in the form of (3). Again, this model can be estimated with ML. Two LM tests, LMERR for spatial error and LMLAG for spatial lag, can be computed on the residuals of the spatial SUR model (Eqs. (2), (3)) in order to discriminate between a spatial lag or a spatial error specification (Anselin 1988). Moreover, the temporal stability of the coefficients ($\alpha_t = \ln a_t$ and/or b_t) and of the spatial coefficients (ρ_t or λ_t) can be tested for in Eqs. (2), (4) and (5). However, due to the presence of spatial autocorrelation, the Wald tests used in this case must be adjusted.

The computation of the LM and LR diagonality tests of the error covariance matrix - as well as the Wald test on the homogeneity of the parameters across equations - points out to the superiority of a SUR specification over 11 individual equations (see Table 1). On the other side, both LM tests on spatial dependence reject the null of no spatial autocorrelation. The higher value of the LM test for spatial-error dependence is an indication that the spatial SUR spatial-error model (5) is more appropriate than a spatial lag one (4). Moreover, it shows a better performance in terms of goodness of fit (higher LIK and lower AIC).

Figure 2 displays the evolution through time of the three estimations of the Pareto exponent. Though the OLS estimators for this parameter are always higher, they follow a similar evolution. From Fig. 2, it is clear that, in general terms, the estimation over time of the b parameter displays a decreasing trend until 1980, from which it starts to increase. As a result, we can deduce two different patterns over the course of the twentieth century: from 1900 to 1980, the size distribution of the set of 722 urban municipalities is increasingly divergent while from the 1980s to the end of the period this distribution becomes progressively even. Looking in more depth, we can also distinguish two other sub-periods in the first moment: from 1900 to 1930 (smaller divergence) and 1930–1981 (steeped divergence). This result is more or less consistent with Lanaspá et al. (2003) that found an inflexion date in the 1970s for the group of 100 Spanish largest cities. In our case, we find evidence of inflexion in the 1980s—instead of the 1970s—due to the composition of the sample. Indeed, when considering the whole set of urban units—not only a group of metropolises—we can capture suburbanization or deconcentration of individual processes, which were common in industrialized countries during the last decades of the twentieth century (Stanback 1991).⁵ In the case of Spain, though larger cities started to lose some population during the 1970s, we find evidence for a general phenomenon of *counter-urbanization* during the 1980s, when the declining process of traditional monocentric

⁵ Fielding (1989) demonstrated that the change from urbanization to counter-urbanization occurred in Spain during the 1980s, some years later than in most countries in Western Europe. Monclús (1997) and Esteve and Devolder (2004) also reached the same conclusion when analyzing urban growth for different sets of municipalities in Catalonia: the inflexion towards convergence in urban areas took place mainly during the 1980s as a consequence of a broad range of political and socio-economic changes.

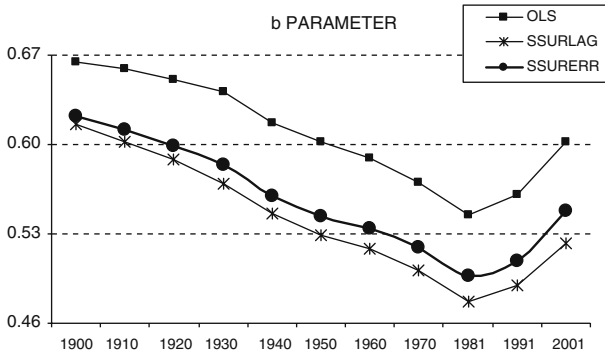


Fig. 2 Evolution of the estimations of the Pareto exponent ($N = 722$). *OLS* OLS estimation; *SSURLAG* Spatial SUR lag model; *SSURERR* Spatial SUR spatial-error model

cities clearly benefited the peripheral cities, leading to the modern multicentric city. This fact can be interpreted as a consequence of a broad range of historical, socio-political and economic changes: end of the Franco dictatorship, generalized increase of family income, return of emigrants, crisis in housing supply inside the core cities, demand of more space for housing and industries, improvement in accessibility and motorization, etc.

Consequently, the analysis of Zipf's law leads to an interesting result, i.e., the existence of two main phases in the evolution of Spanish urban municipalities. The main one, which extends over 80 years, consists in an increase of urban concentration only broken after the 1980s. Inside this first stage, we can distinguish two sub-periods, in which the divergence course between urban municipalities has different speeds:

- From 1900 to the 1930s, the b parameter displays a slower decreasing trend coinciding with a significant industrialization and urbanization expansion that led to progress and social changes in Spain. In the first decade, though most of the active population was located in the countryside, labor force began to migrate to the main industrial cities, e.g. Barcelona and Bilbao, as well as to Madrid and Valencia. Neutrality during World War I and capital stock growth (provided by American and international investments) helped the development of some industrial activities (only located in certain cities) that demanded more workers (Tuñón de Lara et al. 1982). Moreover, during the 1920s, industrialization and urbanization went on growing, especially in the Axis Madrid-North-Barcelona, leading to an incipient development of other satellite towns along the Cantabric Coast (Bilbao Estuary area, Santander, Asturian cities) and the Mediterranean Coast (Valencia and Alicante). However, during the mid 1930s, the economic crisis and the Civil War stopped the urbanization process (Tuñón de Lara and Malerbe 1982).
- From 1940 to the 1970s, the b parameter experiences a quicker decline or, in other words, during this period the largest cities grew at significantly greater rates than the smallest population nuclei, exhibiting an intense divergent growth pattern. Indeed, during the 1940s, Spain lived under an autarkical regime that led to a real ruralization process: the main cities, destroyed after the Civil War, had to be re-built, hunger and poverty expelled a lot of people to the villages and, in general, urban population and active population decreased significantly.

Nevertheless, some big cities grew a lot, such as Madrid (due to the huge centralization and bureaucratization of the *Regimen*), Barcelona and other capitals (Valencia, Saragossa, Alicante and Seville). The incipient political and economic openness during the 1950s stopped the ruralization drive and set the basis for the decisive industrialization and tertiarization process experienced during the 1960s and 1970s (Tuñón de Lara and Viñas 1982). The industrial sector was severely constrained to make it more competitive and many workers had to migrate to Europe or to the Spanish capitals and new economic centers. Development was geographically irregular and affected only the cities located in richer provinces: Guipúzcoa, Biscay, Barcelona, Navarre, Madrid and Álava. Nevertheless, the Development Plans also created new economic poles, such as Vigo, Pontevedra, Coruña and Ferrol (in Galicia), Valladolid and Burgos (in Castile), Huelva, Cádiz, Seville (in Andalusia), Saragossa (in Aragón) and Badajoz (in Extremadura). If in 1960 only 30% of Spanish population lived in cities above 100,000 inhabitants, in 1975, urban population rose to 50%: Spain was no longer rural and became an industrial and urban country (Fusi et al. 1983).

During the two last decades of the twentieth century, the Zipf's parameters change from the 80-year decreasing tendency to a noteworthy increasing one. In other words, the group of 722 urban municipalities displayed a clear convergence growth pattern as the smallest towns grew faster than the largest cities. Actually, Spain went through a strong counter-urbanization process that is not finished yet. By the beginning of the 1980s, there was a peculiar urban structure similar to a star, with its centre in Madrid. In the axis, there were the vast Mediterranean metropolitan areas (Girona-Barcelona-Tarragona, Castellón-Valencia-Alicante-Murcia), Andalusia (Seville and Cádiz), Galicia (A Coruña-Ferrol, Vigo) and the Cantabric Coast (Bilbao-San Sebastián, Santander, Gijón-Oviedo). In addition, inside this big star, there was a vast rural desert, only broken by a few urban oases, like Valladolid, Saragossa, Badajoz, Burgos, Vitoria and Pamplona. In the Islands, there was a similar process due to the huge growth of Palma (the Balearics), Las Palmas and Santa Cruz de Tenerife (the Canary Islands). The cities of Madrid and Barcelona grew towards their respective peripheries as did Valencia and Bilbao, although to a lesser extent. Indeed, the whole Basque Country was declared an "urban area", as well as the Oviedo-Gijón-Avilés triangle (in Asturias) and the cities along the Mediterranean coast from Tarragona (in Catalonia) to Cartagena (in Murcia).

The logical problems of the big cities (with uncontrolled growth in the peripheries and an incipient depopulation process of their historical centers) broke their later expansion in favor of middle-sized even small-cities and certain rural areas. Moreover, this de-urbanization of the largest cities was accompanied by some growth in their neighboring towns: suburban settlements gained many inhabitants and city centers were depopulated, restored and converted into CBD's and/or historical/cultural cores.

3 Mobility within the Spanish urban system 1900–2001

The density functions and Zipf's law allow the characterization of the evolution of the global distribution, but they do not provide any information about the movements of

the urban municipalities within this distribution. For example, they do not say whether the right tail of the initial distribution (year 1900) contains the same municipalities as the right tail in the final distribution (year 2001). A possible way to answer these questions is to track the evolution of each urban municipality's relative size over time by estimating transition probability matrices associated with discrete Markov chains (Kemeny and Snell 1976). This line of analysis has been pursued by Eaton and Eckstein (1997) for Japanese and French urban areas and by Black and Henderson (1999, 2003) for the US urban system.

3.1 Markov chains

The analysis of the evolution in time of an entire cross-section distribution, or distribution dynamics analysis, is a method aimed at describing the law of motion of the distribution as a Markovian stochastic process. In that respect, working in a discrete state–space has several advantages, as argued by Bulli (2001). Indeed, compared to continuous stochastic kernels,⁶ discrete probability distribution and transition matrices are easier to interpret: various descriptive indices and the long-run or ergodic distribution are easier to compute. On the other hand, this methodology raises the problem of arbitrary discretization. We will develop this latter problem when presenting the empirical results.

Formally, denote F_t the cross-sectional distribution of municipal size (population) at time t relative to the Spanish average. Define a set of K different size classes, which provide a discrete approximation of the population distribution. We first assume that the frequency of the distribution follows a first-order stationary Markov process. In this case, the evolution of the municipal size distribution is represented by a transition probability matrix, M , in which each element (i, j) indicates the probability that a municipality that was in class i at time t ends up in class j in the following period.⁷ Formally, the $(K, 1)$ vector F_t , indicating the frequency of the urban municipalities in each class at time t , is described by the following equation:

$$F_{t+1} = M F_t \quad (6)$$

where M is the (K, K) transition probability matrix representing the transition between the two distributions as follows:

$$M = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1K} \\ p_{21} & p_{22} & \dots & p_{2K} \\ \dots & \dots & \dots & \dots \\ p_{K1} & p_{K2} & \dots & p_{KK} \end{bmatrix} \quad (7)$$

⁶ For examples of studies using continuous stochastic kernels, see, among others, Quah (1997), Johnson (2001), Fingleton and Lopez-Bazo (2003) and the references therein.

⁷ The so-called Markov property implies that the future of a process depends only on its present class and not on its history.

where each element $p_{ij} \geq 0$, $\sum_{j=1}^K p_{ij} = 1$. The stationary transition probabilities p_{ij} capture the probability that a municipality in class i in $t - 1$ ends up in class j in t .

The elements of M can be estimated from the observed frequencies in the changes of class from one period to another. Thus, following Amemiya (1985) or Hamilton (1994), the maximum likelihood estimator of p_{ij} is:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} \quad (8)$$

where n_{ij} is the total number of urban municipalities moving from class i in decade $t - 1$ to class j in the immediate following decade t over all the ten transitions and n_i is the total sum of municipalities ever in i over the ten transitions.

If the transition probabilities are stationary, that is, if the probabilities between two classes are time-invariant, then:

$$F_{t+s} = M^s F_t \quad (9)$$

In this framework, one can determine the ergodic distribution (also called the long-term, long-run, equilibrium or steady state distribution) of F_t , characterized when s tends toward infinity in Eq. (9), that is to say, once the changes represented by matrix M are repeated an arbitrary number of times. Such a distribution exists if the Markov chain is regular, that is, if and only if, for some m , M^m has no zero entries. In this case, the transition probability matrix converges to a limiting matrix M^* of rank 1. The existence of an ergodic distribution, F^* , is then characterized by:

$$F^* M = F^* \quad (10)$$

This vector F^* describes the future distribution of the urban municipalities if the movements observed in the sample period are repeated to infinity. Each row of M^t tends to the limit distribution as $t \rightarrow \infty$. According to Eq. (10), this limit distribution is therefore given by the eigenvector associated with the unit eigenvalue of M .

The assumption of a first-order stationary Markov process requires the transition probabilities, p_{ij} , to be of order 1, that is, to be independent of classes at the beginning of previous periods (at time $t - 2, t - 3, \dots$). If the chain is of a higher order, the first-order transition matrix will be misspecified. Indeed, it will contain only part of the information necessary to describe the true evolution of population distribution. Moreover, the Markov property implicitly assumes that the transition probabilities, p_{ij} , depend on i (i.e., that the process is not of order 0).

In order to test this property, Bickelbach and Bode (2003) emphasize the role of the test of time independence. In determining the order of a Markov chain, Tan and Yilmaz (2002) suggest, firstly, to test order 0 versus order 1; secondly, to test order 1 versus order 2; and so on. If the test of order 0 against order 1 is rejected, and the test of order 1 against order 2 is not rejected, the process may be assumed to be of order 1.

To test for order 0, the null hypothesis $H_0 : \forall i : p_{ij} = p_j (i = 1, \dots, K)$ is tested against the following alternative $H_a : \exists i \setminus p_{ij} \neq p_j$. The appropriate likelihood ratio

(LR) test statistic reads as follows:

$$LR^{(O(0))} = 2 \sum_{i=1}^K \sum_{j \in A_i} n_{ij}(t) \ln \frac{\hat{p}_{ij}}{\hat{p}_i} \sim asy\chi^2 \left[(K-1)^2 \right] \quad (11)$$

assuming that $\hat{p} > 0, \forall j(j = 1, \dots, K)$. $A_i = \{j : \hat{p}_{ij} > 0\}$ is the set of nonzero transition probabilities under H_a .

To test for order 1 versus 2, a second-order Markov chain is defined by also taking into consideration the population size classes $k(k = 1, \dots, K)$ in which the municipalities were at time $t - 2$ and assuming that the pair of successive classes k and i forms a composite class. Then, the probability of an urban municipality moving to class j at time t , given it was in k at $t - 2$ and in i at $t - 1$, is p_{kij} . The corresponding absolute number of transitions is $n_{kij}(t)$, with the marginal frequency being $n_{ki}(t - 1) = \sum_j n_{kij}(t - 1)$.

To test $H_0 : \forall k : p_{kij} = p_{ij}(k = 1, \dots, K)$ against $H_a : \exists k : p_{kij} \neq p_{ij}$, the p_{kij} are estimated as $\hat{p}_{kij} = n_{kij}/n_{ki}$, where $n_{kij} = \sum_{t=2}^T n_{kij}(t)$ and $n_{ki} = \sum_{t=2}^T n_{ki}(t - 1)$. The p_{ij} are estimated from the entire data set as $\hat{p}_{ij} = n_{ij}/n_i$. The appropriate LR test statistic reads as follows:

$$LR^{(O(1))} = 2 \sum_{k=1}^K \sum_{i=1}^K \sum_{j \in C_{hi}} n_{kij} \ln \frac{\hat{p}_{kij}}{\hat{p}_{ij}} \sim asy\chi^2 \left[\sum_{i=1}^K (c_i - 1)(d_i - 1) \right] \quad (12)$$

Similar to the notation above, $C_i = \{j : \hat{p}_{ij} > 0\}$, $c_i = \#C_i$, $C_{ki} = \{j : \hat{p}_{kij} > 0\}$ and $d_i = D_i = \#\{k : n_{ki} > 0\}$.

If both Markovity of order 0 and of order 1 are rejected, the tests can be extended to higher orders by introducing additional dimensions for population size at time $t - 3, t - 4$, and so on. However, since the number of parameters to be estimated increases exponentially with the number of time lags, while the number of available observations decreases linearly for a given data set, the reliability of estimates and the power of the test decrease rapidly. Therefore, [Tan and Yilmaz \(2002\)](#) suggest setting an a priori limit up to which the order of the Markov chain can be tested.

3.2 Empirical results

In order to carry out the methodology described above, a discretization of the continuous state-space must be chosen. However, as pointed out by [Magrini \(1999\)](#), [Bulli \(2001\)](#) or [Cheshire and Magrini \(2000\)](#), an improper discretization may have the undesired effect of removing the Markov property and therefore may lead to very misleading results, especially when the computation of ergodic distributions are based on the estimates of the discrete transition probabilities. Some authors ([Quah 1993](#); [Lopez-Bazo et al. 1999](#); [Kawagoe 1999](#) or [Le Gallo 2004](#)) choose to discretize the distribution in such a way that the initial classes include a similar number of individuals. Conversely, [Magrini \(1999\)](#) or [Cheshire and Magrini \(2000\)](#) base their choice between possible

Table 2 Probability transition matrix, 1900–2001: Spain-relative population size

	1 <20%	2 <50%	3 <80%	4 <135%	5 <185%	6 >185%	Number of observations
1	0.944	0.053	0.002	0	0	0	2567
2	0.040	0.879	0.074	0.005	0.002	0.001	1751
3	0	0.162	0.752	0.078	0.005	0.003	1029
4	0	0.001	0.184	0.741	0.066	0.008	852
5	0	0	0	0.273	0.632	0.095	315
6	0	0	0	0.001	0.061	0.938	706

classes in terms of the ability of the discrete distribution to approximate the observed continuous distribution.

In this paper, we have tried numerous ways of discretizing the distribution, with different numbers of classes (5, 6, 7). Finally, the discretization has been chosen by considering the best performance of the test for order one, though we have tried to set up balanced classes even if it comes at some cost to this test. We distinguish between six different classes: (1) population less than 20% of the Spanish average, (2) population between 20 and 50% of the Spanish average, (3) population between 50 and 80% of the Spanish average, (4) population between 80 and 135% of the Spanish average (5) population between 135 and 185% of the Spanish average, and (6) population more than 185% of the Spanish average.

Table 2 contains the first-order transition probability matrix between 1900 and 2001 with the ML estimates \hat{p}_{ij} of the transition probabilities for population. For example, during the century, there were 2,567 instances of an urban municipality having a population size lower than 20 percent of the Spanish average. The majority of these municipalities (94.4%) remained in that size class at the end of the year, while 5.3% moved up one class by the end of the year. Note also that the transition probability matrix is regular.

For the process in Table 2, Markovity of order 0 is tested by comparing each row of the transition matrix to the population distribution at time t using the test statistic (11). The result ($LR = 16602.90$; $prob = 0$; $df = 25$) leaves no doubt that the process strongly depends on the initial condition at time $t - 1$, i.e., that the chain is at least of order 1.

To test Markovity of order 1, six subsamples $k = 1, \dots, 6$ are defined, representing the urban municipalities' size at time $t - 2$. Observations for municipalities that were in the first size class at time $t - 2$ are allocated to the first subsample ($k = 1$) and so on. For each of these subsamples, a separate matrix is estimated for observed transitions from time $t - 1$ to t in the usual way. The general test comparing the matrices for all five subsamples to the matrix for the entire sample simultaneously, similar to Eq. (12) above, results in $LR = 198.12$. This statistic is significant with 63 degrees of freedom ($prob = 0$), indicating that the process under consideration is of a higher order, at least of order 2, if Markovian at all. However, there are a number of classes within subsamples for which we cannot expect reliable estimates of transition probabilities because there are only very few observations available. In addition, Fingleton (1983a,b, 1986, 1999) has argued that tests for Markov chains are inflated in the presence of

spatial autocorrelation, which is proved to exist in our case. Therefore, we decided to keep the assumption of order 1 for the Markov chain.

We can make several comments about this matrix in Table 2, related to interclass movements, mobility speed, convergence pattern and influence of space.

First, the high probabilities on the diagonal show a low interclass mobility, i.e., a high-persistence of urban municipalities to stay in their own class from one decade to another over the whole period. Diagonal elements of the transition approaching 1 have been interpreted as parallel growth by Eaton and Eckstein (1997). However, since these elements are not exactly 1, we can analyze the propensity of cities in each cell to move into other cells. In particular, it appears that the largest and smallest urban municipalities (classes 1 and 6, respectively) have higher persistence while medium-sized cities (categories 3, 4 and 5) have more probability of moving to smaller categories. In addition, in classes 2 and 3, a small number of urban municipalities move up to higher categories more than two steps, even reaching the top (some towns within the Madrid, Barcelona and Bilbao metro areas), while they only move down one cell. Nevertheless, only in class 2 the probability of moving up a class exceeds that of moving down.

This low inter-class mobility of urban municipalities is in line with the results found for other cases such as US MSA's (Black and Henderson 2003) or the largest Spanish cities (Lanaspa et al. 2003): changes in city-size are not drastic over a decade. However, the highest persistence of both the largest and smallest urban municipalities to stay in the initial state highlights the major role of medium-sized towns (10,000 to 70,000 inhabitants) in the processes of urban agglomeration and suburbanization that occurred in Spain during the twentieth century. In other words, on the one hand, a group of cities progressively moved to lower states in the distribution: cities located in the Northwest (Asturias and Galicia), some Castilian provinces inland (mainly Ciudad Real) and in the South of Spain (Andalusia, Alicante and Murcia). On the other hand, another group of medium-sized towns was thriving in Madrid, Barcelona, Bilbao and Seville metro areas, as well as some tourist enclaves in The Balearics, Canary Islands, Comunidad Valenciana, Almería and Málaga. These results demonstrate the existence of some spatial regularity in the main urban changes.

Second, in order to determine the speed with which the urban municipalities move within the distribution, we consider the matrix of mean first passage time M_P , where one element M_{Pij} indicates the expected time for a region to move from class i to class j for the first time. For a regular Markov chain, M_P is defined as (Kemeny and Snell 1976, Chap. 4):

$$M_P = (I_K - Z + ee'Z_{dg}) D \quad (13)$$

where I_K is the identity matrix of order K , Z is the fundamental matrix: $Z = (I_K - M + M^*)^{-1}$, M^* is the limiting matrix, e is the unit vector, Z_{dg} results from Z setting off-diagonal entries to 0, and D is the diagonal matrix with diagonal elements $1/m_j^*$.

Table 3 displays the mean first passage time matrix for population. The mean number of years to reach any class is relatively high: the shortest time passage is 91.9 years and the longest is 3110.7 years. Globally, movements up are slower than movements

Table 3 Mean first passage time matrix in decades, 1900–2001: Spain-relative population

	1 <20%	2 <50%	3 <80%	4 <135%	5 <185%	6 >185%
1	3.93	18.67	39.00	78.88	158.83	311.07
2	51.88	2.82	21.99	62.01	142.50	294.48
3	63.87	11.99	5.53	45.12	127.61	279.61
4	73.55	21.67	9.90	10.20	95.73	251.82
5	82.74	30.86	19.08	9.19	28.87	189.41
6	98.57	46.69	34.92	25.02	18.22	12.89

down, especially for high-size classes, i.e., the expected time to first move from class 5 to class 6 is 1894.1 years. Remember that these calculations account for the fact that starting from class 5, a site might visit classes 4, 3, 2 or 1 before going to class 6. From class 1 it takes 3110.7 years to first visit class 6, with the outstanding upward mobility of some metropolitan cities in the Madrid metro area (Alcobendas, Alcorcón, Coslada, Móstoles, Parla and Torrejón de Ardoz) and the Barcelona metro area (Santa Coloma de Gramenet) as examples of actual moves from state 1 to state 6.

This result of faster declines shows that urban municipalities are more likely to lose population than to gain it, especially inland, in big capitals and old industrial centers (e.g. several cities in Andalusia, Asturias, Comunidad Valenciana and Murcia).⁸ This finding leads to the conclusion of the existence of a general progressive suburbanization process, as in other modern post-industrialized countries (Blakman et al. 1999), which has put an end to the era of big-city growth, favoring the progressive appearance of smaller population nuclei enjoying lower levels of congestion. This conclusion is also compatible with the 80-year phase of divergence, in size, between urban municipalities, only reversed during the last two decades, as pointed out by the Zipf's parameter in Fig. 2.

Third, we consider the ergodic distribution that can be interpreted as the long-run equilibrium urban municipality-size distribution in the urban areas system. Explicitly, given a regular transition matrix, with the passage of many periods, there will be a time where the distribution of urban municipalities will not change any more: that is the ergodic or limit distribution. It is used to assess the form of convergence in a distribution. Concentration of the frequencies in a certain class would imply convergence (if it is the middle class, it would be convergence to the mean), while concentration of the frequencies in some of the classes, that is, a multimodal limit distribution, may be interpreted as a tendency towards stratification into different convergence clubs. Finally, a dispersion of this distribution amongst all classes is interpreted as divergence.

Ergodic distributions are computed for population size in Table 4. It appears that the ergodic distribution is more concentrated in the small-size municipalities (1st and 2nd classes), a result that reveals the existence of convergence towards smaller-size populations. In addition, we find stability of ergodic distribution compared to the initial one, though there is slightly more probability in category 2. This outcome points to a very slight downward convergence, a result compatible with the kernel density

⁸ Again, this result contrasts with the US metro areas behavior (Black and Henderson 2003, pp. 358).

Table 4 Initial versus ergodic distributions 1900–2001: Spain-relative population size

	1 < 20%	2 < 50%	3 < 80%	4 < 135%	5 < 185%	6 > 185%
Initial distribution	0.356	0.243	0.143	0.118	0.044	0.098
Ergodic distribution	0.254	0.355	0.181	0.098	0.035	0.078

function (Fig. 1) and transition matrix (Table 2) results. As we have mentioned earlier, the choice of a discretization may have a heavy impact on the determination of the ergodic distribution. Consequently, we have also computed other ergodic distributions with different discretization methods and different numbers of classes. For example, with 6 classes and the same number of individuals in each class, there is an increase in probability in the ergodic distribution (compared to the initial distribution) for the classes containing the municipalities with population less than 62% of the Spanish average. When the discretization is chosen so that it minimizes the value of the chi-square test of order 1 versus order 0, there is again an increase in probability for the classes containing the municipalities with a population less than 74% of the Spanish average. In fact, in every configuration we have tried, we observe an increase in probability in the ergodic distribution for the classes containing the municipalities with a population less than 60 to 75% of the Spanish average. We can therefore conclude that our result of slightly downward convergence is robust to the choice of discretization.

Fourth, we have analyzed the influence of space on the transition probabilities, as in Le Gallo (2004) and Rey (2001). The relationship between the direction of an urban municipality's transition in the population distribution and the relative populations of its neighbors is considered more generally in Table 5. The probability of a particular transition (Down, None, or Up) conditioned on the populations of the urban municipality's neighbors at the beginning of the year is reported. There is clear evidence that the probability of an upward or downward move is different depending on the urban area context. For example, the probability for an urban municipality of moving up in the hierarchy is 7.1% when its spatial lag contains on average less population whereas it is 8% when it contains on average more population. Conversely, the probability for an urban municipality of moving down in the hierarchy is 18.9% when its spatial lag contains on average less population whereas it is only 3.9% (almost five times lower) when it contains on average more population.

Specifically, the highest degree of spatial autocorrelation is of positive sign and occurs in nuclei with downward movements that are surrounded by smaller-size neighbors (18.9%). This situation takes place in declining areas with high levels of

Table 5 Transition probabilities conditioned on the spatial lag of population

Spatial lag	Move		
	Down	Same	Up
Less population	0.189	0.740	0.071
Same	0.057	0.898	0.045
More population	0.039	0.881	0.080

unemployment and emigration, which is the case of some inland capitals (Córdoba, Jaén) and heads of *comarca*⁹ (Guadix, Úbeda, Orihuela, Lorca), as well as old industrial settlements (Langreo, Mieres, Linares, Puertollano, Tortosa). Otherwise, the lowest degree of spatial autocorrelation is of negative sign and can be found in regressive towns, which are close to a big city absorbing their population (3.9%). This is the case, amongst others, of the urban municipalities of Huelva, which are attracted by the Seville metro area, or the cities of Teruel and Calatayud, which are close to Saragossa. It is also interesting to highlight that the process of counter-urbanization is related to positive spatial autocorrelation (8%), e.g. growing towns surrounded by a big metropolis. In these cases, the big city spills population out to the peripheral towns that benefit from their advantageous location. Most towns in the metro areas of Madrid and Barcelona share this situation, as well as some tourist municipalities along the Mediterranean coast and the Islands. Finally, there is another case of agglomeration characterized by negative spatial autocorrelation when a thriving city grows if it is surrounded by smaller towns (7%). This condition is present in some capitals that are still in an urbanization process (Málaga) or mainly in prosperous middle-sized cities that attract population from dying rural villages. Therefore, the influence of space on the urban municipality transition probabilities seems more important for downward movements.

The influence of neighbors is confirmed by the χ^2 test statistic of independence of direction of move and neighbors population size, with 4 degrees of freedom, which generated a value of 398.087, which is significant at $prob = 0$. In conclusion, direction of movement in the population distribution of urban municipalities is not independent from the geographic environment.

4 Main conclusions

The urbanization process has mainly taken place during the twentieth century producing significant processes of industrialization and economic growth. Specifically in Spain, this process has not been uniform and exhibits different shapes depending on the definition of “urban area”. In our case, we work with a set of 722 municipalities that make up the Spanish urban areas: main cities and their satellite towns.

Zipf’s law shows the existence of two main phases in the evolution of these urban municipalities: 1900–1980 (divergence) and 1980–2001 (convergence). The main one extends over 80 years and consists in an increase in urban concentration, though two different sub-periods should be distinguished: 1900–1940 and 1940–1980. In effect, from 1900 to the 1930s, divergence is not that deep, coinciding with a significant industrialization and urbanization expansion that led to progress and social changes. However, this dynamism is violently broken by the end of the decade due to the Civil War. From 1940 to the 1970s, the largest cities grew much more quickly than the smallest population nuclei, leading to a more intense divergent pattern of growth.

⁹ The *comarca* is a historical -non official- division of Spanish provinces. This area is headed by a city that used to be a historic settlement and traditional center of trade, transportation, administration and cultural activities.

During the last decades of the twentieth century, the Zipf's parameters change from the 80-year decreasing tendency to a noteworthy increasing one. In other words, the group of 722 urban municipalities displayed a clear convergence growth pattern, since the smallest towns grew faster than the largest cities. The logical problems of the big cities (with an uncontrolled growth in the peripheries and an incipient depopulation process of their historical centers) halted their later expansion in favor of middle-sized even small-cities and certain rural areas.

The Markov Chains analysis shows a low interclass mobility, i.e., a high-persistence of urban municipalities to stay in their own class from one decade to another over the whole period. However, the largest and smallest urban municipalities display higher persistence than the medium-sized cities, which have more probability of moving to smaller categories. This proves the major role played by the medium-sized cities in the processes of urban agglomeration and suburbanization occurred in Spain during the twentieth century. In general terms, movements up are slower than movements down, especially for high-size classes. This result of faster declines indicates that urban municipalities are more likely to lose population than to gain it, especially inland, in big capitals and old industrial centers. This conclusion is compatible with the 80-year phase of divergence in size between urban municipalities, only reversed during the last two decades. This is why population convergence is still slight and mainly "downwards" inside the group of urban municipalities.

Finally, the probability of an urban municipality losing population (moving down in the hierarchy) is almost five times higher when it is surrounded by towns that contain, on average, less population. This result confirms the influence of space on urban population dynamism, also being more important for downward movements.

In summary, the Spanish urban municipalities have experienced an agglomeration process throughout practically the whole of the twentieth century, only broken by an incipient suburbanization in the last decades. The main actors of these changes have been the medium-sized towns: on the one hand, some peripheral cities around the metro areas of Madrid, Barcelona, Bilbao and Seville have experienced strong growth whereas on the other hand, a considerable number of industrial nuclei and inland heads of *comarcas* have lost their old influence and size. The influence of space on urban change is certainly conspicuous. It leads to the so-called "two Spains", which are no longer split along the usual North vs South partition. In the case of urban growth, another spatial division is relevant: on the one hand, there is a group of declining towns located in the Northwest, Center and South of Spain, which have progressively moved to lower states in the city-size distribution. On the other hand, there is another group of "winners" formed by most of the metro area cities (Madrid, Barcelona, Bilbao and Seville) and some tourist enclaves in The Balearics, Canary Islands, Comunidad Valenciana, Almería and Málaga.

References

- Amemiya T (1985) *Advanced econometrics*. Harvard University Press, Cambridge
- Anderson G, Ge Y (2005) The size distribution of Chinese cities. *Reg Sci Urban Econ* 35:756–776
- Anselin L (1988) *Spatial econometrics: Methods and models*. Kluwer, Dordrecht

- Anselin L, Le Gallo J, Jayet H (2007) Spatial panel econometrics. In: Matyas L, Sevestre P (eds) *The econometrics of panel data*, 3rd edn. Kluwer, Dordrecht
- Bickenbach F, Bode E (2003) Evaluating the Markov property in studies of economic convergence. *Int Reg Sci Rev* 26:363–392
- Black D, Henderson V (1999) Spatial evolution of population and industry in the United States (AEA Papers and Proceedings). *Am Econ Rev* 89:321–327
- Black D, Henderson V (2003) Urban evolution in the USA. *J Econ Geogr* 3:343–372
- Blakman S, Garretsen H, Van Marrewijk C, Van den Berg M (1999) The return of Zipf: towards a further understanding of the rank-size distribution. *J Reg Sci* 39:183–213
- Bulli S (2001) Distribution dynamics and cross-country convergence: a new approach. *Scott J Polit Econ* 48:226–243
- Cheshire P, Magrini S (2000) Endogenous processes in European regional growth: convergence and policy. *Growth Change* 31:455–479
- Cordoba JC (2003) On the distribution of city sizes. Mimeo, Economics Department, Rice University
- Davis D, Weinstein D (2002) Bones, bombs, and break points: the geography of economic activity. *Am Econ Rev* 92:1269–89
- Dobkins L, Ioannides YM (2000) Dynamic evolution of the US city size distribution. In: Huriot JM, Thisse JF (eds) *Economics of cities*. Cambridge University Press, Cambridge, pp 217–260
- Duranton G (2006) Some foundations for Zipf's law: product proliferation and local spillovers. *Reg Sci Urban Econ* 36:542–563
- Eaton J, Eckstein Z (1997) City and growth: theory and evidence from France and Japan. *Reg Sci Urban Econ* 17:443–474
- 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 de la Asociación de Demografía Histórica, Granada
- Fielding AJ (1989) Migration and urbanization in Western Europe since 1950. *Geogr J* 155:60–69
- Fingleton B (1983a) Independence, stationarity, categorical spatial data and the chi-squared test. *Environ Plann A* 15:483–499
- Fingleton B (1983b) Log-linear models with dependent spatial data. *Environ Plann A* 15:801–814
- Fingleton B (1986) Analyzing cross-classified data with inherent spatial dependence. *Geogr Anal* 18:48–61
- Fingleton B (1999) Estimates of time to economic convergence: an analysis of regions of the European Union. *Int Reg Sci Rev* 22:5–34
- Fingleton B (2001) Theoretical economic geography and spatial econometrics: dynamic perspectives. *J Econ Geogr* 1:201–225
- Fingleton B, Lopez-Bazo E (2003) Explaining the distribution of manufacturing productivity in the EU regions. In: Fingleton B (ed) *European regional growth*. Springer, Heidelberg, pp 375–409
- Fusi JP, Vilar S, Preston P (1983) De la dictadura a la democracia. *Desarrollismo, crisis y transición*. Historia 16, vol. XXV
- Gabaix X (1999a) Zipf's law and the growth of cities (AEA Papers and Proceedings). *Am Econ Rev* 89:129–132
- Gabaix X (1999b) Zipf's law for cities: an explanation. *Quar J Econ* 114:759–767
- Gabaix X, Ibragimov R (2006) Rank $-1/2$: a simple way to improve the OLS estimation of tail exponents. Available for download at: http://econ-www.mit.edu/faculty/index.htm?prof_id=xgabaix&type=paper
- Gabaix X, Ioannides YM (2004) The evolution of city size distributions. In: Henderson V, Thisse JF (eds) *Handbook of regional and urban economics*, vol 4. North Holland, Amsterdam, pp 2341–2378
- Grossman G, Helpman E (1991) *Innovation and growth in the world economy*. MIT Press, Cambridge
- Hamilton JD (1994) *Time series analysis*. Princeton University Press, Princeton
- Hordijk L, Nijkamp P (1977) Dynamic models of spatial autocorrelation. *Environ Plann A* 9:505–519
- Ioannides YM, Overman HG (2003) Zipf's law for cities: an empirical examination. *Reg Sci Urban Econ* 33:127–137
- Johnson PA (2002) A nonparametric analysis of income convergence across the United States. *Econ Lett* 69:219–223
- Kawagoe M (1999) Regional dynamics in Japan: a reexamination of Barro regressions. *J Jpn Int Econ* 13:61–72
- Kemeny J, Snell L (1976) *Finite Markov chains*. Springer, New York
- Krugman P (1996) *The Self-organizing economy*. Blackwell, Cambridge

- Lanaspa L, Perdiguero AM, Sanz F (2004) La distribución del tamaño de las ciudades en España, 1900–1999. *Revista de Economía Aplicada* 34:5–16
- Lanaspa L, Pueyo F, Sanz F (2003) The evolution of Spanish urban structure during the twentieth century. *Urban Stud* 40:567–580
- Le Gallo J (2004) Space-time analysis of GDP disparities among European regions: a Markov Chains approach. *Int Reg Sci Rev* 27:138–163
- Lopez-Bazo E, Vaya E, Mora AJ, Suriñach J (1999) Regional economic dynamics and convergence in the European Union. *Ann Reg Sci* 33:343–370
- Magrini S (1999) The evolution of income disparities among the regions of the European Union. *Reg Sci Urban Econ* 29:257–281
- Mella JM, Chasco C (2006) A spatial econometric analysis of urban growth and territorial dynamics: a case study on Spain. In: Nijkamp P, Reggiani A (eds) *Spatial evolution and modeling*. Edward Elgar, pp 319–360
- Ministerio de Fomento (2000) *Atlas Estadístico de las Áreas Urbanas en España*, Subdirección General de Urbanismo, Madrid
- Monclús FJ (1997) *Planeamiento y crecimiento suburbano en Barcelona: de las extensiones periféricas a la dispersión metropolitana (1897–1997)*. Coloquio sobre El desarrollo urbano de Montréal y Barcelona en la época contemporánea: estudio comparativo, Universidad de Barcelona
- Nitsch V (2005) Zipf zipped. *J Urban Econ* 57:86–100
- Overman HG, Ioannides YM (2001) Cross-sectional evolution of the US city size distribution. *J Urban Econ* 49:543–566
- Pareto V (1897) *Cours d'Economie Politique*. Rouge et Cie, Paris
- Quah D (1993) Empirical cross-section dynamics in economic growth. *Euro Econ Rev* 37:426–434
- Quah D (1996) Empirics for economic growth and convergence. *Euro Econ Rev* 40:1353–1375
- Quah D (1997) Empirics for growth and distribution: stratification, polarization, and convergence clubs. *J Econ Growth* 2:27–59
- Rey S (2001) Spatial empirics for economic growth and convergence. *Geogr Anal* 33:195–214
- Rossi-Hansberg E, Wright M (2004) *Urban structure and growth*. Mimeo, Stanford University, Economics Department
- Soo KT (2005) Zipf's law for cities: a cross-country investigation. *J Urban Econ* 35:239–263
- Stanback T. Jr (1991) *The new suburbanisation: challenge to the central city*. Westview Press, Boulder
- Tan B, Yilmaz K (2002) Markov chain test for time dependence and homogeneity: an analytical and empirical evaluation. *Euro J Oper Res* 137:524–543
- Tuñón de Lara M, Malerbe PC (1982) La caída del rey. De la quiebra de la Restauración a la República (1917–36). *Historia* 16, vol. XXIII
- Tuñón de Lara M, Viñas A (1982) La España de la cruzada. Guerra civil y primer franquismo (1936–1959). *Historia* 16, vol. XXIV
- Tuñón de Lara M, Bahamonde A, Toro, J, Arostegui J (1982) La España de los caciques. Del sexenio democrático a la crisis de 1917. *Historia* 16, vol. XXII
- White E, Hewings GJD (1982) Space-time employment modelling: some results using seemingly unrelated regression estimators. *J Reg Sci* 22:283–302
- Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J Am Stat Assoc* 57:348–368
- Zipf GK (1949) *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge