Ann Reg Sci (2016) 56:369–392
DOI 10.1007/s00168-016-0743-z

ORIGINAL PAPER

Real estate bubble and urban population density: six Spanish metropolitan areas 2001–2011

Joan Carles Martori[1] · **Rafa Madariaga**[1] · **Ramon Oller**[1]

Received: 26 January 2015 / Accepted: 12 January 2016 / Published online: 23 January 2016

Abstract After a period of exorbitant growth in building sector, it is appropriate to measure the impact it has had on the spatial structure of the main Spanish metropolitan areas. Madrid, Barcelona, Valencia, Sevilla, Bilbao and Zaragoza are surveyed, and two periods are compared: 2001 and 2011. The main driving factor behind the evolution of urban density is the huge and fast arrival and settlement of immigrants. This article explores these transformations through exploratory spatial data analysis and classic econometric models that relate urban density with distance to central business district. The presence of spatial effects is tested: spatial lag and error models are considered. Spatial error models are estimated. Two main consequences of real state bubble are as follows: (1) spatial autocorrelation has grown and (2) in Madrid, Barcelona and Zaragoza central areas with high urban density are larger than before. From a methodological point of view, the introduction of spatial effects in the classical urban population density models implies a clear reduction in the explaining power of distance.

JEL Classification C21 · R15 · R21 · R23

✉ Joan Carles Martori
martori@uvic.cat

Rafa Madariaga
rafa.madariaga@uvic.cat

Ramon Oller
ramon.oller@uvic.cat

1 Departament d'Economia i Empresa, Universitat de Vic – Universitat Central de Catalunya, Vic, Barcelona, Spain

1 Introduction

After a long period of growth, the Spanish economy entered in a deep and long crisis. The real estate bubble and burst are maybe the most known features of the Spanish economy in the twenty-first century. The increase in this enormous bubble, its explosion from 2008 on and its relationship with financial stability are in the roots of this crisis. Our work focuses on the effects this bubble has had on the urban structure and especially in urban density.

Spanish population grew from 41 to 47 million between 2001 and 2011, which represents more than 14.6 % (Fig. 1). Along the same period, the stock of dwellings grew more than 20 %. The annual average number of dwellings built between 2001 and 2011 was higher than 440,000; the peak was in 2007 when more than 640,000 new houses were finished. At the same time, the price of houses climbed from 930 euros by square metre in 2001 to more than 2100 in the beginning of 2008 (126 %). Then it has decreased to 1700 at the end of 2011, and it continues decreasing, reaching

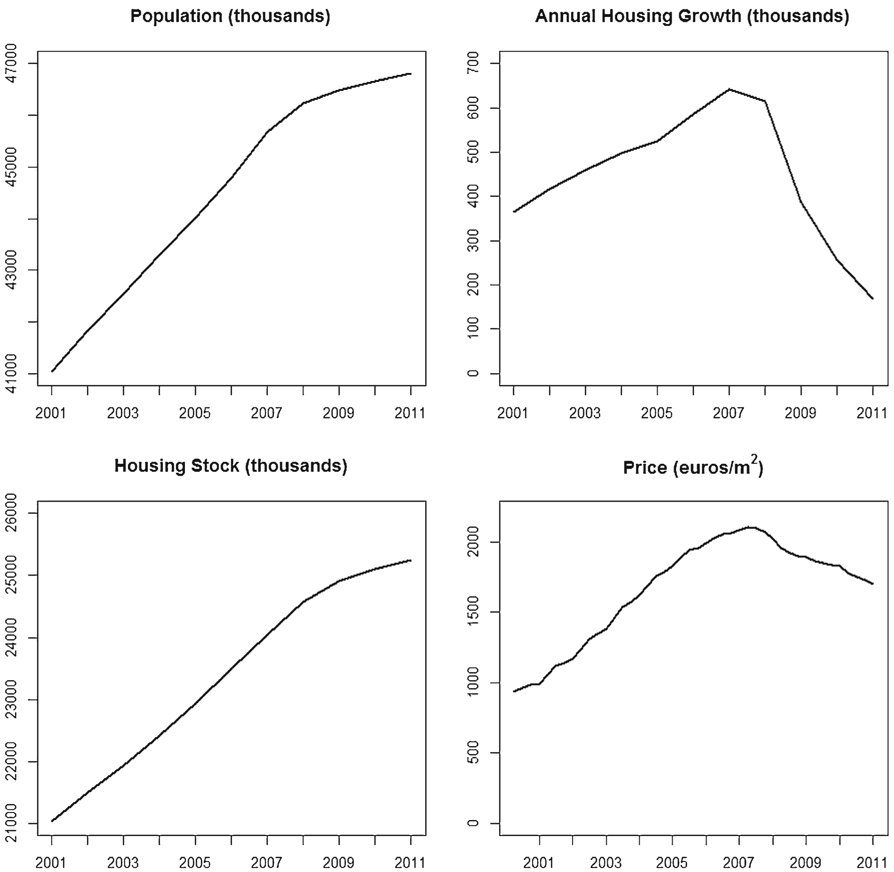

Fig. 1 Annual population, housing growth, housing stock and quarterly house price in Spain 2001–2011

1470 at the end of 2013. Easy access to international credit in financial markets and common monetary policy with low official interest rates, which resulted in negative real interest rates, are one of the main driving forces behind this extraordinary growth in construction. Housing becomes one of the most profitable financial assets. Different international bodies drew attention on the issue: the European Parliament approved the resolution of 26 March 2009 on the impact of extensive urbanisation in Spain on individual rights of European citizens, on the environment and on the application of EU law, based upon petitions received. The Human Rights Council of the United Nations also approved the report on adequate housing as a component of the right to an adequate standard of living, delivered by his mission to Spain in February 2008.

Although the most of this real estate bubble was located on the coast, the phenomenon has affected the structure of cities and metropolitan areas. Two questions about the real estate bubble have to be considered: on the one hand, a traditional emigrant country like Spain became a host country for a large number of immigrants between 2001 and 2011. Hence, the share of immigrant population increased from 3.33 % in 2001 to 12.18 % in 2011. The majority of this immigration flow came from Africa, Latin America and Eastern Europe. The social and economic characteristics of this population, their behaviour in regard to the integration in the country and the presence of labour market and housing market opportunities make larger cities the preferred destination. On the other hand, the speed of this huge immigration flow was especially remarkable. For example, from 2001 to 2008, in Barcelona metropolitan area the percentage of immigrant population grew from 5.05 to 15.16 %.

Moreover, before 1990s, the majority of immigrants were European citizens, pensioners who settled in the coastal regions. In this decade, the immigration flows changed to non-EU immigrants, coming from developing or political unstable countries. The origin of this change can be found in the evolution of the real estate boom and the corresponding increase in job opportunities, which were located in the large metropolitan areas. What is the dynamic of metropolitan areas density when so many people arrives in such a short period? In our view, the dynamic is shaped by two main forces. On the one hand, the majority of immigrants arrive and settle in the centre of metropolitan areas, since this location allows them to find job opportunities and housing facilities at low prices, and even irregular housing facilities are accessible because of family or neighbourhood links with the same origin. Therefore, the growth in immigrant population in the old centres of metropolitan areas would cause an increase in urban density. But this initial dynamic generates that many Spanish born population leave the centres looking for more pleasant locations and less crowded environment. In our opinion, this is the second important factor affecting urban population dynamics. The saturation of these central districts has pushed first native population and then immigrants to peripheral districts, with cheap housing access and more pleasant zones (Martori and Apparicio 2011). If this hypothesis is true, the dynamic of metropolitan areas is shaped by two driving forces: first, intensification in land use, as new dwellings are built in substitution of old buildings in central locations, and second, the extension of the metropolitan areas, as new peripheral areas are built.

Few studies have analysed the evolution of urban structure during this real state bubble and this period of massive arrival of immigrants. The focus of this paper is to analyse the transformation of the urban structure of the main metropolitan areas in

Spain for the period 2001–2011 by means of exploratory spatial data analysis (ESDA) and classical urban population density functions.

The estimation of urban population density functions goes back to the pioneer work of Clark (1951) whose seminal work was empirically driven. Population density in urban areas is a main variable to design health and education public services, to plan public transportation systems and to develop environmental policies. The general objective of the present study is to analyse how population density and its distribution have evolved during the real state bubble, bearing also in mind the large arrival of immigrants in the six main metropolitan areas in Spain. To this purpose, we have constructed two data sets (2001 and 2011). The first method is an application of ESDA analysis to describe and compare the spatial distribution of density. The second method is the estimation of density functions for both periods. Although in Europe and North America econometric analysis and estimation of density functions have a long tradition, up to where we know there are few studies in Spain that use this methodology. Furthermore, modern developments in spatial statistics and econometric methods have improved the understanding of urban structure. These new methods may deal with spatial autocorrelation and allow to consider spatial effects in the econometric model. In this sense, they bring about a more comprehensive explanation of urban population density. Finally, this work intends to apply these methods to improve the estimation of density functions.

This paper is organized as follows: Sect. 2 briefly reviews the development of classical urban density functions. It also presents a general framework of functional forms. Section 3 overviews the metropolitan areas studied and shows the results of ESDA methods, the functional forms selected for each metropolitan area and period and the results of the econometric estimations. Section 4 reviews the literature developed to deal with spatial effects and presents our selected models and the econometric results. Section 5 concludes and sketches future research lines.

2 Classical urban density functions

The classical study of Clark (1951) followed by the microeconomic approach to urban economics has led to an extensive literature on this subject with empirical implementations for a large number of cities and metropolitan areas in different countries and times. Moreover, there are important contributions from different disciplines: spatial economics, regional science, geography, engineering and transport planning.

This section provides an overview of the development of various functional forms, focusing on those works where distance to the city centre (CBD) is the only explanatory variable (see Fig. 2). These models have been named Classical because the first study in Regional Science about this topic used only this explanatory variable. Reviews of this literature can be found in McDonald (1989), Wang and Zhou (1999) and Bunting et al. (2002).

In order to simplify the study of population density in cities and to facilitate comparison of results, Clark (1951) argues that the density of urban population can be described correctly by a negative exponential function:

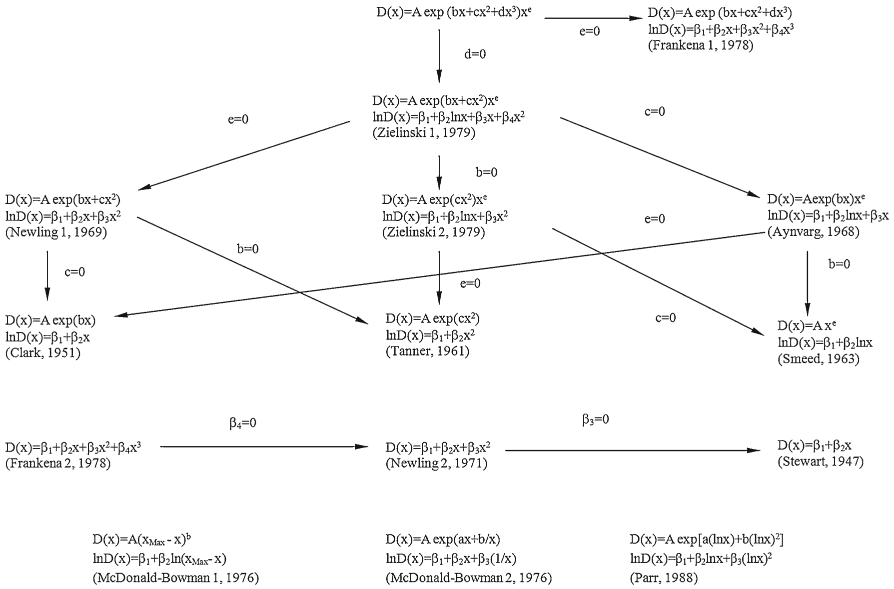

Fig. 2 Functional forms

$$D(x) = Ae^{\beta x} \tag{1}$$

where x is the distance to the central business district (CBD), measured in units of length, $D(x)$ is the resident population density per unit area and $A > 0$ and $\beta < 0$ are parameters. The density gradient can be measured by the coefficient β. A high value of β means that the density decreases rapidly with increasing distance, a pattern that defines a compact city. Low values of β mean density decreases slowly, describing a large city, spread in the territory. The coefficient A is the value of the density when the distance is equal to zero, interpreted as the density at the CBD.

The following work on urban density is due to two geographers, Stewart and Warntz (1958). Their study shows empirical evidence of the relationship between the area occupied by the city and population. In his previous work, Stewart (1947) had already suggested a linear relationship between density and distance.

Tanner (1961), Smeed (1963) and Aynvarg (1969) propose three new functional forms in their studies of traffic in cities. Their contributions are based on two special cases of the quadratic gamma function. Newling (1969), in his study of the spatial variation of urban density, provides a summary of the work of Clark and Tanner and proposes a new functional form. He suggests that the non-residential centre, with low density, is followed by another ring with increasing density until reaching a tipping point at which the gradient takes a negative value and the density decreases. The same author, in later works (Newling 1971), suggests the possibility of a two degree polynomial function.

In the 1960s, Alonso (1964), Muth (1969, 1971) and Mills (1967, 1970, 1972) were the first economists who tried to provide theoretical grounds for the, up to now, empirical work. The monocentric model is still the basic theoretical framework in urban economics. Formal updated developments of the model can be found in Glaeser (2008) and Brueckner (2011).

Empirical research undertaken in the first two decades after the publication of Clark's study is characterized by the accumulation of additional evidence supporting two models: those of Clark and Tanner. Two examples of this line of research are Berry et al. (1963) and Latham and Yeates (1970). Based on the quadratic gamma function, McDonald and Bowman (1976) and Zielinski (1979) set out various new functional forms. Frankena (1978) further generalizes the relationship by proposing two three degree polynomial models. Another notable effort in the functional description of the behaviour of urban density is that of Parr (1987), Parr et al. (1988), Parr and O'Neill (1989), Parr (2012) and Holden and Parr (2013). They have introduced and developed the lognormal function in the analysis of the distribution of both population and density at regional level. Figure 2 provides a summary of the classical functional forms proposed, the mathematical relationships between them and the corresponding econometric model. For an overview, see Martori and Suriñach (2002) and Martori (2010).

The selection of observational units for the statistical estimation of the functional forms of population density is frequently constrained by the availability of data. Most of the works have used census tract data. The census tract areas are taken as a proxy variable for the residential land. Most of the studies quoted use the whole set of census tracts of the metropolitan areas. For example, Frankena (1978), Griffith (1981b) for Toronto, White (1977) for five European cities, Glickman and Oguri (1978) for 71 Japanese cities and Alperovich (1983a, b) for 20 Israeli cities.

The previous paragraphs have shown the mathematical development and econometric estimation of monocentric density functions. However, urban areas evolve and new forms appear requiring a reconceptualization. Thus the debate about urban structure witnessed a new challenge when polycentric forms were proposed and tested. As Griffith and Wong (2007) state, the most noticeable advance has been from monocentric to polycentric structures. The origins of this new approach are Griffith (1981a, b). This more sophisticated conceptualization shifts the modelling of population density to more complex model specifications. Two points have to be made: first, these works focus on larger metropolitan areas where the polycentric structures are more suited. For instance, salient contributions in polycentric conceptualization are Giuliano and Small (1991) for Los Angeles, Cervero and Wu (1997) for San Francisco, McMillen and McDonald (1998), McDonald and McMillen (2000) and McMillen and Lester (2003) for Chicago and Garcia-López and Muñiz (2010) and Garcia-López (2012) for Barcelona Region (this is a different urban area to Barcelona Metropolitan Area that we study). Second, the majority of polycentric structures are focused on employment centres not on population density.

Although this new developments have led to a new research line, the monocentric model literature continues developing and appending new insights and empirical evidence. Moreover, classic urban population density functions are still extensively used. Authors like Anderson (1982, 1985) which test cubic-spline forms, Alperovich and Deutsch (1992, 1994) use them to measure urbanization and to detect the CBD. Alperovich and Deutsch (2002) also apply switching regression techniques and

Tsai (2005) uses them to test compactness versus sprawl. Fallah et al. (2012) propose a new measure of urban sprawl based in urban density and its distribution. It is also worth mentioning that Alperovich and Deutsch (1992) also deal with real estate bubble in Israeli cities and that Zheng and Kahn (2008) estimate a classical density function to study what they call "explosive new development" in Beijing. A recent empirical study by Arribas-Bel and Sanz-Gracia (2014), focused on employment centres, shows that monocentric structures persist in the majority of metropolitan areas in the USA.

3 Metropolitan areas analysed

The Spanish government officially considers 68 metropolitan areas. These include urban areas with at least 50,000 inhabitants defined by administrative rules based on population (total population, density), housing (building typology, land prices), land uses and urban dynamics and transport networks (Ministerio de la Vivienda 2004), and they are conformed by several contiguous municipalities.

Metropolitan areas were selected based on total population, contiguity urban land and homogeneous data availability. Following these criteria, the resulting samples were the six most populated: Madrid, Barcelona, Valencia, Sevilla, Bilbao and Zaragoza, where 13.6 million of inhabitants live (2011) representing almost 30 % of total Spanish population. Since 2001 population in these areas has grown more than 1.5 million, which represents 9 % and more than 1.2 million of dwellings have been built. Growth rate in the number of dwellings ranges from 13 % in Bilbao to more than 20 % in Sevilla.

Metropolitan areas are divided into census tracts; small areas containing between 500 and 2000 residents designed for organizing pooling tables for elections. These are the spatial units that have been used for this study. They provide the most suitable disaggregation level for the study of metropolitan density structure in Spain, as there are available data for them and they are small enough to capture spatial details. Table 1 shows summary statistics: population, average density by census tract and number of dwellings for each urban area and period.

There are some noticeable differences between these six metropolitan areas. Madrid and Barcelona are significantly more populated and have about 3500 and 2100 census tracts, respectively, whereas Bilbao and Zaragoza have less than a million of inhabitants and the number or census tracts is around 500 for the former and 725 for the latter. Valencia and Sevilla are in the middle, exceeding one million of inhabitants and having around 1000 census tracts. Differences in location and orography are also important. Bilbao, Barcelona and Valencia are coastal metropolitan areas, and the first two are limited by mountains that shape the urban development. In contrast, Madrid, Sevilla and Zaragoza are inland cities, surrounded by uninhabited land that allows the extension of the urban area.

3.1 Exploratory spatial data analysis (ESDA)

Before model estimation changes in urban density along the studied period have to be described. To this purpose, several ESDA methods are used. Our aim is to know whether census tracts with similar density show any autocorrelation pattern or whether

Table 1 Summary statistics: population, density (average density by census tract: inhabitants by km^2) and housing (number of dwellings)

Metropolitan area	Population		Density		Housing	
	2001	2011	2001	2011	2001	2011
Madrid	5,132,792	6,054,444	32,665	30,423	2,482,885	2,900,061
Barcelona	2,936,321	3,222,117	31,775	38,439	2,280,334	2,604,934
Valencia	1,349,218	1,379,209	26,446	22,932	1,221,074	1,453,291
Sevilla	1,160,217	1,297,509	21,257	19,454	734,843	884,141
Bilbao	849,122	903,721	35,068	31,694	470,595	534,551
Zaragoza	647,819	742,395	34,094	32,107	432,612	510,754

Table 2 Moran's I and Geary's C coefficients

Metropolitan area	Moran's I		Geary's C	
	2001	2011	2001	2011
Madrid	0.4307	0.5408	0.5509	0.4719
Barcelona	0.4452	0.5244	0.4304	0.4727
Valencia	0.4075	0.4460	0.5911	0.5718
Sevilla	0.3041	0.1606	0.6046	0.6332
Bilbao	0.4194	0.4391	0.5429	0.5356
Zaragoza	0.4235	0.4464	0.5499	0.5267

The null hypothesis of random spatial distribution is rejected in all cases (p value <0.001) except for Moran's I in Sevilla 2011

they are distributed in a random way. A related purpose is to analyse whether the real estate bubble has altered these patterns. For example, Tsai (2005) used Moran's I statistic (Moran 1948) to measure the clustering phenomenon on the distribution of population density in metropolitan areas. As a complementary measure, we also use spatial Geary's C (Geary 1954). While Moran's I statistic ranges between -1 and 1, Geary's C ranges between 0 and 2. Whether census tracts are more clustered (positive spatial autocorrelation), closer to zero the C statistic would be (I would be closer to 1). If urban density shows a dispersed pattern (negative spatial autocorrelation), C statistic would be close to two (I would be closer to -1). A value of one (I equal to 0) indicates a completely random distribution of urban density; there is not any autocorrelation pattern. In general, Moran's I and Geary's C yield similar conclusions. Moran's I is a more global measure, and it is more sensitive to extreme values of urban density, whereas Geary's C is more sensitive to density differences in small areas. However, Moran's I is preferred in most cases since Cliff and Ord (1975, 1981) have shown that Moran's I is consistently more powerful than Geary's C. We use a row standardized contiguity matrix of first-order rook weights. Other forms of weights matrix were tested, yielding similar qualitative and quantitative results. Table 2 presents the results for the six metropolitan areas for 2001 and 2011. Notice that Moran's I and Geary's C vary in opposite directions.

Moran's I for 2001 shows significant positive autocorrelation but always lower than 0.5. All metropolitan areas show an increase in this positive autocorrelation except Sevilla. It shows a spreading of density between 2001 and 2011. Furthermore, in 2011 we cannot reject the hypothesis of a random distribution (see footnote on the table). For the other five metropolitan areas, results show important increases in spatial autocorrelation. Geary's C coefficients show consistent results as the values are lower than 1. Moreover, all coefficients decrease, showing that positive autocorrelation has grown. The exception is Barcelona as the coefficient increases. Summing up, the results of Moran's I and Geary's C coefficients show two patterns. The general pattern is an increase in positive autocorrelation; neighbouring sections of these metropolitan areas tend to converge in urban densities. Sevilla and Barcelona are different since the first positive autocorrelations has decreased. For the second, results are not clear as it depends on the statistic used. This empirical result may be seen as evidence supporting our argument about an intensification of the land use presented in the introduction section.

As is well known, the significant spatial autocorrelation cannot be equal in the whole metropolitan area as it can only be significant in certain locations. Morans *I* and Gearys *C* are global spatial autocorrelation measures and are applied to the entire area of each metropolitan area. Nevertheless, in our case it is also useful to study density local patterns. To analyse this issue, it is relevant to study the local autocorrelation by means of Local indicators of spatial association (LISA). Hot spot analysis based on the measure of local spatial association G_i^* proposed by Getis and Ord (1992) and Ord and Getis (1995) has been used. We prefer the Getis–Ord G_i^* measure for exploratory analysis as it enables clustering of small density values (small values of G_i^*) to be distinguished from clusters of large density values (large values of G_i^*). Anselin's local *I* (Anselin 1995) would be large in both cases, but to identify clusters of small or large density values Moran scatterplot has to be used.

The G_i^* statistic obtained for each census tract of metropolitan areas has been converted in z-scores. The larger the z-score is, the more intense the clustering of high values is. The smaller the z-score is, the more intense the clustering of low values is. Figures 3 and 4 map census tracts with significant (significance level 1 %) positive and negative z-scores (hot and cold spots, respectively). The statistical significance has been adjusted for multiple testing and spatial dependence using the false discovery rate (FDR) correction method (Benjamini and Hochberg 1995) and the optimal fixed distance band. This is, the distance where spatial processes are most active or most pronounced, i.e., reflects the maximum spatial autocorrelation. In our case, the distance band is based on the average distance to the 30 nearest census tracts. The distance ranges between 0.9 and 4.5 km in 2001 and between 1.1 and 5 km in 2011.

The G_i^* statistic allows us to identify the change in clusters of high- and low-density census tracts along the real estate bubble period. We consider the oldest district in the analysed metropolitan areas as the CBD: City Halls, central offices of many large companies and major commercial activities are located in this area, see Madariaga et al. (2014). Hot spot areas have spread around the CBD in Madrid, Barcelona and Zaragoza. They have remained stable in Valencia and Bilbao. In Sevilla, they have decreased. The most populated metropolitan areas (Madrid and Barcelona) show dispersed hot spots far from the central zones. Madrid presents several scattered hot spot areas, whereas in Barcelona there are two big hot spot areas at the northeast of the CBD; the one closer to the CBD has decreased and the other has increased. Variation of cold spot areas follows a similar pattern of variation of hot spot areas in Madrid and Zaragoza, where they have increased, and Sevilla, where they have decreased. In Barcelona, Valencia and Bilbao, while hot spot areas have remained, cold spot areas have increased.

The local indicators tend to reinforce the conclusions of global indicators. Exploratory data analysis shows significant changes in spatial density patterns. The next subsection investigates whether these changes are picked up by classical models of urban density.

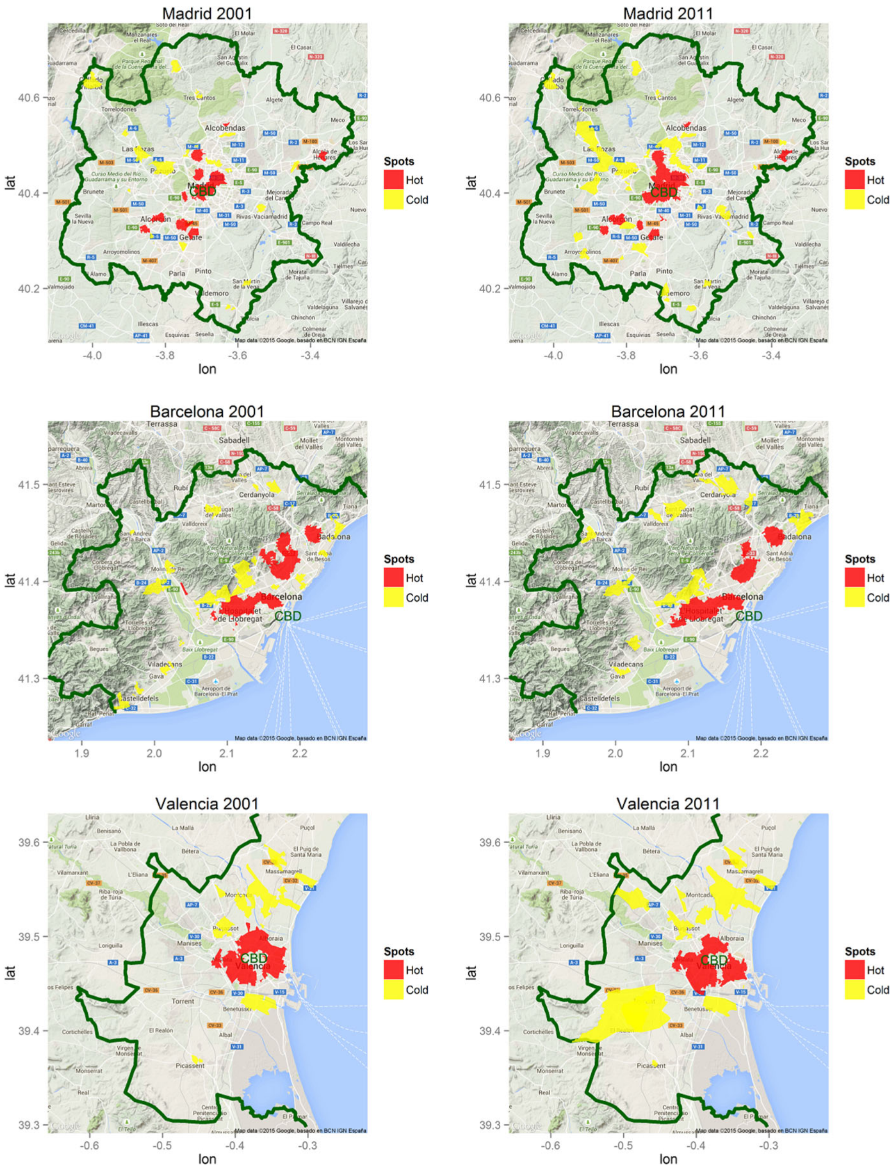

Fig. 3 Madrid, Barcelona and Valencia: hot and cold spot analysis

3.2 Functional forms

Similarly to the work of Martori (2010), we start estimating classical models of urban density by OLS. Our strategy begins with the estimation of fourteen functional forms presented in Fig. 2 and the selection of the best fit model for each metropolitan area and period. This figure includes three types of models: log nested, linear nested and non-

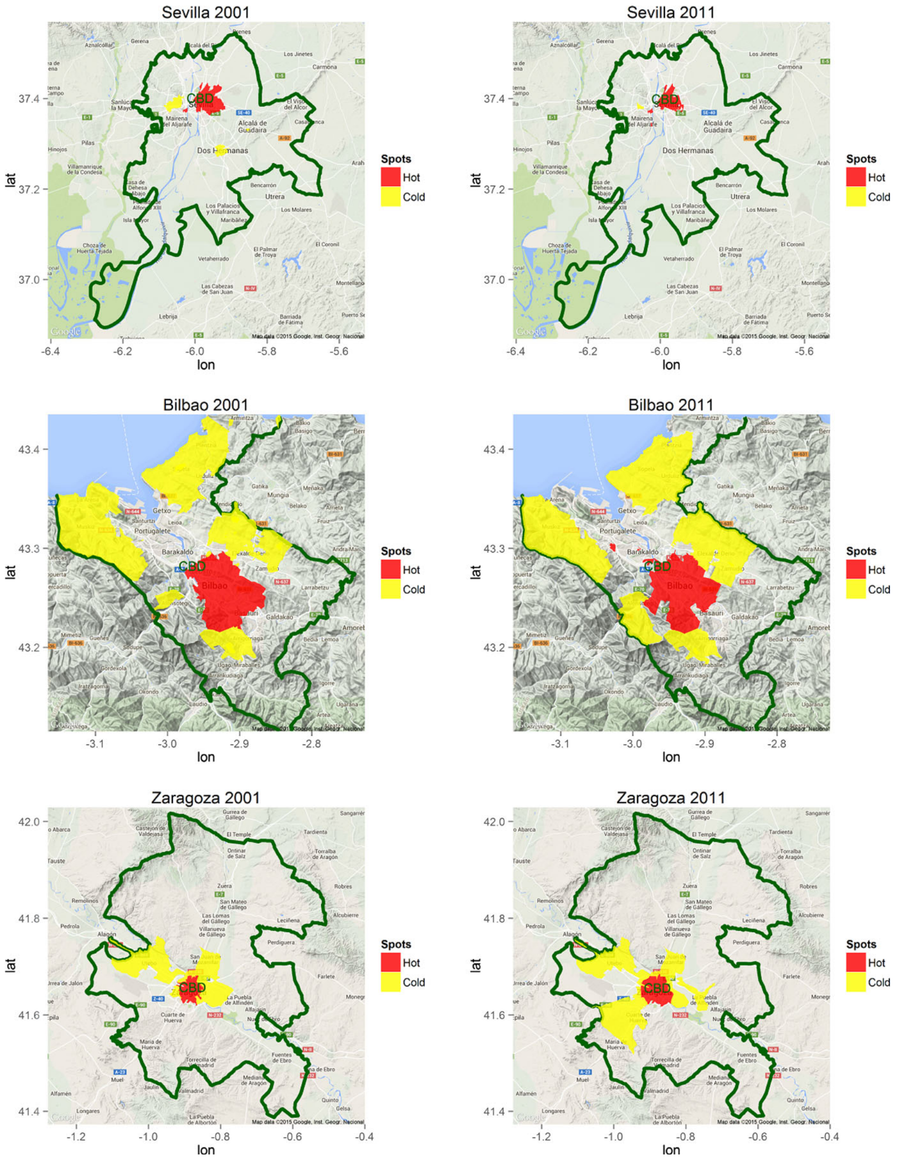

Fig. 4 Sevilla, Bilbao and Zaragoza: hot and cold spot analysis

nested models. For nested models, F test has been used to select the most appropriate functional form for each case. We start with the most general form, and then we test whether it is possible to narrow down the specification. Once this strategy has selected the best log and linear form, a Davidson–MacKinnon–White test (MacKinnon et al. 1983) is performed in order to choose the log or linear form. The selected model has been then compared with non-nested functions using J test (Davidson and MacKinnon

1981). In some cases, this procedure is unable to try just one model. The selection statistic AIC and the usual R^2 measure of goodness of fit are then used to compare results and to finally select the best model. Tables 3 and 4 present the results. Finally, the selected models are: Zielinsky 1 (see Fig. 2) for Madrid and Zaragoza, Newling 1 for Sevilla and Bilbao, and Tanner for Barcelona and Clark for Valencia. As it is usual in this topic, we have taken into account that inferences can be affected by the presence of heteroscedasticity. The results include the Breush–Pagan test that confirms this fact. Since the sample size is large, the consistency property assures that it does not affect the proper estimation of the parameters. However, heteroscedasticity has consequences for the inferences about the parameters unless we use a consistent estimate of the covariance matrix of the coefficients. We have studied the significance of the coefficients using several consistent estimates of the covariance matrix proposed by MacKinnon and White (1985), and results (not reported) lead to the conclusion that distance to the CBD contributes to the explanation of urban density. Finally, it is worth mentioning that R^2 values vary among 0.15 and 0.63. In some cases, they are low, but they move around usual results in this kind of models.

Tables 3 and 4 also give Moran's I test for spatial autocorrelation in the residuals and robust Lagrange Multiplier test for spatial lag (RLM-lag) and error (RLM-error) models. Like in Sect. 3.1, a row standardized contiguity matrix of first-order rook weights has been used. Other forms of weights matrix were tested, yielding similar qualitative and quantitative results. The test statistics confirm the presence of spatial autocorrelation and the need to introduce spatial effects in the model. This issue is studied in next section.

4 Spatial effects

From a theoretical standpoint and according to Anselin and Can (1986), in the studies of urban density with the distance as the only explanatory variable there can be three sources of misspecification: the spatial unit of analysis (irregular census tracts), the possibility of some kind of polycentricism and spatial spillover effects. The result of these misspecifications is the presence of spatial dependence in the error term. In contrast, in urban density models LeSage and Pace (2009) notice that latent unobservable influences related to several factors (urban structure, green areas, amenities or commercial zones, among others) may affect the dependent variable. These two related reasons justify the presence of spatial effects.

Furthermore, from an empirical standpoint, there is another reason for introducing spatial effects. If, for example, the scale and location of the process under study do not correspond exactly with the available data, there will be a mismatch. As a consequence, the off-diagonal elements of the error covariance matrix will not equal zero. The problem with data may imply that there is a systematic spatial pattern.

In summary, we use a strategy for modelling the urban population density in six Spanish Metropolitan Areas that it is based on the distance as explanatory variable. Since this general model has spatial effects, we proceed with the computation of two spatial econometric models. We use spatial lag and spatial error models to control for spatial dependence.

Table 3 Madrid, Barcelona and Valencia: OLS results

	Madrid (2001)	Madrid (2011)	Barcelona (2001)	Barcelona (2011)	Valencia (2001)	Valencia (2011)
Constant	10.4909***	10.5969***	10.6845***	10.7291***	10.5382***	10.3916***
Distance	−0.1458***	−0.1464***	–	–	−0.1759***	−0.1588***
Squared dist	0.0015***	0.0018***	−0.0068***	−0.0069***	–	–
Log distance	0.3131***	0.2335***	–	–	–	–
Moran's *I*	0.5504***	0.5575***	0.4896***	0.5161***	0.3769***	0.3888***
RLM-lag	74.66***	7.03**	22.21***	16.46***	115.63***	122.33***
RLM-error	103.12***	44.23***	55.62***	49.09***	159.92***	169.84***
BP	419.07***	404.48***	147.09***	108.34***	139.35***	145.57***
Sample size	3690	4016	2531	2145	1065	1092
R^2	0.15	0.17	0.25	0.28	0.22	0.21

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4 Sevilla, Bilbao and Zaragoza: OLS results

	Sevilla (2001)	Sevilla (2011)	Bilbao (2001)	Bilbao (2011)	Zaragoza (2001)	Zaragoza (2011)
Constant	10.2635***	10.1648***	9.7644***	9.8070***	11.3273***	11.2176***
Distance	−0.1753***	−0.1611***	0.2405***	0.1936***	−0.8211***	−0.6495***
Squared dist	0.0035***	0.0030***	−0.0352***	−0.0305***	0.0158***	0.0116***
Log distance	–	–	–	–	0.7597***	0.3948***
Moran's *I*	0.3706***	0.3648***	0.5495***	0.5463***	0.4623***	0.4475***
RLM-lag	39.54***	104.33**	0.2555	0.0852	1.49	2.05
RLM-error	53.44***	125.60***	14.36***	13.35***	65.78***	64.72***
BP	60.78***	68.80***	18.49***	20.42***	119.44***	100.63***
Sample size	804	884	726	728	487	533
R^2	0.18	0.16	0.22	0.21	0.64	0.63

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ''.' 0.1 ' ' 1

Spatial lag model is the most frequently encountered specification in spatial econometrics. It may be written as:

$$Y = \rho WY + X\beta + \epsilon \tag{2}$$

where Y is an $(N \times 1)$ vector of observations on a dependent variable measured at each one of the N locations, X is an $(N \times k)$ matrix of exogenous variables, β is a $(k \times 1)$ vector of parameters, ϵ is an $(N \times 1)$ vector of independent and identically distributed disturbances and ρ is a scalar spatial lag parameter. In our case, this means that the population density in each unit (i.e. census tract) is modelled so as to depend on the population density in neighbouring units captured by the spatial lag vector WY.

The spatial error model may be written out as follows:

$$Y = X\beta + u \tag{3}$$

with

$$u = \lambda Wu + \epsilon$$

where λ is a scalar spatial error parameter and u is a spatially autocorrelated disturbance vector. In this model, the spatial influence comes only from the error terms, which means that the density in each unit is modelled so as to depend on the error terms in neighbouring units captured by the spatial error vector Wu.

Specification searches in spatial econometrics are a topic that has been discussed in the urban and regional literature. Florax et al. (2003) and Mur and Angulo (2009) give different strategies for detecting the most appropriate form of spatial autocorrelation. The standard approach in most empirical work is to start with a non-spatial linear regression model (OLS) and then to determine (e.g. by using Moran's I test) whether or not the model needs to be extended with spatial effects. Afterwards, if this is confirmed, the next step is to determine what kind of model should be used.

Following the specific-to-general approach, we computed the robust versions of the Lagrange multiplier tests for the spatially lagged dependent variable (RLM-lag) and for error dependence (RLM-error). The results, presented in Tables 3 and 4, show that RLM-lag and RLM-error are significant for Madrid, Barcelona, Valencia and Sevilla and that RLM-error model is significant for Bilbao and Zaragoza. As a general rule, in this situation, empirical works choose the spatial model with the most significant Lagrange Multiplier test. In our case, these results indicate that the spatial error model is preferable than spatial lag model and provide empirical evidence in the discussion between Anselin and Can (1986) and LeSage and Pace (2009) (see the beginning of this section).

4.1 Spatial error model results

In this section, we present the results of the estimation of error models for the six metropolitan areas, see Tables 5 and 6. Given non-normality of the error terms and the

Table 5 Madrid, Barcelona and Valencia: SEM (GMM) results

	Madrid (2001)	Madrid (2011)	Barcelona (2001)	Barcelona (2011)	Valencia (2001)	Valencia (2011)
Constant	10.9563***	10.7070***	10.9948***	11.1047***	10.7926***	10.6091***
Distance	0.0615	−0.0597*	–	–	−0.1566***	−0.1346***
Squared dist	−0.0027·	−0.0003	−0.0066***	−0.0064***	–	–
Log distance	−0.3438	0.1696***	–	–	–	–
λ	0.6959***	0.6970***	0.6966***	0.7466***	0.5625***	0.5767***
Sample size	3690	4016	2531	2145	1065	1092
R^2	0.14	0.16	0.26	0.29	0.25	0.24

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 6 Sevilla, Bilbao and Zaragoza: SEM (GMM) results

	Sevilla (2001)	Sevilla (2011)	Bilbao (2001)	Bilbao (2011)	Zaragoza (2001)	Zaragoza (2011)
Constant	10.2500***	10.1798***	11.1247***	10.9229***	11.7433***	11.6263***
Distance	−0.0428	−0.0425	0.0597	0.0445	−0.5548**	−0.5602***
Squared dist	−0.0022	−0.0023	−0.0230*	−0.0203*	0.0083	0.0086*
Log distance	–	–	–	–	0.0254	0.1801***
λ	0.5664***	0.5450***	0.7941***	0.7784***	0.7471***	0.7310***
Sample size	804	884	726	728	487	533
R^2	0.16	0.15	0.18	0.18	0.59	0.58

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

presence of heteroscedasticity that may affect the results, we have estimated the error model using generalized methods of moments (GS2SLS) as proposed by Arraiz et al. (2010). The results given in this section were obtained using the spdep (Bivand 2014) and sphet (Piras 2010) R libraries.

The results show that the spatial autocorrelation coefficient (λ) is always significant: There are factors affecting the distribution of density, other than distance, that are not in the functional form of classical models. Following Elhorst (2014), interaction effects among error terms are consistent with situations where omitted factors affecting the dependent variable are spatially autocorrelated. Further research is needed to investigate what other variables related to the housing bubble could explain these results. The evolution of housing prices is a prime candidate, but the lack of reliable data at this disaggregation spatial level has not allowed us to introduce this explanatory variable.

The λ coefficient grows slightly during the period in the most populated metropolitan areas (Madrid, Barcelona and Valencia) and decreases in the less populated ones (Sevilla, Bilbao and Zaragoza). The introduction of spatial effects weakens the distance parameters, which lose signification for four metropolitan areas (Madrid, Sevilla, Bilbao and Zaragoza). These results may suggest that distance is less important than the spatial error effect in order to explain the urban population density distribution. On the other hand, these results may also suggest that the metropolitan area is not monocentric and, consequently, we should look for more adequate models to describe the density distribution.

In Fig. 5, we present profiles for each metropolitan area for both periods. Profiles show the relationship between distance to the CBD and density logarithmic form. Such profiles are based on coefficient estimates by GMM method for SEM presented in Tables 5 and 6. All six metropolitan areas show different size and shape profiles. The size of Madrid, Sevilla and Zaragoza metropolitan areas reaches around 35 km, Barcelona around 25 km and, finally, Valencia and Sevilla around 15 km. Regarding the shape profile, Valencia shows a rapid decrease; in Barcelona and Bilbao the profiles also decrease, but this decrease is slow around the CBD. For 2011, Madrid and Zaragoza show a singular pattern, as they increase their central density in the first 1 km. As the distance parameters for Sevilla are not significant, we only present the profiles, but the shape is not commented. Comparing both periods, for Madrid, we observe an important decrease in central density and a reduction in middle distances (between 10 and 30 km). We note, however, that results in 2001 have to be considered with caution because distance parameters are not significant with a significance level of 5 %. Barcelona shows a slight increase in central density and all along the first 15 km. Further than this distance, the profiles are superposed. Valencia results show a decrease in central density and all along the first 5 km. There are no significant changes further than this length. For Bilbao, an important decrease in central density is shown for the first 10 km. Zaragoza has experienced a reduction in central density up to 2 km. Although the dynamics of every city are different, the procedure has shown that classical models of urban population density are able to capture changes in the patterns of spatial structure of metropolitan areas.

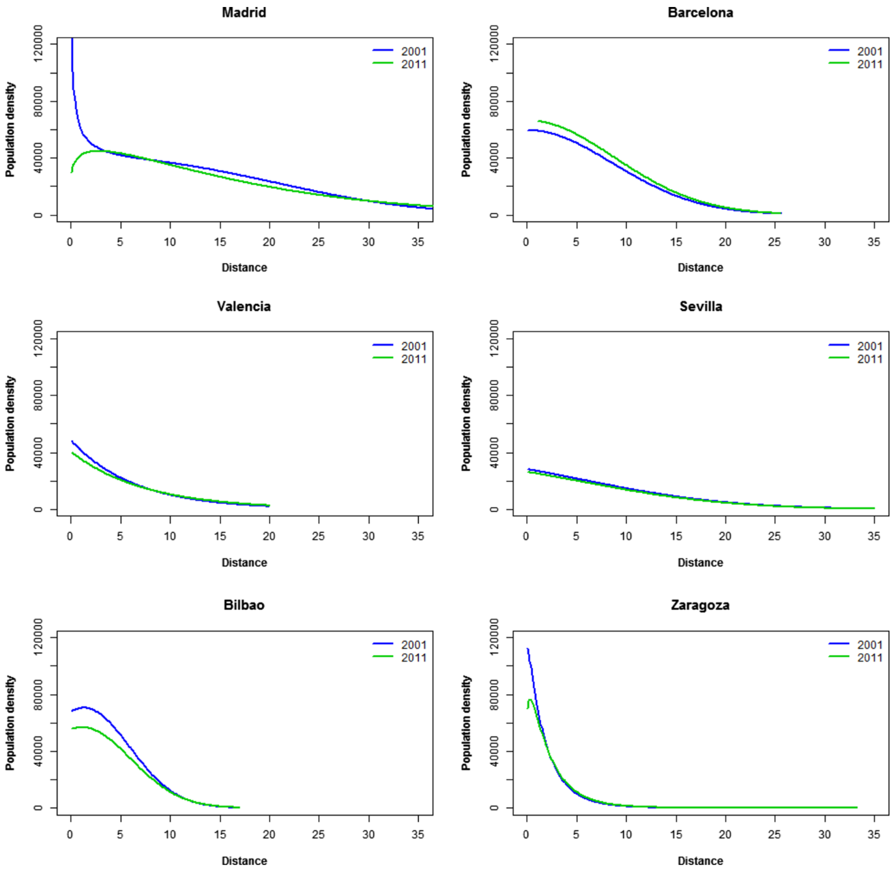

Fig. 5 Metropolitan areas analysed: density profiles with distance

5 Conclusions

In this paper, we have analysed the changes in urban population density that have taken place during the real estate bubble in the six most populated Spanish metropolitan areas. It is also worth noting an important and fast arrival of immigrants coming from developing countries. For this analysis, we have used classical econometric models where the distance to the CBD is the unique explanatory variable. In particular, we have focused on Madrid, Barcelona, Valencia, Sevilla, Bilbao and Zaragoza. Census tracts have been used as the spatial unit to describe and analyse the effects of this intense building process on the urban population density changes. Classical models estimated by OLS show that distance is significant and that density is decreasing with distance. This result points to monocentric structures but is also compatible with some forms of polycentric structures.

Hot spots of high density are detected with LISA analysis; these are local points which nuance the general pattern but do not eliminate the signification of distance. There is a vast literature, theoretical and empirical, which focused on investigating if the monocentric model is still valid as a description of metropolitan areas, which is in line with the results in the present paper. We show that, at least in regard to urban density, metropolitan areas are still well described with monocentric models. When analysing monocentric models with other endogenous variables (e.g. employment centres or land prices), a trend to polycentricity or scatteration may appear. However, our results point that the change in urban density is slower. Hence, while there was a dramatic increase in new dwellings built during the period, the urban density distribution did not change so much.

The ESDA tools have demonstrated their usefulness to describe the variations in urban population density. To study whether the distribution of density shows any autocorrelation pattern and whether it has changed between 2001 and 2011, Moran's I and Geary's C have been used. The results show that for four out of six metropolitan areas, the spatial autocorrelation is positive and has grown. Sevilla is the exception as it shows a spreading of density. The results for Barcelona are not conclusive in terms of both statistics. The growth in positive spatial autocorrelation implies that the urban density of each census tract has tended to converge with neighbour areas. This reinforces the idea that the real estate bubble has boosted urban density.

To complement these results and to pick up more spatial details, we have also applied LISA. Hot spot analysis has been carried out. In 2001, hot spots are concentrated in the central zones. But in Madrid and Barcelona, this is not the pattern. Madrid presents several scattered hot spot areas and Barcelona presents three wide hot spots. It is worth noticing that these are the most populated metropolitan areas. The central hot spot zone of Madrid, Barcelona and Zaragoza has experienced a spreading process. Central areas with high urban density become larger than before. This is not a consistent result for all metropolitan areas, as central hot spots have remained stable in Valencia and Bilbao and decreased in Sevilla.

Our results demonstrate the importance and usefulness of classical models. These models, with a large tradition in different research fields such as geography, economics and transport planning, are still practical devices. Changes in urban population density during a process of intensive building activity may be captured with them. We have extended the classical models to include spatial effects. The use of spatial effects is due both to theoretical reasons and to the nature of available data. Spatial effects were explored using two models: spatial lag and spatial error models. Spatial econometric literature discusses what kind of spatial model is more accurate in urban population density analysis. Our work provides empirical evidence that spatial error model performs better than lag spatial model. The introduction of spatial effects implies a reduction in the explaining power of distance. Nevertheless, this reduction does not affect all cities, neither all distance-related coefficients. This result suggests three ways to get further: first, the inclusion of other variables in addition to distance to obtain more powerful insights on urban density; second, to check the location of the CBD; and finally, our results may also point that the metropolitan areas could be not monocentric and, consequently, more adequate models have to be evaluated, for instance, any kind of polycentric models.

Acknowledgments The authors gratefully thank Jordi Suriñach for his valuable suggestions on previous drafts of the paper. Angels Pinyana proof read fruitfully the paper. We also thank the Data Analysis and

Modelling research group for the fruitful discussions we held together. This research was partially supported by Grant MTM2012-38067-C02 from the Ministerio de Economa y Competitividad. We would like to thank the two reviewers for very helpful improvement suggestions.

References

Alonso W (1964) Location and land use: toward a general theory of land rest. Harvard University Press, Cambridge
Alperovich G (1983a) An empirical study of population density gradients and their determinants. J Reg Sci 23(4):529–540
Alperovich G (1983b) Determinants of urban population density functions: a procedure for efficient estimates. Reg Sci Urban Econ 13(2):287–295
Alperovich G, Deutsch J (1992) Population density gradients and urbanisation measurement. Urban Stud 29(8):1323–1328
Alperovich G, Deutsch J (1994) Joint estimation of population density functions and the location of the central business district. J Urban Econ 36(3):239–248
Alperovich G, Deutsch J (2002) An application of a switching regimes regression to the study of urban structure. Pap Reg Sci 81(1):83–97
Anderson JE (1982) Cubic-spline urban-density functions. J Urban Econ 12(2):155–167
Anderson JE (1985) The changing structure of a city temporal changes in cubic spline urban density patterns. J Reg Sci 25(3):413–425
Anselin L (1995) Local indicators of spatial association. Geogr Anal 27(2):93–115
Anselin L, Can A (1986) Model comparison and model validation issues in empirical work on urban density functions. Geogr Anal 18(3):179
Arraiz I, Drukker DM, Kelejian HH, Prucha IR (2010) A spatial Cliff–Ord-type model with heteroskedastic innovations: small and large sample results. J Reg Sci 50(2):592–614
Arribas-Bel D, Sanz-Gracia F (2014) The validity of the monocentric city model in a polycentric age: US metropolitan areas in 1990, 2000 and 2010. Urban Geogr 35(7):1–18
Aynvarg Y (1969) Zones of influence of middle-size cities, their boundaries and passenger flows. Sov Geogr 10:549–558
Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodol) 57(1):289–300
Berry B, Simmons J, Tennant R (1963) Urban population densities: structure and change. Geogr Rev 12:389–405
Bivand R (2014) spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-87
Brueckner J (2011) Lectures on urban economics. MIT Press Books, Cambridge
Bunting T, Filion P, Priston H (2002) Density gradients in Canadian metropolitan regions, 1971–1996: differential patterns of central area and suburban growth and change. Urban Stud 39(13):2531
Cervero R, Wu K-L (1997) Polycentrism, commuting, and residential location in the San Francisco Bay area. Environ Plan A 29(5):865–886
Clark C (1951) Urban population densities. J R Stat Soc Ser A (Gen) 114:490–496
Cliff A, Ord J (1975) Model building and the analysis of spatial pattern in human geography. J R Stat Soc Ser B (Methodol) 37(2):297–348
Cliff AD, Ord JK (1981) Spatial processes: models and applications, vol 44. Pion, London
Davidson R, MacKinnon JG (1981) Several tests for model specification in the presence of alternative hypotheses. Econom J Econom Soc 49(3):781–793
Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels. Springer, Berlin
Fallah B, Partridge M, Olfert M (2012) Uncertain economic growth and sprawl: evidence from a stochastic growth approach. Ann Reg Sci 49(3):589–617
Florax RJ, Folmer H, Rey SJ (2003) Specification searches in spatial econometrics: the relevance of Hendrys methodology. Reg Sci Urban Econ 33(5):557–579
Frankena M (1978) A bias in estimating urban population density functions. J Urban Econ 5(1):35–45
Garcia-López M-À (2012) Urban spatial structure, suburbanization and transportation in Barcelona. J Urban Econ 72(2):176–190

Garcia-López M-À, Muñiz I (2010) Employment decentralisation: Polycentricity or scatteration? The case of Barcelona. Urban Stud 47(14):3035–3056
Geary RC (1954) The contiguity ratio and statistical mapping. Inc Stat 5(3):115–146
Getis A, Ord JK (1992) The analysis of spatial association. Geogr Anal 24(3):189–206
Giuliano G, Small KA (1991) Subcenters in the Los Angeles region. Reg Sci Urban Econ 21(2):163–182
Glaeser E (2008) Cities, agglomeration, and spatial equilibrium. Oxford University Press, Oxford
Glickman N, Oguri Y (1978) Modeling the urban land market: the case of Japan. J Urban Econ 5(4):505–525
Griffith DA (1981a) Evaluating the transformation from a monocentric to a polycentric city. Prof Geogr 33(2):189–196
Griffith DA (1981b) Modelling urban population density in a multi-centered city. J Urban Econ 9(3):298–310
Griffith DA, Wong DW (2007) Modeling population density across major US cities: a polycentric spatial regression approach. J Geogr Syst 9(1):53–75
Holden D, Parr JB (2013) A note on the average density function in urban analysis. Urban Stud 50(14):3027–3035
Latham R, Yeates M (1970) Population density growth in metropolitan Toronto. Geogr Anal 2(2):177–185
LeSage J, Pace RK (2009) Introduction to spatial econometrics. CRC Press, Boca Raton
MacKinnon J, White H, Davidson R (1983) Tests for model specification in the presence of alternative hypotheses: some further results. J Econom 21(1):53–70
MacKinnon JG, White H (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. J Econom 29(3):305–325
Madariaga R, Martori JC, Oller R (2014) Income, distance and amenities. An empirical analysis. Empir Econ 47(3):1129–1146
Martori JC (2010) Las consecuencias del boom inmobiliario. Cambios en la densidad de las metrópolis españolas, 2001–2007. Scr Nova Rev Electrón Geogr Cienc Soc 14(310–347):1–12
Martori JC, Apparicio P (2011) Changes in spatial patterns of the immigrant population of a southern european metropolis: the case of the barcelona metropolitan area (2001–2008). Tijdschr Econ Soc Geogr 102(5):562–581
Martori JC, Suriñach J (2002) Urban population density functions: the case of the Barcelona region. Doc Recer Univ Vic 6:1–18
McDonald J (1989) Econometric studies of urban population density: a survey. J Urban Econ 26(3):361–385
McDonald JF, Bowman HW (1976) Some tests of alternative urban population density functions. J Urban Econ 3(3):242–252
McDonald JF, McMillen DP (2000) Employment subcenters and subsequent real estate development in suburban Chicago. J Urban Econ 48(1):135–157
McMillen DP, Lester TW (2003) Evolving subcenters: employment and population densities in Chicago, 1970–2020. J Hous Econ 12(1):60–81
McMillen DP, McDonald JF (1998) Suburban subcenters and employment density in metropolitan Chicago. J Urban Econ 43(2):157–180
Mills E (1967) An aggregative model of resource allocation in a metropolitan area. Am Econ Rev 57(2):19–210
Mills E (1970) Urban density functions. Urban Stud 7(1):5–20
Mills E (1972) Studies in the structure of the urban economy. The John Hopkins University Press, Baltimore
Ministerio de la Vivienda (2004) Atlas Estadistico de las Areas Urbanas 2004. Ministerio de la Vivienda, Madrid
Moran PA (1948) The interpretation of statistical maps. J R Stat Soc Ser B (Methodol) 10(2):243–251
Mur J, Angulo A (2009) Model selection strategies in a spatial setting: some additional results. Reg Sci Urban Econ 39(2):200–213
Muth RF (1969) Cities and housing: the spatial pattern of urban land use. University of Chicago Press, Chicago
Muth RF (1971) The derived demand for urban residential land. Urban Stud 8(3):243–254
Newling B (1971) The spatial variation of urban population densities. In: Bourne LS (ed) Internal structure of the city. Oxford University Press, New York
Newling BE (1969) The spatial variation of urban population densities. Geogr Rev 59(2):242–252
Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 27(4):286–306
Parr J (1987) Interaction in an urban system: aspects of trade and commuting. Econ Geogr 63(3):223–240

Parr J (2012) Spatial-structure differences between urban and regional systems. Ann Reg Sci 49(2):293–303
Parr J, O'Neill G (1989) Aspects of the lognormal function in the analysis of regional population distribution. Environ Plan A 21(7):961–973
Parr J, O'Neill G, Nairn A (1988) Metropolitan density functions: a further exploration. Reg Sci Urban Econ 18(4):463–478
Piras G (2010) sphet: Spatial models with heteroskedastic innovations. R package version 0.1-22
Smeed RJ (1963) The effect of some kinds of routing systems on the amount of traffic in central areas of towns. J Inst High Eng 10(1):5–26
Stewart JQ (1947) Empirical mathematical rules concerning the distribution and equilibrium of population. Geogr Rev 37(3):461–485
Stewart JQ, Warntz W (1958) Physics of population distribution. J Reg Sci 1(1):99–121
Tanner JC (1961) Factors affecting the amount of travel. Road Res Tech Pap 51:46–73
Tsai Y-H (2005) Quantifying urban form: compactness versus' sprawl'. Urban Stud 42(1):141–161
Wang F, Zhou Y (1999) Modelling urban population densities in Beijing 1982–1990: suburbanisation and its causes. Urban Stud 36(2):271
White L (1977) How good are two-point estimates of urban density gradients and central densities? J Urban Econ 4(3):292–309
Zheng S, Kahn ME (2008) Land and residential property markets in a booming economy: new evidence from Beijing. J Urban Econ 63(2):743–757
Zielinski K (1979) Experimental analysis of eleven models of urban population density. Environ Plan A 11(6):629–641