Regional Science and Urban Economics 42 (2012) 739–748

Contents lists available at SciVerse ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Change in the distribution of house prices across Spanish cities☆

Catia Nicodemo [a,b,*], Josep Maria Raya [c,d]

[a] Universitat Autònoma de Barcelona, Spain
[b] IZA, Germany
[c] Department of Economics and Business of Universitat Pompeu Fabra, Spain
[d] University School of the Maresme, Spain

ARTICLE INFO

Article history:
Received 9 June 2011
Received in revised form 30 April 2012
Accepted 3 May 2012
Available online 11 May 2012

JEL classification:
C1
R21
R31

Keywords:
Housing price distribution
Housing market
Quantile regression
Counterfactual distribution

ABSTRACT

This paper presents quantile estimates of house prices across two years, 2004 and 2007, in several Spanish cities. We decompose the change in the house price distribution into portions due to changes in the distributions of the explanatory variables and to changes in coefficients over time. We obtain three main results. First, from 2004 to 2007, the change in house prices in Spain is larger at both lower and higher percentiles. Second, most of the difference in the distribution of house prices between 2004 and 2007 is explained by coefficients (with all the variables contributing similarly). Third, we find notable variation in the changes in the house price distribution across cities, with Madrid, Valencia, and Bilbao showing the largest changes.

1. Introduction

Over the period 2004–2009, the price of housing in Spain rose by 31.91%.[1] This period can be divided into two stages: the end of a boom and the beginning of a slump in house prices in Spain. Whereas in the period between 2004 and 2007 the prices rose by about 44.31%, at the end of 2007 house prices entered a downturn in which they fell by 8.60% through the beginning of 2009.[2]

The higher appreciation rate of house prices in Spain is not quite uniform across all major cities. In Table 1 we report the price per square meter in euros for 2004 and its growth rate after three years for the main Spanish provincial capitals. The change in house prices is not the same across cities, and seems to point toward convergence for the main Spanish cities: cities in which prices are relatively low at the beginning of the period analyzed, such as Valencia and Seville, display higher price growth over 2004–2007. Most of these cities are the capitals of the various regions of Spain.

In the face of this evidence, one of the objectives of this paper is to understand the evolution of the house price distribution across time and across regions in Spain. Very few studies have focused on change in the distribution of prices across regions, so the full distribution of house prices over time and its decomposition is a field of housing economics that merits attention. Recent studies on the cross-sectional distribution of house prices include García and Raya (2010), Gyourko et al. (2006), Maza and Pages (2007) and McMillen (2008). Gyourko et al. (2006) and Maattanen and Tervio (2010) studied the relationship between the house price distribution and the income distribution. Maattanen and Tervio (2010) ask whether the recent increases in income inequality in the United States have had any impact on the distribution of house prices. Maza and Pages (2007) suggest that income and nominal interest rates are pivotal explanatory factors in the growth of house prices. In turn, McMillen (2008) focuses on change in the house price distribution over time and asks whether changes in the price distribution come from a change in the distribution of house characteristics such as size, location and age, or from a change in the implicit prices associated with those characteristics. McMillen (2008) shows that the decomposition method is a useful tool to look at the distribution of house prices, and the findings can have important implications (such as for price indexes). Finally, García and Raya (2010) explain the variance in the house price distribution at a point in time rather than over a period by estimating the full impact of the explanatory variables on the Gini index.

☆ We are grateful to the editor and the two anonymous referees for their useful comments. Any remaining errors are ours.
* Corresponding author at: Dept. of Applied Economics, Universitat Autònoma de Barcelona, Edifici B, Campus UAB, 08193 Bellaterra, Barcelona, Spain.
E-mail address: catia.nicodemo@uab.cat (C. Nicodemo).
[1] Source: Ministry of Housing, 2009.
[2] Source: The Ministry of Housing uses appraisal prices. If we use transaction prices instead, the downturn began in early2007 and the drop in prices to date is notably larger, as will be observed in the present paper.

0166-0462/$ – see front matter
doi:10.1016/j.regsciurbeco.2012.05.003

Table 1
House prices and growth rate by cities.
Source: Ministry of Housing.

City	2004[a]	Δ0 (4–07)
Barcelona	2732.2	29.55%
Bilbao	2491.7	34.65%
Madrid	3022.4	28.70%
Malaga	1649.6	36.22%
Seville	1633.4	41.80%
Valencia	1431.3	53.21%
Zaragoza	1804.9	50.37%
Spain	1618.0	44.31%

[a] Price per square meter in euros.

In this context, we are interested in knowing to what extent the difference between house prices become smaller or larger over the period concerned. In this way, we can determine whether the price of cheaper houses increased more or less than the price of more expensive housing. Our analysis addresses these questions for several Spanish cities (Madrid, Barcelona, Valencia, Seville, Zaragoza, Malaga and Bilbao) with the objective of understanding the determinants of the variation of the price across cities.

For this purpose we use a dataset from a private real estate agency that provides very detailed information about the houses sold. Changes in the price distribution can be as a result of more houses or apartments being sold in high price neighborhoods (i.e., changes in the locations of homes sold), different characteristics of housing sold (i.e., changes in the size and quality of the homes sold), or changes in the underlying hedonic price functions. To study the changes in the distribution of house prices from 2004 to 2007, we apply a decomposition method proposed by Melly in 2005, which enables us to isolate these components of house price changes.

The focus of our paper is closely related to the issues discussed in McMillen (2008), but differs in its focus on Spain and because we present a multi-city analysis. In addition, we use Melly's (2005) technique instead of that of Machado and Mata (M–M) (2005), which is the most common quantile regression-based decomposition. Recent research has focused on changes in the whole distribution rather in the average. There are many studies that have extended the Oaxaca (1973) and Blinder (1973) decomposition of differences in the mean to decompositions of the whole distribution. Juhn et al. (1993, JMP) proposed an extension of the Oaxaca decomposition that takes into account the distribution of residuals. DiNardo et al. (1996) and Lemieux (2002) propose an estimator of the counterfactual that makes the distribution of skills constant across time. M–M (2005) propose a method based on the estimation of marginal wage distributions consistent with a conditional distribution estimated by quantile regression. Melly (2005), following M–M, proposed a semi-parametric estimator of distribution functions in the presence of covariates. The method is based on the estimation of the conditional distribution by quantile regression. More information about the method of decomposition and quantile regression is provided by Firpo et al. (2007).

Several authors make use of the M–M decomposition technique in their applications (e.g., Albrecht et al., 2009; Kohn, 2006) to study the wage distribution and recently to estimate the effect of neighborhood diversity (Cobb-Clark and Sinning, 2011), house prices (McMillen, 2008) or immigration and house prices (Degen and Fischer, 2009). In this work we apply the method developed by Melly in 2005. The basic idea of Melly's semi-parametric estimator is to estimate the whole conditional distribution function by parametric quantile regression and integrate it over the range of covariates to obtain an estimate of the unconditional distribution. The quantile regression framework does not need any distributional assumptions and allows the covariates to influence the whole conditional distribution. The first step of the estimation procedure has both a statistical and an economic interpretation, i.e., the quantile regression coefficients can be interpreted as rates of return to house prices at different points of the price distribution (Buchinsky, 1998). This method solves the problem of crossing of different quantile curves and derives the asymptotic distribution. Crossing is an inconsistency which sometimes appears in conditional quantile regressions (as proposed by Koenker and Bassett, 1978). See Chernozhukov et al. (2010) for a further discussion. To fix this problem, Melly, as well as Chernozhukov et al., propose a way to essentially re-order the coefficients based on the individual distribution of predicted values.[3] In addition, it is a fast method for computing the integration of the estimated conditional quantile to obtain the unconditional distribution.

Using this method, we compare differences in house prices and their determinants among the larger Spanish cities to determine whether there are significant differences across regions. We explore the extent to which a housing bubble has different effects on the cross-sectional price distribution in different Spanish cities. As we have observed, the study of house price distributions is a major area for future research and studies of differences across regions are scarce, so with this paper we seek to contribute to the literature on the distribution of house prices at regional level.

We present three main results. First, from 2004 to 2007, the difference in the house price distribution in Spain is larger at lower and higher percentiles. Second, most of the difference in the distribution of house prices between 2004 and 2007 is explained by coefficients (with all the variables contributing similarly). Third, we find notable variation in the changes in the house price distribution across cities. Compared with the general pattern of Spanish cities, Madrid, Valencia, and Bilbao stand out as being significantly different.

The structure of this paper is as follows. In Section 2, we introduce the Spanish cities and point out some differences among them. The next section presents the data and some descriptive statistics. Section 4 includes a discussion of the empirical model and alternative estimation strategies. Section 5 summarizes the results, with a special focus on comparisons across cities. Finally, we will make some concluding remarks.

2. A brief overview of the Spanish cities analyzed

Spain has a federal system of government that leaves a great deal of autonomy to seventeen regions. Because these regions differ significantly across various dimensions — cultural, political, and economic — it is not appropriate to treat Spain as a single, fully integrated housing market. In this section we discuss some of the differences between major Spanish cities.[4]

Table 1 shows initial house prices (2004) and the growth rates for each seven large cities (2004–2007). In this regard, it would be worthwhile to discuss potential differences among the cities that might account for the differences in the price distribution changes. Table 2 shows that economic conditions differ significantly across Spanish cities. Cities with high house prices such as Barcelona, Bilbao and Madrid also have high per capita income and low levels of youth unemployment. The cities with the highest growth in house prices between 2004 and 2007 (Valencia, Zaragoza, Seville, and Malaga)

[3] In M–M the estimated coefficients are randomly assigned and their re-ordering has no notable effect on the counterfactual unconditional distributions.

[4] Some authors have discussed these differences (and their causes) among Spanish cities: (Balmaseda et al., 2002; Taltavull, 2003). In general, the theoretical fundamentals of the determinants of house prices have been exhaustively studied in the housing literature. In these house price models the explanatory variables are basically: population, unemployment (especially youth unemployment), income, interest rates and residential capital (Nellis and Longbottom, 1981; Meen, 1990; DiPasquale and Wheaton, 1994; Johnes and Hyclak, 1994; Potepan, 1994; Abraham and Hendershott, 1996; England and Ionannides, 1997; Hort, 1998; Malpezzi, 1999; Jud and Winkler, 2002; Meen, 2002; Andrew and Meen, 2003; and Taltavull, 2003).

Table 2
Economic differences among Spanish cities.
Source: Spanish Census Bureau and Spanish Housing Ministry, 2009.

City	Popul. 2004	Growth	Unempl. 2004	Growth	P.c. income 2004	Growth	Growth of res. cap. stock
Barcelona	1,578,546	1.05	17.35	−3.46	14,464	14.73	5.32
Bilbao	352,317	0.24	17.52	−0.06	15,988	21.75	4.94
Malaga	547,731	2.47	24.68	−0.35	10,145	17.42	11.71
Madrid	3,099,834	1.05	21.10	−4.66	15,297	15.17	7.17
Seville	704,203	−0.72	24.68	−0.35	10,145	17.42	7.81
Valencia	785,732	1.52	24.13	−3.72	11,802	12.88	7.32
Zaragoza	638,799	2.44	13.65	0.22	13,762	18.78	4.69

also tend to have the highest rates of population, the highest reduction in youth unemployment rates, the highest per capita income growth, and the highest growth in the residential capital stock.

3. Data

The database used in this research comprises 21,517 houses sold over the period 2004–2007 by a real estate agency for which the transaction details and the physical and location characteristics of each property are known. The data were collected bi-annually from 2004 to the last semester in 2007. In our paper, we use for the pooled data set for the hedonic model and the subset of sales from 2004 and 2007 for the decomposition analysis. The variables included in the analysis are described in the Appendix (see Table A.1).

In Tables 3 and 3a we report descriptive statistics for the variables used in estimation for our sample of Spanish cities for two different years (2004 and 2007). Madrid and Barcelona have the highest average house prices compared to other Spanish cities in 2004. By 2007, the price growth in house prices was at least 20% for all cities (with the exception of 17% for Barcelona). The growth of average prices varies significantly across cities. In cities such as Valencia and Zaragoza, the increase is above 40%, which is in line with the official price published by the Ministry of Housing (see Table 1). The characteristics of the buildings sold do not change much from one year to the other; the most notable change is that the structural variable "lift" is more common in 2007.

In Figs. 1 and 2 we show kernel density estimates of the price per square meter for the full sample of Spanish cities and for the each city studied. For Spain as a whole, the 2007 price distribution became less skewed (thicker on the right), shifted to the right, and had lower kurtosis when compared with the distribution in 2004.

The same patterns are evidence for the kernel density estimates for individual Spanish cities. By 2007, their house price distributions had shifted to the right. However, the shift was larger for smaller cities such as Valencia or Bilbao than for cities such as Barcelona or Madrid.

4. Empirical models

4.1. Hedonic regression

Our objective is to determine whether the change in the distribution of house prices in Spanish cities can be explained by the size, quality and locational characteristics of the homes sold, and if there are systematic differences between different cities within a country. A widely used approach to deal with product heterogeneity in terms of quality is hedonic analysis. The economic literature on hedonic prices arose in the context of the car market. This was the framework for the classical work by Griliches (1971), who popularized these models. Once the technique had been popularized in the 1950s (Tinbergen, 1951), it took over a decade to provide it with a theoretical foundation. The classic work was that of Rosen (1974),[5] who shows how marginal prices are determined implicitly for the characteristics of heterogeneous products.

In this section, we estimate a hedonic function model by OLS. Housing fits perfectly into the framework of hedonic price models. Good examples of hedonic studies of the housing market include Palmquist (1984), Mendelsohn (1984), Bartik (1987), Mills and Simenauer (1996), Bover and Velilla (2001), Ekeland et al. (2002, 2004), Bin (2005), and García and Raya (2010). Quantile regression has also recently been used in the literature on housing economics (see McMillen and Thorsnes, 2006; Coulson and McMillen, 2007; and Zietz et al., 2008).

The model that will be considered has the following specification:

$$p_i = \delta + X'_i + \sum_j \alpha_j D_i^j + \sum_t \tau_t T_i^t + u_i \tag{1}$$

where p represents the price of the house (measured as the logarithm of the price per square meter), X is a vector of physical characteristics, D^j is a dummy variable corresponding to the area j in which the house is located, T is a dummy variable representing the semester of sale, and u is an error term. The parameters to be estimated are δ, β, α, and τ. The results are reported in Table 4, where we present the results of the hedonic pricing model by quantile for the total sample and for the four years. The OLS estimates imply that an increase of 1 year in the age of the dwelling produces a 0.04% decrease in the price per square meter, while an increase of 1 m^2 in the useful floor area decreases the price per square meter by 0.48%. The availability of a lift increases the price per square meter by 14.93%, while having a separate kitchen in the dwelling decreases the price per square meter by 1.91%. Finally, bedrooms and floor are not significant in the OLS equation.

The quantile estimates imply that the age of a dwelling has a diminishing (less negative) quantitative effect on ale prices as we advance toward higher percentiles. The impact of floor area also is lower at higher percentiles. In the case of the variables "type of kitchen", "floor", and "bedrooms", the sign of the estimated coefficients changes from negative to positive as we advance to higher percentiles. In other words, an increase in the number of bedrooms, living on a higher floor and having a separate kitchen in the dwelling are all positively valued at lower percentiles but negatively valued at higher percentiles. Finally, all other variables being equal, the having a lift increases sale price, and the effect appears to increase as we advance towards higher percentiles.

4.2. Quantile decomposition

How does the distribution of house prices evolve over time? Did the distribution change because homes sold in 2004 have different characteristics (size, neighborhoods, etc.) from homes sold in 2007? Or is the distribution change unrelated to the size, quality, and location of the house, being caused by changes in the basic hedonic price functions?

To answer these questions, we follow the literature on the distribution of earnings and examine the effects of covariates on different quantiles of the log price distribution. We can decompose changes in the distribution of house prices into the part due to changes in the distribution of the explanatory variables and the part induced by changes in the coefficients of the quantile regression.

Machado and Mata (2005) presented an approach using quantile regression to decompose differences in log wages between two

[5] The hedonic technique rests on modern consumer choice theory, according to which the consumer derives utility not directly from the good but from its characteristics. See Lancaster (1966).

Table 3
Descriptive statistics, 2004.

Year 2004	Madrid	Barcelona	Malaga	Seville	Zaragoza	Bilbao	Valencia
Log of sale price	5.17	5.24	4.80	4.49	4.82	5.14	4.44
Log of floor area	4.07	4.10	4.22	4.15	4.09	4.18	4.24
Age	42.43	50.72	29.91	33.53	35.82	42.52	39.07
No. of bedrooms	2.55	2.77	2.88	2.89	2.57	2.58	3.01
Floor	2.71	3.08	3.27	2.59	3.09	3.32	3.38
Type of kitchen	0.96	0.90	0.98	0.98	0.92	0.96	0.94
Lift	0.29	0.47	0.64	0.30	0.43	0.60	0.45
Tot. obs.	1731	791	177	567	590	214	424

Table 3a
Descriptive statistics, 2007.

Year 2007	Madrid	Barcelona	Malaga	Seville	Zaragoza	Bilbao	Valencia
Log of sale price	5.34	5.48	5.00	4.87	5.17	5.35	4.99
Log of floor area	4.01	4.06	4.270	4.16	4.11	4.09	4.23
Age	44.31	50.49	32.50	42.17	36.73	47.61	37.47
No. of bedrooms	2.47	2.68	2.69	2.69	2.78	2.46	2.90
Floor	2.61	3.30	3.32	2.29	3.02	3.33	3.23
Type of kitchen	0.94	0.83	0.94	0.98	0.98	0.91	0.93
Lift	0.37	0.52	0.66	0.23	0.52	0.47	0.49
Tot. obs.	874	403	246	294	209	163	151

groups, which allows the differences at various quantiles of the distributions to be analyzed. Assuming linearity between the quantiles of the dependent variable Y and the covariates X, then the τ^{th} conditional quantile of Y is given by the following equation:

$$Q_y(Y|X) = X_i\beta(\tau)\forall\tau, i\in(0,1)$$

than Koenker and Bassett (1978) solve by minimizing in $\beta(\tau)$:

$$\hat{\beta}(\tau) = \min n^{-1}\left[\sum_i^n \rho_\tau(Y_i - X_i\beta)\right], (i = 1, \ldots n),$$

with the check function ρ_t weighting the residuals μ_i asymmetrically:

$$\rho_\tau(\mu_i) = \begin{cases} \tau\mu_i & \text{if } \mu_i \geq 0, \\ (\tau-1)\mu_i & \text{if } \mu_i < 0, \end{cases} 0.$$

In an approach analogous to the Blinder (1973) and Oaxaca (1973) decomposition for OLS estimates, Machado and Mata (2005) proposed an estimator of counterfactual unconditional wage distributions based on quantile regressions. The difference of the θ^{th} unconditional quantile between two groups' distributions can be decomposed as:

$$\hat{F}_{Y1}^{-1}(\theta|T=1) - \hat{F}_{Y0}^{-1}(\theta|T=0) = \underbrace{\hat{F}_{Y1}^{-1}(\theta|T=1) - \hat{F}_{Y1}^{-1}(\theta|T=0)}_{Characteristics} + \underbrace{\hat{F}_{Y1}^{-1}(\theta|T=0) - \hat{F}_{Y0}^{-1}(\theta|T=0)}_{Coefficients} + residual$$

where $\hat{F}_{Yt}^{-1}(\theta|T=t)$ denotes the θ^{th} unconditional quantile of wage Y for group t, while $\hat{F}_{Yt}^{-1}(\theta|T=0)$ is the counterfactual unconditional wage distribution. Usually it is easy to estimate the conditional distribution function by inverting the conditional quantile function. However, the estimated conditional quantile function is not necessarily monotonic, which means that it may be difficult to invert. Moreover, no asymptotic results or variance estimators were provided in M–M 2005.

Melly (2005) extended the M–M method to help circumvent these problems. Melly (2005) proposed to estimate first the whole conditional wage distribution by quantile regression. Then, the conditional distribution is integrated over the range of covariates to obtain an estimation of the unconditional distribution. Melly's method can be represented in the following expressions:

$$F_{yt}(q|X_i) = \int_0^1 1\left(F_{yt}^{-1}(\tau|X_i) \leq q\right) = \int_0^1 1\left(X_i\beta_t(\tau) \leq q\right) d\tau \quad \forall\tau, i\in(0,1)$$

where $F_{yt}\,(q|X_i)$ is the conditional quantile function of the dependent variable Y evaluated at quantile q given a set of characteristics $X = X_i$, and the conditional quantiles of Y are linear in X.

An estimator of the conditional distribution of Y_t given X_i at q is:

$$\hat{F}_{yt}(q|X_i) = \int_0^1 1\left(X_i\hat{\beta}_t(\tau) \leq q\right) d\tau = \sum_{j=1}^n \left(\tau_j - \tau_{j-1}\right) 1\left(X_i\hat{\beta}_t(\tau) \leq q\right)$$

A natural estimator of the θ_{th} quantile of the unconditional distribution of Y is given by:

$$\hat{F}_{yt}(q|T=t) = \frac{1}{n_t}\sum \hat{F}_{yt}(q|X_i)$$

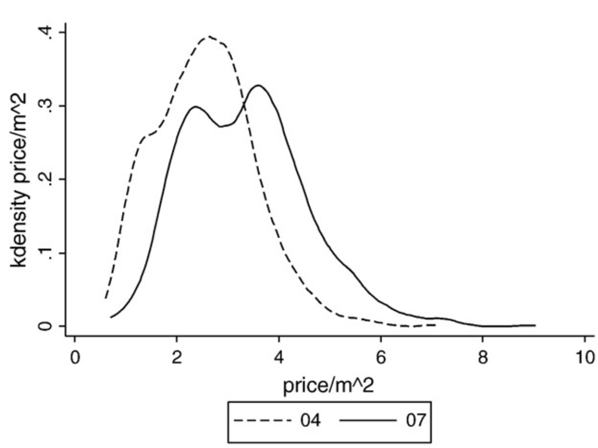

Fig. 1. Kernel density estimation for Spain.

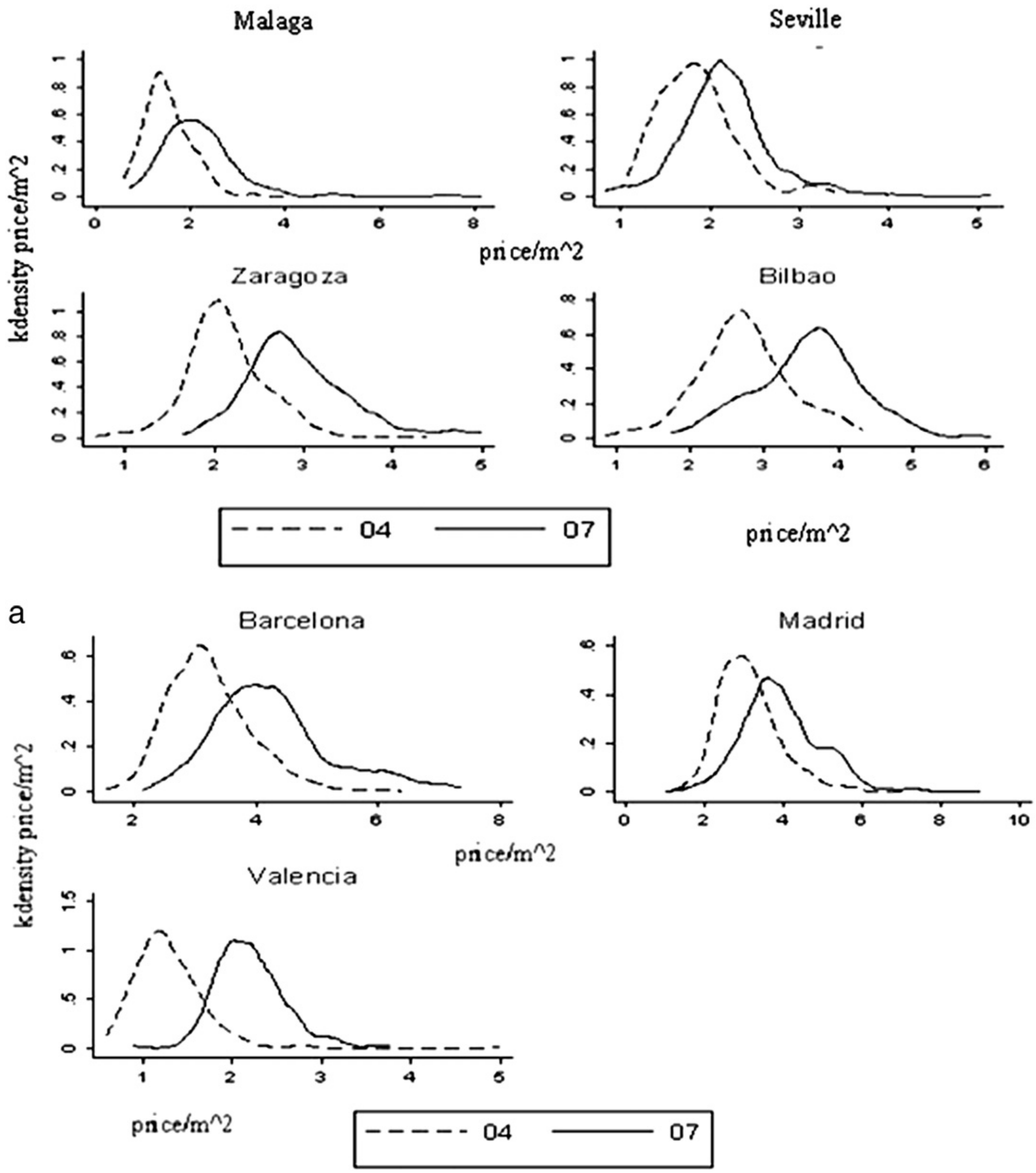

Fig. 2. Kernel density estimation for Spanish cities.

Table 4
Estimation results of the hedonic pricing model.

	OLS	Q10	Q25	Q50	Q75	Q90
Age	−0.0004**	−0.0025**	−0.0022**	−0.0018**	−0.0012**	−0.0004
Floor area	−0.0048**	−0.0067**	−0.0061**	−0.0054**	−0.0045**	−0.0039**
Lift	0.1493**	0.1109**	0.1065**	0.1136**	0.1301**	0.1578**
Type of kitchen	−0.0191**	0.0299**	−0.0017	−0.0221**	−0.0368**	−0.0541**
Floor	0.0003	0.0041**	0.0011**	0.0001	−0.0017**	−0.0042**
Bedrooms	0.0031	0.0337**	0.0106**	−0.0031*	−0.0149**	−0.0230**
Semester (ref.: 2004:I)						
2004:II	0.0632**	0.0800**	0.0646**	0.0618**	0.0554**	0.0460*
2005:I	0.1406**	0.1517**	0.1448**	0.1417**	0.1263**	0.1092**
2005:II	0.2166**	0.2368**	0.2212**	0.2060**	0.1944**	0.1715**
2006:I	0.2836**	0.2915**	0.2940**	0.2793**	0.2713**	0.2443**
2006:II	0.3105**	0.3293**	0.3200**	0.3124**	0.3005**	0.2738**
2007:I	0.3124**	0.3340**	0.3290**	0.3165**	0.2994**	0.2691**
2007:II	0.279**	0.2651**	0.2578**	0.2683**	0.2723**	0.2528**
2008:I	0.1495**	0.1875**	0.1835**	0.1836**	0.1758**	0.1596**
2008:II	0.0544**	0.0227	0.0269*	0.0318**	0.0464**	0.0552**
Intercept	0.6025**	0.4166**	0.6439	0.7827**	0.8534**	0.9165**
R-squared	0.79	0.61	0.63	0.63	0.6	0.6
No. obs.	21,517					

** Significance at 5%.
* Significance at 10%.

From this expression we can calculate the unconditional and counterfactual quantile distributions, which are respectively:

$$\hat{q}_t(\theta) = \inf\left\{q: \frac{1}{n_t}\sum_t \hat{F}_{yt}(q|X_i) \geq \theta\right\}$$

$$\hat{q}_t(\theta) = \inf\left\{q: \frac{1}{n_t}\sum_0 \hat{F}_{yt}(q|X_i \geq \theta)\right\}.$$

Following Oaxaca–Blinder (1973), the difference between house prices in 2004 and 2007 can be decomposed into the difference in characteristics and coefficients. The differences in characteristics represent changes over time in the types of houses that are in the sample, e.g., differences in the size or age of the homes that were sold in 2004 versus 2007. The differences in coefficients are, in effect, the residual or unexplained part of the change in house prices between 2004 and 2007. The decomposition is given by the following equation:

$$\hat{q_1}(\theta) - \hat{q_0}(\theta) = \underbrace{\hat{q_1}(\theta) - \hat{q_{c1}}(\theta)}_{characteristics} + \underbrace{\hat{q_1}(\theta) - \hat{q_{c1}}(\theta)}_{coefficients}.$$

5. Results

Quantile decompositions using Melly's approach are reported in the following figures for house prices in the whole of Spain and in Spanish cities. In Fig. 3, we report estimates of the log sale price decomposition between the years 2004 and 2007 in Spain, along with confidence intervals. Unlike the simple Oaxaca (1973) decomposition, the method developed by Melly (2005) shows changes across the entire distribution of house prices. In the quantile decomposition, we divide the overall difference in house prices across the two years into the change explained by differences in the characteristics of the sample of houses sold and the change caused by the differences in the estimated coefficients. The differences in characteristics between 2004 and 2007 are in terms of aspects such as size, parking, number of rooms, etc. Changes in coefficients imply changes in the return to these variables; for example, a neighborhood may have become more expensive (land prices have raised).

Fig. 3 shows that the overall price difference between 2004 and 2007is positive in all cases: house prices increased during this period. Most of the difference in price is due to changes in coefficients rather than differences in the characteristics of the homes that have sold. This result is not surprising: houses themselves cannot change quickly, so the differences in characteristics can only explain the change in house prices if there has been a significant change in the types of homes that have sold over time or there in the neighborhoods in which the homes are located. The difference in coefficients is larger at lower percentiles than at higher percentiles. One plausible explanation for this pattern is that the demand for low-priced houses increased due to the sharp increase in house prices. In this respect, most buyers decided to buy low priced houses because they are more affordable (especially those who enter for the first time in the market). Finally, medium priced and high priced houses also increased but less dramatically. These results are different from those obtained in McMillen (2008) for Chicago house prices between 1995 and 2005.

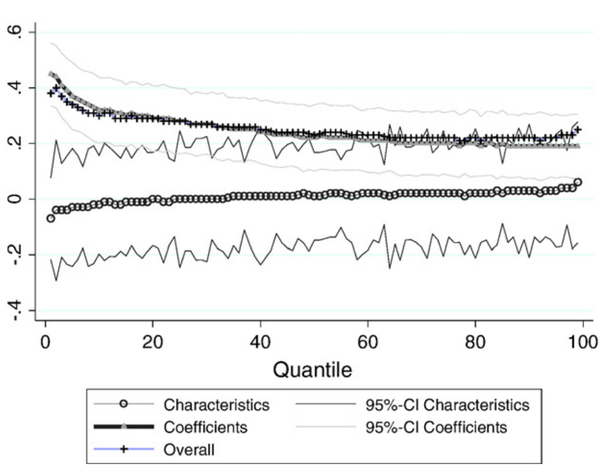

Fig. 3. Decomposition of log house prices, 2004–2007.

In Fig. 4, we repeat the decomposition method for the effects of individual explanatory variables on change in house prices. There may be some variable that affect the price more than others, such as the neighborhood in which the sales are located. The decomposition isolates the effects of individual variables and determines the extent to which they can explain the difference across years.

Fig. 4 shows that the differences in characteristics for individual variables explain very little of the change in the distribution of sales prices. In contrast, the returns on each characteristic (the coefficients) are positive and affect house prices between 2004 and 2007 in the same direction. The implicit prices of nearly all the variables contribute considerably to the change in the distribution (the peak of the change in densities for all variables is about 0.60, with the exception of log size and age, for which the peak is 0.5). Changes in the distributions of the explanatory variables (characteristics) themselves have little or no effect on changes in the house price distribution. In general, the coefficients changed in a way that would tend to reduce the number of medium and high priced sales. In this sense, the forces affecting the distribution of house prices in Spain can be considered different from those of Chicago as reported by McMillen (2008).

5.1. Comparison among cities

In this section, we repeat the decomposition analysis using Melly's method to explain the variation of the prices across cities. Cities in Spain differ for many reasons — geographical history, economic structure, and federal or council government policies. In Section 2, we presented many economic determinants of potential differences among Spanish cities that might account for the differences in the price distribution changes. However, it is difficult to achieve a representative sample at the city level. Our dataset allows us to explore some differences in the distribution of house prices across regions in Spain.

The analysis is reported in Figs. 5 and 6. The majority — but not all — of the cities have a similar pattern to the one we have seen in Fig. 3 for Spain. As we can see in Figs. 5 and 6, price differences are larger at lower and higher percentiles.

Analyzing the results by city, Zaragoza presents a uniform increase of the price along the whole distribution, and is also the city that is least affected by the boom in our dataset. Very similar patterns are observed in Seville and Malaga. In these cases, increasing differences in sales prices at lower percentiles are larger than at higher percentiles. In Barcelona, differences in sales prices are almost constant along the price distribution. The main exceptions to the general pattern are in Madrid and Valencia. In Madrid, price differences increase along the distribution from lower at lower percentiles to higher at higher percentiles. Valencia has the opposite pattern. Bilbao presents the same pattern as Valencia but much less pronounced.

The maximum variation in price differences among percentiles is observed in Valencia, while the minimum variation in price differences is observed in Barcelona. In general, variation in price differences is small for all cities (lower than 0.3), with the exception of Valencia, for which price differences vary from 0.9 to 0.3 approximately. In sum, in Valencia (especially, but also in Bilbao), some convergence in house prices is achieved during the 2004–2007 period,

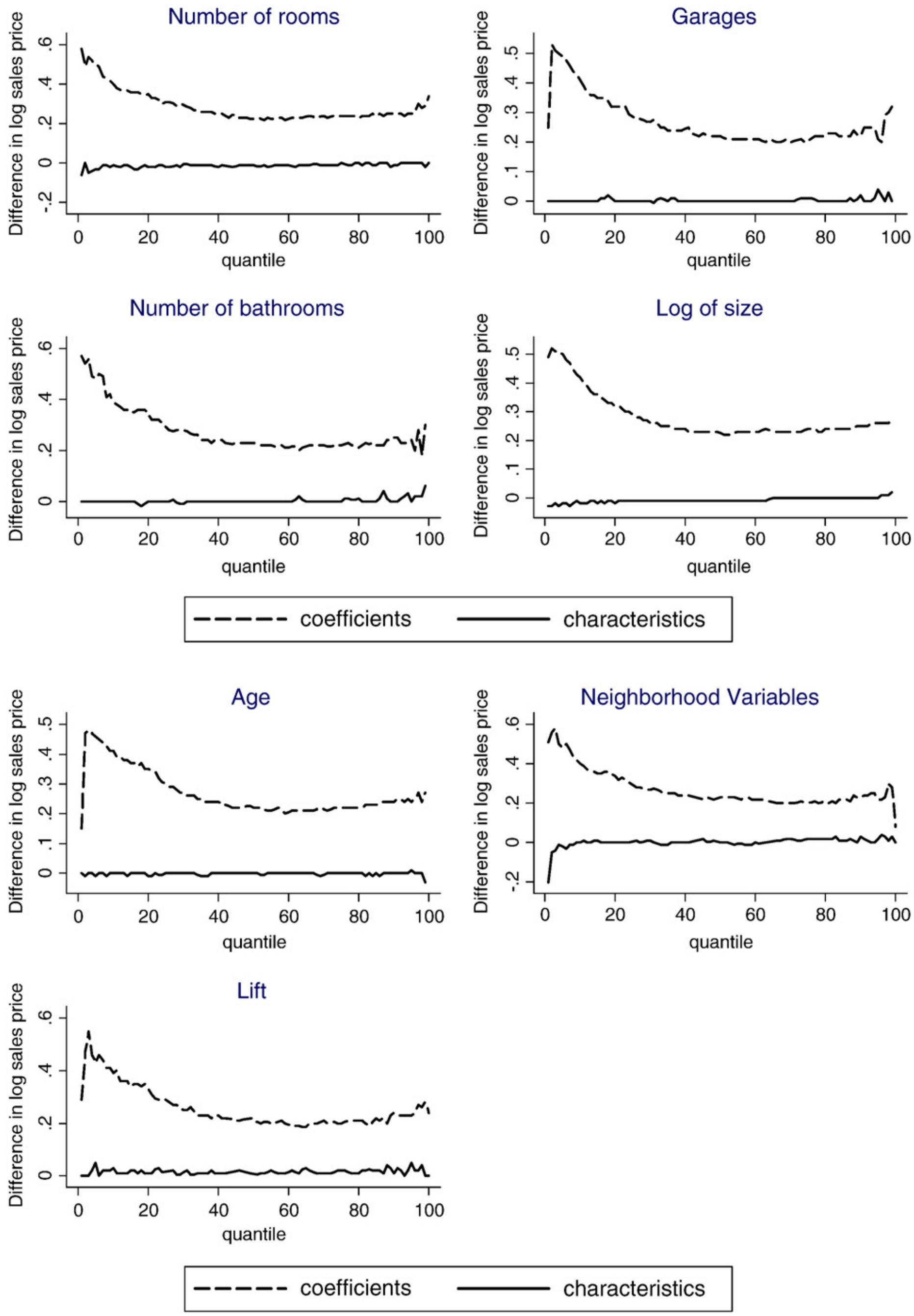

Fig. 4. Decomposition of the effect of explanatory variables on house prices, 2004–2007.

insofar as price growth is higher at lower percentiles. Thus, prices in 2007 are more similar along the distribution. However, some divergence is observed in Barcelona and Madrid. In the two largest cities in Spain, prices in 2007 are less similar along the distribution because price growth is higher at higher percentiles. This smaller price growth at lower percentiles (in relative terms with respect to price growth at higher percentiles) in Barcelona and Madrid may be due to lower demand for homes at these percentiles. This lower housing demand is mainly motivated by the lower population growth observed in Table 2. In conclusion, the distribution shift farther to the right at high prices in big cities (this result is in line with McMillen, 2008), but by more at low prices in smaller cities.

Again, at lower and medium percentiles, the characteristics of the house do not seem to explain the change in the price. However, at higher percentiles, in Madrid, Barcelona, Bilbao, Malaga, Seville and Zaragoza, the change in the price distribution between 2004 and 2007 can also be explained by the change in the characteristics. This moderate effect begins at around the 80th percentile in all these cities with the exception of Zaragoza, where it starts at the 90th percentile. In these cases, it is the change in the characteristics of the sample of houses sold in 2004 and 2007 (see Table 1) that explains the change in the price distribution between 2004 and 2007. Looking at the characteristics we find that in Madrid, Barcelona and Bilbao in 2007, the samples of houses are smaller, in Malaga houses in 2007 are older, and finally, in Zaragoza, in 2007 the percentage of houses with a lift has increased considerably.

In summary, a notable result of this paper is that, although most of the difference in the distribution of house prices between 2004 and

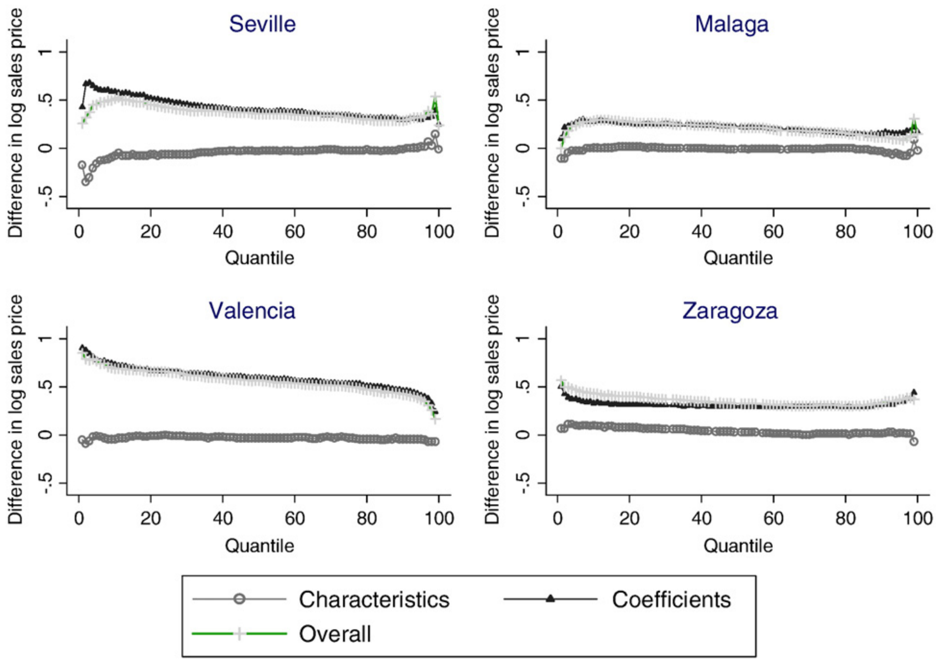

Fig. 5. Decomposition of log house prices by Spanish cities, 2004–2007.

2007 is explained by coefficients in all cases, there are different patterns among Spanish cities. These different patterns are not only a consequence of different patterns in the change in the coefficients, but also in the change of characteristics.

6. Conclusions

In this paper, we have conducted a study of the determinants of changes in the full distribution of prices between 2004 and 2007 (boom period) and the differences in various Spanish cities' behavior. This type of analysis using decomposition methods is useful as an exploratory tool and has important policy implications in terms of price indexes and the study of inequality in house prices. Thanks to these methods, it is possible to observe the internal distribution of the change of the price over time, something that a simple index does not consider. The distribution of house prices for sales of homes in Spain in 2007 became less skewed (thicker on the right), moved to the right, and had lower kurtosis compared to the distribution in 2004. All the Spanish cities experienced a shift in their house price distribution. However, the shift was larger for small cities such as Valencia and Bilbao than for bigger cities such as Barcelona and Madrid.

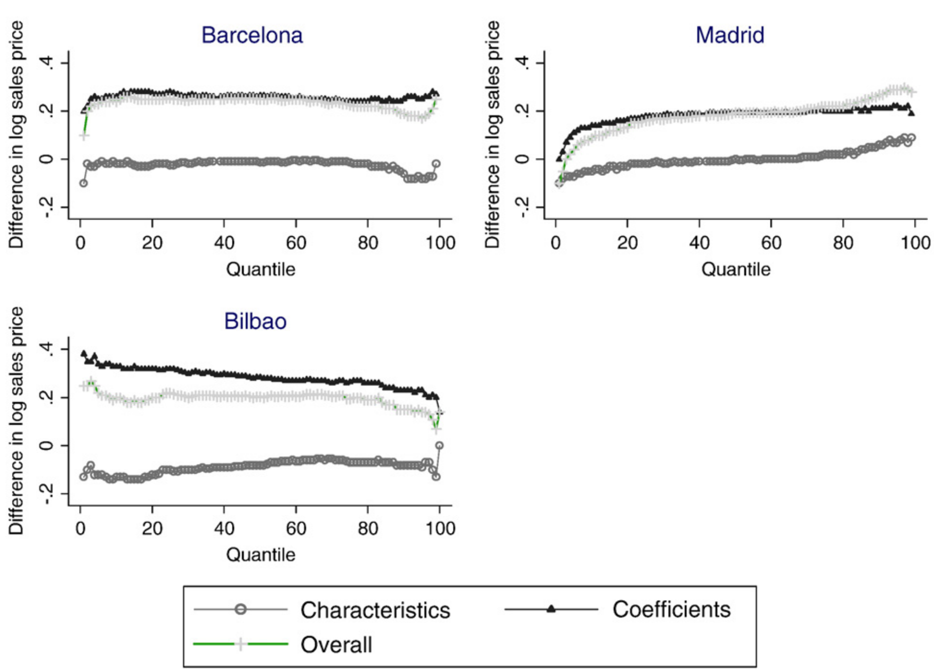

Fig. 6. Decomposition of log house prices by Spanish cities, 2004–2007.

We used a procedure developed by Melly (2005) to decompose this change in the house price distribution into the portion due to changes in the distributions of the explanatory variables and the portion due to changes in their coefficients. This method solves the problem of the crossing of different quantile curves and derives the asymptotic distribution.

We find that the difference is larger at lower percentiles and higher percentiles. Furthermore, most of the difference in the distribution of house prices between 2004 and 2007 is explained by coefficients (with all the variables contributing similarly).

Among cities, we find notable variation in the changes in house price distribution. Although most of the difference in the distribution of house prices between 2004 and 2007 is explained by coefficients in all cases, characteristics also explain some of it. Compared with all of Spain, Madrid, Valencia and Bilbao are the cities whose patterns are most different. Thus, in Valencia (especially, but also in Bilbao), some convergence in house prices is achieved during the 2004–2007 period, insofar as price growth is higher at lower percentiles. Therefore, prices in 2007 are more similar along the distribution. However, some divergence is observed in Barcelona and Madrid. In the two largest cities in Spain, prices in 2007 are less similar along the distribution because price growth is higher at higher percentiles. This smaller price growth at lower percentiles (in relative terms with respect to price growth at higher percentiles) in Barcelona and Madrid may be due to lower demand for homes at these percentiles. This lower housing demand is mainly motivated by the lower population growth observed.

Thus, the results presented here suggest (as pointed out in McMillen, 2008) that a single price index is not an appropriate measure of house price evolution, since variance of the house price distribution increased over time (with, in our case, greater appreciation rates at lower and higher quantiles). Allowing appreciation rates to vary across quantiles provides a more accurate picture of the change in the full distribution of house prices than an estimator that focuses on only the mean or median. And this result is reinforced by the fact that variation across quantiles of the appreciation rates, and its components, are different among the cities which make up the index (see Deng et al., 2012 and McMillen, forthcoming).

Thus, in terms of future research, it would be interesting to test findings in the bust period which started in 2008. In addition, it would also be interesting to use the framework of García and Raya (2010) to analyze which characteristic contributes to equality or inequality in house prices.

Appendix A

Table A.1
Description of the variables.

Variable	Description
Sale price	Price at which each house was sold (in thousands of euros)
Useful floor area	Gross square meters of the dwelling
Age	Number of years that have passed since it was built
Floor	Storey on which the dwelling is located
Rooms	Number of rooms in the dwelling
Lift	Dummy variable taking the value 1 if the property has a lift and 0 otherwise
Type of kitchen	Dummy variable taking the value 1 if the dwelling has a separate kitchen and 0 if it has an open kitchen
Location	Dummy variables corresponding to the postal district of each house

Table A.2
Detailed descriptive statistics.

Year 2004	Madrid					Barcelona					Malaga				
Percentile	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
Log of sale price	0.8	0.9	1.1	1.3	1.4	0.9	1.0	1.1	1.3	1.4	0.3	0.4	0.6	0.7	0.9
Log of floor area	30	48	60	72	88	41	50	60	71	85	47	58	69	80	95
Age	24	33	39	44	69	27	33	39	74	94	8	24	30	37	44
No. of bedrooms	2	2	3	3	3	2	2	3	3	4	2	3	3	3	4
Floor	0	1	2	4	5	1	1	3	4	6	1	1	3	4	6
Type of kitchen	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Lift	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1

Year 2004	Seville					Zaragoza					Bilbao					Bilbao				
Percentile	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
Log of sale price	0.2	0.2	0.4	0.6	0.8	0.5	0.6	0.7	0.9	1.0	0.7	0.9	1.0	1.1	1.3	0.2	0.2	0.4	0.6	0.7
Log of floor area	42	54	64	75	87	42	50	60	70	80	50	56	65	75	90	50	60	70	80	95
Age	17	26	32	40	54	10	20	36	40	53	24	32	36	45	74	22	30	34	40	50
No. of bedrooms	2	3	3	3	4	2	2	3	3	3	2	2	3	3	3	2	3	3	3	4
Floor	0	1	2	4	5	1	2	3	4	6	1	1	3	5	6	1	2	3	5	5
Type of kitchen	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Lift	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1

Year 2007	Madrid					Barcelona					Malaga				
Percentile	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
Log of sale price	1.0	1.2	1.3	1.5	1.6	1.2	1.3	1.4	1.5	1.8	0.4	0.6	0.8	0.9	1.0
Log of floor area	35	45	58	70	88	35	49	60	75	85	50	60	73	85	100
Age	20	33	39	45	67	20	30	39	57	89	10	25	30	35	50
No. of bedrooms	1	2	3	3	3	1	2	3	3	4	2	2	3	3	4
Floor	0	1	2	4	5	1	1	3	5	6	0	1	3	5	7
Type of kitchen	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Lift	0	0	0	1	1	0	0	1	1	1	0	0	1	1	1

Year 2007	Seville					Zaragoza					Bilbao					Valencia				
Percentile	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
Log of sale price	0.3	0.5	0.7	0.9	1.2	0.8	0.9	1.0	1.2	1.3	0.9	1.1	1.3	1.4	1.5	0.6	0.7	0.8	0.9	1.0
Log of floor area	40	55	65	75	90	47	54	63	70	80	35	42	50	70	85	50	60	70	84	92
Age	17	29	35	40	45	21	30	35	40	48	30	35	40	49	78	25	30	35	42	47
No. of bedrooms	2	2	3	3	3	2	2	3	3	4	2	2	2	3	3	2	3	3	3	4
Floor	0	1	2	3	4	1	1	3	4	6	0	1	3	4	6	1	2	3	5	7
Type of kitchen	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Lift	0	0	0	0	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	1

References

Abraham, J.M., Hendershott, P.H., 1996. Bubbles in metropolitan housing markets. J Housing Research 7-2, 191–207.

Albrecht, J., Van Vuuren, A., Vroman, S., 2009. Counterfactual distributions with sample selection adjustments: econometric theory and an application to the Netherlands. Labour Economics 16-4, 383–396.

Andrew, M., Meen, G., 2003. House price appreciation, transactions and structural change in the British housing market: a macroeconomic perspective. Real Estate Economics 31-1, 99–116.

Balmaseda, M., San Martin, I., Sebastian, M., 2002. Una aproximación cuantitativa a la burbuja inmobiliaria. Situación Inmobiliaria. Servicio de Estudios del Banco Bilbao Vizcaya.

Bartik, T.J., 1987. The estimation of demand parameters in hedonic price models. Journal of Political Economy 95, 81–88.

Bin, O., 2005. A semiparametric hedonic model for valuing wetlands. Applied Economic Letters 12, 597–601.

Blinder, A.S., 1973. Wage discrimination: reduced form and structural estimates. The Journal of Human Resources 8 (4), 436–455.

Bover, O., Velilla, P., 2001. Hedonic House Prices without Characteristics: The Case of New Multiunit Housing. Banco de España, Estudios Económicos, p. 73.

Buchinsky, M., 1998. The dynamics of change in the female wage distribution in the USA: quantile regression approach. Journal of Applied Econometrics 13, 1–30.

Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78 (3).

Cobb-Clark, D., Sinning, M., 2011. Neighborhood diversity and the appreciation of native- and immigrant-owned homes. Regional Science and Urban Economics 41, 214–226.

Coulson, N., McMillen, D., 2007. The dynamics of intraurban quantile house price indexes. Urban Studies 44 (8), 1517–1537.

Degen, K., Fischer, M., 2009. Immigration and Swiss house prices. Centre for Economic Policy Research 7583 (dp).

Deng, Y., McMillen, D., Sing, T., 2012. Private residential price indices in Singapore: a matching approach. Regional Science and Urban Economics 42, 485–494.

DiNardo, J.E., Fortin, N.M., Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. Econometrica 65, 1001–1046.

DiPasquale, D., Wheaton, W., 1994. Housing market dynamics and the future of housing prices. Journal of Urban Economics 35-1, 1–27.

Ekeland, I., Heckman, J.J., Nesheim, L.P., 2002. Identifying hedonic models. American Economic Review 92, 304–309.

Ekeland, I., Heckman, J.J., Nesheim, L.P., 2004. Identification and estimation of hedonic models. Journal of Political Economy 1, 60–109.

England, P., Ionannides, Y.M., 1997. House price dynamics: an international empirical perspective. Journal of Housing Economics 6, 119–136.

Firpo, S., Fortin, N., Lemieux, T., 2007. Unconditional quantile regressions. NBER Technical Working Papers 0339.

García, J., Raya, J., 2010. Price and income elasticities of demand for housing characteristics in the city of Barcelona. Regional Studies 45 (5), 597–608.

Griliches, Z., 1971. Price Indexes and Quality Change: Studies in New Methods of Measurement. Harvard University Press.

Gyourko, J., Mayer, C., Sinai, T., 2006. Superstar cities. NBER Working Paper No. 12355.

Hort, K., 1998. The determinants of urban house price fluctuations in Sweden 1968–1994. Journal of Housing Economics 7, 93–120.

Johnes, G., Hyclak, T., 1994. House prices, migration, and regional labor markets. Journal of Housing Economics 3, 312–329.

Jud, D., Winkler, D., 2002. The dynamics of metropolitan housing prices. Journal of Real Estate Research 23 (1–2), 29–45.

Juhn, C., Murphy, K.M., Pierce, B., 1993. Wage inequality and the rise in returns to skill. Journal of Political Economy 101, 410–442.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 6 (1), 3350.

Kohn, K., 2006. Rising wage dispersion, after all! The German wage structure at the turn of the century. Discussion Paper 2098. IZA, Bonn.

Lancaster, K.J., 1966. A New Approach to Consumer Theory. Journal of Political Economy 74, 132–157.

Lemieux, T., 2002. Decomposing changes in wage distributions: a unified approach. Canadian Journal of Economics 35, 646–688.

Maattanen, N., Tervio, M., 2010. Income distribution and housing prices: an assignment model approach. CEPR Discussion Papers 945. C.E.P.R.

Machado, J.A.F., Mata, J., 2005. Counterfactual decomposition of changes in wage distribution using quantile regression. Journal of Applied Econometrics 20, 445–465.

Malpezzi, S., 1999. A simple error correction model of house prices. Journal of Housing Economics 8, 27–62.

Maza, L., Pages, J., 2007. Analysis of house prices in Spain. Working Paper No. 0307. Bank of Spain, Spain.

McMillen, D., 2008. Changes in the distribution of housing prices over time: structural characteristics, neighborhood, or coefficients? Journal of Urban Economics 64, 573–589.

McMillen D. (forthcoming): "Repeat sales as a matching estimator", *Real Estate Economics*.

McMillen, D., Thorsnes, P., 2006. Housing renovations and the quantile repeat sales price index. Real Estate Economics 34 (4), 567–584.

Meen, G.P., 1990. The removal of mortgage market constraints and the implications for econometric modelling of UK house prices. Oxford Bulletin of Economics and Statistics 52-1, 1–23.

Meen, G.P., 2002. The time-series behavior of house prices: a transatlantic divide? Journal of Housing Economics 11, 1–23.

Melly, B., 2005. Decomposition of differences in distribution using quantile regression. Labour Economics 12 (4), 577–590.

Mendelsohn, R., 1984. Estimating the structural equations of implicit markets and household production functions. The Review of Economics and Statistics 66, 673–677.

Mills, S.E., Simenauer, R., 1996. New hedonic estimates of regional constant quality house prices. Journal of Urban Economics 39, 209–215.

Nellis, J.G., Longbottom, J.A., 1981. An empirical analysis of the determination of house prices in the United Kingdom. Urban Studies 18-1, 9–21.

Oaxaca, R., 1973. Male–female wage differentials in urban labor markets. International Economic Review 14, 693–709.

Palmquist, R.B., 1984. Estimating the demand for the characteristics of housing. The Review of Economics and Statistics 64, 394–404.

Potepan, M.J., 1994. Intermetropolitan migration and housing prices: simultaneously determined? Journal of Housing Economics 3, 77–91.

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. Journal of Political Economy 1, 34–55.

Taltavull, P., 2003. Determinants of housing prices in Spanish cities. Journal of Property Investment & Finance 21-2, 109–135.

Tinbergen, J., 1951. Some remarks on the distribution of labour incomes. International Economic Papers 1, 195–207.

Zietz, J., Zietz, E., Sirmans, G., 2008. Determinants of house prices: a quantile regression approach. The Journal of Real Estate Finance and Economics 37 (4), 317–333.