IZA DP No. 6503

Change in the Distribution of House Prices across Spanish Cities

Catia Nicodemo
Josep Maria Raya

April 2012

Change in the Distribution of House Prices across Spanish Cities

Catia Nicodemo
*Universitat Autònoma de Barcelona
and IZA*

Josep Maria Raya
*Universitat Pompeu Fabra
and University School of the Maresme*

ABSTRACT

Change in the Distribution of House Prices across Spanish Cities

This paper presents the quantile estimation of house price between two years, 2004 and 2007 (a boom house price period) in several Spanish cities. We decompose the change in house price distribution into portions: changes in the distributions of the explanatory variables and changes in coefficients over time. Our main results are three. Firstly, from 2004 to 2007, the difference in housing price in Spain is larger at lower and higher percentiles. Secondly, the most important part of the difference in the distribution of housing prices between 2004 and 2007 is explained by coefficients (with all the variables contributing similarly). Thirdly, among cities, we can find a lot of variation in change of house price distribution. With respect to Spain's cities pattern, Madrid, Valencia and Bilbao, are the cities which big difference among them.

Corresponding author:

Catia Nicodemo
Dept Applied Economics
Universitat Autònoma de Barcelona
Edifici B - Campus UAB
08193 Bellaterra (Barcelona)
Spain
E-mail: catia.nicodemo@uab.cat

Change in the distribution of house prices across

Spanish cities

1 Introduction

Over the period 2004–2009, the price of housing in Spain rose by 31.91%[1]. However, this period can be broken down into two stages: the end of a boom and the beginning of a slump in housing prices in Spain. Thus, while in the period between 2004 and 2007 prices was around 44.31%, at the end of 2007 housing prices entered a downturn in which they were to fall 8.60% by the beginning of 2009[2].

The higher appreciate rate of house price is Spain is not quite common at all the most important cities. In Table 1 we report the price of squared meter in euro for 2004 and his growth rate after three years for the main Spanish provincial capitals. We can observe that the change of the house price is not the same across cities; however what seems to point towards house price convergence for the main Spanish capitals. Cities in which prices are relatively low at the beginning of the period analysed such as Valencia or Seville, a higher price growth is observed, getting closer to the price of the cities with higher price in 2004. Most of these cities are the capital of the several regions of the Spain.

Table 1: House price and growth rate by cities

City	2004*	$\Delta 0(4-07)$
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%

Source: Ministry of Housing, *Price of squared meter in euro

---

[1] Source: Ministry of Housing, 2009.

[2] Source: The Ministry of Housing uses appraisal prices. If we use transaction selling prices instead, the downturn took place at the beginning of 2007 and the drop in prices to date is notably larger, as will be observed in the present paper.

In front at this evidence one of the objectives of this paper is understand the evolution of the house price distribution across time and across regions in Spain. Very few studies have focused on the change in the distribution of price across region, so the full distribution of housing prices over time and its decomposition is a field of housing economics that merits development. Recent studies on the cross-sectional distribution of house prices include Garcìa and Raya (2010), Gyourko et al. (2006), Maza and Pages (2007), 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) suggested that income and nominal interest rates are pivotal explanatory factors in the growth of the house price. On the other hand, McMillen (2008) focuses on the change in the house price distribution over time and asks whether the change in the price distribution comes from a change in the distribution of house characteristics such as size, location, age, or from a change in the implicit prices associated with those characteristics. McMillen (2008) shows that the decomposition method is an useful tool to look at the distribution of the house price, and the findings can have important implications (such as for price indexes). Finally, Garcìa and Raya (2010) explain the variance in the house price distribution in a point on the time rather than over a period by estimating the full impact of the explanatory variables on the Gini index.

In this context, we are interested in knowing to what extent the difference between house prices became smaller or larger over the period concerned. In this way we can find out whether the price of cheaper houses increase more or less in relative terms than that of more expensive housing. Our analysis tends to study these questions for various Spanish cities (Madrid, Barcelona, Valencia, Seville, Zaragoza, Malaga and Bilbao) with the objective to understand the determinants of the variation of the price across cities.

For this purpose we use a dataset from a private real estate agency that give very detailed information about the houses sold. Changes in the distribution of price can be as a result of: more houses or apartment being sold in high price neighbourhoods (change in locations of homes sold), different characteristics of housing sold (change in size and quality of the homes sold) or changes in the underlying hedonic price functions.

In this respect, to study the changes in the distribution of house price from 2004 to 2007 and its decomposition we apply a decomposition method proposed by Melly in 2005, that permit us to isolate these components of housing price changes.

The focus of our paper is closely related to the issues discussed in McMillen (2008), but differs from it because is related to Spain and we present an analysis across cities. In addition we using the technique of Melly's (2005) 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 on the whole distribution rather in the average. There are many studies that have extended Oaxaca (1973) and Blinder (1973) decomposition of differences at the mean to decomposition of the whole distribution. Juhn et al. (1993, JMP) have proposed an extension of the Oaxaca decomposition by taking account of the distribution of residuals. DiNardo et al. (1996) and Lemieux (2002) proposed an estimator of counterfactual makes the distribution of skills constant across time. M-M (2005) proposed a method based on the estimation of marginal wage distributions consistent with a conditional distribution estimated by quantile regression. Melly (2005) following M-M proposes 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 are provided by Firpo, Lemieux, and Fortin (2007).

Several authors make use of 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 neighbourhood diversity (Cobb-Clark and Sinning (2011)), house price (McMillen (2008)) or immigration and house price (Degen and Fisher (2009)) .In this work we apply the method developed by Melly in 2005. The basic idea of Melly's semi-parametric estimator, is to estimating 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 estimation has a statistical and an economic interpretation, i.e. the quantile regression coefficients can be interpreted as rates of return to house price at different points of the price distribution (Buchinsky, 1998). This

method solving the problem of crossing of different quantile curves and derives the asymptotic distribution. This is an inconsistency which sometimes appears in conditional quantile regressions (as proposed by Koenker and Basset, 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 add if fast in compute 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 most important Spanish cities, in order to check if we find significant differences across regions. In other words we explore if the effect of a housing bubble on the cross-sectional price distribution affect at the Spanish cities in the same measure. As we have observed, the study of housing price distribution is a main area of future research and the papers of differences across regions is limited, so with this paper we want to contribute to the literature among the distribution of house price at region level.

We find three important results. Firstly, from 2004 to 2007, the difference in housing price distribution in Spain is larger at lower and higher percentiles. Secondly, the most important part of the difference in the distribution of housing prices between 2004 and 2007 is explained by coefficients (with all the variables contributing similarly). Thirdly, among cities, we can find a lot of variation in change of house price distribution. With respect to general Spain's cities pattern, Madrid, Valencia and Bilbao, are the cities which pattern are more different.

The structure of this paper is as follows. In section 2, we introduce Spanish cities and point out some differences among them. The next section studies the data and the descriptive statistics. Section 4 studies the empirical model and estimation strategies. Section 5 summarizes results with especial focus in comparison across cities. Finally, we will conclude with some remarks.

---

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

2. A little overview of the Spanish cities analyzed

In this section we look at the distribution of house price in the Spanish cities. Spain has seventeen regions, it's a kind of federal state, i.e. each region has a local government. These regional governments are responsible for the administration of schools, universities, health, social services, culture, urban and rural development and, in some cases, policing. Under this framework we can find very huge difference between one region to other, such as among cities between region. The differences could come for several reasons: cultural, political or economics. To explore these differences is very interesting if we want to understand better the policies that the central government need to make. This is particularly true when we want to analyze shocks at regional level, because for most of them the reaction could be different, such as in the house price distribution during the boom period.

As we report in Table 1 we observe not only initial house prices (2004), but also gap price (2004-2007) is different among them. In this sense, it would be worthwhile to add some discussion of potential differences among the cities that might account for the differences in the price distribution changes.

Some Spanish authors have discussed these differences (and the causes) among cities: (Balmaseda et al. 2002; Taltavull and 2003; Hiebert and Roma, 2010). In general, theoretical fundamentals about the determinants of housing prices have been exhaustively studied in the housing literature. Explanatory variables in these housing prices models are basically: population, unemployment (especially youth unemployment), income, interest rates and residential capital (Nellis and Longbottom 1981, Meen 1990, Di Pasquale and Wheaton 1994, Johnes and Hyclak 1994, Potepan 1994, Abraham and Hendershott 1996, England and Ioannides 1997, Hort 1998, Malpezzi 1999, Jud and Winkler 2002, Meen 2002, Andrew and Meen 2003 and Taltavull 2003).

In Table 2 we show information about some of these variables[4] (population, youth unemployment, per capita income and residential capital stock) not only in terms of its level in 2004 but also in terms of its growth rate during 2004-2007 per each cities

---

[4] We do not report information about interest rate because it's equal in all Spanish cities.

consider in this study. As we can observe, cities in which housing prices were higher in 2004 (Barcelona, Bilbao and Madrid) are also cities in which per capita income is higher, population is higher (in the case of Barcelona and Madrid) and less youth unemployment is observed (in the case of Barcelona and Bilbao). Between 2004 and 2007, cities in which housing price growth was higher (Valencia, Zaragoza, Seville and Malaga), were also cities with large population growth (Malaga and Zaragoza), higher reduction in the youth unemployment rate (Valencia), higher per capita income growth (Zaragoza, Malaga and Seville), and higher growth in the residential capital stock (Malaga, Seville and Valencia).

Table 2: Economic differences among Spanish cities

City	Popul. 2004	growth	Unempl. 2004	growth	P.c. income 2004	growth	Growth of Res. cap. stock
Barcelona	1578,546	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	3099,834	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

Source: Spanish Census Bureau and Spanish Housing Ministry 2009

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 locations characteristics of each property are known. The data was collected each semester from 2004 to the last semester in 2007, and pooled in a unique dataset. In our paper, we use the whole sample for the hedonic model and the subset of those from 2004 and 2007 for the decomposition analysis. The variables included in the analysis are described in appendix (see Table A.1).

In Tables 3 and 3a we report the descriptive statistics of our sample for Spanish cities in two different years (2004 and 2007) and the variables that we take into account for our estimation. As we can see Madrid and Barcelona have the highest average house prices compared to other Spanish cities in 2004. In 2007 the price growth was, at least 20% for all cities (with the exception of the 17% for Barcelona), the growth of the average price is quite heterogeneous among cities, therefore in cities such as Valencia and Zaragoza the increase is above 40%, this is in line with the official price of the Ministry of Housing (see Table 1). The characteristics of the building sold not change a lot from one year to another; we just can observe that the structural variable "lift" is more present in the 2007.

Table 3: Descriptive statistics: 2004

Year 2004	Madrid	Barcelona	Malaga	Sevilla	Zaragoza	Bilbao	Valencia
Log of sales 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
Num. 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	Sevilla	Zaragoza	Bilbao	Valencia
Log of sales 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
Num. 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

In Figures 1, 2 and 2a we show the kernel density estimation of the price per square meter for Spain (pooled cities) and for the each city studied. As we can see, in 2007

price distributions became less skewed (thicker on the right), it has moved to the right, and it has low kurtosis compared to distributions in 2004. A general observation is that the rate of appreciation was particularly rapid for homes with higher prices.

This fact is remarked when we observed the Spanish cities kernel density estimation. In 2007 they have experimented a shift on right in their housing price distribution. However, the shift was larger for small cities such as Valencia or Bilbao, than for bigger cities such as Barcelona or Madrid.

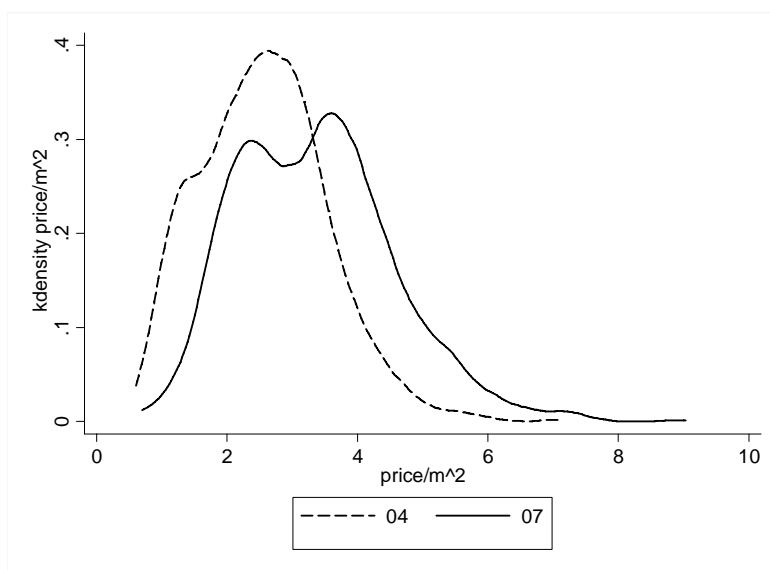**Figure 1: Kernel density estimation for Spain**

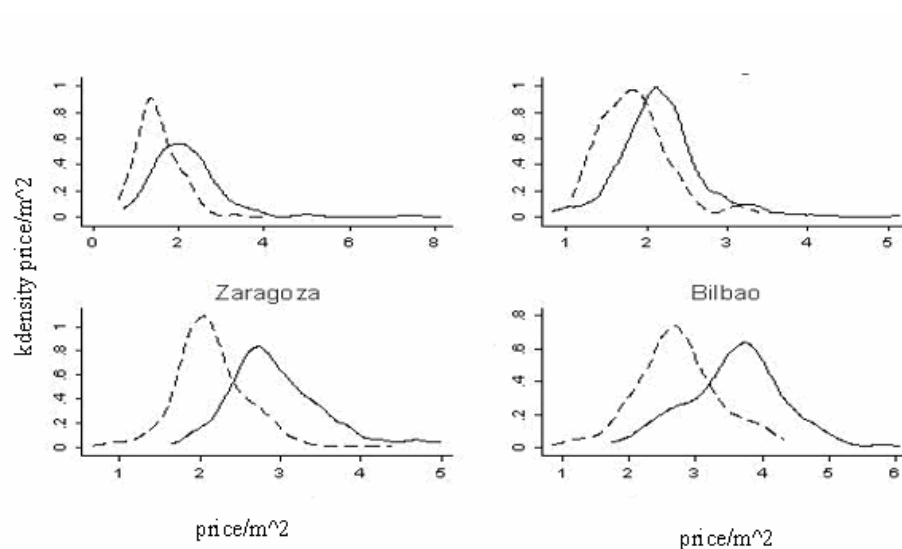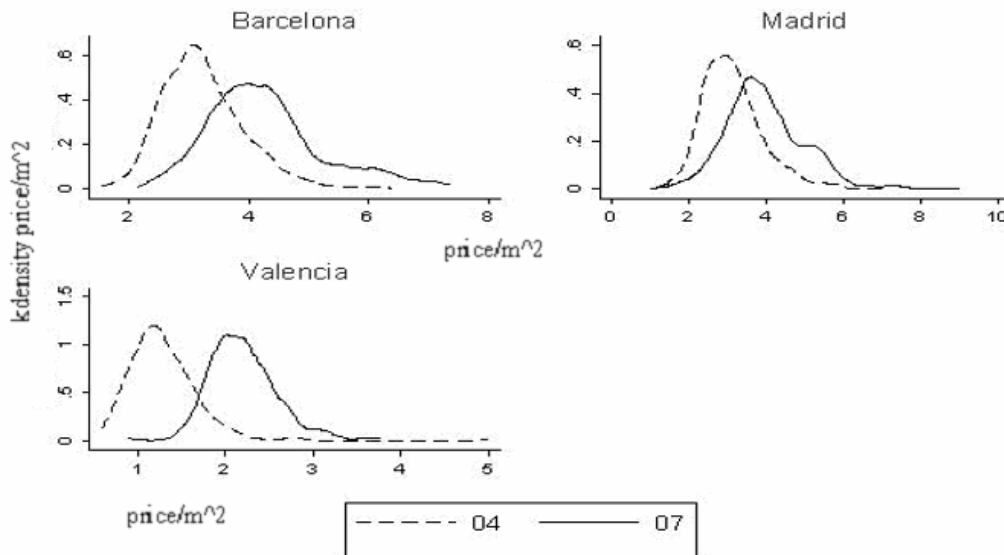**Figure 2: Kernel density estimation for Spanish Cities**

Figure 2a: Kernel density estimation for Spanish Cities

4 Empirical Models

4.1 Hedonic regression

At this point, we are interested in determining if the distributional change can be explained by size, quality and location characteristics of the homes sold and/or the coefficients account for this changes, and if there are systematic differences between different cities within a country. A large used approach to deal with product heterogeneity in terms of quality is the 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. In this case the classical work was that of Rosen (1974)[5]. In it he showed how heterogeneous products are composed of various characteristics and how the marginal price is implicit in these. In this section we

---

[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).

estimate a hedonic function model by OLS. Clearly, housing is a good that fits perfectly into the framework of hedonic price models. Some outstanding work in the literature on hedonic prices applied to the housing market includes articles by Palmquist (1984), Mendelsohn (1984), Bartik (1987), Mills and Simenauer (1996), Bover and Velilla (2001), Ekeland et al (2002 and 2004), Bin (2005) and Garcia and Raya (2010). In relation to quantile estimation, in the context of housing, it is also easy to appreciate that the evaluations that individuals make of the physical characteristics of their home differ according to whether homes have higher or lower price per square meter. 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:

$$p_i = \delta + X_i'\beta + \sum_j \alpha_j D_i^j + \sum_t \tau_t T_i^t + u_i \qquad (1)$$

where p is the price of the house (we express it in its logarithm of the price per squared meter), X is a vector of physical characteristics of the house, Dj is a dummy variable corresponding to the area j in which the house is located, T is a dummy variable corresponding to the semester t to which the observation of the house i corresponds, and finally, u and δ are respectively the corresponding error term and the initial price per square meter of a house with the reference characteristics (and β, α and τ, are estimated parameters).

This model is estimated and reported in Table 4, where we present the results of the hedonic pricing model by quantile for the total sample and for four years. With regard to OLS estimation, 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 square meter in the useful floor area decrease the price per squared meter in 0.48%. The availability of a lift increases the price per square meter by 14.93%, while to have a separate kitchen into the dwelling decrease the price per square meter in 1.91%. Finally, bedrooms and floor are not significant in the OLS equation. In terms of quantile estimation, we can see that, with regard to the effect of the age of a dwelling on quantile regressions, it appears to have a diminishing quantitative impact (less negative) as we advance towards higher percentiles. The impact of the floor area also falls at higher percentiles. In the case of

type of kitchen, floor and bedrooms variables, the sign of the impact changes from negative to positive as we advance to higher percentiles. That is, an increase in the number of bedrooms, to live in a higher floor and to have a separate kitchen into the dwelling is positive valued at lower percentiles but negative valued at higher percentiles. Finally, all other variables being equal, the impact of having a lift is positive and appears to have an increasing quantitative impact as we advance towards higher percentiles.

Table 4: Estimation results of the hedonic pricing model

	OLS	Q 10	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
Nm. Obs.			21,517			

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

4.2 Quantile Decomposition

How does the distribution of house price evolve over time? Did the distribution change because homes sold in 2005 have different characteristics (size, neighbourhoods, etc.) than home sold in 2007? On second thought, is the distributional change unrelated to the size, quality, and location of the house, and has caused by changes in the basic hedonic price functions?

To answer these questions we use the different literature to estimate the distribution of earnings or prices based on the quantile regression, by looking at the effects of covariates on different quantiles of log price distribution and not only at the average of variables. We can decompose changes in the distribution of house prices into the part due by changes in the distribution of the explanatory variables and the part induced by changes in the coefficients of quantile regression.

Machado and Mata (2005) present an estimator using quantile regression to decompose differences in log wages between two groups where the differences at various quantiles of distributions can be analysed.

Assuming linearity between the quantiles of the dependent variable Y and the covariates X, then the τth conditional quantile of Y is given by the following equation:

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

Koenker and Basset (1978) proposed a method of calculating the quantile regression that can be estimated by minimizing $\beta(\tau)$ of the following expression:

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

with the check function ρ_t weighting the residuals μ_i asymmetrically:

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

Based on quantile regression and the Blinder and Oaxaca (1973) decomposition, Machado-Mata (2005) have proposed an estimator of counterfactual unconditional wage distributions based on quantile regressions, where the difference of the θth unconditional quantile between two groups' distributions can be decomposed as:

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

where $\hat{F}_{Yt}^{-1}(\theta\,|\,T=t)$ denotes the θth unconditional quantile of wage Y for groups t's price, 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 and so it may not be easy to invert it. Moreover, no asymptotic results and method to estimate the variance consistently have been provided in M-M 2005, so Melly (2005) decided to extend the M-M method. 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. Four our study we use the Melly method instead of Machado and Mata due some advantage in the calculation of the quantile decomposition, however the some results is can achieve using M-M. Melly's method can be represented in the following expressions:

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

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 is linear in X.

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

$$\widehat{F_{yt}}(q|X_i) = \int_0^1 1(X_i\widehat{\beta}_t(\tau) \leq q)d\tau = \sum_{j=1}^n (\tau_j - \tau_{j-1})1(X_i\widehat{\beta}_t(\tau) \leq q)$$

Following the convention of taking the infimum of the set, a natural estimator of the θ_{th} quantile of the unconditional distribution of Y is given by:

$$\widehat{F_{yt}}(q|T = t) = \frac{1}{n_t}\sum \widehat{F_{yt}}(q|X_i)$$

From this expression we can calculate the unconditional and counterfactual quantiles distributions that are, respectively:

$$\widehat{q_t}(\theta) = inf\{q : \frac{1}{n_t}\sum_t \widehat{F_{yt}}(q|X_i) \geq \theta\}$$

$$\widehat{q_{c1}}(\theta) = inf\{q : \frac{1}{n_t}\sum_0 \widehat{F_{yt}}(q|X_i) \geq \theta\}$$

According with Oaxaca-Blinder (1973) the difference between the year 2007 and 2004 of the house price can be decomposing as the difference in characteristics and coefficients. Where the characteristics effects represent the several aspects of the house such as: size, number of bathrooms, age, etc, thatexplain part of the difference across these two years and coefficients are the residual part or unexplained part of the change of household price between 2004 and 2007.

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

5 Results

Quantile decompositions using the Melly's approach are reported in the following figures for the whole Spain house price and for Spanish cities. In Figure 3 we report the estimation of the log sales price decomposition between the years 2004 and 2007 in Spain and the confidence intervals. Unlike the simple Oaxaca (1973) decomposition, the method developed by Melly (2005) give us the opportunity to look at the entire price distribution of house prices. In quantile decomposition we split the difference in the houses price (overall) between two years, $\Delta(07 - 04)$, in change due to the difference in the characteristics of sample of house sold and, change caused by coefficients (i.e. returns). The differences in characteristics between 2004 to 2007 means that houses sold have certain characteristics different from one year to another, such as: size, parking, numbers of rooms, etc. Change in coefficients means change in return of these characteristics; for example a neighbourhood which has became expensive (land price has risen).

Looking the Figure 3 we find that the price difference between 2007 and 2004 (overall) is always positive, house prices have increased during this period. Most of difference in price is due at change in coefficients instead of characteristics effects. This is should be not unexpected; the houses cannot be changed quickly such that if prices increased in a short period of time, this has to be due to coefficients. We can see also that the difference is larger at lower percentiles than higher percentiles. A plausible explanation is that due to the high increase in housing prices, the demand for low-priced houses increased (income restrictions). In this sense, most buyers decided to buy low-priced houses because the price is more affordable. Finally, house of medium-priced and high-price have increased but in less proportion. Similar results have obtained in McMillen (2008) for Chicago housing prices between 1995 and 2005.

Figure 3: Decomposition of log house prices 2004-2007
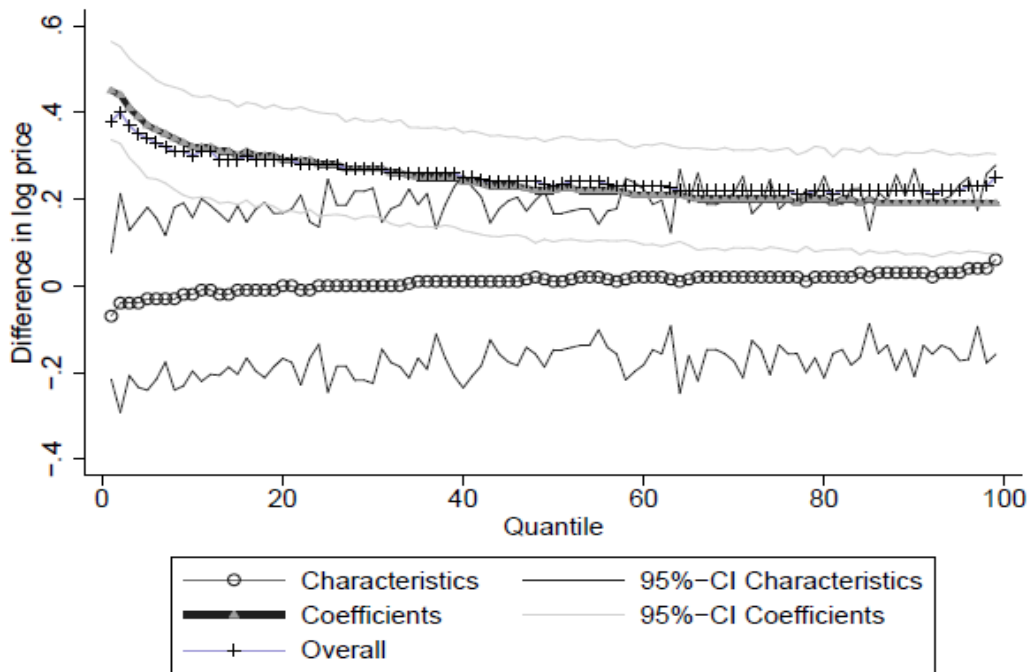
In the Figure 4, we repeat the decomposition method considering the effects of single explanatory variable on change in the house price. May be there is some variable that could have a weight on the price high with respect others, such as for example neighbourhood. In this way we isolate their effects and are able to look at the weight of each variable to explain difference across years.

As we can see on the Figure 4, seems that the characteristics effects is very close to zero, while the returns of each characteristics (coefficients effects) are positive and affect at the house price between 2004 and 2007 in the some direction. Almost all variables' implicit prices 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 in which the peak is 0.5) while any 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 middle and high priced sales. In this sense, forces affecting the distribution of house prices in Spain can be considered similar to those of Chicago such as reported by McMillen 2008

Figure 4: Decomposition of the effect of explanatory variables on house price, 2004–2007
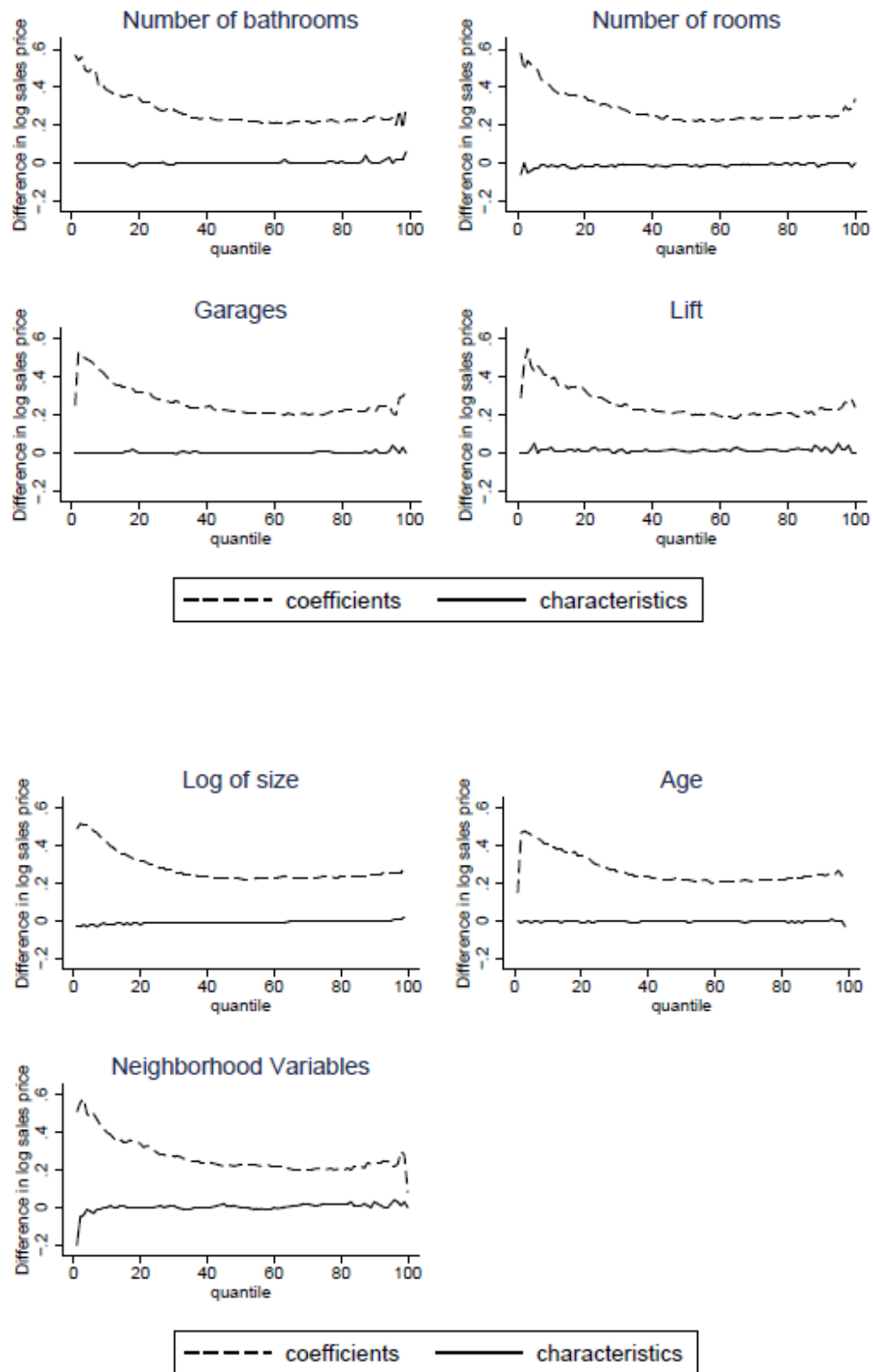
5.1 Comparison among cities

In this section we reported the same analysis of decomposition using Melly's method, to understand the variation of the price among cities. Explore how the price has evolved among Spanish cities is a very important topic. Cities in Spain can be different for many reasons: geographical history, economic structure and federal or council government policies. In fact, in section 2, we have presented many economic determinants of potential differences among Spanish cities that might account for the differences in the price distribution changes. However, having a representative sample at city level is a very difficult duty. Thanks our dataset we can explore some difference in the distribution of house across regions in Spain.

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

Analyzing city by city, Zaragoza presents a uniform increasing of the price along all the distribution, and it's also the city less affected by the boom in our dataset. Also, very similar patterns are observed in Sevilla and Malaga. In these cases, increasing differences in sales prices at lower percentiles represent the main change with respect to the general pattern. In Barcelona, differences in sales prices are almost constant along price distribution. Finally, the main exceptions to the general pattern are 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.

Otherwise, the maximum variation in price differences among percentiles is observed in Valencia, while 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 varies from 0.9 to 0.3 approximately. To sum up, in Valencia (specially, but also in Bilbao), some convergence in housing prices is achieved during 2004-2007 period, in the sense that price growth has been higher at lower percentiles. Thus, prices in 2007 are more similar along the distribution. Instead, some divergence is observed in Barcelona and Madrid. In the two larger cities of Spain, prices in 2007 are less similar along the distribution because price growth has

been higher at higher percentiles. This smaller price growth (in relative terms with respect to price growth at higher percentiles) at lower percentiles in Barcelona and Madrid may be due to lower demand for homes in these percentiles. This lower housing demand has been motivated by, principally, lower population growth observed in Table 2.

Again, at lower and medium percentiles, the characteristics of the house seem don't explain the change in the price. However, at higher percentiles, in Madrid, Barcelona, Bilbao, Malaga, Sevilla and Zaragoza, the change in the price distribution between 2004 and 2007 can be also explained by the change in the characteristics. This little 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) which 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.

To sum up, a remarkable result of this paper is that, although the most important part of the difference in the distribution of housing prices between 2004 and 2007, is always explained by coefficients, there are different patterns among Spanish cities.

These different patterns are not only consequence of different patterns in the change in the coefficients, but also in the change of characteristics.

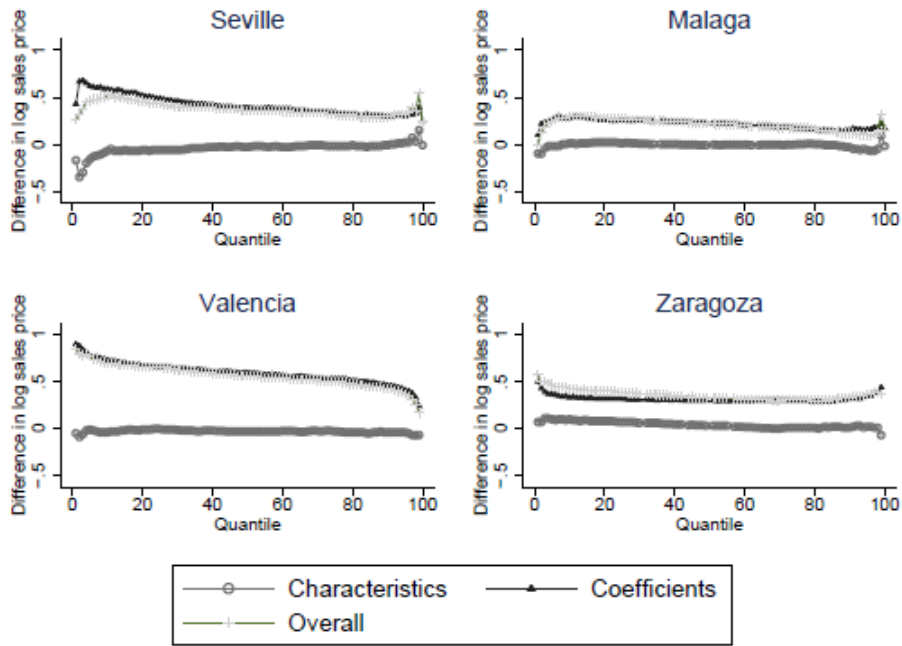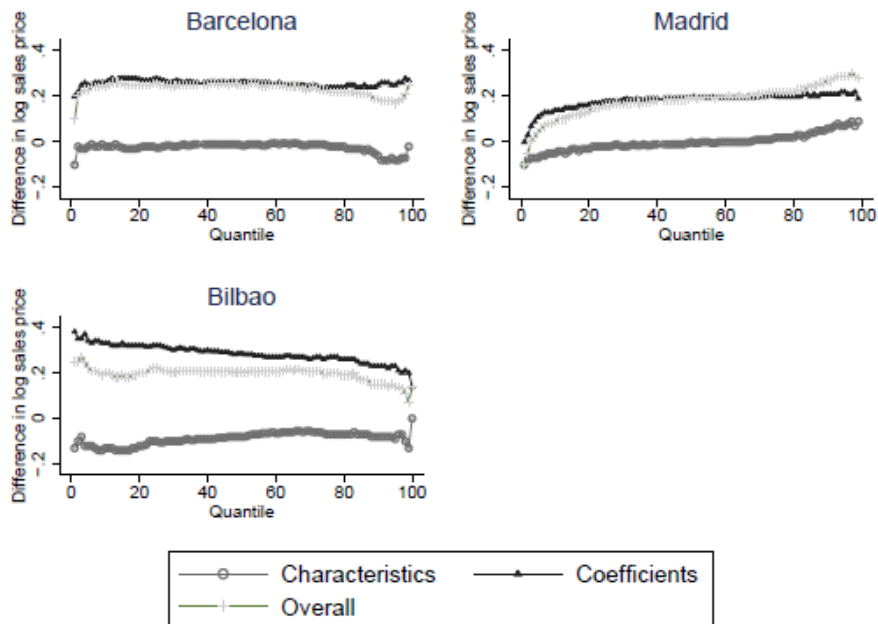**Figure 5: Decomposition of log house price by Spanish Cities, 2004-2007**

Figure 6: Decomposition of log house price by Spanish Cities, 2004-2007

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' behaviour. 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 housing prices. Thanks to these methods it is possible to observe the inter distribution of the change of the price over the time, thing that a simple index not consider. The distribution of housing prices for sales of homes in Spain in 2007 became less skewed (thicker on the right), moved to the right, and has lower kurtosis compared to distribution in 2004. A general explanation is that the rate of appreciation was particularly rapid for homes with higher prices. All the Spanish cities have experimented a shift in their housing price distribution. However, the shift was larger for small cities such as Valencia or Bilbao, than for bigger cities such as Barcelona or Madrid.

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 crossing of different quantile curves and derives the asymptotic distribution.

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

Among cities, we can find a lot of variation in change of house price distribution. Although the most important part of the difference in the distribution of housing prices between 2004 an 2007, is always explained by coefficients, characteristics also explained some part. With respect to Spain's pattern, Madrid, Valencia and Bilbao, are the cities which pattern are more different. Thus, in Valencia (specially, but also in Bilbao), some convergence in housing prices is achieved during 2004-2007 period, in the sense that price growth has been higher at lower percentiles. Thus, prices in 2007 are

more similar along the distribution. Instead, some divergence is observed in Barcelona and Madrid. In the two larger cities of Spain, prices in 2007 are less similar along the distribution because price growth has been higher at higher percentiles. This smaller price growth (in relative terms with respect to price growth at higher percentiles) at lower percentiles in Barcelona and Madrid may be due to lower demand for homes in these percentiles. This lower housing demand has been motivated by, principally, lower population growth observed.

Thus, results presented here suggest (as pointed out in McMillen, 2008) that a single price index is not an appropriate measure of housing 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 for the fact that variation across quantiles of the appreciation rates, and its components, are different among cities which make up the index.

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 Garcia and Raya (2010) to analyze which characteristic contributes to the equality or inequality in housing prices.

References

Abraham J. M. and Hendershott P. H. (1996): "Bubbles in Metropolitan Housing Markets*", J. Housing Research, 7-2, pag. 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. and Meen G. (2003): "House price appreciation, transactions and structural change in the British housing market: a macroeconomic perspective*", Real State Econ. 31-1, 99-116.*

Balmaseda M., San Martin I. and 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. and Velilla P. (2001): "Hedonic house prices without characteristics: the case of new multiunit housing", Banco de España, Estudios Económicos 73.

Buchinsky M. (1998a): "The Dynamics of Change in the Female Wage Distribution in the USA: Quantile Regression Approach", Journal of Applied Econometrics 13:1–30.

Chamberlain, G. (1980): "Analysis with qualitative data", *Review of Economic Studies 47:225–238.*

Chernozhukov V., Fernández-Val I. and Galichon A. (2010): "Quantile and Probability Curves Without Crossing", *Econometrica, 78(3).*

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

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

Degen K. and Fischer M. (2009): "Immigration and Swiss house prices", C*entre for Economic Policy Research, dp 7583*

DiPasquale D. and Wheaton W. (1994): "Housing Market Dynamics and the Future of House Prices", *Journal of Urban Economics, 35-1, 1–27.*

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

Firpo S., Fortin N. and Lemieux T. (2007): "Unconditional Quantile Regressions", *NBER Technical Working Papers 0339*

Ekeland, I., Heckman, J. J. and Nesheim, L. P. (2002): "Identifying hedonic models", American Economic Review 92, 304-309.

Ekeland, I., Heckman, J. J. and Nesheim, L. P. (2004): "Identification and estimation of hedonic models", Journal of Political Economy 1, 60-109.

England, P., and Ionannides, Y. M. (1997): "House price dynamics: an international empirical perspective", Journal of Housing Economics, 6, 119-136.

Hiebert P. and Roma M. (2010): "Relative house price dynamics across euro area and US cities: convergence or divergence?", European Central *Bank, WP 1206*

Garcia J., Hernandez, P. and Lopez-Nicolas, A. (2001): How Wide is the Gap? An Investigation of Gender Wage Differences Using Quantile Regression", *Empirical Economics 26(1):14967.*

Garcìa J. and Raya, J. (2010): "Price and income elasticities of demand for housing characteristics in the city of Barcelona", *Regional Studies 45(5): 597-60*8

Garcìa J. and Raya, J. (2012): "Analysis of inequality in housing prices: a quantile approach", *Journal of Housing Economics (forthcoming)*

Griliches, Z. (1971): "Price indexes and quality change: studies in new methods of measurement": *Harvard University Press.*

Gyourko J., Mayer C. and Sinai T. (2006): "Superstar Cities", *NBER Working Paper No. 12355*

Hall P. Sheather, S.J. (1988): "On the distribution of a studentized quantile", *Journal of the Royal Statistical. Society, Series B, 50, 381391*

Hendricks W., Koenker, R. (1992): "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity", *Journal of the American Statistical Association 87(417): 5868.*

Hort K. (1998): "The determinants of urban house price fluctuations in Sweden 1968-1994", *Journal of Housing Economics, 7, 93-120.*

Johnes G. and Hyclak, T. (1994): "House prices, migration, and regional labor markets", Journal of Housing Economics, 3, 312-329.

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

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

Koenker R. and 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.*

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

Maattanen N. and Tervio M. (2010): "Income Distribution and Housing Prices: An Assignment Model Approach·, CEPR Discussion Papers 945, C.E.P.R.

Machado J. A. F. and 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. and Pages J. (2007): "Analysis of house prices in Spain", *Working paper num. 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., Thorsnes, P. (2006) "Housing renovations and the quantile repeat sales-price index", *Real State Economics 34(4):567–584.*

Meen G. P. (1990): "The removal of mortgage market constraints and the implications for econometric modeling 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. and Simenauer R. (1996): "New hedonic estimates of regional constant quality house prices", Journal of Urban Economics 39, 209-215.

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

Oaxaca R and Ransom M. (1998): "Calculation of approximate variance for wage decomposition differentials", Journal of Economic and Social Measurement 24:55–61.

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.*

Timbergen J.   (1951): "Some remarks on the distribution of labour incomes.", *International Economic Papers 1:195–207.*

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

Appendix

TABLE A1: Description of the variables	
Variable	*Description*
Sales price	Price at which each house was sold (in thousands of euros)
Useful floor area	Gross square meter 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.2a:Detailed descriptive statistics

Year 2004	Madrid					Barcelona					Malaga					Sevilla					Zaragoza					Bilbao					Valencia				
Percentile	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
Log of sales 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	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	30	48	60	72	88	41	50	60	71	85	47	58	69	80	95	42	54	64	75	87	42	50	60	70	80	50	56	65	75	90	50	60	70	80	95
Age	24	33	39	44	69	27	33	39	74	94	8	24	30	37	44	17	26	32	40	54	10	20	36	40	53	24	32	36	45	74	22	30	34	40	50
Num. of bedrooms	2	2	3	3	3	2	2	3	3	4	2	3	3	3	4	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	1	3	4	6	1	1	3	4	6	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	0	1	1	1	1	1	1	1	1	1	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	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1

TABLE A.2b:Detailed descriptive statistics

Year 2007	Madrid					Barcelona					Malaga					Sevilla					Zaragoza					Bilbao					Valencia				
Percentile	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90	10	25	50	75	90
Log of sales 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	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	35	45	58	70	88	35	49	60	75	85	50	60	73	85	100	40	55	65	75	90	47	54	63	70	80	35	42	50	70	85	50	60	70	84	92
Age	20	33	39	45	67	20	30	39	57	89	10	25	30	35	50	17	29	35	40	45	21	30	35	40	48	30	35	40	49	78	25	30	35	42	47
Num. of bedrooms	1	2	3	3	3	1	2	3	3	4	2	2	3	3	4	2	2	3	3	3	2	2	3	3	4	2	2	2	3	3	2	3	3	3	4
Floor	0	1	2	4	5	1	1	3	5	6	0	1	3	5	7	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	0	1	1	1	1	1	1	1	1	1	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	1	1	1	0	0	1	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	1	1	0	0	0	1	1