Journal of Economics and Econometrics Vol. 56, No. 1, 2013 pp. 61-73

ISSN 2032-9652 E-ISSN 2032-9660

Real estate appraisals, hedonic models and the measurement of house price dispersion

GAETANO LISI,† MAURO IACOBINI

ABSTRACT

This paper aims to provide a simple method for measuring the price dispersion in the housing market controlling for the differences in attributes or qualities of the residential real estate units. Precisely, the paper proposes an extended hedonic pricing model which incorporates standard market situations (where a better good is sold at a higher price) as well as non-standard market situations (in which the opposite is true). The extended model is able to take into account the variance in house prices which can not be attributed to the heterogeneous nature of real estate goods. The main result of this analysis is that the extended model explains a greater proportion of the variability of selling price, thus giving an important contribution for the application of the hedonic method to the real estate appraisals.

JEL Classification: C51, C78, D83, E30, R31.
Keywords: Housing price dispersion, residual price volatility, hedonic pricing models.

---

†Corresponding author. Creativity and Motivations (CreaM) Economic Research Centre, Department of Economics and Law, University of Cassino and Southern Lazio, via S. Angelo, I-03043 Cassino, Italy; Phone: +39 0776 996150; E-mail: gaetano.lisi@unicas.it. The authors thank two anonymous referees for their comments and helpful suggestions.

1 INTRODUCTION

The empirical anomaly known as 'price dispersion' is probably the most important distinctive feature of housing markets. It refers to the phenomenon of selling two houses with very similar attributes and in near locations at the same time but at very different prices (for a review on this topic see Leung, Leong and Wong, 2006). Thus, an important part of housing price dispersion can not be attributed to the heterogeneous nature of real estate goods. Remaining price differentials are in fact empirically non negligible and basically due to the heterogeneity of the parties, in particular the bargaining with search costs. Indeed, housing markets are characterised by a decentralised exchange framework in which search/matching frictions and bargaining power of the parties play a key role (see e.g. Quan and Quigley, 1991; Vukina and Zheng, 2010; Leung and Zhang, 2011). However, measuring the heterogeneity of the parties is not an easy task. Furthermore, if important housing characteristics are omitted from the hedonic price function, the correlation between those characteristics and buyer-seller attributes will lead to biased estimates (Harding *et al.*, 2003a). Therefore, the main aim of this paper is to provide a simple method for measuring the house price differentials which can not be attributed to the heterogeneous nature of the real estate goods. Precisely, the paper proposes an extended hedonic pricing model which incorporates standard market situations (where a better house is sold at a higher price) as well as non-standard market situations (in which the opposite is true). The model is able to explain a greater proportion of the variability of selling price, thus taking into account the variance in house prices which can not be attributed to the heterogeneous nature of real estate goods.

In particular, the large increase in the adjusted R-squared of the estimates in the extended hedonic pricing model may pose an important contribution for the application of the hedonic method to the real estate appraisals. According to the exhaustive survey by Lentz and Wang (1998, p. 19), in fact, "the real problem" involved with the application of the regression method to the real estate appraisals is the large standard error of the regression estimates which might render the fitted/estimated values of residential properties useless.

Furthermore, the use of extended hedonic models which take into account variables other than housing characteristics is of crucial importance in the real estate appraisals. The selling prices, in fact, implicitly reflect the effects of such variables. Consequently, the marginal prices obtained without taking into account such effects overestimate or underestimate the actual implicit prices of the housing characteristics. In turn, it will be reflected in an overestimation or underestimation of the "price adjustment" if the hedonic price coefficients estimated from a regression equation are utilised as adjustment factors in the method widely used in real estate appraisals, namely the sales comparison approach.[1]

The rest of the paper is organised as follows: section 2 briefly presents the dataset used in the empirical analysis; section 3 presents the theoretical foundations of the empirical model; section 4 extends the standard hedonic pricing model to non-standard market situations; while section 5 shows the results of the analysis; finally, section 6 concludes the work.

2 DATASET

In this empirical analysis, we employ data from the Canadian housing market. The Canadian housing market is a market sufficiently 'thick', i.e. a market with a sufficient amount of trading. Given the positive relation between trading volume and house price (Leung, Lau and Leong, 2002), it is in fact necessary to choose a 'thick' market to compute the price dispersion in the housing market.

[1] An alternative to reduce the heterogeneity in the multiple regression analysis is to restrict sale observations to properties with similar attributes and market characteristics. However, the lower the number of observations included in the estimate, the higher the variance of the estimates. Therefore, there is a trade-off between minimizing the variance of the estimate and reducing the bias caused by inadequacy of the model to take into account the heterogeneity (Lipscomb and Gray, 1995).

The dataset used in this empirical analysis is characterised by many binary variables. Data on housing characteristics, in fact, typically consists of one continuous regressor (the lot size) and many ordered and unordered categorical variables (Henderson, Parmeter and Kumbhakar, 2007; Haupt, Schnurbus and Tschernig, 2010). The dataset contain 546 observations on sales prices of houses sold during July, August and September, 1987, in the city of Windsor, Canada. The following variables are available (Source: Anglin and Gencay, 1996):

Variable	Description
price	sale price of a house
lotsize	the lot size of a property in square feet
bedrooms	number of bedrooms
bathrms	number of full bathrooms
stories	number of stories excluding basement
driveway	dummy, 1 if the house has a driveway
recroom	dummy, 1 if the house has a recreational room
fullbase	dummy, 1 if the house has a full finished basement
gashw	dummy, 1 if the house uses gas for hot water heating
airco	dummy, 1 if there is central air conditioning
garagepl	number of garage places
prefarea	dummy, 1 if located in the preferred neighbourhood of the city

Further details about this dataset are reported in Table 1.

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
price	546	68121.60	26702.67	25000	190000
lotsize	546	5150.266	2168.159	1650	16200
bedrooms	546	2.965201	0.7373879	1	6
bathrms	546	1.285714	0.5021579	1	4
stories	546	1.807692	0.8682025	1	4
driveway	546	0.8589744	0.3483672	0	1
recroom	546	0.1776557	0.3825731	0	1
fullbase	546	0.3498168	0.4773493	0	1
gashw	546	0.0457875	0.2092157	0	1
airco	546	0.3168498	0.465675	0	1
garagepl	546	0.6923077	0.8613066	0	3
prefarea	546	0.2344322	0.4240319	0	1

Source: Anglin and Gencay (1996).

We also report some descriptive statistics for selling price data (see Table 2). Price dispersion is typically measured by the standard deviation of prices, the coefficient of variation and the price skewness, since the distribution of prices is typically asymmetric and the standard deviation may be insufficient to capture the degree of price dispersion (Leung, Leong and Wong, 2006).

Table 2: Descriptive statistics on selling price

	Percentiles	Smallest		
1%	26500	25000		
5%	35000	25000		
10%	40500	25000	Obs	546
25%			Sum of Wgt.	546
50%	62000		Mean	68121.60
		Largest	Std. Dev.	26702.67
75%	82000	174500		
90%	105000	175000	Variance	7.13e+08
95%	120000	175000	Skewness	1.206503
99%	155000	190000	Kurtosis	4.930871

3 THEORETICAL BASIS

Basically, the selling price of housing is the outcome of pairwise negotiations. Also, the housing market (like the labour market) clears not only through price but also through time and money that the parties spend on the market. Hence, we derive our empirical model from the bargaining between a seller and a buyer in a (decentralised) market with search and matching frictions.

Let the selling price be P and the market value of the good be V. The market value of housing depends positively on the housing characteristics X, i.e. $V = f(X)$, with $\partial V / \partial X > 0$ *ceteris paribus* (*hedonic hypothesis*). Furthermore, the search process in the housing market is costly in terms of time and money. Precisely, the search cost of seller is c^s and the search cost of buyer is c^b. Hence, when the parties meet each other, they always decide to bargain rather than continue the search.

For risk-neutral buyers and sellers, the selling price can be expressed as the (generalized) Nash bargaining solution for given bargaining parameters. Using this result, the selling price can be expressed as the weighted average of the seller and buyer net gains:

$$P = \arg\max\{(P + c^s - V)^\alpha \cdot (V + c^b - P)^{1-\alpha}\}$$

where $0 \leq \alpha \leq 1$ represent a measure of the bargaining power of seller (thus, 1 – α is a measure of the bargaining power of buyer). Note that c^s and c^b represent saved and/or retrieved costs, i.e. gains. The equilibrium price is thus given by:

$$\frac{\partial\{(P + c^s - V)^\alpha \cdot (V + c^b - P)^{1-\alpha}\}}{\partial P} = 0 \Rightarrow P = V + \alpha \cdot c^b - (1-\alpha) \cdot c^s \qquad (1)$$

with $\frac{\partial P}{\partial \alpha} > 0$ and $\frac{\partial P}{\partial\ 1-\alpha} < 0$. Precisely, if $\alpha = 1$, $P = V + c^b$; whereas, if $\alpha = 0$, $P = V - c^s$. Therefore, in this simple model P can higher or lower than V. Indeed, market value and selling price coincide only in case of perfect markets, namely in absence of search costs: $c^s = c^b = 0$.

The empirical counterpart of equation (1), i.e. the *"extended" hedonic price model* is the following:

$$P = f(X) + \Gamma - \Omega \qquad (2)$$

where $\Gamma \equiv \alpha \cdot c^b$ and $\Omega \equiv (1-\alpha) \cdot c^s$. Thus, the selling price depends not only on the housing characteristics, as in the standard hedonic model, but also on Γ and Ω. In short, Γ represent all those factors that increase the selling price; whereas, Ω are all those factors that reduce the selling price. However, in order to estimate equation (2), we need to construct proxies for the aggregate variables Γ and Ω.[2] We will do this in the next section.

---

[2] It would be extremely difficult to measure the individual components, namely the search costs of buyer and seller, and the bargaining parameters.

4 RESIDUAL PRICE VOLATILITY

Unlike previous works which make use of the characteristics of buyers and sellers (Harding *et al.*, 2003a; Harding *et al.*, 2003b; Cotteleer and Gardebroek, 2006), we measure the variance in house prices which can not be attributed to the heterogeneous nature of real estate goods by exploiting the available information regarding real estate units, thus avoiding the important problem of correlation between (omitted) housing-characteristics and buyer-seller attributes, which leads to biased estimates in the hedonic models. Furthermore, the empirical strategy used in this paper takes into account the difference in attributes or qualities of the residential real estate units, as suggested by Leung, Leong and Wong (2006).

In order to measure (*ex post*) the residual price volatility, for each real estate unit i we calculate:

1. The unit price, or price per square meter (p_i), in order to compare real property with different floor areas;
2. The number of "advantages" (a_i). An advantage refers to the presence of a desired housing characteristic (for example, the presence of an elevator or location in a valuable area is an advantage, *ceteris paribus*). Given the presence of many binary variables, the number of "advantages" can be simply calculated as the sum of degree or intensity of housing characteristics. For example, the number of "advantages" of a real estate unit with two bedrooms, two bathrooms, the presence of an elevator and a driveway is six.
3. The simple average of both the unit prices (p_{mean}) and the number of "advantages" (a_{mean}).

In markets for heterogeneous goods, such as a home, standard market situations take place when the property with higher (lower) advantages (mix of desired housing characteristics) is sold at a higher (lower) price, namely if $(a_i - a_{mean}) > 0$ and $(p_i - p_{mean}) > 0$ or $(a_i - a_{mean}) < 0$ and

$(p_i - p_{mean}) < 0$; otherwise, if $(a_i - a_{mean}) > 0$ and $(p_i - p_{mean}) < 0$, or $(a_i - a_{mean}) < 0$ and $(p_i - p_{mean}) > 0$, the selling price is affected by factors other than the heterogeneous nature of real estate. Therefore, we construct three independent dummy variables. Precisely,

- The first dummy variable refers to the situation where $(a_i - a_{mean}) > 0$ and $(p_i - p_{mean}) < 0$. We call this dummy *"negative residual price volatility"*;
- The second dummy variable refers to the situation where $(a_i - a_{mean}) < 0$ and $(p_i - p_{mean}) > 0$. We call this dummy *"positive residual price volatility"*;
- Finally, the third dummy variable refers to the standard market situations (the reference dummy variable).

Obviously, the latter is excluded from the analysis and it is used to evaluate the coefficients associated with the two other dummy variables. Eventually, we include the first two dummy variables in the hedonic price model, thus estimating for each real estate unit i the following "extended" hedonic price function:

$$P_i = f(X_{i,j}, \beta_j) + \gamma \cdot Neg_i + \rho \cdot Pos_i + \varepsilon_i \tag{3}$$

where: $P =$ overall selling price; $X =$ set of j-housing characteristics; $f(X, \beta)$ = standard hedonic price function which captures the variance in house prices due to the heterogeneous nature of real estate goods; $Neg =$ dummy variable *"negative residual price volatility"*; $Pos =$ dummy variable *"positive residual price volatility"*; β, γ and $\rho =$ regression coefficients; $\varepsilon =$ stochastic error term (white noise).

Since the dummy variables are determined *ex post*, there is no simultaneity in the model between selling price and residual price volatility. Also, the effects of a change in the selling price on the construction of the dummy variables are qualitatively insignificant, since the unit price and the average unit price vary in the same direction.

As regards the functional form of the hedonic price function, this model basically extends Anglin and Gencay (1996) by adding two new explanatory variables. Hence, the relationship between the dependent variable (selling price), the continuous regressor (the lot size) and the discrete variables (except the variable ‘number of garage places’ which may take the zero value) is represented in terms of relative changes (i.e. elasticity). Indeed, Haupt, Schnurbus and Tschernig (2010) show that the linear parametric model proposed by Anglin and Gençay (1996) predicts better than the nonparametric specification proposed by Parmeter, Henderson and Kumbhakar (2007).

The residual price volatility could be explained by the bargaining of the parties. A strong buyer, in fact, can pay a lower price for a good house; similarly, a strong seller can charge a higher price for a bad house. Furthermore, the process of gathering information, even when it is publicly available, is costly and time consuming, thus buyers and sellers may enter the market with insufficient or incomplete information. Hence, this residual price volatility is also compatible with the presence of asymmetric information. In fact, if buyers are not fully informed of the lowest price available in the market, they end up paying an incomplete information “tax” which raises the price they pay. Similarly, if sellers are not fully informed about the highest price they could charge, they too suffer an incomplete information “tax” that lowers the price they receive (Kumbhakar and Parmeter, 2008).

5 ESTIMATION RESULTS

The estimate of equation (3) is performed using Ordinary Least Squares (OLS). Three main empirical results are obtained from this analysis:

1. The dummy variables created as proxies of the residual price volatility, and incorporated into the hedonic price function, are statistically significant and their signs are as expected, namely

negative for the dummy *"negative residual price volatility"* and positive for the dummy *"positive residual price volatility"* (Table 3).

Table 3: Estimation results of hedonic models (HM)

	Standard HM	Extended HM
Explanatory variables	ln_price	ln_price
ln_lotsize	0.3030148 (11.59)***	0.4392664 (17.48)***
ln_bedrooms	0.0829402 (1.96)*	0.1721897 (4.57)***
ln_bathrms	0.2613809 (8.66)***	0.2698705 (10.23)***
ln_stories	0.1659226 (6.86)***	0.1562551 (7.39)***
driveway	0.1038256 (3.76)	0.1069126 (4.44)***
recroom	0.0569133 (2.25)**	0.0773661 (3.49)***
fullbase	0.0975449 (4.64)***	0.1121337 (6.09)***
gashw	0.172675 (4.05)***	0.1547949 (4.16)***
airco	0.1754013 (8.50)***	0.1754636 (9.72)***
garagepl	0.0500099 (4.46)***	0.0705442 (7.11)***
prefarea	0.1357077 (6.13)***	0.159891 (8.21)***
Neg		- 0.2103235 (-9.77)***
Pos		0.1886027 (8.15)***
cons	8.014121 (37.63)***	6.756817 (32.19)***
Obs.	542	542
Prob > F	0.0000***	0.0000***
Adjusted R-square	0.6906 (69.06%)	0.7642 (76.42%)
Mean VIF (variance inflation factors)	1.28	1.32
Ramsey RESET test. Ho: No omitted variables – Prob > F	0.3425	0.7972
Breusch-Pagan / Cook-Weisberg Test for heteroskedasticity. Ho: Constant variance – Prob > chi2	0.8345	0.2599
Skewness/Kurtosis tests, normal data. Ho: Normal data – Prob > chi2 / z	0.1708	0.2327
Shapiro-Wilk test for normal data Ho: Normal data – Prob > z	0.03825**	0.08971*

Notes: t-statistics in parentheses; * denotes significance at 10% level, ** at 5% level, and *** at 1% level (* Prob < 0.10; ** Prob < 0.05; *** Prob < 0.01). We deleted four severe outliers where the studentized residuals in absolute value were higher than 3. Details are available upon request.

2. Compared to a traditional hedonic model without such proxies, the adjusted R-squared is significantly higher (see Table 3); whereas the mean and the standard deviation of the prediction error (i.e. the percentage difference between predicted and observed selling prices) is lower (see Table 4).

3. In the extended hedonic model, the variables 'bedrooms' and 'recroom' are statistically significant at any level of confidence and (unlike the standard model) the Shapiro-Wilk test for normal data shows that the null hypothesis of 'normal distribution of residuals' is not rejected at the confidence level of 5% (see Table 3).

Therefore, the extended hedonic pricing model explains a greater proportion of the variability of selling price, thus taking into account the variance in house prices which can not be attributed to the heterogeneous nature of real estate goods.

Table 4: Prediction Error (PE)

Variable	Obs	Mean	Std. Dev.	Min	Max
PE_standard model	542	0.1650264	0.1462958	0.0000601	0.8269243
PE_extended model	542	0.1415149	0.1237820	0.0004868	0.7190421

6 CONCLUSIONS

Price dispersion is probably the most important distinctive feature of housing markets. It refers to the phenomenon of selling two houses with very similar attributes and in near locations at the same time but at very different prices. Thus, the house price is affected by factors other than the housing characteristics. This paper provides a simple method for measuring the variance in house prices which can not be attributed to the heterogeneous nature of real estate goods. Precisely, the paper proposes and estimates an extended hedonic pricing model which takes into account the residual price volatility. The main result of this analysis is that the extended model is able to explain a greater proportion of the variability of selling price, thus giving an important contribution for the application of the hedonic method to the real estate appraisals.

REFERENCES

Anglin, Paul M., and Gençay Ramazan (1996), "Semiparametric Estimation of a Hedonic Price Function", Journal of Applied Econometrics, 11(6), 633-648.

Cotteleer, Geerte, and Cornelis Gardebroek (2006), "Bargaining and market power in a GIS-based hedonic pricing model of the agricultural land market", American Agricultural Economics Association, 21255 (2006 Annual meeting, July 23-26, Long Beach, CA).

Harding, John P., John R. Knight, and C. F. Sirmans (2003b), "Estimating Bargaining Effects in Hedonic Models: Evidence from the Housing Market", Real Estate Economics, 31(4), 601-622.

Harding, John P., Stuart S. Rosenthal, and C. F. Sirmans (2003a), "Estimating Bargaining Power in the Market for Existing Homes", Review of Economics and Statistics, 85 (1), 178-188.

Haupt, Harry, Joachim Schnurbus, and Rolf Tschernig (2010), "On Nonparametric Estimation of a Hedonic Price Function", Journal of Applied Econometrics, 25(5), 894-901.

Henderson, Daniel J., Subal C. Kumbhakar, and Christopher F. Parmeter (2007), "Nonparametric estimation of a hedonic price function", Journal of Applied Econometrics, 22(3), 695-699.

Kumbhakar, Subal C., and Christopher F. Parmeter (2008), "Estimation of Hedonic Price Functions with Incomplete Information", Virginia Tech Working Paper, 2008-04, January.

Lentz, George H., and Ko Wang (1998), "Residential Appraisal and the Lending Process: A Survey of Issues", Journal of Real Estate Research, 15(1), 11-40.

Leung, Charles Ka Yui, and Zhang, Jun (2011), "Fire Sales in Housing Market: Is the House–Search Process Similar to a Theme Park Visit?", International Real Estate Review, 14(3), 311-329.

Leung, Charles Ka Yui, Garion C. K. Lau, and Youngman Chun Fai Leong (2002), "Testing alternative theories of the property price-trading volume correlation", Journal of Real Estate Research, 23(3), 253-264.

Leung, Charles, Youngman Leong, and Siu Wong (2006), "Housing Price Dispersion: An Empirical Investigation", The Journal of Real Estate Finance and Economics, 32(3), 357-385.

Lipscomb Joseph B., and J. Brian Gray (1995). "A Connection between Paired Data Analysis and Regression Analysis for Estimating Sales Adjustments", Journal of Real Estate Research, 10(2), 175-183.

Quan, Daniel C., and John M. Quigley (1991), "Price Formation and the Appraisal Function in Real Estate Markets", Journal of Real Estate Finance and Economics, 4(2), 127-146.

Vukina, Tomislav and Xiaoyong Zheng (2010), "Bargaining, Search and Price Dispersion: Evidence from the Live Hogs Market", Agricultural and Resource Economics Review, 39(3), 534–546.