Economics Bulletin

Volume 32, Issue 1

Measuring the Housing Price Dispersion in Italy

Gaetano Lisi
Agenzia del Territorio

Mauro Iacobini
Agenzia del Territorio

Abstract

This paper has two interconnected goals. The first is to provide a simple method for measuring the variance in house prices which can not be attributed to the heterogeneous nature of real estate goods. The second goal is to show the strong statistical significance of this residual volatility in the Italian housing market.

Acknowledgements: We are grateful for very helpful comments and suggestions received by an anonymous referee. A previous version of this paper was presented at the 41st Meeting of Ce.S.E.T. (Centro Studi di Estimo e di Economia Territoriale) entitled "Appraisals: Evolving Proceedings in Global Change", November 14-15, 2011, Rome (Italy). The usual disclaimer applies.

Citation: Gaetano Lisi and Mauro Iacobini, (2012) "Measuring the Housing Price Dispersion in Italy", *Economics Bulletin*, Vol. 32 No. 1 pp. 811-822.

Contact: Gaetano Lisi - gaetano.lisi@unicas.it, Mauro Iacobini - Mauro.Iacobini@agenziaterritorio.it.

Submitted: January 19, 2012. **Published:** March 04, 2012.

1. Introduction

According to the literature (for a review see Leung, Leong and Wong, 2006), an important part of housing price dispersion can not be attributed to the heterogeneous nature of real estate goods. Indeed, the empirical anomaly known as 'price dispersion' 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. Remaining price differentials are in fact empirically non negligible and basically due to the heterogeneity of the parties (see e.g. 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.

The main aim of this paper is to provide a simple method to measure the house price differentials which can not be attributed to the heterogeneous nature of the real estate goods and show the strong statistical significance of this residual volatility in the Italian housing market.

2. The strategy

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.

Our (simple) strategy is the following. 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 a characteristic with maximum degree or intensity (for example, the presence of an elevator or location in a valuable area is an advantage).[1] Furthermore, for each sample analysed, we calculate the

[1] This calculation does not include the attributes with maximum degree but present in homogenous manner in the sample.

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 is sold at a higher (lower) price; 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 a dummy variable which assumes value 1 if non-market situations take place and 0 in all other cases. Finally, we include this dummy variable in the hedonic price function, thus estimating for each real estate unit i the following "extended" model:

$$P_i = f\left(X_{i,j}, \beta_j\right) + \gamma \cdot D + \varepsilon_i \qquad [1]$$

where: P = overall selling price; X = set of *j*-housing characteristics; $f(X,\beta)$ = hedonic price function (which captures the variance in house prices due to the heterogeneous nature of real estate goods);[2] D = dummy variable created as a proxy of the residual price volatility; β, γ = regression coefficients; ε = stochastic error term (white noise).

3. Dataset

In the empirical analysis, we use information regarding the market survey conducted by the Provincial Offices of the Territorial Agency (one of the four Italian Tax Agencies). This market survey regards the Italian residential properties that were sold during (the end of) 2009 and (the beginning of) 2010. For each real estate unit, the Provincial Offices reported:

1. The nominal selling price (as indicated in the bills of sale);[3]
2. The housing characteristics considered most influential in the price formation process and the corresponding score or unit of measure (from which we calculated the number of advantages).

By the stepwise method (10% significance level for addition to the model, and 20% significance level for removal from the model) we select the most influential housing characteristics. In short, we perform a preliminary regression analysis with all the available attributes and only those statistically significant for each sample analysed are included in the hedonic price function. Because of the strong correlation with the lot size, we exclude the

[2] In order to choose the most suitable functional form, we compared the models frequently adopted in the empirical estimation (linear, logarithm and logarithm-linear) by using three methods of comparison: the so-called *PE test* suggested by MacKinnon, White and Davidson (1983); the Box-Cox transformation suggested by Davidson and MacKinnon (1981); and the partial regression analysis used by Brown and Ethridge (1995).

[3] Because of the limited time period, we do not need to convert nominal prices into real prices.

number of bathrooms and the number of balconies from the regression analysis. Precisely, the housing characteristics included in the regression analysis are the following (see *Table 1*):

Table 1. Housing characteristics

Variable	*Type*
Lot size	*continuous quantitative variable*
State of real estate unit	*binary variable (1 if excellent; 0 otherwise)*
Location	*binary variable (1 if valuable area; 0 otherwise)*
Architectural style	*binary variable (1 if elegant; 0 otherwise)*
Quality of view	*binary variable (1 if excellent; 0 otherwise)*
Elevator	*binary variable (presence or absence)*
New construction	*binary variable (presence or absence)*

We use one continuous regressor (lot size) and six binary regressors. Indeed, we transform all the ordinal qualitative variables (*State of real estate unit, Location*, *Architectural style*, *Quality of view*) into binary regressors in order to save degrees of freedom. For example, the variable "Location" can assume the following conditions: i) the property is located in a degraded area; ii) the property is located in a normal area; iii) the property is located in a valuable area. Rather than to assign arbitrary scores or create an independent dummy variable for each condition, it is possible to distinguish between a *valuable area* (which takes the value 1 if the property is located in a valuable area) and a "*non valuable" area* (which assumes the value 0 in all other cases). This strategy is used for all the ordinal qualitative variables and may help to determine the most appropriate functional form, since dummy variables, by definition, cannot be transformed.

We focus on 7 Italian cities and 12 OMI zones.[4] The Italian housing market is particularly "thin" (the trading number is often insufficient to carry out a regression analysis). Hence, we also construct a (almost) thick market (named "Italy") by combining the house price data of the 7 Italian cities (see *Table 2*).

Table 2. Cross-section observations

City – Zone OMI	*Observations (bills of sale)*
Alessandria – B1	*100*
Cosenza – B1	*60*
Cosenza – D1	*61*
Crotone – D1	*64*
Genova – D43	*60*
Taranto – B1	*62*
Taranto – C1	*89*
Taranto – D1	*63*
Taranto – E2	*65*
Venezia Mestre – E23	*79*

[4] The Italian acronym OMI stands for "Osservatorio Mercato Immobiliare" and it refers to the Italian Real Estate Market Observatory. The OMI zone reflects a homogenous sector of the local property market, in which there is substantial uniformity of appreciation for environmental and socio-economic conditions.

Vercelli – B1	*100*
Vercelli – B2	*80*
Italy	*883*

Since it can be problematic to apply OLS regression to panel data, we carry out a cross-section analysis. In fact, in order to simplify the analysis, we accept a 2-year time period as a 1-year time period.

In the *Appendix*, we report some descriptive statistics for selling price data. 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). Unfortunately, our data availability does not permit the study of the time paths of these measures. However, the empirical strategy developed in section 2 is able to compute price dispersion "controlling" for the difference in attributes (or "qualities"), as suggested by Leung, Leong and Wong (2006).

4. Results and conclusions

The estimate of equation [1] is performed using Ordinary Least Squares (OLS). Two main empirical results are obtained from this analysis:

1) The dummy variable created as a proxy of the residual price volatility, and incorporated into the hedonic price function, is always statistically significant;

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

Table 4. Comparison between models

Cities / zones OMI	***Extendend Hedonic Model***		***Standard Hedonic Model***	
	R2-adjusted	***Standard deviation of Prediction Error****	***R2-adjusted***	***Standard deviation of Prediction Error****
Alessandria / B1	*85,22%*	*21,31%*	*72,88%*	*29,35%*
Cosenza / B1	*76,51%*	*32,31%*	*67,66%*	*37,41%*
Cosenza / D1	*90,02%*	*15,03%*	*81,06%*	*21,96%*
Crotone / D1	*82,19%*	*11,06%*	*73,06%*	*12,96%*
Genova / D43	*69,94%*	*17,47%*	*56,80%*	*22,98%*
Taranto / B1	*96,06%*	*11,61%*	*91,08%*	*17,90%*
Taranto / C1	*91,63%*	*8,00%*	*85,55%*	*10,69%*

Taranto / D1	*93,63%*	*5,93%*	*84,99%*	*8,30%*
Taranto / E2	*85,55%*	*6,49%*	*79,94%*	*7,85%*
Venezia Mestre / E23	*75,17%*	*19,68%*	*65,16%*	*24,65%*
Vercelli B1	*94,96%*	*10,10%*	*89,93%*	*14,24%*
Vercelli B2	*85,83%*	*18,67%*	*78,74%*	*23,86%*

* *Prediction Error = (predicted prices – observed selling prices) / observed selling prices.*

In a nutshell, this extended hedonic pricing model allows a major drawback of the standard hedonic pricing theory to be overcome, namely the assumption of competitive markets. Precisely, two key assumptions are usually adopted: 1) buyers and sellers, acting alone, cannot influence market prices; 2) buyers and sellers have full information regarding the market prices (Pope, 2008a, 2008b). However, in the actual housing markets this is hardly true. Housing markets are "thin" (i.e., markets with an insufficient amount of trading), local and decentralized, and thus buyers and sellers may have some market power (Harding et al., 2003a, 2003b; Cotteleer and Gardebroek, 2006). Therefore, this residual price volatility could be explained by the bargaining of the parties. 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, our model 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).

Indeed, we also calculate three independent dummy variables:

- If $(a_i - a_{mean}) > 0$ and $(p_i - p_{mean}) < 0$, it is assumed that this negative difference is due to the power of the buyer (we call this dummy *"buyer"*).
- If $(a_i - a_{mean}) < 0$ and $(p_i - p_{mean}) > 0$, it is assumed that this positive difference is due to the power of the seller (we call this dummy *"seller"*).
- Again, in markets for heterogeneous goods, such as a home, standard market situations take place when the property with higher (lower) advantages is sold at a higher (lower) price (we call this dummy *"standard market situation"*).

Hence, we construct a dummy for each category and include the first two dummies (namely, *"buyer"* and *"seller"*) in the estimation of the hedonic price model for the Italian housing market (obviously, the *"standard market situation"* reference dummy is excluded from the analysis). We find that the two dummy variables are statistically significant and their signs are as expected, namely negative for the dummy *"buyer"* and positive for the dummy *"seller"* (see *Table 5*).

Table 5. Role of residual price volatility in the formation process of housing prices in Italy

Cities / zones OMI	*Dummy variable "seller"*		*Dummy variable "buyer"*	
	Expected sign	***Statistically significant ****	***Expected sign***	***Statistically significant ****
Alessandria / B1	*Yes*	*Yes*	*No*	*No*
Cosenza / B1	*Yes*	*Yes*	*Yes*	*Yes*
Cosenza / D1	*Yes*	*Yes*	*Yes*	*No*
Crotone / D1	*Yes*	*Yes*	*Yes*	*Yes*
Genova / D43	*Yes*	*Yes*	*Yes*	*Yes*
Taranto / B1	*Yes*	*Yes*	*Yes*	*Yes*
Taranto / C1	*Yes*	*Yes*	*Yes*	*Yes*
Taranto / D1	*Yes*	*Yes*	*No*	*No*
Taranto / E2	*Yes*	*Yes*	*Yes*	*Yes*
Venezia Mestre / E23	*Yes*	*Yes*	*Yes*	*Yes*
Vercelli B1	*Yes*	*Yes*	*Yes*	*No*
Vercelli B2	*Yes*	*Yes*	*Yes*	*No*
Italy	*Yes*	*Yes*	*Yes*	*Yes*

** 5% significance level.*

Therefore, we are able to provide an economic explanation for the existence of residual price volatility.

References

Brown Jeff E., and Don E. Ethridge (1995), "Functional Form Model Specification: An Application to Hedonic Pricing", *Agricultural and Resource Economics Review*, 24(2), 166-173.

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).

Davidson, Russell, and James G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, 49(3), 781-793.

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.

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

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

MacKinnon, James G., Halbert White, and Russell Davidson (1983), "Tests for Model Specification in the Presence of Alternative Hypotheses: Some Further Results", *Journal of Econometrics*, 21(1), 53-70.

Pope, Jaren C. (2008a), "Buyer Information and the Hedonic: The Impact of a Seller Disclosure on the Implicit Price for Airport Noise", *Journal of Urban Economics*, 63(2), 498-516.

Pope, Jaren C. (2008b), "Do Seller Disclosures Affect Property Values ? Buyer Information and the Hedonic Model", *Land Economics*, 84(4), 551-572.

APPENDIX

Descriptive statistics on selling price

ITALY (all cities and OMI zones)

prezzo

	Percentiles	Smallest		
1%	27500	14000		
5%	55000	15000		
10%	67028	15000	Obs	883
25%	99202	15050	Sum of Wgt.	883
50%	132000		Mean	141192.9
		Largest	Std. Dev.	69479.72
75%	170000	500000		
90%	218000	550050	Variance	4.83e+09
95%	250250	620400	Skewness	2.022144
99%	420000	636349	Kurtosis	11.78734

Alessandria - zone OMI B1

price

	Percentiles	Smallest		
1%	14500	14000		
5%	46000	15000		
10%	50000	30000	Obs	100
25%	68000	35000	Sum of Wgt.	100
50%	92500		Mean	104208
		Largest	Std. Dev.	60071.09
75%	122750	250000		
90%	166500	287000	Variance	3.61e+09
95%	207500	310000	Skewness	2.298197
99%	365000	420000	Kurtosis	10.98994

Cosenza - zone OMI B1

price

	Percentiles	Smallest		
1%	15000	15000		
5%	28750	15050		
10%	60000	27500	Obs	60
25%	97550	30000	Sum of Wgt.	60
50%	125500		Mean	134130.8
		Largest	Std. Dev.	66697.95
75%	167500	220000		
90%	202000	230000	Variance	4.45e+09
95%	225000	268000	Skewness	1.291707
99%	425000	425000	Kurtosis	7.499519

Cosenza - zone OMI D1

price

	Percentiles	Smallest		
1%	35000	35000		
5%	84100	52000		
10%	88075	73000	Obs	61
25%	105900	84100	Sum of Wgt.	61
50%	138000		Mean	143267.8
		Largest	Std. Dev.	58863.37
75%	165000	240400		
90%	214770	264500	Variance	3.46e+09
95%	240400	298700	Skewness	1.650936
99%	394000	394000	Kurtosis	7.510921

Crotone - zone OMI D1

price

	Percentiles	Smallest		
1%	52500	52500		
5%	70000	60000		
10%	75000	65000	Obs	64
25%	89000	70000	Sum of Wgt.	64
50%	108500		Mean	108808
		Largest	Std. Dev.	27503.22
75%	125000	150000		
90%	142300	150000	Variance	7.56e+08
95%	150000	185000	Skewness	.5866066
99%	198000	198000	Kurtosis	4.011414

Genova - zone OMI D43

price

	Percentiles	Smallest		
1%	95000	95000		
5%	100000	96000		
10%	105500	100000	Obs	60
25%	133500	100000	Sum of Wgt.	60
50%	170000		Mean	178744.2
		Largest	Std. Dev.	60928.44
75%	211500	275000		
90%	252500	305000	Variance	3.71e+09
95%	290000	330000	Skewness	1.09893
99%	400000	400000	Kurtosis	4.762783

Taranto - zone OMI B1

price

	Percentiles	Smallest		
1%	19000	19000		
5%	30000	20000		
10%	40000	23000	Obs	62
25%	60000	30000	Sum of Wgt.	62

50%	88000		Mean	105455.7
		Largest	Std. Dev.	62779.3
75%	145023	240000		
90%	190000	240000	Variance	3.94e+09
95%	240000	250000	Skewness	.9869875
99%	298000	298000	Kurtosis	3.530719

Taranto - zone OMI C1

price

	Percentiles	Smallest		
1%	75000	75000		
5%	85000	80000		
10%	98000	80000	Obs	89
25%	113000	80000	Sum of Wgt.	89
50%	135000		Mean	145659.6
		Largest	Std. Dev.	43832.97
75%	170000	230000		
90%	215000	230000	Variance	1.92e+09
95%	225000	245000	Skewness	.7087849
99%	280000	280000	Kurtosis	2.898878

Taranto - zone OMI D1

price

	Percentiles	Smallest		
1%	70000	70000		
5%	110000	88400		
10%	130000	92000	Obs	63
25%	150000	110000	Sum of Wgt.	63
50%	165000		Mean	168863.5
		Largest	Std. Dev.	37185.39
75%	185000	235000		
90%	220000	240000	Variance	1.38e+09
95%	235000	240000	Skewness	.0269188
99%	260000	260000	Kurtosis	3.418758

Taranto - zone OMI E2

price

	Percentiles	Smallest		
1%	55000	55000		
5%	93000	70000		
10%	119000	72000	Obs	65
25%	135000	93000	Sum of Wgt.	65
50%	145000		Mean	143359.9
		Largest	Std. Dev.	27129.34
75%	158000	185000		
90%	175000	187000	Variance	7.36e+08
95%	185000	188750	Skewness	-.6822701
99%	218000	218000	Kurtosis	5.132266

Venezia – zone OMI E23

price

	Percentiles	Smallest		
1%	75000	75000		
5%	100000	95600		
10%	118000	97950	Obs	79
25%	137800	100000	Sum of Wgt.	79
50%	172000		Mean	206958.8
		Largest	Std. Dev.	111337
75%	241250	500000		
90%	306500	550050	Variance	1.24e+10
95%	500000	620400	Skewness	2.158337
99%	636349	636349	Kurtosis	7.91447

Vercelli - zone OMI B1

price

	Percentiles	Smallest		
1%	34000	30000		
5%	45500	38000		
10%	60500	40000	Obs	100
25%	81250	40000	Sum of Wgt.	100
50%	118750		Mean	122064.7
		Largest	Std. Dev.	55422.36
75%	150000	245000		
90%	200000	260000	Variance	3.07e+09
95%	220000	280000	Skewness	.8501437
99%	295000	310000	Kurtosis	3.791025

Vercelli - zone OMI B2

price

	Percentiles	Smallest		
1%	27000	27000		
5%	55000	48000		
10%	62500	52000	Obs	80
25%	93500	55000	Sum of Wgt.	80
50%	127500		Mean	147025.7
		Largest	Std. Dev.	83282.02
75%	183750	345000		
90%	277500	350000	Variance	6.94e+09
95%	342500	363000	Skewness	1.488779
99%	450000	450000	Kurtosis	5.135608