



Measuring the Effect of Location on House Prices in Italy

Gaetano Lisi^{1*}, Mauro Iacobini²

¹Department of Economics, e-Campus University, Via Isimbardi, 10, I-22060 Novedrate (CO), Italy

²The Revenue Agency, Taxpayers' Services (Estimate and Observatory on the Real Estate), I-00185 Roma

E-mail: gaetano.lisi@uniecampus.it

Abstract: The Italian housing market is characterised by both a strong heterogeneity of real estate assets and a reduced number of property sales. These features, indeed, hamper the use of the hedonic price method, namely, the method that is mostly used for assessing the house prices and for estimating the monetary value of housing characteristics. In this paper, therefore, a hedonic model with dummy variables that identify housing submarkets is used to achieve two important results: enabling greater use of multiple regression analysis in the study of the Italian real estate market, and catching, in the simplest possible manner, the effect of location on house price. Indeed, the house's location is, together with the area in square metres, the housing characteristic that most influences the house price.

Keywords: location, housing submarkets, house price, hedonic pricing models

1. Introduction

According to the well-known hedonic price theory (see the seminal works by Rosen^[1] and Epple^[2], and the influential surveys by Sheppard^[3] and Malpezzi^[4]), the price of a composite good (such as housing) crucially depends on its intrinsic and extrinsic characteristics, each of which can be evaluated independently. Indeed, the characteristics of a composite good are known as “hedonic”, because their monetary values can only be evaluated indirectly through the overall price of the good. Precisely, the monetary values of the housing characteristics (the hedonic prices) can be obtained by the estimation of the coefficients of a regression model, where the house price is a function of the main housing characteristics possessed by the property (the so-called “hedonic price models”).

It is shared opinion that the characteristic “location” is, together with the area in square metres, the housing characteristic that most influences the house price [Many factors can influence the value of a home, but location is one of the most influential (see, e.g., New York State Department of Taxation and Finance. “How to Estimate the Market Value of Your Home”, publication 1115, November 2012)]. The reason is quite clear and intuitive: the house's location is the only housing characteristic that one cannot change. To capture this feature, very sophisticated statistical / econometric models are often used, the so-called “spatial models” [For an overview of these models, see, e.g., Ward and Gleditsch^[5]]. Alternatively, in order to capture the effect of location, there is a considerably simpler method to implement: the use of binary variables to identify the various real estate sub-markets that exist in a given territory or city [However, the special specification of the hedonic pricing model is used to cope with problems of non-random distribution of the errors in the space of investigation, heteroscedasticity of the variables, etc.]. Specifically, for the characteristic “location”, one will have a binary variable for each possible submarket, which will assume value 1 if the dwelling is located in that particular submarket and 0 otherwise [A binary variable or dummy variable is a variable that is used in empirical models to represent a non-quantitative characteristic, such as gender, race or precisely location (see, e.g., Hill^[6])]. Indeed, Bourassa et al.^[7] show that the gain in terms of correctness and accuracy—that is derived from the inclusion in a standard hedonic model of binary variables that refer to the various sub-markets—is not inferior to the gain which is derived from the use of spatial methods. This result has enormous empirical importance, since a hedonic model with binary variables is considerably simpler to implement and to interpret economically than spatial models.

This approach can be particularly useful for the Italian real estate market, which is characterised by little dynamism (in terms of number of trades) and by subdivision into OMI zones. The “OMI zone” defines a homogeneous sector of the local real estate market (of a particular city), in which there is a substantial uniformity of appreciation for economic and socio-environmental conditions [The acronym OMI refers to the institute that takes care of the data and statistics relating

to the Italian real estate market, i.e., the Observatory on the Market of Immovable property (OMI) of the Italian Revenue Agency]. The OMI zone, therefore, identifies a real estate submarket, and it is able to distinguish between different locations; in the sense that two similar dwellings, with respect to characteristics, could have very different prices due to the fact that they belong to two different OMI zones. Consequently, by creating binary variables that refer to the different OMI zones, it is possible to grasp the effect that the location of the property in a particular real estate submarket (OMI zone) has on its overall house price, simply by including such binary variables in a standard hedonic model, as if they were further housing characteristics. Furthermore, if each OMI zone identifies a particular real estate submarket, then by aggregating more OMI zones that refer to the same city it is possible to obtain a more “extended” real estate market, with a dynamism that will be amplified, since the total number of trades will be given by the sum of the sales that are realised in all the sub-markets (OMI zones). It follows that, compared to a small increase in the number of parameters to be estimated (the regression coefficients of the binary variables), there will be a considerable increase in the number of observations and, therefore, a net increase in the degrees of freedom of the regression model, with a consequent increase in the (economic) reliability and significance (statistics) of the analysis. Eventually, therefore, two important purposes can be achieved by the present analysis: allowing greater use of multiple regression analysis in the study of the Italian real estate market, and grasp, in the simplest possible manner, the key role of location in the estimate of house prices.

The rest of this paper is organised as follows. Section 2 outlines the model, while Section 3 shows the empirical results. Finally, Section 4 concludes the work

2. The model

The econometric model that was used in the present work is relatively simple to implement and it consists of including, in a standard hedonic model, the binary variables which refer to the OMI zones of the city of Taranto (a provincial capital of Italy) [The city of Taranto, in fact, represents an Italian real estate market that is characterised by a relative dynamism, which is understood to mean the number of trades compared to the national average. The OMI zones of the municipality of Taranto are 26 in number, of which 13 zones are mainly for residential use. Among the 13 zones that are predominantly residential, those zones that characterised by greater market liveliness, in terms of number of trades, are as follows: OMI zone B1 (central zone), OMI zone C1 (semi-central zone), OMI zone D1 (peripheral zone), and OMI zone E2 (suburban zone). In these zones, in fact, there is an average of 50% of the transactions which are recorded in the capital city that are residential (villas included) and it was, therefore, possible to find a sufficient amount of data (exclusively sales deeds), although the survey was extended by more than one semester (to be precise, real estate units that were bought and sold between the second half of 2008 and the first half of 2010 have been considered)]. Specifically, the four binary variables that were constructed and considered in the present analysis are in Table 1.

Table 1. Location, OMI zones and binary variables

Location	Variable (symbol)	Description
Zone OMI D1	d1	= 1 if the house is located in zone D1 and 0 otherwise
Zone OMI C1	c1	= 1, if the house is located in zone C1 and 0 otherwise
Zone OMI B1	b1	= 1, if the house is located in zone B1 and 0 otherwise
Zone OMI E1	e2	= 1, if the house is located in zone E2 and 0 otherwise

Obviously, the dummy variables that identify more than one mode or category must be interpreted with respect to the mode / category that was chosen as a reference. For example, when considering the four OMI zones in Table 1, it is necessary to define the reference OMI zone (the choice is completely subjective and does not affect the results of the analysis) and, once the regression model is estimated, it is also necessary to interpret the results that have been obtained for the other OMI zones, with the aim of comparing them with the reference OMI zone, i.e., one must answer the following question: *How does the price change when considering a house' location in an OMI zone that is different from the reference zone?* The answer to this question is given by the estimation of the regression coefficients that are associated with the binary variables “OMI zones”. Concisely, in this model, the spatial structure is by-passed as information on the OMI zones is available. This solution can be applied to the estimation of house prices in all Italian cities. Of course, the power of the model relies on the quality of the clustering offered by the OMI zones.

The OMI zone that was chosen as a reference is the C1. This implies that the binary variable “c1” is not included in the hedonic model. Considering the simplest of the hedonic models, the linear type, the hedonic price function including the binary variables “OMI zones” is as follows:

$$P_i = a + \sum_{j=1}^n (\beta_j \cdot X_{j,i}) + \gamma_1 \cdot b1 + \gamma_2 \cdot d1 + \gamma_3 \cdot e2 + \varepsilon_i \quad (1)$$

where P_i is the selling price of the i -th house, a is the intercept, $X_{j,i}$ represents the degree or intensity of the j -th characteristic owned by the i -th house (such as the number of bathrooms or the surface area in square metres); n is the number of housing characteristics that are considered in the analysis; β_j is the coefficient that expresses the relationship between P_i and $X_{j,i}$; while, γ_1 , γ_2 and γ_3 are the regression coefficients that are associated with the binary variables OMI zones. Finally, ε_i is the stochastic error term. Under the usual hypotheses of the specification of the classical regression model (linear relation in the parameters, exogenous explanatory variables and errors with zero mean, constant variance and non-correlated with each other), all the parameters of the model (β_j , γ_1 , γ_2 and γ_3) can be estimated by using the Ordinary Least Square (OLS) estimator. Precisely, under these assumptions, the linear OLS estimator is the best possible solution, namely it is correct and with minimum variance or alternatively, the Best Linear Unbiased Estimator (BLUE).

From (1), by omitting the error term, it is possible to obtain the expected price (\hat{P}_i) which will be different according to the OMI zone in which the property is located. Precisely,

$$\hat{P}_i = a + \sum_{j=1}^n (\beta_j \cdot X_{j,i}) \quad (2)$$

(2) is the expected price in the case where the i -th house is located in the reference OMI zone, i.e., $b1 = d1 = e2 = 0$. If, instead, the i -th house is located in the OMI B1 zone, i.e., $b1 = 1$ and $d1 = e2 = 0$, the expected price is equal to:

$$\hat{P}_i = a + \sum_{j=1}^n (\beta_j \cdot X_{j,i}) + \hat{\gamma}_1 \quad (3)$$

where $\hat{\gamma}_1$ is the estimation of the regression coefficient γ_1 . In the case where the i -th house is located in the OMI zone D1, i.e., $d1 = 1$ and $b1 = e2 = 0$, the expected price is given by:

$$\hat{P}_i = a + \sum_{j=1}^n (\beta_j \cdot X_{j,i}) + \hat{\gamma}_2 \quad (4)$$

where $\hat{\gamma}_2$ is the estimation of the regression coefficient γ_2 . Finally, if the i -th house is located in the OMI zone E2, i.e., $e2 = 1$ and $b1 = d1 = 0$, then the expected price is equal to:

$$\hat{P}_i = a + \sum_{j=1}^n (\beta_j \cdot X_{j,i}) + \hat{\gamma}_3 \quad (5)$$

where $\hat{\gamma}_3$ is the estimation of the regression coefficient γ_3 . As previously stated, the coefficients γ_1 , γ_2 and γ_3 must be interpreted with a view to comparing them with the reference OMI zone C1. Precisely, under the condition of *ceteris paribus*, namely, considering constant the other housing characteristics $X_{j,i}$, the economic meaning of the parameters γ_1 , γ_2 and γ_3 is the following:

- If $\hat{\gamma}_1 > 0$, the location of the property in the OMI zone B1 compared to the location in the reference OMI zone (the OMI zone C1), increases the price of $\hat{\gamma}_1$; whereas, if $\hat{\gamma}_1 < 0$, the location in that submarket decreases the price by $\hat{\gamma}_1$;
- If $\hat{\gamma}_2 > 0$, the location of the property in the OMI zone D1 compared to the location in the reference OMI zone, increases the price of $\hat{\gamma}_2$; whereas, if $\hat{\gamma}_2 < 0$, the location in that submarket decreases the price by $\hat{\gamma}_2$;
- If $\hat{\gamma}_3 > 0$, the location of the property in the OMI zone E2 compared to the location in the reference OMI zone, increases the price of $\hat{\gamma}_3$; whereas, if $\hat{\gamma}_3 < 0$, the location in that submarket decreases the price by $\hat{\gamma}_3$.

3. The empirical analysis

The housing characteristics that are included in the analysis are reported in Table 2 [For the sake of simplicity, we exclude from the model the housing characteristic “energy efficiency”. However, there is a huge literature probing how “energy efficiency” is becoming a relevant characteristic contributing to the overall value of the property (in both the rental market and the sales market). For the Italian housing market, e.g., see Bisello, et al. ^[8] and Bottero, et al. ^[9]].

Table 2. Housing characteristics

Variable	Acronym	Type	Score
Lot size	sup	Continuous variable	square metres
Maintenance status of whole building	s_ed	Categorical variable	from 1 to 5
Maintenance status of real estate unit	s_ui	Categorical variable	from 1 to 5
Distributive quality of real estate unit	q_dis	Categorical variable	from 1 to 5
Safety of the whole building	sic	Categorical variable	from 1 to 5
Brightness of real estate unit	lum	Categorical variable	from 1 to 5
Window	aff	Discrete variable	number
Quality of the building's location [This housing characteristic aims at catching the key role of location in addition to that of the OMI zone]	loc	Categorical variable	from 1 to 5
Quality of landscape of real estate unit	q_aff	Categorical variable	from 1 to 5
Architectural style of building	t_arc	Categorical variable	from 1 to 5
Bathroom	bagn	Discrete variable	number
Presence of lift	asc	Binary variable	1 = yes; 0 = no
New construction	n_cos	Binary variable	1 = yes; 0 = no

The first operation to be performed is to integrate the dataset with the observations that are related to the four OMI zones under consideration in the analysis. The number of overall observations is 277 (62 for the OMI zone B1, 88 for the OMI zone C1, 62 for the OMI zone D1, and 65 for the OMI zone E2) [These are transaction prices referring to the same period (2009-2010). Source: Observatory on the Market of Immovable property (OMI) of the Italian Revenue Agency].

With respect to the *categorical variables* (i.e., qualitative variables that can be ordered), the procedure which was adopted in this analysis consists of transforming the categorical variables into binary variables. Precisely, given the range from 1 to 5 with which they are classified, viz.: 1 = *very poor*; 2 = *poor*; 3 = *normal*; 4 = *good*; 5 = *excellent*, the binary variable takes on value 1 (one) if the score that expresses its quality is greater than 3 (= *normal* = threshold value) and 0 otherwise. This strategy is used for all the *categorical variables* and has the advantage of estimating only one parameter. Moreover, using a model that is rich in binary variables simplifies the problem related to the choice of the functional form. In fact, the well-known Box-Cox transformation can only be applied to variables with strictly positive values. Indeed, the choice of the best functional form for the hedonic model is perhaps the main problem to be solved [1-4], [10].

In testing the hypothesis of normality of the distribution of the price variable, both in level and in natural logarithm, it is possible to obtain a first indication with respect to the functional form to be used for the hedonic model. The results show that only the hypothesis of the normality of the price distribution in the level is not rejected at the usual levels of significance (of 5% and 10%) [The statistical software that was used in the analysis is STATA 11]:

Shapiro-Wilk W test for normal data (variable: Price) *Prob > z = 0.12036*
Skewness / Kurtosis tests for Normality (variable: Price) *Prob > chi2 = 0.7057*
Shapiro-Wilk W test for normal data (variable: natural logarithm of Price) *Prob > z = 0.0000*
Skewness / Kurtosis tests for Normality (variable: natural logarithm of Price) *Prob > chi2 = 0.0000*

Therefore, these results seem to suggest that the best hedonic model should be sought among linear models in the dependent variable (price).

Furthermore, we use the Box-Cox transformation which is only applied to the four non-binary variables, i.e., price, lot size (X_1), bathrooms (X_2), and windows (X_3):

$$\frac{P_i^\lambda - 1}{\lambda} = \beta_1 \cdot \frac{X_1^\lambda - 1}{\lambda} + \beta_2 \cdot \frac{X_2^\lambda - 1}{\lambda} + \beta_3 \cdot \frac{X_3^\lambda - 1}{\lambda} + \sum_{j=4}^n (\beta_j \cdot X_{j,i}) + \gamma_1 \cdot b1 + \gamma_2 \cdot d1 + \gamma_2 \cdot e2 + \varepsilon_i \quad (6)$$

where λ is the parameter that is used for the transformation, while for the error term the usual assumptions are valid. The Box-Cox transformation suggests that the best model is linear: in fact, only the null hypothesis that the lambda reference parameter is equal to 1 is not rejected at the usual levels of confidence:

lambda = -1, with p-value / Prob > chi2 = 0.000; lambda = 0, with p-value / Prob > chi2 = 0.000; lambda = 1, with p-value / Prob > chi2 = 0.155 [For simplicity, we have chosen the same parameter, lambda, for the transformation of the price dependent variable and the independent variables. It is possible to show that for $\lambda = 0$, the transformation implies a

logarithmic model, whereas for $\lambda = 1$, the Box-Cox transformation implies a linear model. In the case, instead, of $\lambda = -1$, the correct transformation to be applied to the variables would have been the reciprocal transformation].

Among the hedonic models that are usually used in empirical studies in real estate, in this specific instance, the linear function is the best hedonic model. Indeed, in the real estate sector, the linear model is often preferred, since there is the obvious advantage of directly estimating the monetary values of the housing characteristics^{[10], [11]}.

Therefore, the hedonic model to be estimated is exactly the model that was previously introduced, namely, the model (1). The results of the estimate are shown in Table 3 [The housing characteristics that are not included in the estimate, because they are statistically non-significant at the usual 10% confidence level are: quality of landscape (p-value = 0.9389), number of windows (p-value = 0.7334) and architectural style (p-value = 0.4003)].

Table 3. Estimation results

Variable (dependent: price)	Coefficient	Standard error	t-statistic	p-value
sup	1207.57	43.64	27.67	0.000
s_ed	15092.01	2648.62	5.70	0.000
s_ui	16027.46	2403.15	6.67	0.000
q_dis	11440.18	2516.58	4.55	0.000
sic	10918.89	4022.26	2.71	0.007
lum	9541.81	2519.54	3.79	0.000
loc	16618.82	2733.33	6.08	0.000
bagn	10436.44	2820.95	3.70	0.000
asc	5382.10	2641.41	2.04	0.043
n_cos	9368.91	3703.76	2.53	0.012
d1	-12322.47	3323.48	-3.71	0.000
b1	-38546.41	3204.24	-12.03	0.000
e2	-44568.22	4058.38	-10.98	0.000

Adj R² = 0.8759
F(13, 263) = 150.79
Prob > F = 0.0000

However, before commenting on the results, it is necessary to check the statistical correctness of the model that is estimated, by using the most important tests that are usually used in empirical analyses. In particular, the test on the correct specification of the model (the Ramsey RESET test) confirms the goodness of the outcome of the preliminary analysis on the choice of the functional form to be used for the hedonic price function. The null hypothesis (H_0) of the correct specification of the model is not rejected at the usual levels of confidence: the p-value associated with the test (0.7402), in fact, is above the two thresholds of 5% and 10%. Instead, the Breusch-Pagan / Cook-Weisberg test for heteroscedasticity (p-value = 0.0317) and the two tests on the normal distribution of residuals, i.e. the Skewness / Kurtosis test (p-value = 0.0001) and the Shapiro-Wilk W test (p-value = 0.00164), are not exceeded. Furthermore, we perform the analysis of residuals. Basically, in order for the estimated model to be statistically satisfactory, the residuals should be causally arranged, thus forming a sort of “cloud” around their average value, i.e., zero (see Figure 1).

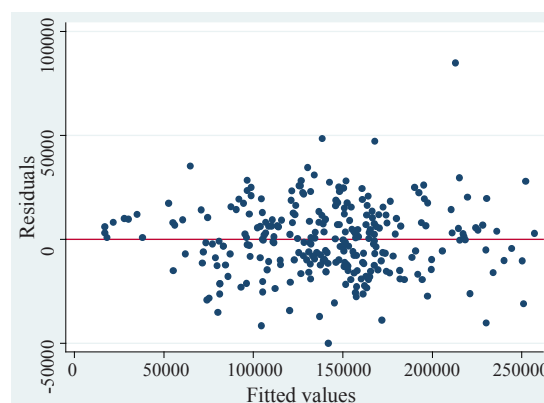


Figure 1. Analysis of residuals

The analysis of residuals clearly shows the presence of a “severe outlier” (a serious anomalous value) in the upper right-hand corner of Figure 1. In fact, it is an observation that is quite different (distant) from the distribution of the other residuals. The severe outlier can be precisely identified using standardised residuals and a well-known “rule of thumb”. In brief, once the standardised residuals have been calculated, we highlight those residuals that exceed the tolerance range which is defined between +3 and -3.

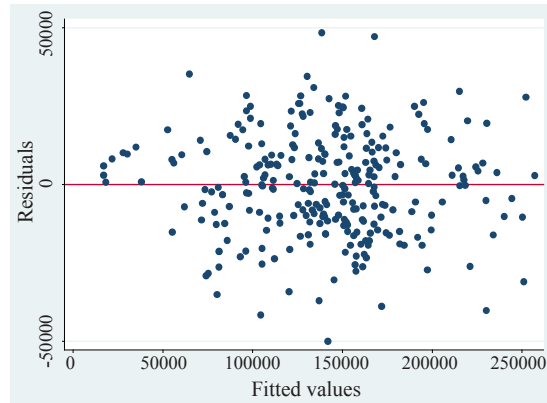


Figure 2. Analysis of residuals (with no severe outliers)

Since there is only one observation that greatly exceeds the tolerance range ($rstud = 5.455607$), we simply aim at eliminating the observation corresponding to the serious anomalous value and then we repeat the analysis of the residuals (see Figure 2). In Figure 2, there are no other severe outlier; indeed, with the elimination of the severe outlier, the “cloud” that the residuals form around their average value is much more visible when compared to the graph in Figure 1 (within which the “cloud” is more “flattened”, and greatly so). Then, we proceed again to the estimate of the linear model, net of the serious anomalous value.

The results of the estimation of the linear model without the severe outlier are reported in Table 4 (now the observations are 276 in number, instead of 277).

Table 4. Estimation results (with no severe outliers)

Variable (dependent: price)	Coefficient	Standard error	t-statistic	p-value
sup	1198.97	41.46	28.92	0.000
s_ed	14000.06	2522.63	5.55	0.000
s_ui	14378.18	2301.56	6.25	0.000
q_dis	10324.16	2398.06	4.31	0.000
sic	7151.89	3880.77	1.84	0.066
lum	8990.14	2394.26	3.75	0.000
loc	17766.16	2603.61	6.82	0.000
bagn	9064.16	2690.08	3.37	0.001
asc	6484.54	2515.96	2.58	0.011
n_cos	7378.09	3535.34	2.09	0.038
d1	-10671.39	3169.89	-3.37	0.001
b1	-39157.65	3044.26	-12.86	0.000
e2	-41742.69	3887.79	-10.74	0.000
Adj R ² = 0.8843				
F(13, 262) = 162.62				
Prob > F = 0.0000				

As in the previous estimate, the housing characteristics, that are statistically insignificant and which not included in the analysis, are the same: quality of landscape, architectural style, and number of windows. However, there is a slight improvement in the goodness of fit of the estimated model: the adjusted R² is in fact slightly higher (0.8843 against the 0.8759 of the previous model with all 277 observations). Moreover, without that one but severe outlier, the linear model now overcomes all the main tests and can, therefore, be defined as statistically correct. Precisely, Ramsey RESET test (p-value = 0.7635); Breusch-Pagan / Cook-Weisberg test for heteroscedasticity (p-value = 0.7019); Skewness / Kurtosis test (p-value = 0.9390) and Shapiro-Wilk W test (p-value = 0.97430). Moreover, multicollinearity is not a problem (mean

VIF-Variance Inflation Factor = 1.52). Therefore, the removal of the severe outlier had an important positive effect on the statistical correctness of the model (increase in the adjusted R^2 and the overcoming of all the main statistical tests).

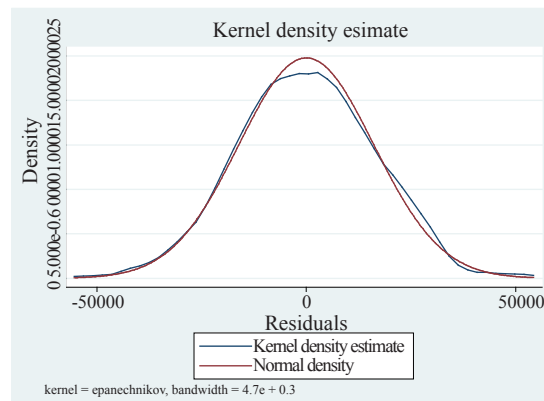


Figure 3. Distribution of residuals and Normal distribution

Also from the graphical perspective, it is immediately evident that the fundamental hypothesis of the normal distribution of residuals can hardly be rejected (see Figure 3) [The elimination of the severe outlier also contributes to eliminating the (previous) slightly leptokurtic form (i.e., a form that is more elongated than a normal distribution) of the distribution of residuals; in fact, the relative index of kurtosis is now significantly lowered (Kurtosis = 3.017356)].

At this point, once ascertained that the model is statistically correct, it is possible to comment on the results of the estimates from an economic point of view. The monetary values (hedonic prices) of the housing characteristics are shown in Table 5.

Table 5. Hedonic prices of housing characteristics

Housing characteristic	Hedonic price
sup	€ 1,198.96
bagn	€ 9,064.16
s_ed	€ 14,000.06
s_ui	€ 14,378.18
q-dis	€ 10,324.16
sic	€ 7,151.88
lum	€ 8,990.14
loc	€ 17,766.16
asc	€ 6,484.54
n_cos	€ 7,378.09

The hedonic price of the lot size is approximately 1,200 euros (per square metre), while an extra bathroom has an important effect on an increase in the house price (of about 9,000 euros). On average, under the *ceteris paribus* condition, a state of maintenance (of both the building and the real estate unit) above the norm (the threshold value of 3) determines a price increase of approximately 14,000 euros. Under the same conditions, “distributive quality of real estate unit”, “safety of the whole building” and “brightness of real estate unit” that are above the norm will increase the price by approximately 10,000; 7,000 and 9,000 euros, respectively. Further, in some respects the contribution of the “quality of the building’s location” is surprising, since it is solely a feature of “detail” within the same OMI zone (a location of detail above the norm determines an increase in price that is equal to almost € 18,000). This result represents a further demonstration of the crucial importance of the characteristic “location” in the explanation of the house prices. Finally, the presence of the lift implies, *ceteris paribus*, an increase in the price of about 6,500 euros, while a new construction increases the price of about 7,400 euros.

Focussing, instead, on the binary variables of the OMI zones, the parameters γ_1 , γ_2 and γ_3 express the change in the price of the property that emerges from the comparison between the location in one of the three OMI zones under consideration in the analysis (B1, D1 and E2) and the reference OMI zone, C1. The estimates of the hedonic price of the OMI zones are shown in Table 6.

Table 6. Hedonic prices of submarket dummy variables

Submarket (OMI zone)	coefficient	Hedonic price
B1	γ_1	- € 39,157.65
D1	γ_2	- € 10,671.39
E2	γ_3	- € 41,742.69

All the parameters have a negative sign. This means that, on average, the location in the OMI zone C1 always determines an increase in price. Precisely,

- *ceteris paribus* (namely, subject to the degree and intensity of the housing characteristics considered in the analysis), the location of the property in the OMI zone E2, compared to the location in the reference OMI zone (the OMI zone C1), reduces the price of € 41,742.69;

- *ceteris paribus*, the location of the property in the OMI zone B1 compared to the location in the reference OMI zone (the OMI zone C1), reduces the price of € 39,157.65;

- *ceteris paribus*, the location of the property in the OMI one D1 compared to the location in the reference OMI zone (the OMI one C1), reduces the price of € 10,671.39.

The results obtained appear to be consistent with the characteristics of the OMI zones that are considered in the analysis. In fact, the reference OMI zone C1 is, commercially, the most active zone of Taranto, where the highest number of transactions is being recorded and to where the economic centre of the city is progressively being transferred.

Also, these results are potentially useful in the real estate appraisal. For example, given two properties, property A and property B, that are identical with respect to all housing characteristics except for location, namely $\sum_{j=1}^n (\beta_j \cdot X_{j,i})$ is the same in the model (1), with property A that is located in the OMI zone C1 (the OMI zone of reference) and which sold for € 110,000.00, it is possible to estimate the house price of property B that has not yet been bought and is located in the OMI zone D1 as follows:

$$\widehat{P}_B = P_A + \widehat{\gamma}_2$$

$$\widehat{P}_B = 110,000.00 + (-10,671.39) = 99,328.61$$

that is, we add to the house price of the property located in the OMI zone C1, the estimated regression coefficient (the hedonic price) that refers to the OMI zone D1. Obviously, this simple estimation procedure can be used in the case of similar properties (which is a very particular and often unrealistic case). Nevertheless, this extremely simplified example has highlighted the effectiveness of the simple model developed in this paper. The OMI zones, in fact, are able to discriminate between different locations, in the sense that two similar properties, by way of their characteristics, could also have very different prices due to the fact that they belong to two different OMI zones.

Of course, it is possible to use the same procedure in case where we know the house price of property B and want to estimate the house price of property A, viz.:

$$\widehat{P}_A = P_B - \widehat{\gamma}_2$$

$$\widehat{P}_A = 99,328.61 - (-10,671.39) = 110,000.00$$

4. Conclusions

The Italian housing market is characterised by little dynamism (in terms of number of trades) and by subdivision into OMI zones. The “OMI zone” defines a homogeneous sector of the local real estate market (of a city), in which there is a substantial uniformity of appreciation for economic and socio-environmental conditions. The OMI zone, therefore, identifies a real estate submarket, and it is able to distinguish between different locations; in the sense that two similar dwellings, with respect to characteristics, could have very different prices due to the fact that they belong to two different OMI zones. Ultimately, by creating binary variables that refer to the different OMI zones, it is possible to grasp the

effect that the location of the property in a particular real estate submarket (OMI zone) has on its house price, simply by including such binary variables in a standard hedonic model, as if they were further housing characteristics. Indeed, location is, together with the area in square metres, the housing characteristic that most influences the house price. The house's location is, in fact, the only housing characteristic that one cannot change.

Consequently, it is not always so necessary and so obvious to resort to the so-called spatial models-which are often difficult to interpret from an economic perspective-in order to grasp the important effect that the characteristic "location" has on the house price. Indeed, it is possible to obtain equally accurate and correct results simply by introducing, in a standard hedonic model, the binary variables that refer to the different sub-markets.

Eventually, the empirical problem which is derived from a limited number of trades-which mainly characterises the Italian context and which strongly limits the use of multiple regression analysis in the estimation of hedonic prices-can somehow be overcome by "merging" the various sub-markets that exist in a particular city.

References

- [1] Rosen S. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*. 1974; 82(1): 34-55.
- [2] Epple D.R. Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy*. 1987; 95(1): 59-80.
- [3] Sheppard S. Hedonic Analysis of Housing Markets, in *Handbook of Regional and Urban Economics Volume 3: Applied Urban Economics*, edited by Paul Cheshire and Edwin Mills, Amsterdam: North Holland. 1999; 41: 1595-1635.
- [4] Malpezzi S. Hedonic pricing models: A selective and applied review, in T.O. Sullivan, K. Gibbs (eds.). *Housing Economics and Public Policy: Essay in Honour of Duncan MacLennan*. Oxford: Blackwell Science; 2003.
- [5] Ward M. D., Gleditsch K. S. *Spatial Regression Models (2nd Edition)*. SAGE Publications; 2009.
- [6] Hill R.C., Griffiths W.E., Lim G.C. *Principles of Econometrics*. New York: John Wiley & Sons, Inc.; 2011. 4th ed.
- [7] Bourassa S.C., Cantoni E., Hoesli M. Spatial dependence, housing submarkets, and house price prediction. *Journal of Real Estate Finance and Economics*. 2007; 35(2): 143-160.
- [8] Bisello A., Antonucci V., Marella G. Measuring the price premium of energy efficiency: A two-step analysis in the Italian housing market. *Energy and Buildings*. 2020; 208(1): 109670.
- [9] Bottero M., Bravi M., Federico Dell'Anna F., Mondini G. Valuing buildings energy efficiency through hedonic prices method: are spatial effects relevant? *Valori e valutazioni*. 2018; 21: 27-39.
- [10] Eurostat. *Handbook on Residential Property Prices Indices (RPPIs), Methodologies & Working papers*. 2013. Available from: <http://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF>.
- [11] Maurer R., Pitzer M., Sebastian S. Hedonic price indices for the Paris housing market. *Allgemeines Statistisches Archiv (Journal of the German Statistical Society)*. 2004; 88(3): 303-326.