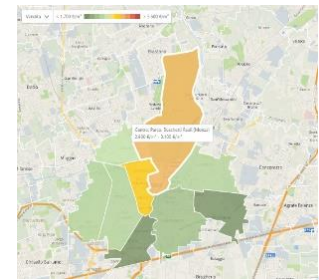


ANALYSIS OF HOUSE PRICES IN MONZA

Introduction

For my econometric project I decided to value the prize of the houses I selected in my birthplace, Monza. To complete the dataset, first of all, I decided the aim of my research; substantially I'm trying to figure out which one of the main "attraction" in Monza can influence the price of the houses the most. To complete my dataset I used immobiliare.it, an internet website, in which I found some important and relevant value that can influence the price, too. Secondly, I decided to circumscribe the area where I would have done my detection. Monza is divided into eight price zone and, in order to compare the most important attraction of the city, I decided to take into account just the most expensive area, called "Centro", "Parco" and "Boschetti Reali" situated into the north east of the city, given the fact that this is also the most important, touristic and tidy area of the city. Then, I selected my point of interest to sort the three most important price variable of this area; I chose "duomo" as a point of interest that could efficaciously represent the variable "centre", the "villa reale" to define "boschetti reali" and then I selected the central point of the park to evaluate the "parco" component.



The point is to give a significant regression model in order to being able to estimate the price of the houses in the orange area in the picture above, making an important focus on some distance variable within it. Due to a urban redevelopment policy, in the last 10 years, the city center has undergone a restyling that could have, certainly, increased the price of the houses in that area.

The dataset is composed of 105 cases randomly selected in the previously mentioned area, using the cross-section method to create a randomly selected sample. At this point I started selecting some variables that could influence the explained variable Y. The variables I selected are divided into three categories.

Variables

House data variables

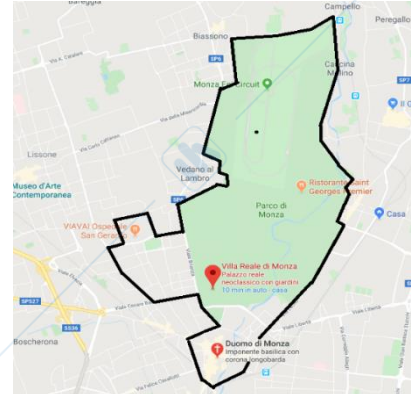
I chose as variable that could influence the price of the houses in monza first of all the number of locals for each observation (ROOM), the dimension of the house measured in mq (SQRDSIZE), the number of bathrooms (BATHRM) and the floor (STAIR). These variable where selected in order to define a first linear regression that can predict a price value (PRICE). It's expected that number of locals, bathroom and the dimension of the lot are in a certain way correlated and that each one will be positively correlated with the explained variable, while the floor is expected to be inversely proportional.

House features variables

Then I decided to introduce some qualitative variables in order to complete a more accurate estimation of the price equation; I intorduced the dummy variable (BOX) that take value 1 if there is a car parking lot and 0 otherwise. After that I introduced another dummy variable (FLAT) that takes value 1 if this is a flat and 0 otherwise and, in the end, another dummy called (ELEVTR) that takes value 1 if there is an elevator and 0 otherwise. I introduced this type of qualitative variables, with this value, because i think that the presence of an elevator, the garden or the kind of house can influence positivetely the price. To decide if a garden can be considered relevant i decided to count only gardens that are equals or bigger than 500 mq.

Distance variable

Selected the area, the variable and the point of interest, I calculated the three variable that i want to analyze deeply. I compared each evaluation's address with the point of interest I found before, obtaining three new variables that could analyze PRICE in given some distances. The first one is the variable DuomoDIST, the value of the distance between the observation and the dome, selected as symbol of the city centre. Then i found the variable VillaDIST, that evaluates the distance between the house and the Villa reale, and in the end I created the variable ParcoDIST as the distance between an house and a point in the center of the park, the viale Cavriga. With these variable i want to analyze which of this point of interest can influence the most the value of an house into this area.



Analysis

First of all we compute on our software, evIEWS, the regression line given our selected variable. As we can see in the successive image, in our regression line we can see a dependent explained variable, Yprice, that's composed of an intercept, C, and eleven explanatory, previously selected, variables.

Dependent Variable: YPRICE__X1000\$_
 Method: Least Squares
 Date: 12/13/19 Time: 23:09
 Sample: 1 105
 Included observations: 105

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-130.5362	88.06273	-1.482309	0.1416
X01ROOM	31.07876	19.72484	1.575616	0.1185
X02SQRDSIZE	1.398401	0.252362	5.541241	0.0000
X03BATHRM	98.93512	31.41696	3.149099	0.0022
X04STAIR	-12.43670	11.76293	-1.057279	0.2931
X05BOX	16.16428	42.89337	0.376848	0.7071
X06GARDN	-57.25473	36.32856	-1.576025	0.1184
X07FLAT	61.64013	53.83076	1.145073	0.2551
X08ELEVTR	114.6855	41.54213	2.760705	0.0069
X09DUOMODIST__X1000M_	-9.027023	23.51380	-0.383903	0.7019
X10PARCODIST__X1000M_	37.66374	25.75042	1.462646	0.1469
X11VILLADIST__X1000M_	-71.39982	34.45141	-2.072479	0.0410
R-squared	0.748395	Mean dependent var		451.1000
Adjusted R-squared	0.718635	S.D. dependent var		306.8430
S.E. of regression	162.7612	Akaike info criterion		13.12966
Sum squared resid	2463682.	Schwarz criterion		13.43297
Log likelihood	-677.3069	Hannan-Quinn criter.		13.25256
F-statistic	25.14793	Durbin-Watson stat		2.220082
Prob(F-statistic)	0.000000			

As we can see our line can explain very well the model, given the fact that we selected a lot of variables. Although we can observe an high R-squared value, 0.748395, analyzing the p-value of our explanatory variable, we can see that only a few of them can be considered relevant with a p-value under 0.05: the variable X02SQRDSIZE, X03BATHRM, X08ELEVTR and X11VILLADIST, so this can't be considered a good regression equation for YPRICE. That happens because our estimation method is the least squares that tend to maximize the R-squared, and, as we chose eleven variable, it's normal we get an high value. Even taking into account the adjusted value of R-squared we get a relatively high goodness of fit.

However can still be relevant running the ramsey reset test, to verify if our model is not correctly specified. As we can see in the two pictures below, the ramsey test runned with fitted values 1 is significant, even caputring some of the non linearity variables. In the ramsey test we can see that the fitted variables $FITTED^2$ have a probability equal to 0.8911.

As we can see also from the tables, we can check the F value to understand if our model is well specified, and it's confirmed by our value > 0.05 .

Ramsey RESET Test
Equation: UNTITLED
Omitted Variables: Squares of fitted values
Specification: YPRICE__X1000\$ C X01ROOM X02SQRDSIZE
X03BATHRM X04STAIR X05BOX X06GARDN X07FLAT X08ELEVTR
X09DUOMODIST__X1000M_ X10PARCODIST__X1000M_
X11VILLADIST__X1000M_

	Value	df	Probability
t-statistic	0.137294	92	0.8911
F-statistic	0.018850	(1, 92)	0.8911
Likelihood ratio	0.021511	1	0.8834

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	504.6760	1	504.6760
Restricted SSR	2463682.	93	26491.20
Unrestricted SSR	2463177.	92	26773.66

LR test summary:

	Value
Restricted LogL	-677.3069
Unrestricted LogL	-677.2962

Unrestricted Test Equation:
Dependent Variable: YPRICE__X1000\$_
Method: Least Squares
Date: 12/13/19 Time: 23:51
Sample: 1 105
Included observations: 105

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-138.5965	106.2281	-1.304707	0.1952
X01ROOM	31.24414	19.86627	1.572723	0.1192
X02SQRDSIZE	1.460153	0.516395	2.827590	0.0058
X03BATHRM	101.3848	36.27532	2.794869	0.0063
X04STAIR	-12.87897	12.25639	-1.050796	0.2961
X05BOX	15.46815	43.41851	0.356257	0.7225
X06GARDN	-59.39178	39.70040	-1.496000	0.1381
X07FLAT	63.36189	55.55102	1.140607	0.2570
X08ELEVTR	116.5019	43.80827	2.659358	0.0092
X09DUOMODIST__X1000M_	-8.762098	23.71745	-0.369437	0.7127
X10PARCODIST__X1000M_	38.92649	27.47265	1.416918	0.1599
X11VILLADIST__X1000M_	-73.51181	37.89710	-1.939774	0.0555
FITTED^2	-2.76E-05	0.000201	-0.137294	0.8911

R-squared	0.748447	Mean dependent var	451.1000
Adjusted R-squared	0.715635	S.D. dependent var	306.8430
S.E. of regression	163.6266	Akaike info criterion	13.14850
Sum squared resid	2463177.	Schwarz criterion	13.47708
Log likelihood	-677.2962	Hannan-Quinn criter.	13.28165
F-statistic	22.81064	Durbin-Watson stat	2.222623
Prob(F-statistic)	0.000000		

In any case this can't be considered a good linear regression because there are not so many significant values that can explain the dependent variable.

In order to find a good regression line that could allow us to test our thesis, we transformed all the variables in which there are no 0 results in order to being able to do the logarithm, to see if in this case we can observe some relevant variables. The only variables not considered as logarithms are the dummy variables for type of house, elevator, garden and box, in addition to the floor variable, that has many 0 observations.

Dependent Variable: LOGY
Method: Least Squares
Date: 12/14/19 Time: 00:51
Sample: 1 105
Included observations: 105

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.649422	0.382119	9.550495	0.0000
LOGX01	0.185910	0.160966	1.154966	0.2511
LOGX02	0.243944	0.076561	3.186281	0.0020
LOGX03	0.646163	0.134411	4.807369	0.0000
X04STAIR	-0.018271	0.025147	-0.726587	0.4693
X05BOX	0.182013	0.096314	1.889779	0.0619
X06GARDN	-0.052422	0.080827	-0.648577	0.5182
X07FLAT	0.273783	0.115130	2.378030	0.0195
X08ELEVTR	0.207827	0.092303	2.251589	0.0267
LOGX09	-0.106691	0.055028	-1.938858	0.0556
LOGX10	0.252629	0.173973	1.452112	0.1498
LOGX11	-0.397096	0.089834	-4.420317	0.0000

R-squared	0.792211	Mean dependent var	5.866163
Adjusted R-squared	0.767633	S.D. dependent var	0.737783
S.E. of regression	0.355644	Akaike info criterion	0.877436
Sum squared resid	11.76287	Schwarz criterion	1.180746
Log likelihood	-34.06541	Hannan-Quinn criter.	1.000343
F-statistic	32.23352	Durbin-Watson stat	1.921724
Prob(F-statistic)	0.000000		

As we can see this model can explain much better the regression, by the fact that there are more variable in wich p-value take value below 0.05. The considered variables are C, LOGX02, LOGX03, X07FLAT, X08FLAT AND LOGX11, with value below 0.05 and considering also variables X05BOX and LOGX09 adjusting the p-value after rising the critical point to 0.06.

Dependent Variable: LOGY
Method: Least Squares
Date: 12/14/19 Time: 01:49
Sample: 1 105
Included observations: 105

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.615544	0.379814	9.519256	0.0000
LOGX02	0.275090	0.070567	3.898254	0.0002
LOGX03	0.703378	0.110155	6.385355	0.0000
X05BOX	0.222475	0.091708	2.425900	0.0171
X07FLAT	0.290506	0.109050	2.663972	0.0091
X08ELEVTR	0.173076	0.088019	1.966353	0.0521
LOGX11	-0.423847	0.083729	-5.062142	0.0000
LOGX09	-0.118744	0.051345	-2.312655	0.0229
LOGX10	0.275406	0.167910	1.640198	0.1042

R-squared	0.787286	Mean dependent var	5.866163
Adjusted R-squared	0.769560	S.D. dependent var	0.737783
S.E. of regression	0.354167	Akaike info criterion	0.843718
Sum squared resid	12.04167	Schwarz criterion	1.071201
Log likelihood	-35.29522	Hannan-Quinn criter.	0.935899
F-statistic	44.41372	Durbin-Watson stat	1.930574
Prob(F-statistic)	0.000000		

Using only the relevant variables, and the park one due to our thesis, we can estimate a new regression line of significative variables.

As we can see, in this model, all the explanatory variables are significant except one, the variable X08ELEVTR, that's slightly over the critical p-value. In this model the R-squared has a relatively high value, 0.787286, so the regression is well represented by this estimators.

Now we run again the Ramsey reset test with fitted value 1 to see if there are non linear parts not captured by the equation. Our ramsey test has a F value of 1.380457 < 4.6052, so we do not reject null hypothesis δ_1 .

Ramsey RESET Test

Equation: UNTITLED

Omitted Variables: Squares of fitted values

Specification: LOGY C LOGX02 LOGX03 X05BOX X07FLAT X08ELEVTR
LOGX11 LOGX09 LOGX10

	Value	df	Probability
t-statistic	1.174929	95	0.2430
F-statistic	1.380457	(1, 95)	0.2430
Likelihood ratio	1.514789	1	0.2184

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	0.172473	1	0.172473
Restricted SSR	12.04167	96	0.125434
Unrestricted SSR	11.86920	95	0.124939

LR test summary:

	Value
Restricted LogL	-35.29522
Unrestricted LogL	-34.53783

Unrestricted Test Equation:

Dependent Variable: LOGY

Method: Least Squares

Date: 12/14/19 Time: 01:53

Sample: 1 105

Included observations: 105

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.442258	0.799239	5.558112	0.0000
LOGX02	0.576548	0.266066	2.166935	0.0327
LOGX03	1.494747	0.682459	2.190236	0.0310
X05BOX	0.450477	0.214558	2.099560	0.0384
X07FLAT	0.631173	0.309699	2.038016	0.0443
X08ELEVTR	0.340243	0.167212	2.034799	0.0447
LOGX11	-0.920824	0.431160	-2.135690	0.0353
LOGX09	-0.237064	0.112992	-2.098055	0.0386
LOGX10	0.612100	0.331967	1.843856	0.0683
FITTED^2	-0.095146	0.080980	-1.174929	0.2430

R-squared	0.790332	Mean dependent var	5.866163
Adjusted R-squared	0.770469	S.D. dependent var	0.737783
S.E. of regression	0.353467	Akaike info criterion	0.848340
Sum squared resid	11.86920	Schwarz criterion	1.101098
Log likelihood	-34.53783	Hannan-Quinn criter.	0.950762
F-statistic	39.78870	Durbin-Watson stat	1.953641
Prob(F-statistic)	0.000000		

Then we analyse some of the variables, to see if there is some correlation between them. First of all we analyzed the first group of variable that could probably be correlated to each other, the logarithms of the first three variable X01ROOM, X02SQRSIZE and X03BATH.

Covariance Analysis: Ordinary

Date: 12/14/19 Time: 01:59

Sample: 1 105

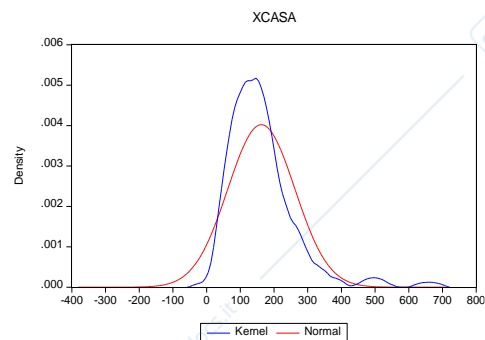
Included observations: 105

Correlation	LOGX01	LOGX02	LOGX03	LOGY
LOGX01	1.000000			
LOGX02	0.679181	1.000000		
LOGX03	0.769742	0.669404	1.000000	
LOGY	0.733304	0.677667	0.782087	1.000000

As we can see there is a strong correlation between LOGX01, LOGX02 and LOGX03, and between these variables and the logY one. This was expected as it was logical that rising the number of rooms or bathrooms would have increased also the square size of the houses. So we decided to compute this three variables into another one, a logXCASA variable, taking the value seen below, that could be used to draw a new regression line.

	LOGX01	LOGX02	LOGX03	LOGY
Mean	1.249633	4.872865	0.573458	5.866163
Median	1.386294	4.990433	0.693147	5.966147
Maximum	2.079442	6.476972	1.386294	7.313220
Minimum	0.000000	0.693147	0.000000	4.234107
Std. Dev.	0.396111	0.683637	0.457956	0.737783
Skewness	-0.530844	-2.056265	-0.085056	-0.262930
Kurtosis	3.496192	14.84769	1.712242	2.239014
Jarque-Bera Probability	6.008572 0.049574	688.1024 0.000000	7.381758 0.024950	3.743367 0.153864
Sum	131.2115	511.6508	60.21312	615.9472
Sum Sq. Dev.	16.31801	48.60543	21.81125	56.60962
Observations	105	105	105	105

New variable Xcasa data



So we computed our definitive regression line after all our consideration. With this regression line we would be able to understand and analyze the effect of the distance on the price of the houses. As we can see in our model all our variables are significant; the only one that's not considered significant is the logarithm of the distance from the park. We can finally assume that, given the fact that this variable never was relevant in our model, that does not really explain the price of the houses.

Dependent Variable: LOGY
 Method: Least Squares
 Date: 12/14/19 Time: 10:19
 Sample: 1 105
 Included observations: 105

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.138529	0.281171	11.16234	0.0000
LOGXCASA	0.336022	0.031129	10.79440	0.0000
X05BOX	0.216595	0.092642	2.337977	0.0214
X07FLAT	0.293614	0.110299	2.661976	0.0091
X08ELEVTR	0.211387	0.088629	2.385071	0.0190
LOGX09	-0.135427	0.051995	-2.604601	0.0106
LOGX10	0.222445	0.169776	1.310224	0.1932
LOGX11	-0.385133	0.084924	-4.535043	0.0000
R-squared	0.778220	Mean dependent var	5.866163	
Adjusted R-squared	0.762215	S.D. dependent var	0.737783	
S.E. of regression	0.359766	Akaike info criterion	0.866407	
Sum squared resid	12.55487	Schwarz criterion	1.068613	
Log likelihood	-37.48635	Hannan-Quinn criter.	0.948345	
F-statistic	48.62437	Durbin-Watson stat	1.783517	
Prob(F-statistic)	0.000000			

$$\text{LogY} = \beta_0 + \beta_1 \text{LogXCASA} + \beta_2 \text{X05BOX} + \beta_3 \text{X07FLAT} + \beta_4 \text{X08ELEVTR} + \beta_5 \text{LogX09DUOMO} + \beta_6 \text{LogX10PARCO} + \beta_7 \text{LogX11VILLA}$$

$$\text{LogY} = 3.138 + 0.336 \text{LogXCASA} + 0.216 \text{X05BOX} + 0.293 \text{X07FLAT} + 0.211 \text{X08ELEVTR} - 0.135 \text{LogX09DUOMO} + 0.222 \text{LogX10PARCO} - 0.385 \text{LogX11VILLA}$$

Our R-squared is 0.778220 so this mean that the regression line is less precise than the previous ones, due to the fact that we eliminated some of our variables. The r-squared value is given by subtracting to 1 the unexplained variance divided by the total variance, and introducing new variable can alterate the value of the coefficient of determination, given the fact that we are using OLS method that tend to maximize R-squared. However it's not so important that our value is less than before, because our regression can be considered still well explained.

Again we run a Reset test with our final equation, to analyze the effect of non linear component on our regression. As before we choose a ramsey test with fitted value \wedge^2 and \wedge^3 , so we can explain most of non linear effect of our variables. The two fitted variable $\text{FITTED}\wedge^2$ and $\text{FITTED}\wedge^3$ take value respectively 3.911882 and -0.224479, and both of them can be considered significant due to their p-value. However as we can see in our table, the value of our ramsey test can't be considered correctly specified because $0.0136 < 0.05$, probably due to some non linearity of our parameters.

Ramsey RESET Test				
Equation: EQ02BUONA				
Omitted Variables: Powers of fitted values from 2 to 3				
Specification: LOGY C LOGXCASA X05BOX X07FLAT X08ELEVTR LOGX09 LOGX10 LOGX11				
	Value	df	Probability	
F-statistic	4.500484	(2, 95)	0.0136	
Likelihood ratio	9.504949	2	0.0086	
F-test summary:				
	Sum of Sq.	df	Mean Squares	
Test SSR	1.086586	2	0.543293	
Restricted SSR	12.55487	97	0.129432	
Unrestricted SSR	11.46829	95	0.120719	
LR test summary:				
	Value			
Restricted LogL	-37.48635			
Unrestricted LogL	-32.73388			
Unrestricted Test Equation:				
Dependent Variable: LOGY				
Method: Least Squares				
Date: 12/14/19 Time: 10:28				
Sample: 1 105				
Included observations: 105				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-24.81293	9.320915	-2.662070	0.0091
LOGXCASA	-7.219323	2.525926	-2.858090	0.0052
X05BOX	-4.677069	1.637812	-2.855682	0.0053
X07FLAT	-6.263546	2.196154	-2.852052	0.0053
X08ELEVTR	-4.524232	1.584158	-2.855922	0.0053
LOGX09	2.910757	1.019036	2.856383	0.0053
LOGX10	-4.791385	1.686717	-2.840657	0.0055
LOGX11	8.263470	2.894738	2.854652	0.0053
FITTED \wedge^2	3.911882	1.304210	2.999426	0.0035
FITTED \wedge^3	-0.224479	0.074861	-2.998608	0.0035
R-squared	0.797414	Mean dependent var		5.866163
Adjusted R-squared	0.778222	S.D. dependent var		0.737783
S.E. of regression	0.347446	Akaike info criterion		0.813979
Sum squared resid	11.46829	Schwarz criterion		1.066737
Log likelihood	-32.73388	Hannan-Quinn criter.		0.916401
F-statistic	41.54864	Durbin-Watson stat		1.907157
Prob(F-statistic)	0.000000			

Heteroskedasticity test

Once we find out which is our best regression model we started computing some heteroskedasticity tests for our last found equation, in order to find if the variance of the explained variable is constant or not.

Breusch-pagan-godfrey test

First of all we run the Breusch-Pagan-Godfrey test to see if there is heteroskedasticity. As we can see in our table, the probability of our F statistic is below 0.05, so we can assume that there is not heteroskedasticity, as we do not reject the null hypothesis.

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	2.235374	Prob. F(7,97)	0.0377
Obs*R-squared	14.58531	Prob. Chi-Square(7)	0.0417
Scaled explained SS	11.77989	Prob. Chi-Square(7)	0.1080

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/14/19 Time: 10:30
Sample: 1 105
Included observations: 105

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.338955	0.124123	2.730800	0.0075
LOGXCASA	-0.031989	0.013742	-2.327811	0.0220
X05BOX	-0.019548	0.040897	-0.477976	0.6337
X07FLAT	-0.034050	0.048692	-0.699298	0.4860
X08ELEVTR	-0.030943	0.039125	-0.790859	0.4310
LOGX09	0.012235	0.022953	0.533041	0.5952
LOGX10	0.049979	0.074948	0.666847	0.5065
LOGX11	-0.076873	0.037490	-2.050518	0.0430

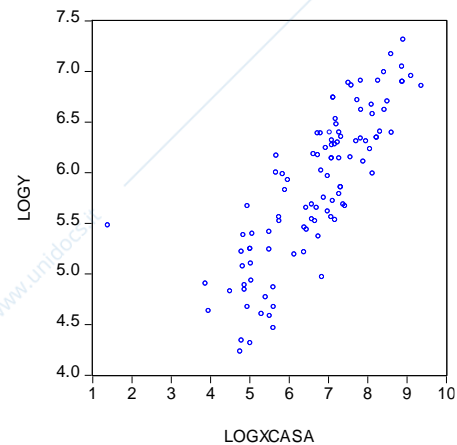
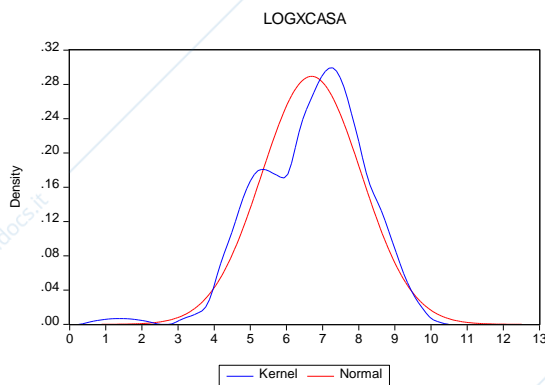
R-squared	0.138908	Mean dependent var	0.119570
Adjusted R-squared	0.076767	S.D. dependent var	0.165290
S.E. of regression	0.158819	Akaike info criterion	-0.768973
Sum squared resid	2.446673	Schwarz criterion	-0.566767
Log likelihood	48.37109	Hannan-Quinn criter.	-0.687035
F-statistic	2.235374	Durbin-Watson stat	1.923442
Prob(F-statistic)	0.037661		

Conclusion

Now we can start analyzing the variables included in our model to find out some information, highlighting the effects of the last three ones based on the distance.

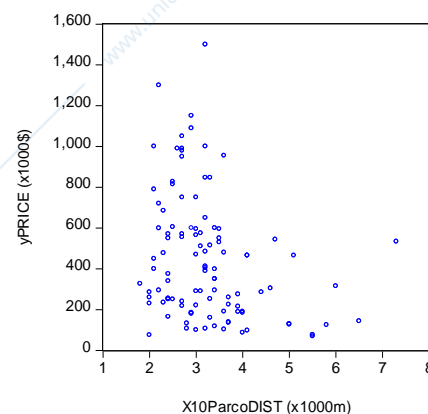
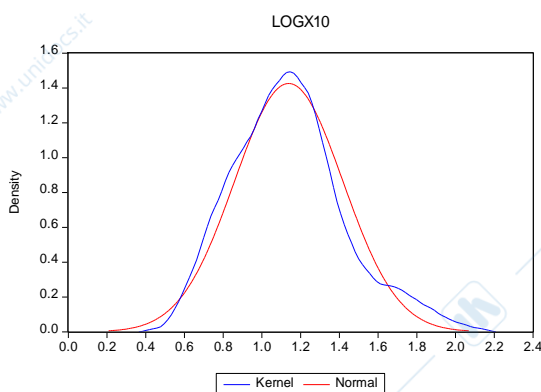
$$\text{LogY} = 3.138 + 0.336 \text{LogXCASA} + 0.216 \text{XO5BOX} + 0.293 \text{XO7FLAT} + 0.211 \text{XO8ELEVTR} - 0.135 \text{LogXO9DUOMO} + 0.222 \text{LogX10PARCO} - 0.385 \text{LogX11VILLA}$$

LogXCASA

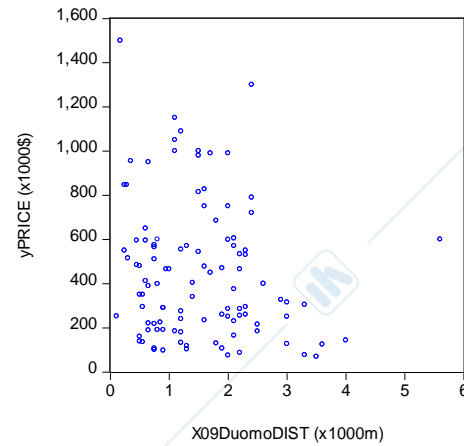
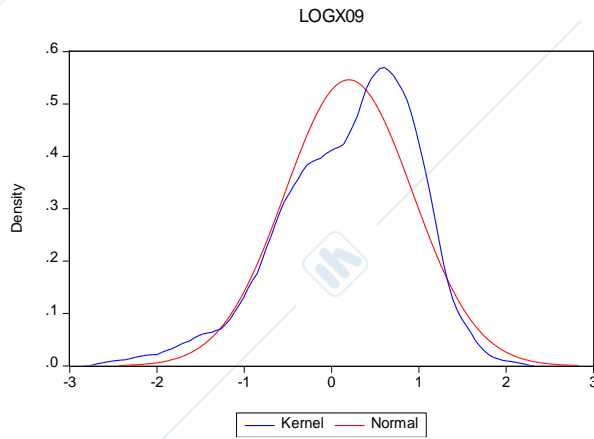


As we can see our first variable, a group composed of three house data variable (number of rooms, size and bathrooms) is directly correlated to the price of the houses. As expected, an increase of the value of this variables is correspond to an increase of the price of the houses. With more precision we can say that, assuming elasticity between the explained and explanatory variable, an increase of 1% of logXCASA will increase the price of the house of the 0.336% given all the other variables. In addition as we can see in the second graph, the distribution of this variables is similar to a normal distribution.

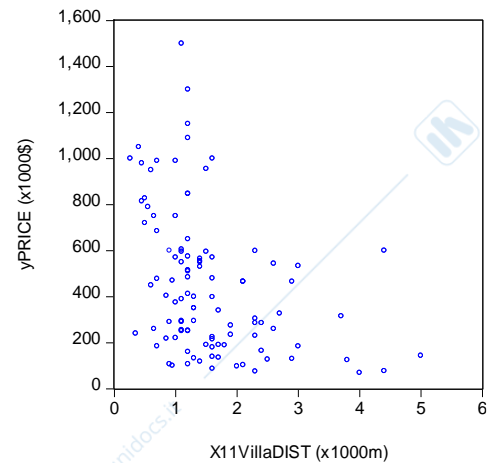
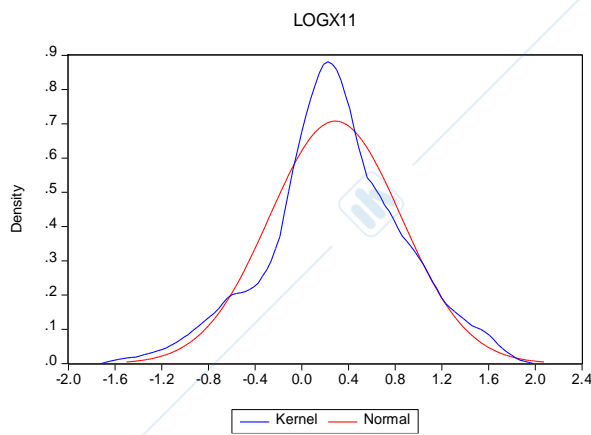
Distance variables



As we previously stated, our variable that measure the distance between the houses and the park is not relevant, so we can start stating that this is not highly correlated to the price. This variable doesn't explain well the direction of the price model so we can't make any assumption or prevision based on this variable. However we can say that, based on the data we collected, an increase of 1% of distance can increase the price of the house of 0.222% given all the other variables.



The distance between our houses and the dome is relevant. As we can come closer to the church, we see a rising of the prices, as previously stated by our regression. Our variable is also significant, so we can say that our results can confirm our theories about distance and price. Given our regression line, we can say that an increase of 1% of the distance will decrease the price of 0.135%



In the end we can see how the distance from the Villa reale is significant too. We can confirm our hypothesis, as we can see into the scatter plot graph that the prices rise as we came closer to the point of interest. From our regression we can say that an increase of the distance from the point will be converted into an increase of 0.385 of the price.