MWSUG 2016 - Paper RF09

Quantifying the relative importance of crime rate on Housing prices

Aigul Mukanova, University of Cincinnati, Cincinnati, OH

ABSTRACT

As a part of Urban and Regional Economics class at University of Cincinnati students were required to have a small empirical project for hedonic house price model. Using SAS® software author attempts to measure an effect of the crime on the house value in Ohio.

INTRODUCTION

Crime is a non-market good that comes with the bought house. It directly impacts everyone by changing the neighborhood quality. Even intuitively we know that in neighborhoods with high crime rate the house would be sold for a lower price compared to the similar house which could be sold for a higher price in a safe neighborhood.

I want to build a hedonic price model in which the explanatory variables have information about crime in the neighborhood and therefore I could explicitly quantify the effect of crime rates on the house value.

EMPIRICAL PROJECT

Data

Ohio housing data set for the project was provided by Professor David Brasington, University of Cincinnati.

Since the hedonic pricing method should include not only house characteristic but also the characteristics of the surrounding neighborhood and public goods and buyer characteristics (Brasington and Hite, 2008), I needed to add those characteristics into the model in order to have more robust results.

The house price was chosen as a dependent variable. One of the challenges for me was to choose the independent variables. The provided data had 434 variables. Since I wanted to focus on a crime effect, and I also needed to use environmental, neighborhood and buyers characteristics, I chose only those variables that I thought had an effect on house value.

For house description I took basic features that were mentioned on every real estate website: number of bedrooms, bathes, square footage of the house, age, lot size, previous sale price and additional structures like fireplace, garage, pool, air conditioner, deck.

For the environmental quality I chose the total air pollution in census block group the house is in, measured in short tons (a short ton is 2000 pounds).

Neighborhood quality: As a proxy for school quality I used the expenditure for pupil (*squality2)* (Brasington, 1999). Also added the variable for the average commute time to work in minutes, because, I think, time to work is also important factor when we choose a place to live.

Tax rate, density and ethnic heterogeneity could also affect the price of the house. Therefore, these variables were also added into the model.

In the data provided there was not any information about particular buyers. According to Tiebout (1956) hypothesis, households sort themselves into local jurisdictions based on their preferences for public goods and services given their budget constraint. The households reveal their preferences by moving into matching neighborhoods, “voting with their feet” (Tiebout, 1956). And therefore, in order to get information about potential buyers, for my model I used demographics of people who already have chosen a particular neighborhood assuming that buyers have similarities in income level, level of education, marital status, whether they have kids or not. Percentage of unemployed labor force in the neighborhood tells us if this neighborhood attracts people with a stable job or without.

Some variables, like tax rate and unemployment, could feature both neighborhood quality and buyers

characteristics.

As it was mentioned earlier, I wanted to investigate the crime effect on the house value. I choose all available variables that were explanatory about crime in the neighborhood. *Totalcrime* is a variable for grand total of actual offenses in police district per thousands of persons in police district. *Clearratio3* is a variable for the percent of actual offenses in police district cleared by arrest. *Policeratio3* is a variable for the number of police officers per 1000 residents in police district. *Policeemprat3* – total number of police agency employees per 1000 residents in police district. I also choose variable *droprte_sd*, dropout rate for schools. I justified it with next logic: the higher the dropout rate, then the crime rate is also higher.

At the beginning I had 127009 observations in the data. After removing the missing values of following variables: the house price (hp_cbg), unique house identifier code (j) and census block group for each house (blkgrp), I ended up with 120658 observations. Below is the SAS code for cleaning the data.

```
data project.two;
set project.one;
if j=. then delete;
if blkgrp=. then delete;
if hp_cbg=. then delete;
run;
```

Once the variables were chosen required changes were made.

I took logarithms of house price (hp_cbg) and average income of households (Avginc_cbg). Now these variables (*lHp and lincome)* are normalized and are easier to interpret. I also reduced levels of education for percentage of persons 25 years or older in census block group into three groups: low education (*loweduc,* group with less than a Bachelor's degree education), medium education (*mediumeduc*, group with a Bachelor's degree) and high education (*higheduc,* group with either Master's, Doctorate, or professional school degree). Have created new variable agehouse2, the square of the actual house age. This action was necessary to capture very old houses the prices for which only grow with time, for example castles built in 1800s. For easier coding I renamed many variables from their original names in the data set. Overall, there were 33 independent variables.

MODEL

The model was chosen as a single linear regression:

Log(house price) = f(House Characteristics, Environment characteristics, Neighborhoods characteristics, Buyer Characteristics) + e

Below is the code of regression.

```
proc reg data=project1.hedonic2 plots=none;
model lHp = bedrooms fullbath partbath buildingsqft lotsize agehouse2
agehouse Prevsaleamt
            garageD onestoryD airD fireplace deckD poolD
            AirQuality squality2 Commute_cbg taxrate density_cbg Unemp_cbg
ethnics
            lincome havekids married separated loweduc mediumeduc higheduc
            totalcrime clearratio3 policeratio3 Policeemprat3 droprte_sd
    ;
      run;
```

The overall F-test shows that this regression is significant:

F-value = 15390.4 with p-value <.0001 at any significance level.

R^2= 0.8569

After running the regression I checked for heteroscedasticity. The White test detected the heteroscedasticity in the variance of residuals.

The graphical method using a plot statement in the REG procedure also indicates a mild heteroscedasticity. As it visible on the Figure 1 the pattern of the data points is getting a little narrower on the right end.

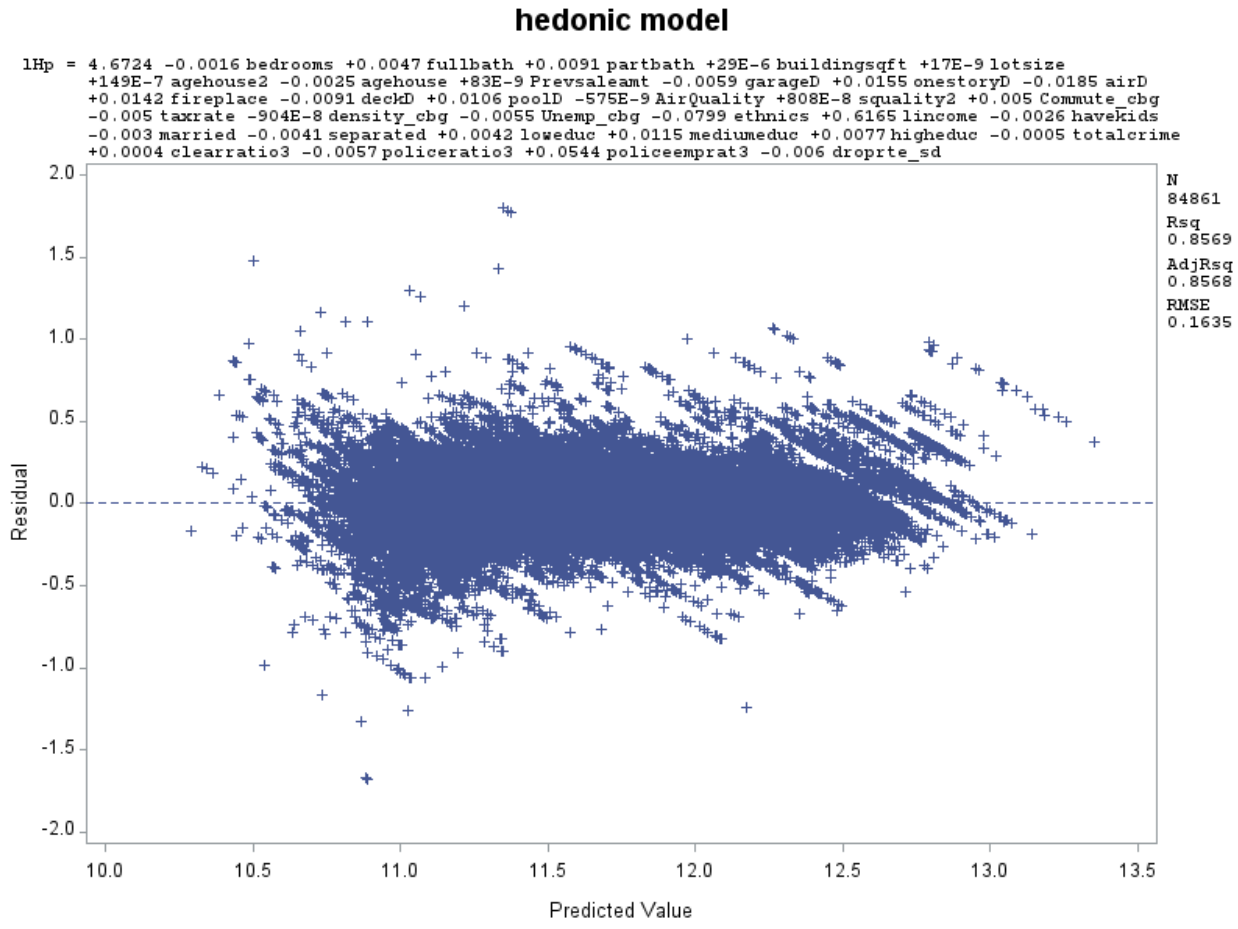

Figure 1.Residuals and Predicted Values. The plot for this paper was generated using SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Below is the code of testing for heteroscedasticity:

White test:

```
proc model data=project1.hedonic2;
parms b0 b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16 b17 b18 b19
b20 b21 b22 b23 b24 b25 b26 b27 b28 b29 b30
b31 b32;
lHp=b0 +b1*bedrooms +b2*fullbath +b3*partbath +b4*buildingsqft +b5*lotsize
+b6*agehouse2 +b7*agehouse +b8*Prevsaleamt
```

```
+b9*garageD +b10*onestoryD +b11*airD +b12*fireplace +b13*deckD +b14*poolD
+b15*AirQuality
+b16*squality2 +b17*Commute_cbg +b18*taxrate +b19*density_cbg
+b20*Unemp_cbg+b21*ethnics +b22*lincome +b23*havekids
+b24*married +b25*separated +b26*loweduc +b27*mediumeduc +b28*higheduc
+b29*totalcrime +b30*clearratio3 +b31*policeratio3
+b32*Policeemprat3 + b33*droprte_sd;
fit lHp/white;

run;
```

Graphical method:

```
     proc reg data=project1.hedonic2 plots=none;
model lHp = bedrooms fullbath partbath buildingsqft lotsize agehouse2
agehouse Prevsaleamt
            garageD onestoryD airD fireplace deckD poolD
            AirQuality squality2 Commute_cbg taxrate density_cbg Unemp_cbg
ethnics
            lincome havekids married separated loweduc mediumeduc higheduc
            totalcrime clearratio3 policeratio3 Policeemprat3 droprte_sd
    ;
      plot r.*p.;
      run;
      quit;
```

I fixed for heteroscedasticity and included heteroscedastic consistent standard errors and p-values into interpretation of the results by using option /HCC with the MODEL statement.

The VIF and TOL options with the MODEL statement revealed collinear variables.

The variable *Policeemprat3,* total number of police agency employees per 1000 residents in police district, was removed from the model because it was strongly correlated with *policeratio3,* number of police officers per 1000 residents in police district. The removal of the variable did not affect the significance of the parameter estimates and overall F-test.

Pearson Correlation Coefficients, N = 108864 Prob > |r| under H0: Rho=0		
	policeratio3	**policeemprat3**
policeratio3	1.00000	0.98936 <.0001
policeemprat3	0.98936 <.0001	1.00000

Table 1. Correlation Coefficients. The output for this paper was generated using SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

As expected variables *agehouse* and *agehouse2* are collinear. Last and also expected, all education

variables have a very strong multicollinearity.

INTERPRETATION

As the results in the table 2 show, based on the estimates of the parameters, most of the variables in the model are statistically significant.

Since my interest lies in quantifying crime into a house value, I want to focus on explaining related parameter estimates.

B(Totalcrime): increase in totalcrime by 1 unit (extra crime per 1000 persons) is associated with a **-** 0.00049092*100 = -0.049 percent change in house price, all else constant. (se=0.00002168, t-value = -22.65, p-value <.0001)

B(clearratio3): increase in clearratio3 by 1 unit increases the house price by 0.050295 percent, holding all else equal. (se=0.00003605, t-value= 13.95, p-value <.0001)

B(policeratio3): increase in policeratio3 by 1 unit increases the house price by 0.102 percent, all else equal. (se=0.00004989, t-value= 20.50, p-value <.0001)

B(droprte_sd): increase in dropout rate by 1 unit decreases the house price by 0.635 percent, all else equal. (se=0.00018089, t-value= -35.12, p-value <.0001)

Table 2 contains parameter estimates for all variables used in the model.

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent		
						Standard Error	t Value	Pr > |t|
Intercept	1	4.67357	0.06708	69.67	<.0001	0.08186	57.10	<.0001
bedrooms	1	-0.00169	0.00091649	-1.84	0.0652	0.00125	-1.35	0.1776
fullbath	1	0.00460	0.00137	3.36	0.0008	0.00170	2.70	0.0069
partbath	1	0.00875	0.00134	6.51	<.0001	0.00146	6.01	<.0001
buildingsqft	1	0.00002979	0.00000138	21.57	<.0001	0.00000338	8.80	<.0001
lotsize	1	1.319526E-8	6.490871E-9	2.03	0.0421	7.763282E-9	1.70	0.0892
agehouse2	1	0.00001475	5.068501E-7	29.10	<.0001	5.816849E-7	25.36	<.0001
agehouse	1	-0.00249	0.00006974	-35.71	<.0001	0.00007750	-32.14	<.0001
Prevsaleamt	1	8.633783E-8	4.636989E-9	18.62	<.0001	8.18098E-9	10.55	<.0001
garageD	1	-0.00494	0.00127	-3.88	0.0001	0.00128	-3.86	0.0001
onestoryD	1	0.01432	0.00140	10.21	<.0001	0.00159	9.01	<.0001
airD	1	-0.02037	0.00147	-13.90	<.0001	0.00137	-14.82	<.0001
fireplace	1	0.01361	0.00120	11.32	<.0001	0.00136	10.02	<.0001
deckD	1	-0.00886	0.00199	-4.46	<.0001	0.00206	-4.30	<.0001
poolD	1	0.00968	0.00446	2.17	0.0301	0.00466	2.08	0.0379

Parameter Estimates								
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Heteroscedasticity Consistent		
						Standard Error	t Value	Pr > |t|
AirQuality	1	-5.31577E-7	2.24115E-7	-2.37	0.0177	2.372295E-7	-2.24	0.0250
squality2	1	0.00001042	6.634485E-7	15.70	<.0001	6.850641E-7	15.20	<.0001
Commute_cbg	1	0.00534	0.00015294	34.90	<.0001	0.00016792	31.79	<.0001
taxrate	1	-0.00521	0.00012854	-40.56	<.0001	0.00014167	-36.80	<.0001
density_cbg	1	-0.00000901	2.523108E-7	-35.71	<.0001	3.061717E-7	-29.43	<.0001
Unemp_cbg	1	-0.00549	0.00017828	-30.77	<.0001	0.00024666	-22.24	<.0001
ethnics	1	-0.08099	0.00647	-12.52	<.0001	0.00837	-9.67	<.0001
lincome	1	0.61683	0.00413	149.47	<.0001	0.00540	114.28	<.0001
havekids	1	-0.00261	0.00006545	-39.83	<.0001	0.00007781	-33.51	<.0001
married	1	-0.00299	0.00008505	-35.15	<.0001	0.00010733	-27.85	<.0001
separated	1	-0.00398	0.00016282	-24.47	<.0001	0.00020816	-19.14	<.0001
loweduc	1	0.00396	0.00050957	7.76	<.0001	0.00055519	7.13	<.0001
mediumeduc	1	0.01125	0.00054100	20.80	<.0001	0.00058878	19.11	<.0001
higheduc	1	0.00765	0.00034717	22.03	<.0001	0.00038468	19.88	<.0001
totalcrime	1	-0.00049092	0.00001547	-31.73	<.0001	0.00002168	-22.65	<.0001
clearratio3	1	0.00050295	0.00003526	14.27	<.0001	0.00003605	13.95	<.0001
policeratio3	1	0.00102	0.00004480	22.83	<.0001	0.00004989	20.50	<.0001
droprte_sd	1	-0.00635	0.00016502	-38.50	<.0001	0.00018089	-35.12	<.0001

Table 2. Parameter Estimates. The output for this paper was generated using SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

CONCLUSION

As the results show crime indeed affects the housing prices. All estimates made sense to me, as the crime grows, prices of houses fall. As the police increases the clearing of the offences by arrest and the number of police grows, which means the police presence is effective and the neighborhood is becoming safer, the prices for houses increase. And last, assuming that dropping out from a school might increase the crime rate, the house in the area with high dropout rate will be sold for less than in a lower school dropout rate area.

It is easy to notice that the changes are very small. It is probably because we cannot find full information about crime in neighborhood because not all crimes are reported. Every person looks at crime differently. If a potential buyer was a victim of a crime, the importance of safe neighborhood would be more significant.

The results for the other variables: the better quality of a school positively affects the house price, the increase in air pollution has negative effect on housing prices. An interesting result is for commute time: with increasing the commute time the house price actually increases, all else equal. A longer commute time suggests the house is located in the suburbs where the housing prices are usually higher.

Tax rate and unemployment variables are negatively related to the house price which was expected.

This work is one of the many applications of hedonic pricing method discussed in class. Hedonic pricing method can be used for the capitalization of taxes and public services, the measurement of relative importance and demand of non-market goods, the evaluation of policy alternatives, real estate application and more.

REFERENCES

1. Brasington, David M. and Hite, Diane, A Mixed Index Approach to Identifying Hedonic Price Models (May 2008). Regional Science and Urban Economics, Vol.38, No. 3, 2008. Available at SSRN: http://ssrn.com/abstract=1162012
2. David Brasington (*1999*) Which Measures of School Quality Does the Housing Market Value? Journal of Real Estate Research: 1999, Vol. 18, No. 3, pp. 395-413.
3. Tiebout, C. (1956), "A Pure Theory of Local Expenditures", *Journal of Political Economy* 64 (5): 416–424.

ACKNOWLEDGMENTS

The author would like to thank Professor David Brasington for generously sharing his Ohio housing data set.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Aigul Mukanova
Student in MA Applied Economics
Lindner College of Business,
University of Cincinnati, Cincinnati, OH
mukanoat@mail.uc.edu