Spring 3-26-2010

Determinants of Housing Prices in Pittsylvania County

Jason Hall
Lynchburg College

Determinants of Housing Prices in Pittsylvania County

Defended by Jason Hall

3/26/2010

Messerschmidt, chair ______________________

Schnur, committee ______________________

Prinzinger, committee ______________________

Determinants of Housing Prices in Pittsylvania County

Jason Hall

Senior Research Project

Submitted in partial fulfillment of the graduation requirements for the Economics major

School of Business and Economics

December 2009

Abstract

The purpose of this research is to determine the implicit factors influencing housing prices in Pittsylvania County, a rural county in Southside Virginia. Currently in Pittsylvania County there is a debate over whether or not to mine what is believed to be the largest deposit of uranium in the United States, containing approximately 119 million pounds of uranium ore at an estimated value of $7 billion to $10 billion. The Virginia Coal and Energy Commission issued its final approval for the National Academy of Sciences to study the debate from a socio-economic standpoint. The $1.2 million study is currently taking place on Coles Hill Farm in Chatham, VA and is expected to last 18 months. In addition to the uranium issue, much of the previous work on housing valuation models has been done in urban areas, so it is of interest to see which factors are important in a rural community as opposed to a metropolitan area.

The sample consists of 163 transactions obtained from the Navica Multiple Listing Service system for the date range of September 1st, 2008 through August 31st, 2009. A hedonic pricing model was used to estimate the regressions because of the heterogeneous nature of housing market. The hedonic model allows us to measure the marginal effect of a one unit change in any of the independent variables on the dependant variables. Continuous and dummy variables were used in the regressions. The continuous variables estimated were: acres of the property, age of the house, age squared, number of bathrooms, number of bedrooms, square footage of the basement, number of days the house was on the market, miles away from Cole Hill Farm, and square footage of the house. The dummy variables that were estimated include: whether or not the house had central air, if the basement was finished or not, if the house was stick built or not, and location dummies for Chatham High School, Dan River High School, and Tunstall High School.

The major findings were that across the board location within Pittsylvania County does not seem to matter, at least based on the four high school districts. The presence of the current uranium situation also doesn't seem to have any significant influence on prices in the current time period. The most significant variables are the physical characteristics that make up the house and property itself such as acres, age, age squared, bedrooms, bathrooms, basement square footage, house square footage, and whether the house was stick built or not.

Tables of Contents

I. Introduction

This research intends to determine exactly what the important factors are influencing housing prices in Pittsylvania County, Virginia. Extensive research has been done on factors influencing housing prices in urban areas like Atlanta, Chicago, New Haven, and Toronto, but hedonic models are not typically robust, or able to be applied to other geographical areas and still give an accurate estimate. To accomplish this we must look at variables that encompass physical characteristics of the structure, lot characteristics or size, and neighborhood factors influenced by location.

Current market conditions after the 2007 housing market collapse have left an adverse affect on the market that is currently being felt by potential sellers, banks, and realtors. This makes the time period of this research a unique time because what was important before the collapse may not be important now. This leads us back to the robustness issue again because cross sectional data only tells us what is going on in the market at a particular time, and events like the housing burst lead researchers to need to calculate new equations. There is no way to tell the actual affects of the housing bubble burst on the market based on the model due to the cross sectional nature and not having a pre-bubble analysis to compare it to, but it still needs to be noted that market conditions are constantly changing, and this will change the estimated coefficients.

The debate over whether or not the Coles Hill Farm uranium deposit should be mined continues to grow day by day, especially with the national attention the issue is getting from TIME Magazine, The Economist, and the CBS Evening News with Katie Couric. A 1982 study done by the Coal and Energy Commission found that mining the uranium was safe, but as commercial interest and uranium prices diminished a moratorium was placed on the mining. Rising demand for uranium has placed new interest in mining the $7 billion to $10 billion deposit as energy prices continue to look more and more volatile every day (The Cole Hill Progress, 2009). With all of the health concerns and future income concerns for area farmers, the hedonic pricing model can be used to determine if property and housing value fall closer the proximity to the mine.

Not only can this research and model help determine the socio-economic cost of mining uranium, but it can also help determine how efficient the market is working. The right investor could use a model to take advantage of an underpriced property on the market. Realtors could use this research to help their clients determine fair market value for their houses. Tax policy makers could also use this model to determine how much tax revenue will be collected based on the appraised value of the houses in the county.

II. Literature Review

There have been a significant number of studies done on housing markets in the United States. While these studies explore many of the same variables they all attempt to explain something slightly different that is going on in a particular model. Houses are heterogeneous goods with heterogeneous consumers so each market has its own unique characteristics requiring its own model to be specified in order to explain how even the most basic structural characteristics can affect real estate values. Obviously a swimming pool is going to be a much greater desire in sunny California than rural Southside Virginia, or a basement may hold more value in Kansas due to the protection it brings from tornadoes. The type of model typically used in explaining housing prices is called a hedonic model because it uses measures of the quality of a product as independent variables instead of measures of the market for that product, or things like quantity demanded, income, etc (Studenmund 2006).

Two of the pioneers of this research were Kain and Quigley with their research being done in the early 1970's. Their results showed that qualitative measures affected housing price about as much quantitative measures such as square footage, acres, and number of bedrooms. While research to measure qualitative bundles had been done previously by economists, Ridker and Henning (1967), they used aggregate census tract data to measure air pollution, school quality, and accessibility to downtown and these affects on housing prices. The issue here is with the deficiency in using aggregate census data and lack of importance it places on each individual observation.

Kain and Quigley (1970) sought to correct this by having interviewers rate the physical characteristics of each house in the sample in a number of areas such as condition of the walls and flooring on a scale of one to five. They also had building inspectors look at aspects of the outside of the property and adjacent properties. They were to note any adverse factors such as high noise levels, smoke, or heavy traffic. The result was that qualitative measures of the actual house itself had a significant bearing on price, but so did the quality of adjacent houses and environmental issues (Kain and Quigley 1970).

While qualitative measures are important, quantitative measures can't be ignored. Probably the most obvious impacts on housing value would seem to be the square footage of the house, how many bedrooms and bathrooms there are, how much land is around the house, and whether or not there is a basement or not. Other features that usually add value include the types of heating and air conditioning, as well as access to public utilities. A seven year sample done on the New Haven Metropolitan area, by Grether and Mieszkowski (1973), gave conclusive evidence that the size of the house plays a significant role in the selling price. Not only did size matter, but so did how that size was divided up, meaning more bedrooms and bathrooms increased housing value. Also of interest was that a one-car garage appeared to be worth $800 and a two-car garage worth another $500. In the study the estimates indicate that an additional 1000 square feet of land is worth about $800. While all of these variables were significant at the .05 level, not all of variables performed as expected. Among the notable variables that were either insignificant or had the wrong sign were: basement, more than two car garages, and storm windows. Grether and Mieszkowski noted the importance of data reporting and its affect on their results as in the case with the basement variable. The four reported conditions were: no basement, a partial basement, full basement, and "yes". With no indication of actual size and having to distinguish between a full basement and an answer of "yes" the researcher must take certain liberties in attempting to use a variable like this. (Grether and Mieszkowski, 1973).

Grether and Mieszkowski (1978) continued their research further to see how nonresidential land uses affected the prices of adjacent housing. Nonresidential land uses include industrial, commercial, high-density dwellings, and highways. The nonresidential land use problem is another attempt to measure neighborhood quality, but as it relates to specific externalities. Expectations are that a desirable externality has a positive impact on real estate values while a negative externality will do the opposite. The findings were in general that most of the zoning or externality issues measured were so localized that they did not seem to have a significant impact on any housing or rent values aside from the ones that were located right next door (Grether and Mieszkowski, 1978). This is different from the findings of Kain and Quigley in their aforementioned article, where they found that the presence of industrial and commercial uses did have a statistically significant negative impact upon rent values and single-family house values.

A later study was done by Mieskowski and Saper in 1977 to determine if airport noise, the obvious byproduct of close proximity to an airport, had effects on urban property values in the Toronto area. The authors point out the important policy implications of such studies. For example, the results could be used to determine if compensation is due for homeowners who are impacted by noise or even aid in the design of new airports that have less of a social impact. What this means is that the social disamenities could be weighed against the cost of construction of a bigger more internalized airport that reduces the noise level for surrounding houses. In dealing with heterogeneous consumers it must be noted that every individual has a different aversion to any externality in question. Some people are less affected by noise than others. For Mieskowski and Saper there is systematic evidence that houses located in various noise contours do sell at a discount, although the magnitude of the discount varies (Mieskowski and Saper, 1977). A recent study by Cohen and Coughlin confirmed these findings on their sample done in the Atlanta metropolitan area. They found that a reduction in noise levels in proximity to the airport caused a spike in housing prices (Cohen and Coughlin, 2009).

While no formal studies could be found on adverse effects of uranium mining on housing prices, the best option is to look at externalities like airport noise and air pollution. Ridker and Henning (1967) found that in their sample of the St. Louis metropolitan area that there was a statistically significant willingness to pay for cleaner air. They obtained an air pollution coefficient that was used to show the marginal value of a change in air pollution levels on an individual's willingness to pay a premium for cleaner air.

To reiterate and elaborate on a point mentioned earlier, special attention needs to be paid to the make-up of the housing market itself. Houses are heterogeneous in nature, so aside from mass produced mobile homes there is almost no identical house on the market and even then there are lot and location variables to differentiate the houses. As also mentioned earlier, hedonic regression analysis is used in housing markets to measure the marginal effects of an additional unit of a particular variable on the price of a house. Sirmans, MacPherson, and Zietz (2005) completed a study analyzing the heterogeneous nature of housing markets and consumers by looking at how the quantity of bedrooms affects housing prices in forty different studies. Twenty-one of the studies found a positive, significant relationship. Nine of them found a negative, significant relationship, and the remaining ten had no statistical significance. This shows that consumers do not demand identical characteristics at the same level in different geographical areas (Zietz, Zietz, & Sirmans, 2007). The demand and supply functions are very difficult to identify empirically for any housing market. The general belief is that for hedonic pricing models the underlying supply curve is vertical due to the fixed nature of housing stock at any given time (Rosen, 1974) and (Zietz, Zietz, and Sirmans, 2007). The fixed nature of housing stock implies that supply variables have little to do with housing prices in a hedonic regression.

It's also worth mentioning the importance of choosing the appropriate functional form as well as choosing the appropriate sample selection. Many of the early articles on valuing implicit housing values used aggregate census tract data (Kain and Quigley) and (Linneman). While these data sets can be rich in information they can also leave room for bias when compared to samples derived from individual transactions. Linneman (1978) points out that little attention is usually paid to selection sample although it also has the potential for large amounts of bias if you are not careful. Linneman points out that samples that only include owner-occupied housing can lead to selection bias due to the fact that lower income groups tend to rent instead of buy. However, his analysis was for Chicago and Los Angeles and the income distributions there are not like in Pittsylvania County where there is no true "upper-class". This issue can be further developed after looking at the descriptive statistics for Pittsylvania County.

Many studies have already been covered that analyze how the quality of a neighborhood affects housing prices using such variables as crime rates, distance from the central business district, and SOL test scores, but there are other ways of implicitly measuring these effects as well. First it is worth analyzing how the time a house stays on the market affects price. One would expect that the longer a house stays on the market the more likely those potential buyers become inclined to believe that something is wrong with the property, and that in turn requires the seller to sell at a discount (Taylor 1999). Theoretically this is known as a trade-off in economics. There is a trade-off between selling price and the time-on-the-market. However, one recent study found that there was no direct correlation between actual time-on-the-market and selling price, but that when you include list price and use a two stage process for empirical analysis there does become a trade-off between list price and time-on-the-market (Anglin, Rutherford, and Springer, 2003). Possible explanations for why selling price wasn't significant could come from what Taylor found in his research. The first reason is that the consumer could just be one of the first people to look at the house despite its time-on-the-market, and the second is that the list price could just simply be way too high for the attributes that the house has to offer the consumer.

While location can be measured in a variety of ways such as high school districts, presence of public services, or distance from central business district, there is no consensus on the best way to measure this. Chiodo, Hernandez-Murillo, and Owyang (2003) sought to expand on the pioneering work of Tiebout (1956), which found significant benefits of the presence of public services on housing prices. The work of Chiodo, Hernandez-Murillo, and Owyang sought to determine if parents are willing to pay a premium to move into the best school districts in the St. Louis area based on standardized testing scores. Their belief was that in the areas with the best testing scores competition would increase for the limited supply of housing and thus increase the selling prices. They found that there is a strong positive correlation between school quality and housing prices in their linear model where an increase in test scores by half a standard deviation results in an 11% increase in housing price.

The other variable that implicitly contains some degree of quality is the age variable. Naturally, it is expected that houses depreciate in value over time. This doesn't mean that the house is a low quality house if it is older, but that a buyer wants a discount for the fact that the older a house is the more chance that something in the house can break down and need repairing. Sirmans, MacPherson, and Zietz (2005) looked at more than one hundred past studies on the topic of housing prices and found that across the board age, had a negative impact upon the price.

III. Theoretical Model and Methodology

The regression technique used for this analysis is ordinary least squares (OLS), which calculates ^βs so as to minimize the sum of the squared residuals. OLS was chosen due to its ease and practical use in fitting a linear regression line through the sample of housing data that has been obtained. In effect OLS is minimizing $\sum (Y_i - Y^{\wedge}_i)^2$, or minimizing the squared difference between the actual housing prices and the estimated housing prices.

The type of OLS model used for the housing data in this study is called a hedonic pricing model. A hedonic pricing model uses measures of the qualities of a product as independent variables instead of measures of the market for that product. Hedonic models are most useful in examples like the housing market because of the heterogeneous nature of houses. Hedonic models have been used by many of the pioneers of the housing research including Grether and Mieszkowski, Kain and Quigley, Linneman, and Ihlanfeldt and Martinez-Vasquez. All of the regressions ran on the housing data for Pittsylvania County are hedonic pricing models with the exception of the one including variables for the different types of financing and its affect on price.

The specified model is $Y=\beta_0+\beta_1S+\beta_2L+\beta_3N+€$, where Y is the dependant variable for housing price. The structural variables are designated by the variable, S, which includes age, square footage, basement square footage, percentage of basement finished, number of bedrooms, number of bathrooms, days on the market, inclusion of central air, and whether the house is stick built or a mobile home. Certainly there are many more possibilities of structural variables, but these are the ones that were included for the initial regression. The β_2 is the coefficient for the lot variables which is marked as L. The only lot variable taken from the sample was one for the measurement of acreage. Finally neighborhood effects are measured by the variable N. The different generalized "neighborhoods" were broken down by dummy variables for Chatham High School, Dan River High School, Tunstall High School, and Gretna High School, with Gretna being the default variable. Finally, the last variable measuring any sort of neighborhood effect is a distance variables used to measure the mileage from each particular observation to the proposed uranium mine on Cole Hill Farm in Chatham, VA. The reason for the inclusion of more structural variables rather than neighborhood variables comes from the fact that Pittsylvania County is an extremely rural county with an estimated population of 61,123 people as of 2008 (U.S. Census Bureau) spread out over roughly 971 square miles of land. This means there are only about 63 persons for each square mile of land. The general difference of doing housing valuation in a rural area as opposed to an urban area is that there is not as significant a drop off from the different ends of the wealth spectrum with the majority of these homes belonging to middle class families.

Looking first at the independent variables for the structural characteristics the majority of the hypothesized signs will have a positive bearing on the overall price. Theory suggests that the more bedrooms and bathrooms a house has that the price will increase because of the extra utility presented by the additional space and less crowding. This goes hand in hand with the overall square footage variable in that more overall square footage provides less crowding in all rooms such as the kitchen, dining room, den, living room, etc. We expect the addition of a basement in general to add value to the house, and as that basement's square footage increases so will value. Perhaps an equally important point about a basement is whether or not it is finished.

The size variables are important, but the measurements of structural quality must also be considered. The addition of a finished basement, whether partial or full, adds to the overall quality of the basement and presents additional living space, although this square footage is not measured in the overall square footage of the house. In previous studies the addition of central air was included as a measurement of quality that provided information about how up-to-date the house was, so with many of the houses in the sample being of significant age the expectation is that central air would add to the overall value of the house. One important thing to note about Pittsylvania County is the significant number of single wide and double wide mobile homes, as well as modular homes, which doesn't come as a shock for an area where median annual household income is approximately $38,000. A dummy variable for whether or not the house was stick built or some form of mobile/modular home was included with the belief that a stick built house would be significantly higher in value.

Due to the nature of the housing market at the time period the sample was taken, September 1st, 2008 to August 31st 2009, a proxy variable for overall quality was included that measured the days the property was listed on the market. For the Dan River Region of the Virginia Association of Realtors from 4th quarter of 2008 through the 3rd quarter of 2009 there were 622 sales as opposed to 708 from the year before, showing a decline of 12.15% in home sales. However, at the same time the median selling price for the time period the data was taken was actually up to $96,804 from $88,796 the year before, a 9.02% increase. Under normal market conditions the hypothesis would be that housing prices will fall the longer the house stays on the market because the buyer begins to perceive that something must be wrong with the property. The current burst in housing prices does make this a unique market so it is possible that the variable may not even be significant, but theory still suggests that it is worth including in the equation. Previous studies using qualitative measures to determine housing prices have used realtor assessments of the properties on a scale of one to five with five being excellent quality and one being very poor. Such data was not available for Pittsylvania County and would certainly seem to allow a great deal of bias into the sample. For these reasons the proxy variable for days on the market was used.

The variable for age of the structure was included under the theoretical belief that the older the house is the less money it will bring on the market. It is worth arguing that age may be one of the most important variables in distinguishing two similar properties and their values on the market. Two structurally identical properties will sell at different prices if one is 15 years old and the other 30 years old. The 30 year old house should sell at a discount, ceteris paribus, due to the nature of housing and the problems that tend to arise with age. Such problems include updating bathrooms, switching to central air, or installing a new heating system.

While there certainly appears to be a trade off between age and price, theory suggests it should be at a diminishing rate the older the house becomes. In some instances, like in a historical district, the price actually may increase with age, but that is not the hypothesis for this regression. In order to estimate the diminishing nature of age and price a variable for age squared was included with the hypothesis that the coefficient is positive.

A second model will be specified that needs to be independent of the first one because it is measuring what is in essence a different dependant variable. The second model will look at estimating the actual selling price on the market by using the important relevant variables from the first model, but also including a series of dummy variables regarding different types of financing and loans. This is better suited by specifying a new model because the first model is really designed to look at housing price determinants explained by traits the house has itself and in its location, where the second looks at the impact on type of financing chosen by the buyer on actual price at closing.

The different dummy variables used in the model include conven, fha, other, va, and vhda. All of these variables are being compared to houses paid for in cash. Conven stands for a conventional fixed rate mortgage and houses in the sample were given a value of "1" if they were financed this way and a "0" if otherwise. This is the most basic loan available and typically comes in 15, 20, and 30-year terms with a rate that is locked in for the duration of the loan. These loans also now require only small down payments, but because of the convenience of being able to receive payment in cash the expectation is that any loan will make the house sell for more. The hypothesis is that conven will have a positive impact on price.

The Federal Housing Administration (FHA) can insure loans made by private lenders resulting in the issuance of an FHA loan. This means that if the buyer defaults on the loan then the lender can still receive payment from the FHA. These loans are usually low to middle size loans that are ideal for first time home buyers because of the low down payment requirement of roughly 3%-5%. The hypothesis here is still that houses financed through and FHA loan will sell at a premium to cash financed houses.

A second type of government loan in addition to the FHA loan is the VA loan. This loan is a long-term low or no down payment loan that is insured by the Department of Veterans Affairs and can only be obtained by qualified military veterans. Once again the hypothesis is that VA financed houses will sell at a higher price than those done by cash. The final type of loan as indicated in the multiple listing service databases for our sample includes a VHDA loan. This stands for the Virginia Housing Development Authority and actually can include any of the types of loans already mentioned as well as many more. The only difference is that the loan is obtained through the Virginia Housing Development Authority. This could lead to complications and bias in the coefficients and overall significance level and it one of the main problems in econometrics because the researcher can only make decisions based on the information provided in the reporting of the data set. If we knew the specific types of VHDA loans we could eliminate that variable and label the other observations as conventional, fha, or va.

IV. Data Analysis and Results

The results for the first model, PRICE=f(ACRES, AGE, AGESQ, AIR, BATH, BED, BMSQ, FINBM, SQFT, STICK, DAYS, MILES, CHS, DRHS, THS), provide a great deal of insight into what is happening in the

Pittsylvania County housing market. Not all of the expectations for all of the dependant variables panned out as expected, but a look at Table 1 below may help explain the situation better.

Table 1

Dependent Variable: PRICE
Sample: 1 163
Included observations: 163

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-95546.08	25425.54	-3.757878	0.0002
ACRES	1671.697	447.5044	3.735599	0.0003
AGE	-2226.910	428.7945	-5.193420	0.0000
AGESQ	18.37882	4.718015	3.895457	0.0001
AIR	2548.242	10908.56	0.233600	0.8156
BATH	20791.84	7233.448	2.874402	0.0046
BED	15031.00	6363.824	2.361944	0.0195
BMSQ	29.18918	5.605585	5.207160	0.0000
CHS	3202.129	10765.66	0.297439	0.7666
DAYS	17.11739	25.66267	0.667015	0.5058
DRHS	849.9413	13368.88	0.063576	0.9494
FINBM	-15427.48	15565.53	-0.991131	0.3233
MILES	204.2813	611.9772	0.333805	0.7390
SQFT	56.20127	8.248860	6.813217	0.0000
STICK	64935.28	10106.02	6.425409	0.0000
THS	5037.944	13277.38	0.379438	0.7049

R-squared	0.839028	Mean dependent var	135355.3
Adjusted R-squared	0.822602	S.D. dependent var	87934.42
S.E. of regression	37036.77	Akaike info criterion	23.97021
Sum squared resid	2.02E+11	Schwarz criterion	24.27389
Log likelihood	-1937.572	F-statistic	51.08010
Durbin-Watson stat	1.671355	Prob(F-statistic)	0.000000

All of the highlighted variables in Figure 1 indicate the variables that are significant. The far right column, prob, stands for the p-value of the independent variables which is calculated based on the t-statistic. A larger t-statistic results in a higher degree of significance, or a lower p-value. The p-values being reported in Table 1 are the values for a two-tailed significance test, but all of the variables are being hypothesized as a one-tailed test, so the p-value must be divided in half. After doing this all of the significant highlighted variables are significant at even the .01 level.

The overall adjusted R^2 is 0.822602, which means that the model explains roughly 82% of the variation in the dependant variable. This is a desirable level for the adjusted R^2 since this is a cross sectional data set and they are by nature more difficult to obtain high R^2 values. While a high adjusted R^2 is good it is not the only important aspect of the results. It means little if the direction of the coefficient of a significant variable is opposite from the expected sign. This is where theory combines with statistical analysis to determine the goodness of fit of the model.

The constant term, or the y-intercept, carries a value of -95546.08 and is significant at the .01 level. The constant term must not be suppressed because it serves the purpose of absorbing any nonzero mean that the observations of the error term may have (Studenmund). The constant term must not be omitted because otherwise the regression would be forced through the origin and this is almost never the case in practical application. In order to achieve OLS estimates with minimum variance the constant term must be included because forcing the regression through the origin would provide inaccurate slope coefficients. The constant term also remains in place to absorb the impact of any omitted variables.

The independent variable number of acres of the property has a positive coefficient with a value of 1671.697. This means that an additional acre of land appears to be worth $1671.70. A look at the descriptive statistics in Figure 4 shows that the minimum lot size for the sample was .2 acres while the maximum lot was 66 acres. The mean lot size was 2.94 acres meaning that the average lot is worth $4915 in Pittsylvania County. This doesn't necessarily mean that this is what a lot of land with no house would sell for, only what it would sell for being bundled with a house.

As mentioned earlier, age should impact selling price at a diminishing rate, so for age we must look at the combined effects of age and agesq. The coefficient for age is -2226.91 and agesq's coefficient value is 18.38. The best way to explain how these two variables work together is to look at a few examples from the sample and show how much an extra year depreciates the value of a house. The newest house from the sample was one year, and 1 x -2226.91=-2226.91 for the age variable. On the other hand to calculate agesq we must take 1^2 x 18.38=18.38. The combined effect of the two variables would be -2226.91+18.38=-2208.53, meaning that the first year depreciates a house's price by $2208.53. The median age of all the houses in the sample is 22 years old. Using the same calculation we can figure out the depreciation of the 22nd year of a house in Pittsylvania County. Age is calculated as 22 x -

2226.91=-48892.02. Agesq's value is 22^2 x 18.38= 8895.92. The combined effect shows that 22 years has a negative effect of $39996.10. Finally taking the oldest house in the sample, 109 years old, we can do the same calculations. Age is calculated as 109 x -2226.91= -242733.19. Agesq equals 109^2 x 18.38= 218372.78. The combined effect actually shows age as having a negative effect of $24361.19. This shows that due to the inclusion of the agesq variable that at a certain age an additional year actually adds value to the house. This is not common in the sample because a 109 year old house is certainly considered an outlier.

Aside from age, square footage of the house probably come to mind as being the most obvious variable for influencing the selling price despite the location or market. It is only natural to assume that the larger the house, other things equal, the more it will cost. The square feet of the houses in the sample only included above ground rooms that were finished, so this excludes any garage or basement space. For this particular sample the coefficient for sqft is 56.2. This means that an additional square foot of above grade living space in Pittsylvania County adds $56.20 to the overall value of the house. The t-statistic for this variable is the largest of any of the variables in the model with a value of 6.81, also meaning it has the smallest p-value of 0.0000. Needless to say excluding this variable would lead to a significant drop in the overall explanatory power of the model. Table A-3 shows that the smallest house in the sample was only a mere 738 square feet while the largest was 4423. This means that the range of square footage value would be $41,475.60 to $248,572.60. The most accurate representation of the area is to look at the median square footage of 1512 with a value of $89,974.40. Keep in mind this doesn't mean that we can simply add all the variables value together because the constant term is an extremely negative value that must be cancelled out.

The next two variables should be mentioned together and are also consistent with the sqft variable. The number of bedrooms, bed, has a coefficient of 15,031 implying that an additional bedroom adds $15,031 to the house. The number of bathrooms, bath, has a larger coefficient of 20791.84 or a value of $20,791.84. It is reasonable to accept these results based on theory because the median number of bathrooms was 2 and the median number of bedrooms was 3. The addition of the second bathrooms carries more importance to the average family of 3 than the addition of a third bedroom. The t-statistic for bathrooms was 2.87 while bedrooms were only 2.36, but each is still highly significant at the .05 level.

The basement square footage variable was added after initially treating the addition of a basement as a dummy variable. There were complications with treating it as a dummy variable because listing for type of basement came in a variety of forms such as: partial unfinished, partial finished, full unfinished, full finished, cellar, or none. It would be very inaccurate to treat a fully finished basement in the same manner as a 200 square foot cellar. For this reason basement was measured by the reported square feet in the Navica Database, and a separate variable was added for whether the basement was finished or not based on the percentage of the basement that was finished. There were some uncertainties in the way the data was reported about the make up of the basement and some assumptions had to be made, but as a whole it was more accurate than putting everything into one category.

The coefficient for basement square footage is 29.19 making an additional square foot worth $29.19. Obviously this isn't as high as an actual square foot of finishing above ground living space, but it is still a significant coefficient for a sample with a median basement size of 728 square feet. The p-value, like with sqft, is also 0.0000 making this a variable that is crucial to the model. On the other hand the significance for the percentage of the basement that is finished, finbm, was neither significant nor had the appropriate sign. The lack of significance could be due to the fact that only 14.2% of the total basement square feet in the entire sample were finished, so most of the value was already being captured by bmsq. A look at the correlation matrix in Table A-4 shows that finbm and bmsq do share some degree of multicollinearity with the value of .51. After calculating the variance inflation factor for finbm there doesn't appear to be a severe degree of multicollinearity because the value is only 1.96. Only if this were greater than 5 would there be significant problems in the model.

The final significant variable is the stick variable. Once again this is a dummy that measures the difference in a value that a stick built house has from a mobile or modular built house. The coefficient for stick is 64935.28. This shows that being stick built adds $64,935.28 to the selling price and does a great deal to help overcome the constant terms value of -95546.08. The t-statistic for stick is 6.425 making this the second most significant variable to the overall equation. This variable is very important when looking at Table A-3 and seeing that the sample consisted of nearly 20% mobile or modular homes. This just goes to show the nature of the housing market in the county and that there is a great demand for lower income housing in the area.

The biggest surprise from the results is that not a single one of the location dummy variables, chs, drhs, or ths came back significant. Most people in the real estate business preach the importance of location, location, location, but at least by breaking down the location by dummy variables based on the four high schools with Gretna High School being the control variable it doesn't seem to make a difference. The only real sound hypothesis was that Tunstall High School, ths, would be significantly higher in value that Gretna High School. It isn't necessarily surprising that the remaining two variables came back as insignificant. The majority of all of the upper priced houses of $300,000 came from the Tunstall school district. For this reason and its close proximity to Danville, the central business district, the hypothesis was made. However, once looking at the observations included in the Tunstall district there becomes some evidence of why problems occurred. First over 37% of all the houses sold were sold in Tunstall, and only a handful of these were greater than $300,000 in value. Tunstall is also made up of much of the lower valued houses which offset the value of the other houses. Because of the county's large size this becomes a difficult variable to measure. One way to measure this based on previous research would have been to measure distance from the central business district, but the job market is so depressed in Danville right now and unemployment is so high that the dummy variable method was chosen. Later in the paper we will look at an alternative method of measuring location significance specifically within the Tunstall High School district.

Another surprise from the results was that the days on the market, our proxy for quality, came back as insignificant with the coefficient also having the wrong sign. The minimum amount of time on the market was one day while the maximum was 542 days. The average time on the market was 159.57 days. There is no clear answer for why this variable wasn't significant other than perhaps the unique situation of the current recovery after the housing bubble burst of 2007. The VIF for days was only 1.06 so there are no signs of severe multicolinearity at all. Perhaps many of the people that were planning on placing their house on the market were holding out because of fear and resulting in a change in market conditions.

The last insignificant variable is the one for measuring distance from the Cole Hill Farm uranium deposit. This isn't necessarily surprising given that so far it is only speculation as to whether or not the uranium will be mined. The results of this could change if the decision is made to mine and people's fears become a reality. There was also some degree of difficulty in determining how to measure this variable. The only feasible way to measure it was to use Mapquest and measure the shortest distance from each house to 1040 Coles Rd. Chatham, VA 24531. The biggest concern is that once the uranium is mined it will get in the Bannister River that leads into Franklin County and contaminate the water supply. If the deposit is also open pit mined then the radiation can get into the air and blow downstream contaminating the air supply.

While it certainly doesn't appear that the presence of the uranium mine has any effect on prices, the model does not allow for any way of measuring lost sales due to fears of all the negatives that come with living near the mine. One realtor that was interviewed on the topic and is very familiar with the area said that she had lost numerous sales due to the health concerns and fears of the uranium devaluing housing prices in the future[1]. Some of these lost sales came on houses that were up to 15 miles away for the proposed uranium mine. This opens up an interesting avenue for further research into the topic, by looking at a time series analysis to determine if the presence of the revived uranium debate leads to a decline in sales.

The multicollinearity problem has been mentioned to some degree already, but needs to be further investigated and explained to full understand why it can lead to problems. Perfect multicollinearity is not the issue here because none of the variables are measuring the exact same thing, however imperfect multicollinearity should be considered. Imperfect multicollinearity is defined as a linear functional relationship between two or more independent variables that is so strong that it can significantly affect the estimation of the coefficients of the variables. The correlation matrix in Table A-4 and the VIF table in Table A-5 can each be used to detect signs of multicollinearity.

The highest degree of multicollinearity is seen between age and agesq with a value of 0.936565. This is no surprise as they are both measuring something similar. What this high value means is that age increases then agesq also increases at a very similar rate, which is true since agesq is simply the squared value of age. This is confirmed by the variance inflation factor value of 10.87 for age and 9.09 for agesq. There is also a fairly strong correlation between the number of bedrooms and number of bathrooms, which is suggested by theory. These two variables share a correlation value of 0.618064. All that this means is that the results show that as the number of bedrooms increase so too will the number of

[1] Janet Hogan with Aaron Johnson Auction and Realty Co. 2009

bathrooms in the house. Sqft shares a moderate amount of correlation with both bed and bath with values of 0.615790 and 0.715370. Obviously as the number of bedrooms and bathrooms increases the overall amount of square feet of the house will also increase. Problems of multicollinearity also exist with the ths dummy variable because of the VIF value of 4.9 and it shares a high correlation of 0.541169 and -0.536792 with miles and chs respectively. All of the previous examples of multicollinearity have been supported by theory, but it can also be random within a particular sample and that could be the case here. The association between ths and miles means that the farther away from the Cole Hill uranium deposit we move the closer the location to Tunstall, which based on the majority of the observations being in the southern part of the county would make this true. This is basically saying the same things as the -.53679 correlation with chs because the farther away from Chatham the more likely that the house will be located in Tunstall since that's where the majority of the observations are.

Aside from multicollinearity, heteroskedasticity is one of the most significant problems faced in cross-sectional data analysis like this study. Heteroskedasticity is a violation of Classical Assumption V for obtaining the best linear unbiased estimator. Classical Assumption V states that the observations of the error term are drawn from a distribution that has a constant variance (Studenmund). One cause of heteroskedasticity generally comes from improper functional form, such as an omitted variable or some other specification error. The Ramset RESET test is a formal test that can be ran to alert the existence of specification error. The Ramsey RESET results in Table A-6 in the appendix show that there does appear to be some form of specification error in the model based on the extremely low probability value of the f-statistic of 0.000056. The downfall of the Ramsey RESET is that it does nothing to show what the specification error is. Specification error does not necessarily mean there is heteroskedasticity, so to measure this one formal test is the White Heteroskedasticity Test. The results of this test are presented in Table A-7 in the Appendix and fortunately there does not appear to be any signs of heteroskedasticity in the model. The probability value of the f-statistic from the White Test is 0.224710, significantly higher than the trigger value of .05. The presence of heteroskedasticity causes OLS to no longer be the minimum-variance estimator, and leads to bias in the standard errors of the estimated coefficients, which means unreliable estimates from the model.

Another way to measure the overall fit of the model is to look at the f-statistic and p-value for the overall equation. What this tells us is that the overall combined slopes of all the coefficients are significantly from zero. The F-statistic is equal to 51.08 with a p-value of 0.0000. A second use of the F-statistic is to look at the chow test in Figure 2. The chow test is used to tell whether or not the independent variables that are important in observations 1-81 are just as important as the ones in observations 82-163. To calculate this, the observations were sorted in ascending order by price. The results of the chow test show a p-value of 0.0000, meaning that there is a significant difference in the important coefficients between the two sub-samples. This is not surprising because the lower priced houses are demanded by lower income groups who can't afford the same traits that are demanded by someone with more income. Running a separate regression on observations 1-81 shows that the significant variables at the alpha .1 level are: age, agesq, sqft, bmsq, and stick. The significant variables on the second observation subsample include: acres, age, agesq, bmsq, bath, bed, sqft, stick, and chs. This shows that the upper-middle class person will value all of the same variables as the lower income person, but also will desire more bedrooms, bathrooms, and acres of land. The only surprise is high significance of living in Chatham for the second subsample.

After evaluating everything that has been discussed the model of best fit is shown in Table 2 below:

Table 2

Dependent Variable: PRICE
Sample: 1 163
Included observations: 163

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-77235.68	16901.57	-4.569735	0.0000
ACRES	1749.446	431.0006	4.059035	0.0001
AGE	-2345.937	406.1920	-5.775439	0.0000
AGESQ	18.94398	4.519816	4.191316	0.0000
BATH	18339.33	6467.489	2.835618	0.0052
BED	13398.78	5777.984	2.318936	0.0217
BMSQ	26.91362	5.045986	5.333668	0.0000
SQFT	59.57292	7.354095	8.100646	0.0000
STICK	65246.51	9770.592	6.677846	0.0000

R-squared	0.836854	Mean dependent var	135355.3
Adjusted R-squared	0.828379	S.D. dependent var	87934.42
S.E. of regression	36428.72	Akaike info criterion	23.89773
Sum squared resid	2.04E+11	Schwarz criterion	24.06855
Log likelihood	-1938.665	F-statistic	98.74269
Durbin-Watson stat	1.690111	Prob(F-statistic)	0.000000

The purpose was to drop what were seemingly irrelevant variables and only focus on the ones that mattered. The model in Figure 10 includes seven additional degrees of freedom from dropping these extra variables. The adjusted R^2 increased from 0.822602 to 0.828379, but the major increase is seen in the increase in the overall F-statistic from 51.08010 to 98.74269. Another important factor to look at is a decrease in the Akaikie info criterion from 23.97021 to 23.89773 and a decrease from the Schwarz criterion from 24.27389 to 24.06855. These two criteria are used to determine whether or not specification errors such as omitted or irrelevant variables exist. There is nc way to tell what the specific errors are, but these criteria can simply help determine the better fit comparing two models side by side. The lower the value between the two is the most desirable.

As mentioned earlier, due to the apparent lack of significance of location based on high schoo district in the county another model was specified looking only at the observations within the Tunstall I

School district. The expert opinion of most realtors in the area is that due to higher amounts of wealth the part of the county near and in Tunstall most of the housing there tends to sell at a premium to the ı of the county, so this is the basis for this regression. The variable ths was replaced with dummy varia for the elementary schools that lead up to Tunstall High School. The variables include bros for Brosvil Elementary School and twin for Twin Springs Elementary school. The default location was Stony Mill Elementary School. This was a two-tailed test because the hypothesized signs of the coefficients wer unknown. The model specified was price=f(acres, age, agesq, bath, bed, bmsq, sqft, stick, bros, twin) The results can be seen in Figure 11. This model appeared to yield a very good fit with an adjusted R 0.925257 and an f-statistic of 74.03710. The problem is that while all of the variables from the model (best fit are significant at the .01 level, once again neither of the location variables are significant. At le based on school districts within Pittsylvania County there doesn't appear to be any area that will sell a premium relative to the rest of the county.

The final regression ran comes from the second model explained at the end of the Theoretica Model and Methodology section of this paper. The purpose was to see whether the type of loan financing on a house caused it to sell at a premium to cash financing, as theory would suggest. The specified model was: price=f(acres, age, agesq, bath, bed, bmsq, sqft, stick, conven, fha, other, va, vh This shows that the other variables specified were taken from the model of best fit in Table A-10. Th results for the financing model can be seen in the appendix in Table A-12. As expected based on the adjusted R^2, the f-statistic, and the expected signs and significance of the variables from the first mo(this model also appears to be a good fit. However, looking specifically at the financing dummy variał none of them are significant at any reasonable level except for vhda. This variable is significant at th level for a one-tailed test with a p-value of 0.0958. The coefficient shows that a house financed by a vhda loan should sell for an additional $24,285.87

V. Conclusions and Policy Implications

The housing market in Pittsylvania County does not appear quite as differentiated as first anticipated, at least based on the high school districts in the county. However, theory suggests that si of the property, the house, and the age should play a huge role in determining price, and that is exactl what the results show. There are still opportunities for further research on this topic still to be done immediately and into the future.

There is still reason to believe days on the market certainly plays into the selling price, but it is bit more complicated to measure than previously attempted by just adding days on the market into a model with just selling price. Days on the market really affects the discount the seller gives from their price to the selling price. Without question a $400,000 house on the market for two years is going to s for more than a $100,000 house that is only on the market for two months, but there should also be a much deeper discount offered for the former. It would be interesting to see how a model running the difference between listing and selling price against days on the market to study the overall significanc especially in a market like the one that has been seen as of late.

There is certainly plenty of research left to be done on the effects of uranium mining on housir prices, but also on sales in general. While price may not be affected until the question of mining in Pittsylvania County is an afterthought, sales are being affected all of the time. Once there are enough observations a time-series analysis could be used to measure any drop off in sales once the presence mining came into question.

Finally, the model of best fit can serve as a useful tool for realtors and appraisers alike in help their clients appropriately value their houses. It would take a lot of the guess work out of the job and g them a reliable model that could also help owners determine what the individual aspects of their hous are worth, or how an additional bedroom or bath would increase the house's value. This model can al be useful to tax assessors and policy makers for determining housing values and how a change in property taxes will increase or decrease revenues.

VI. Appendix

Table A-1

Dependent Variable: PRICE
Sample: 1 163
Included observations: 163

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-95546.08	25425.54	-3.757878	0.0002
ACRES	1671.697	447.5044	3.735599	0.0003
AGE	-2226.910	428.7945	-5.193420	0.0000
AGESQ	18.37882	4.718015	3.895457	0.0001
AIR	2548.242	10908.56	0.233600	0.8156
BATH	20791.84	7233.448	2.874402	0.0046
BED	15031.00	6363.824	2.361944	0.0195
BMSQ	29.18918	5.605585	5.207160	0.0000
CHS	3202.129	10765.66	0.297439	0.7666
DAYS	17.11739	25.66267	0.667015	0.5058
DRHS	849.9413	13368.88	0.063576	0.9494
FINBM	-15427.48	15565.53	-0.991131	0.3233
MILES	204.2813	611.9772	0.333805	0.7390
SQFT	56.20127	8.248860	6.813217	0.0000
STICK	64935.28	10106.02	6.425409	0.0000
THS	5037.944	13277.38	0.379438	0.7049

R-squared	0.839028	Mean dependent var	135355.3
Adjusted R-squared	0.822602	S.D. dependent var	87934.42
S.E. of regression	37036.77	Akaike info criterion	23.97021
Sum squared resid	2.02E+11	Schwarz criterion	24.27389
Log likelihood	-1937.572	F-statistic	51.08010
Durbin-Watson stat	1.671355	Prob(F-statistic)	0.000000

Table A-2

Chow Breakpoint Test: 82

F-statistic	4.549118	Probability	0.000000
Log likelihood ratio	72.02523	Probability	0.000000

Table A-3

	PRICE	ACRES	AGE	AGESQ	BATH	BED
Mean	135355.3	2.936669	27.52761	1256.853	2.033742	3.15337
Median	110000.0	1.000000	22.00000	484.0000	2.000000	3.00000
Maximum	459900.0	66.00000	109.0000	11881.00	5.500000	6.00000
Minimum	23000.00	0.200000	1.000000	1.000000	1.000000	2.00000
Std. Dev.	87934.42	7.013545	22.40902	1886.864	0.772060	0.68116
Skewness	1.535725	6.223374	1.065550	2.697447	1.269350	0.74107
Kurtosis	5.526598	48.95618	3.880515	11.96027	6.022796	4.7630
Jarque-Bera	107.4272	15395.98	36.11054	742.9497	105.8298	36.029
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
Sum	22062914	478.6770	4487.000	204867.0	331.5000	514.00
Sum Sq. Dev.	1.25E+12	7968.749	81350.63	5.77E+08	96.56442	75.165
Observations	163	163	163	163	163	163

	BMSQ	CHS	DAYS	DRHS	FINBM	MILES
Mean	706.0491	0.325153	159.5706	0.196319	0.142176	19.873
Median	728.0000	0.000000	141.0000	0.000000	0.000000	20.280
Maximum	3037.000	1.000000	542.0000	1.000000	0.952000	36.800
Minimum	0.000000	0.000000	1.000000	0.000000	0.000000	4.7600
Std. Dev.	748.3776	0.469876	116.7107	0.398437	0.261776	7.03322
Skewness	0.528795	0.746519	1.151241	1.529060	1.523130	-0.2379
Kurtosis	2.193815	1.557290	3.949427	3.338025	3.842508	2.5127
Jarque-Bera	12.01060	29.27597	42.12759	64.29236	67.84548	3.1501
Probability	0.002466	0.000000	0.000000	0.000000	0.000000	0.2069
Sum	115086.0	53.00000	26010.00	32.00000	23.17470	3239.4
Sum Sq. Dev.	90731180	35.76687	2206664.	25.71779	11.10133	8013.5
Observations	163	163	163	163	163	163

	SQFT	STICK	THS
Mean	1638.890	0.797546	0.374233
Median	1512.000	1.000000	0.000000
Maximum	4423.000	1.000000	1.000000
Minimum	738.0000	0.000000	0.000000
Std. Dev.	632.3584	0.403067	0.485416
Skewness	1.386697	-1.480959	0.519779
Kurtosis	5.485066	3.193240	1.270170
Jarque-Bera	94.18184	59.83664	27.66240
Probability	0.000000	0.000000	0.000001
Sum	267139.0	130.0000	61.00000
Sum Sq. Dev.	64780090	26.31902	38.17178
Observations	163	163	163

Table A-4

	PRICE	ACRES	AGE	AGESQ	AIR	BATH
PRICE	1.000000	0.130183	-0.302668	-0.240854	0.236249	0.716331
ACRES	0.130183	1.000000	0.061764	0.055820	-0.034306	-0.071305
AGE	-0.302668	0.061764	1.000000	0.936565	-0.246139	-0.442561
AGESQ	-0.240854	0.055820	0.936565	1.000000	-0.220706	-0.384991
AIR	0.236249	-0.034306	-0.246139	-0.220706	1.000000	0.248340
BATH	0.716331	-0.071305	-0.442561	-0.384991	0.248340	1.000000
BED	0.579182	-0.084375	-0.290031	-0.282097	0.103158	0.618064
BMSQ	0.585771	0.051757	-0.154124	-0.191391	0.173520	0.362034
CHS	-0.089099	-0.045167	-0.035739	-0.012736	0.039749	-0.047446
DAYS	-0.019337	0.000132	-0.037381	-0.041575	-0.039846	-0.024843
DRHS	-0.107809	0.039140	0.091340	0.052662	-0.216678	-0.162134
FINBM	0.337662	-0.127044	-0.083644	-0.146150	0.035266	0.423497
MILES	0.158939	0.163974	-0.158420	-0.181177	0.085813	0.112386
SQFT	0.759375	0.085771	-0.165766	-0.113284	0.189957	0.715370
STICK	0.336715	-0.078500	0.379576	0.295627	-0.001944	0.012169
THS	0.264144	0.034720	-0.142541	-0.130707	0.158480	0.246105

	BED	BMSQ	CHS	DAYS	DRHS	FINBM
PRICE	0.579182	0.585771	-0.089099	-0.019337	-0.107809	0.337662
ACRES	-0.084375	0.051757	-0.045167	0.000132	0.039140	-0.127044
AGE	-0.290031	-0.154124	-0.035739	-0.037381	0.091340	-0.083644
AGESQ	-0.282097	-0.191391	-0.012736	-0.041575	0.052662	-0.146150
AIR	0.103158	0.173520	0.039749	-0.039846	-0.216678	0.035266
BATH	0.618064	0.362034	-0.047446	-0.024843	-0.162134	0.423497
BED	1.000000	0.253138	-0.079630	0.026457	-0.043396	0.412703
BMSQ	0.253138	1.000000	-0.064733	-0.094346	-0.142149	0.513383
CHS	-0.079630	-0.064733	1.000000	0.068974	-0.343069	-0.070928
DAYS	0.026457	-0.094346	0.068974	1.000000	0.019612	0.001204
DRHS	-0.043396	-0.142149	-0.343069	0.019612	1.000000	-0.129707
FINBM	0.412703	0.513383	-0.070928	0.001204	-0.129707	1.000000
MILES	0.086561	0.046772	-0.475246	-0.065216	0.193586	0.076637
SQFT	0.615790	0.252077	-0.100346	0.044903	0.025517	0.175572
STICK	0.023862	0.476799	-0.073984	-0.124287	-0.058481	0.274485
THS	0.236053	0.153134	-0.536792	-0.106213	-0.382212	0.189842

	MILES	SQFT	STICK	THS
PRICE	0.158939	0.759375	0.336715	0.264144
ACRES	0.163974	0.085771	-0.078500	0.034720
AGE	-0.158420	-0.165766	0.379576	-0.142541
AGESQ	-0.181177	-0.113284	0.295627	-0.130707
AIR	0.085813	0.189957	-0.001944	0.158480
BATH	0.112386	0.715370	0.012169	0.246105
BED	0.086561	0.615790	0.023862	0.236053
BMSQ	0.046772	0.252077	0.476799	0.153134
CHS	-0.475246	-0.100346	-0.073984	-0.536792
DAYS	-0.065216	0.044903	-0.124287	-0.106213
DRHS	0.193586	0.025517	-0.058481	-0.382212
FINBM	0.076637	0.175572	0.274485	0.189842
MILES	1.000000	0.159545	-0.026192	0.541169
SQFT	0.159545	1.000000	0.097342	0.178308
STICK	-0.026192	0.097342	1.000000	0.105682
THS	0.541169	0.178308	0.105682	1.000000

Table A-5

VARIABLES	VIF'S
ACRES	1.16
AGE	10.87
AGESQ	9.09
BATH	3.69
BED	2.22
BMSQ	2.08
CHS	3.03
DAYS	1.06
DRHS	3.33
FINBM	1.96
MILES	2.17
SQFT	3.23
STICK	1.96
THS	4.9

Table A-6

Ramsey RESET Test:

F-statistic	8.015953	Probability	0.000056
Log likelihood ratio	25.17299	Probability	0.000014

Table A-7

White Heteroskedasticity Test:

F-statistic	1.240435	Probability	0.224710
Obs*R-squared	130.7083	Probability	0.322510

Table A-8

Dependent Variable: PRICE
Sample: 1 81
Included observations: 81

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	47671.70	21658.97	2.201014	0.0313
STICK	35749.42	7762.966	4.605124	0.0000
FINBM	10068.60	16316.22	0.617092	0.5393
ACRES	557.4190	869.4725	0.641100	0.5237
AGE	-1589.016	369.9857	-4.294803	0.0001
AGESQ	10.94828	3.632759	3.013766	0.0037
BATH	-2420.132	5456.198	-0.443556	0.6588
BED	-2314.082	5187.002	-0.446131	0.6570
SQFT	29.33578	7.896563	3.715007	0.0004
THS	-3820.303	9255.447	-0.412763	0.6811
AIR	4986.904	6180.615	0.806862	0.4227
BMSQ	8.990187	6.281874	1.431131	0.1572
CHS	-4911.829	7463.595	-0.658105	0.5128
DAYS	5.259358	18.46765	0.284788	0.7767
DRHS	-589.6194	8845.678	-0.066656	0.9471
MILES	43.17693	416.4127	0.103688	0.9177
R-squared	0.493950	Mean dependent var		72445.23
Adjusted R-squared	0.377169	S.D. dependent var		22854.62
S.E. of regression	18036.78	Akaike info criterion		22.61321
Sum squared resid	2.11E+10	Schwarz criterion		23.08619
Log likelihood	-899.8352	F-statistic		4.229717
Durbin-Watson stat	0.921750	Prob(F-statistic)		0.000022

Table A- 9

Dependent Variable: PRICE
Sample: 82 163
Included observations: 82

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-114361.9	65639.88	-1.742263	0.0861
STICK	44506.68	25304.05	1.758876	0.0832
FINBM	-25757.37	20875.47	-1.233858	0.2216
ACRES	1527.153	557.6774	2.738417	0.0079
AGE	-2810.480	649.9897	-4.323885	0.0001
AGESQ	31.03463	8.659251	3.583986	0.0006
BATH	26337.08	11999.23	2.194897	0.0317
BED	34412.00	10236.22	3.361788	0.0013
SQFT	41.08652	12.31773	3.335559	0.0014
THS	14334.29	22081.45	0.649155	0.5185
AIR	-26165.26	43519.23	-0.601235	0.5497
BMSQ	34.52714	7.328176	4.711559	0.0000
CHS	38764.07	18792.71	2.062718	0.0431
DAYS	-10.54879	42.32956	-0.249206	0.8040
DRHS	-3656.498	22829.28	-0.160167	0.8732
MILES	810.8495	994.5405	0.815301	0.4178

R-squared	0.810897	Mean dependent var	197498.2
Adjusted R-squared	0.767920	S.D. dependent var	84154.33
S.E. of regression	40541.12	Akaike info criterion	24.23120
Sum squared resid	1.08E+11	Schwarz criterion	24.70080
Log likelihood	-977.4792	F-statistic	18.86779
Durbin-Watson stat	1.667032	Prob(F-statistic)	0.000000

Table A-10

Dependent Variable: PRICE
Sample: 1 163
Included observations: 163

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-77235.68	16901.57	-4.569735	0.0000
ACRES	1749.446	431.0006	4.059035	0.0001
AGE	-2345.937	406.1920	-5.775439	0.0000
AGESQ	18.94398	4.519816	4.191316	0.0000
BATH	18339.33	6467.489	2.835618	0.0052
BED	13398.78	5777.984	2.318936	0.0217
BMSQ	26.91362	5.045986	5.333668	0.0000
SQFT	59.57292	7.354095	8.100646	0.0000
STICK	65246.51	9770.592	6.677846	0.0000

R-squared	0.836854	Mean dependent var	135355.3
Adjusted R-squared	0.828379	S.D. dependent var	87934.42
S.E. of regression	36428.72	Akaike info criterion	23.89773
Sum squared resid	2.04E+11	Schwarz criterion	24.06855
Log likelihood	-1938.665	F-statistic	98.74269
Durbin-Watson stat	1.690111	Prob(F-statistic)	0.000000

Table A- 11

Dependent Variable: PRICE
Sample: 1 60
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-119706.9	22926.78	-5.221269	0.0000
ACRES	1492.898	489.7376	3.048363	0.0037
AGE	-2248.468	682.4679	-3.294614	0.0018
AGESQ	28.23010	9.464049	2.982878	0.0044
BATH	22413.96	8768.867	2.556084	0.0137
BED	23307.93	8360.656	2.787811	0.0075
BMSQ	40.29219	6.992363	5.762314	0.0000
BROS	-13817.28	13025.39	-1.060796	0.2940
SQFT	58.76802	9.841991	5.971151	0.0000
STICK	57718.29	14963.34	3.857314	0.0003
TWIN	-11884.33	12016.97	-0.988963	0.3275

R-squared	0.937925	Mean dependent var	166995.2
Adjusted R-squared	0.925257	S.D. dependent var	106867.9
S.E. of regression	29216.80	Akaike info criterion	23.56702
Sum squared resid	4.18E+10	Schwarz criterion	23.95098
Log likelihood	-696.0105	F-statistic	74.03710
Durbin-Watson stat	1.961751	Prob(F-statistic)	0.000000

Table A-12

Dependent Variable: PRICE
Sample: 1 163
Included observations: 163

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-83316.46	17739.24	-4.696731	0.0000
ACRES	1760.249	437.7567	4.021067	0.0001
AGE	-2319.818	411.8714	-5.632386	0.0000
AGESQ	19.28651	4.559961	4.229533	0.0000
BATH	19963.34	6624.946	3.013360	0.0030
BED	15220.55	5986.645	2.542417	0.0120
BMSQ	27.84545	5.206618	5.348087	0.0000
SQFT	56.01138	7.718370	7.256891	0.0000
STICK	63243.18	10398.08	6.082199	0.0000
CONVEN	6032.936	7833.622	0.770134	0.4424
FHA	-1689.465	8689.476	-0.194427	0.8461
OTHER	-11982.51	14995.63	-0.799066	0.4255
VA	8101.775	13451.87	0.602279	0.5479
VHDA	24285.87	18512.19	1.311885	0.1916

R-squared	0.841323	Mean dependent var	135355.3
Adjusted R-squared	0.827479	S.D. dependent var	87934.42
S.E. of regression	36524.11	Akaike info criterion	23.93131
Sum squared resid	1.99E+11	Schwarz criterion	24.19703
Log likelihood	-1936.402	F-statistic	60.77055
Durbin-Watson stat	1.655650	Prob(F-statistic)	0.000000

VII. References

Anglin, P. M., Rutherford, R., & Springer, T. M. (2003). The Trade-off Between the Selling Price of Residential Properties and Time-on-the-Market: The Impact of Price Setting. *The Journal of Real Esta Finance and Economics* , 95-111.

Chiodo, A. J., Hernandez-Murillo, R., & Owyang, M. T. (2009). Nonlinear Hedonics and the Search for School Quality. *Research Department St. Louis Federal Reserve Bank* .

Cohen, J. P., & Coughlin, C. C. (2008). Airport-Related Noise, Proximity, and Housing Prices in Atlanta *Federal Reserve Bank of St. Louis* .

Grether, D. M., & Mieszkowski, P. (1973). Determinants of Real Estate Values. *Journal of Urban Economics* , 127-146.

Grether, D. M., & Mieszkowski, P. (1980). The Effects of Nonresidential Land Uses on the Prices of Adjacent Housing: Some Estimates of Proximity Effects. *Journal of Urban Economics* , 1-15.

Kain, J. F., & Quigley, J. M. (1970). Measuring the Value of Housing Quality. *Journal of the American Statistical Association* , 532-548.

Linneman, P. (1980). Some Empircal Results of the Nature of the Hedonic Price Function for the Urba Housing Market. *Journal of Urban Economics* , 47-68.

Mieszkowski, P., & Saper, A. M. (1978). An Estimate of the Effects of Airport Noise on Property Value *Journal of Urban Economics* , 425-440.

Ridker, R. G., & Henning, J. A. (1967). The Determinants of Residential Property Values with Special. *Review of Economics and Statistics* , 246-257.

Rosen, S. M. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competitio *Journal of Real Estate Literature* , 34-55.

Sirmans, G. S., MacPherson, D. A., & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature* , 3-46.

Studenmund, A. H. (2006). *Using Econometrics: A Practical Guide.* Boston: Pearson Education, Inc.

Taylor, C. R. (1999). Time on the Market as a Signal of Quality. *Review of Economic Studies* , 555-578.

Zietz, J., Zietz, E., & Sirmans, G. S. (2007). Determinants of House Prices: A Quantile Regression Approach. *Department of Economics and Finance Working Paper Series 2007* .