Real Estate Home Pricing: Hedonic model variables and Community Amenities Roles

by

Tim Dorr

University of Bridgeport

Abstract

Forecasting a home sale asking price using traditional hedonic pricing models is problematic: buyers have lost confidence in the market's response to such forecasted asking prices. With the restriction of federal and state funding sources for community services home buyers may face increased property taxes depending on the community they select. In turn this may affect the prices and neighborhoods that they search in. This study examines whether hedonic model variables remain viable for sales price forecasting and whether the level of available community services affects town selection decisions for home buyers. All data was developed from databases in the public domain. The study focuses on an examination of variance as measured within a linear regression model.

Keywords: Real Estate asking prices, Hedonic modeling, Regression, ANOVA

Introduction

Although the residential real estate market has stabilized in the five -- seven years following the financial crisis of 2008 the market remains generally depressed with price growth stagnant (National Association of Realtors [NAR], 2014). In turn forecasting an initial asking price using traditional hedonic pricing models[i] is problematic: buyers have lost confidence in the market's response to such forecasted asking prices (NAR, 2014). Additionally, community amenities such as access to schools, libraries and transportation may be viewed by buyers as a potential tax liability given reduced availability of federal and state funding for local services.

These conditions raise two questions:

1. Are hedonic variables such as square footage, number of: bathrooms, bedrooms, etc. good predictors of the variance observed in asking prices? The expectation is that the sold price will vary directly with changes in square footage of properties. Although there are several variables to choose from, square footage was chosen as the other variables are dependent or alter the total square footage of a property by their presence.
2. Given that community residential real estate tax rolls have contracted as a result of the 2008 Recession, have home buyers as measured by sold prices retreated from communities that have high community services offerings available to residents? The expectation is that although the communities offer a significantly different range of services there will be no significant difference among the average sold prices per square foot

among those communities. To evaluate this possibility, the sales price per square foot of properties was compared across three adjacent communities in Fairfield County, Connecticut. Given that the communities vary in the number and size of homes available for sale, creating this sold price per square foot variable minimized community size and housing stock availability effect.

The variables that will be examined are the publicly available town recorded sold price of homes in thousands, the square footage of the homes sold and the town in which the homes were sold. The first two variables are numeric and the third variable is categorical. This data will be drawn from Listingbook (Listingbook Website, n.d.). An additional numerical variable: sold price per square foot, will be created for analyzing the second question.

By definition a Hedonic model offers several variables that can be unbundled to support an asking price forecasting model. These variables were examined in a prior study relating to this geographic area (Dorr, 2015). Modeling the data using a step-wise multi-regression model revealed that square footage leads other variables with its explanatory power approaching 63%. The risk in hedonic modeling is the amount of inter-variable correlation: if lot size increases it will be reflected in increased taxes paid. While these added variables increase the model's explanatory power that increase does not offset the danger of overfitting the training data (Frost, 2015).

The data for this study will be drawn from the contiguous communities of Easton, Fairfield, and Weston located in Fairfield County, Connecticut. Easton and Weston are exurban communities located approximately 65 miles northeast of Manhattan, New York (see Appendix A). They are considered within a commutable distance of Manhattan. Development in Easton is restricted as 75% of the town land is a part of the Southern Connecticut watershed. Weston is similar to Easton in that it exurban in nature however, commercial development is not restricted. In contrast Fairfield is a large suburb of Manhattan with higher density and significant commercial development. Given the wide range of median home values indexing the value to square footage creates a relative measure across the three towns.

Table 1

Selected Demographic Data for Sample Towns

	Median HH Value	Median HH Income	Cost of Living Index
Easton	$729,919	$125,557	151.8
Fairfield	$493,741	$105,059	154.3
Weston	$1,000,000	$184,547	163.6

Source: (United States Census Bureau [Census], 2010)

Data Collection

Study data was drawn from information that is available through Listingbook®. This company complies publicly available data on the homes available for sale as well as the amount of time that homes are available prior to selling along with the prices that homes have sold for. Information on the homes available such as square footage, number of room etc. is copied from local town records that are reviewed and submitted to Listingbook by real estate agents listing homes for sale. This information is added to the active listings for the agent.

A systematic sampling scheme was used to gather the data. A systematic sample was chosen as it reflects the logical homogenous nature of the housing market. In effect houses for sale should be uniformly distributed over the entire housing stock of the towns. Given the total number of homes in each town sold during the past 36 months (Listingbook Website, n.d.) a *K*-interval was developed using the formula $k = \frac{N}{n}$ where *n* represents the sample size and *N* the population size. Starting at a random point in the list approximately 40 properties were selected for each town. The data for each sold home selected was entered into an a password protected Excel spreadsheet. The data included sold price, square footage, and days on market [DOM]. A fourth data column was created for sold price per square foot.

Sampling bias. Applying a systematic sampling method ensures that each home sold has an equal probability of selection. As a result this sampling approach is similar to a simple randon sample with the benefit of being more efficient; given an equal probabilty for any one home sold to be selected sampling bias is minimized.

Statistical Analysis Design

The development of the hypotheses reflects our challenge to a null claim. In one sense we challenge the status quo or null when we suspect that it is not correct. The following hypotheses for each question reflect this position:

1. **Question One**.
    a. **Ho**: There is no linear relationship between the sold price and the square

    footage of a home in the three Fairfield County markets selected.

b. **Ha**: There is a linear relationship between the sold price and the square
footage of a home in the three Fairfield county markets selected.

2. **Question Two**:
   a. **Ho**: There is no difference in the average sold price per square foot based on the home's town location.
   b. **Ha**: There is a difference in the average sold price per square foot based on the home's town location.

At the outset it is predicted that the following hypotheses will be supported as a result of the tests:

3. **Question One**.
   a. **Ha**: There is a linear relationship between the sold price and the square
   footage of a home in the three Fairfield county markets selected.
4. **Question Two**:
   b. **Ho**: There is no difference in the average sold price per square foot based on the home's town location.

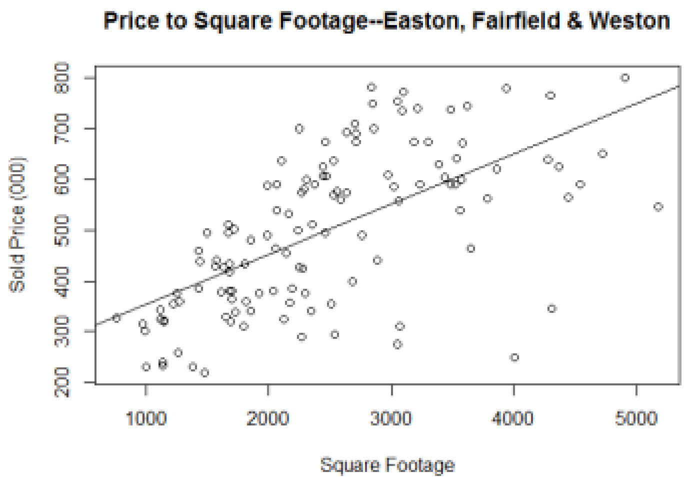
Each set of hypotheses will be tested at the $\alpha = 0.05$ significance level.

Statistical Analysis Results

Question One

In order to test whether a linear relationship exists between the sold price of a house and its square footage a linear regression test will be used. An initial inspection of the data suggests that there is a linear relationship however it is not strong based on the variability in the data as shown in Figure 1:

Figure 1. Data Summary



When the results of the test are examined there is a relationship as demonstrated by the coefficient of determination $r^2 = 0.3879$; however, this value suggests that only 38.8% of the variance observed in the dependent variable price is explained by the variance in the corresponding square footage of the property that was sold. In effect, approximately 61% of the variance observed in the sold price data remains unexplained.

As a measure of significance an analysis of variance test was conducted which resulted in an F-value of 76.688 and a resulting p-value of 1.451e-14 (see Figure 2). This small p-value supports a conclusion that we have strong evidence to reject the null hypothesis.

Figure 2. ANOVA Table

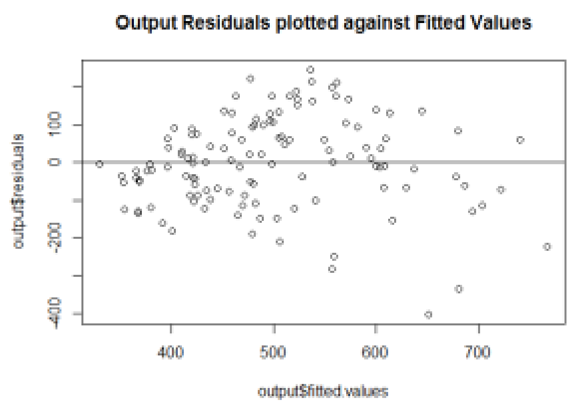
Response: Price

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SqFt	1	1110894	1110894	76.688	1.451e-14
Residuals	121	1752788	14486		

Assumption tests. When conducting a linear regression test there are five assumptions that require examination to validate the regression results:

1. Assumption One—Data-pairs Relationship -- Distribution
    a. A data scatter plot shows a relationship that is suggested by figure 1.
    b. The residuals are normally distributed with a mean of zero about the fitted values:

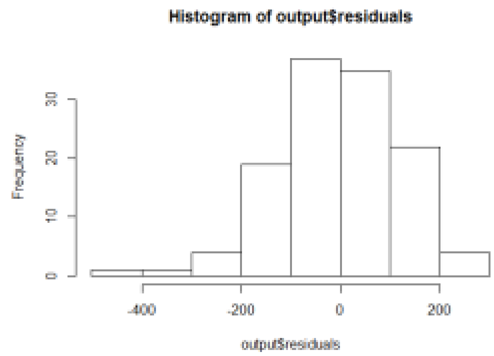
    Figure 3.

    In this case the scatter diagram and the residuals plot meet the criteria: a linear model is appropriate.

2. Assumption Two—Residuals Normal Distribution
    a. The residuals are normally distributed with a mean of zero (Figure 4):

Figure 4.



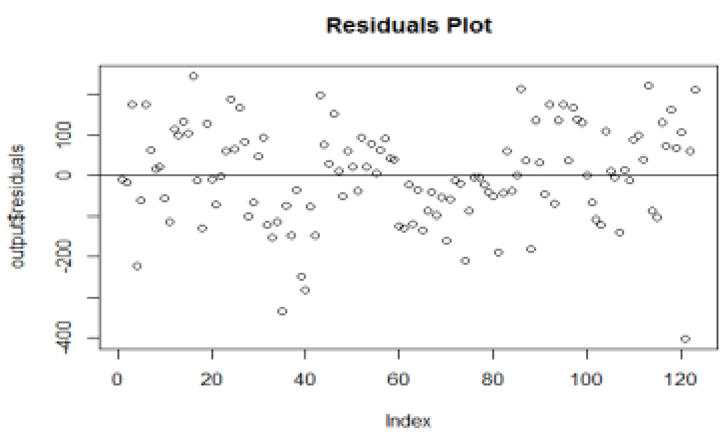
data: output$residuals
W = 0.9847, p-value = 0.1796

While the residual histogram shows a left skew the Shapiro-Wilks test confirms normal distribution—the p-value is greater than α = 0.05 and we fail to reject the null hypothesis: residuals are normally distributed..

3. Assumption Three—Residual Variance

   a. The residuals have constant variance:

Figure 5.



In this instance there are no apparent triangles or *bow ties* so the residuals have constant variance.

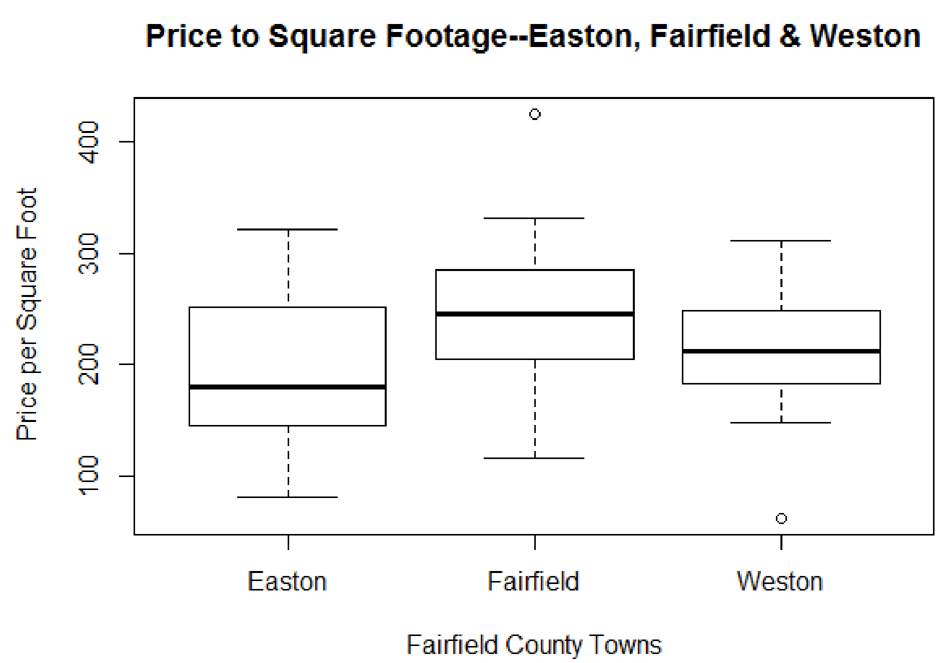
4. Assumption Four—Residual Independence
    a. Figure 5 shows that the residuals are independent based on the lack of a pattern in their distribution about zero.

Based on a p-value of 1.451e-14 we have enough data to support rejecting the null hypothesis that there is no linear relationship between the sold prices (000) of homes and their respective square footage at the α = 0.05 significance level.

Question Two

In order to test whether there is no difference in the average pricing per square foot of homes sold in the three communities—Easton, Fairfield and Weston—at the α = 0.05 significance level an ANOVA test will be conducted. An initial inspection of the data using boxplots suggested that there is no difference among the average sold price per square foot as shown in Figure 6.

Figure 6. Town Data Summary



The results for the ANOVA test are presented in Table 2.

Table 2.

Analysis of Variance

	df	Sum of Squares	Mean SS	F-value	p-value
Town	2	65538	32769	10.24	7.85e-05
Residual	120	384127	3201		

Having conducted the test, the resulting small p-value suggested that validity be tested using a series of post hoc tests.

1. **Levene's Test for Homogeneity of Variance**

   This test assesses the equality of the variances. If the p-value is sufficiently high such that it exceeds the $\alpha = 0.05$ significance level, then we cannot reject the null hypothesis that the sample variances are equal. In this instance, the p-value is 0.689 (greater than $\alpha = 0.05$) indicating that we cannot reject the null; the variances are equal.

2. **Shapiro-Wilks Normality Test**

   The null hypothesis for this test is that the population is normally distributed and that the data taken will be normally distributed as well.

Table 3

Shapiro-Wilks Normality Test (α = 0.05)

Easton	W = 0.9634	p-value = 0.1947
Fairfield	W = 0.9734	p-value = 0.4582
Weston	W = 0.9675	p-value = 0.2849

Table 3 summarizes the p-values for each town's sample data; in each case the p-value exceeds the α = 0.05 significance level and the null hypothesis that the populations and in turn the samples are normally distributed cannot be rejected.

3. **Pair-wise Comparison Test with Pooled Standard Deviation—Bonferroni**

   In this test we evaluate the pairs of data to determine which pair results in a statistically significant difference. In this instance the pair with a significant difference (p-value less than α = 0.05) is Easton and Fairfield with a p-value of 4.3e-05 as reported in table 4.

Table 4

P-value Adjustment Method--Bonferroni

Data: Price, Foot & Town		
	Easton	Fairfield
Fairfield	4.3e-05	--
Weston	0.061	0.095

4. **Tukey Honestly Significant Difference [HSD] Test**

   The Tukey HSD test confirms the pair-wise comparison test. Table 5 reports that there is a significant difference at the α = 0.05 significance level between the sample data for the towns of Easton and Fairfield.

Table 5

Tukey HSD Test

95% family-wise Confidence				
Town	Diff	Lower	Upper	P Adjusted
Fairfield-Easton	56.527	26.864	86.191	0.0000428
Weston-Easton	29.205	-0.2725	58.683	0.0527370
Weston-Fairfield	-27.322	-57.162	2.517	0.0800222

Based on a p-value of 7.85e-05 and the results of the post hoc tests we have enough data to support rejecting the null hypothesis at the $\alpha = 0.05$ significance level that there is no difference in the average sold price per square foot based on the town that the property is located in.

Findings

This study was undertaken to examine two questions:

1. Are hedonic variables such as square footage, number of: bathrooms, bedrooms, etc. good predictors of the variance observed in asking prices and,
2. given that community residential real estate tax rolls have contracted as a result of the 2008 Recession, have home buyers as measured by sold prices retreated from communities that have high community services offerings available to residents.

Based on the tests used to address these questions at the $\alpha = 0.05$ significance level, clearly within the contiguous towns of Easton, Fairfield and Weston located in Fairfield County Connecticut the square footage of homes offered for sale can be used to explain differences in the sold prices of those homes. Further, although the real estate market for these towns has changed following the 2008 Recession such factors as the services offered by towns will still play a positive role in home purchase selection.

Discussion

As originally suggested variables such as the square footage of a property can be used to explain changes in the selling prices of homes in these markets. While the finding was significant at the $\alpha = 0.05$ level the coefficient of determination was not as robust as anticipated. While the coefficient of determination may be strengthened by the addition of other variables such as the number of bedrooms and bathrooms this leads to correlation which may give a false sense of variance explanation. These results suggest that further study is needed on these variables to define their correlation to changes in selling prices of homes in these towns.

The findings for question two were not anticipated. It would appear that high levels of community services are attractive to home buyers in this county. This seems apparent when considering that a significant difference exists between Fairfield with substantial community services offerings and Easton where community services offerings are limited. If we consider the three towns, there may be variables that create attractiveness other than community services. Given the high median home price in Weston status factors may play a significant role in attracting buyers to that market.

Table 1 data shows that Weston had higher values for each of the three measures vis-à-vis Fairfield and Easton. These differences may be masking another variable that has a higher explanatory value in relation to the variance observed in selling prices of homes in these communities.

Implications

This study is limited by the geography selected; while the feel of each town is significantly different the three towns are within the economic sphere of Manhattan. This influence is felt in terms of average Fairfield County household income—ranked 39th out of 3,144 counties in the U'S. (Census, 2010) and the concurrent real estate values versus other areas of the United States.

While this study demonstrates that an alternative home sale price forecasting model is available, the data selected may include or mask some alternative explanatory variables (Borsboom, 2008). Specifically, intrinsic latent variables may guide buyers, sellers and agents in home sales pricing and these variables are not contained within the variables used in hedonic pricing models. For realtors this alternative model offers a way to work with familiar variables within a new price forecasting model.

Appendix A

Raw Data Used

Town	Price	SqFt	PriceFoot	DOM
Easton	600.0	3575	167.8	76
Easton	620.0	3856	160.8	96
Easton	675.0	2464	273.9	102
Easton	545.0	5172	105.4	334
Easton	625.0	4363	143.3	372
Easton	692.0	2636	262.5	70
Easton	673.0	3583	187.8	49
Easton	590.0	3233	182.5	84
Easton	510.0	2358	216.3	113
Easton	424.9	2276	186.7	60
Easton	590.0	4539	130	76
Easton	599.0	2303	260.1	66
Easton	580.0	2285	253.8	184
Easton	637.5	2530	252	289
Easton	675.0	3188	211.7	62
Easton	781.0	2842	274.8	96
Easton	590.0	3520	167.6	172
Easton	565.0	4438	127.3	83
Easton	625.0	2445	255.6	176
Easton	590.0	3483	169.4	150
Easton	650.0	4717	137.8	189
Easton	418.0	1676	249.4	71
Easton	800.0	4900	163.3	88
Easton	710.0	2698	263.2	88
Easton	570.0	2530	225.3	80
Easton	690.0	2717	254	86
Easton	765.0	4298	178	67
Easton	440.0	2890	152.3	223
Easton	562.5	3788	148.5	129
Easton	559.0	2584	216.3	176
Easton	575.0	2269	253.4	218
Easton	400.0	2688	148.8	354
Easton	465.0	3653	127.3	87
Easton	357.3	2175	164.3	85
Easton	345.0	4304	80.16	169
Easton	360.0	1813	198.6	71
Easton	340.0	2345	145	116
Easton	315.0	978	322.1	225
Easton	310.0	3068	101	83
Easton	275.0	3047	90.25	99
Easton	380.0	2036	186.6	202
Easton	355.0	2513	141.3	93

Town	Price	SqFt	PriceFoot	DOM
Fairfield	754.0	3048	247.4	62
Fairfield	502.0	1722	291.5	60
Fairfield	440.0	1572	279.9	111
Fairfield	675.0	2717	248.4	76
Fairfield	427.5	1632	262	130
Fairfield	427.0	2250	189.8	62
Fairfield	531.0	2167	245	88
Fairfield	499.0	2238	223	101
Fairfield	490.0	2764	177.3	67
Fairfield	675.0	3298	204.7	227
Fairfield	429.9	1560	275.6	45
Fairfield	539.0	2066	260.9	56
Fairfield	465.0	2058	226	126
Fairfield	460.0	1428	322.1	49
Fairfield	494.5	1492	331.4	56
Fairfield	480.0	1855	258.8	120
Fairfield	437.5	1440	303.8	46
Fairfield	230.0	1008	228.2	255
Fairfield	239.9	1144	209.7	64
Fairfield	355.0	1229	288.9	54
Fairfield	260.0	1266	205.4	77
Fairfield	377.5	1611	234.3	293
Fairfield	232.1	1142	203.3	167
Fairfield	339.0	1728	196.2	35
Fairfield	325.0	1124	289.2	56
Fairfield	340.0	1854	183.4	106
Fairfield	300.0	998	300.6	114
Fairfield	230.0	1382	166.4	210
Fairfield	365.0	1698	215	47
Fairfield	385.0	1432	268.9	59
Fairfield	360.0	1278	281.7	89
Fairfield	295.0	2536	116.3	99
Fairfield	330.0	1652	199.8	198
Fairfield	375.0	1260	297.6	99
Fairfield	326.3	768	424.9	73
Fairfield	343.0	1116	307.4	71
Fairfield	380.0	1677	226.6	97
Fairfield	318.5	1152	276.5	108
Fairfield	290.0	2266	128	162
Fairfield	381.0	1702	223.9	64

Town	Price	SqFt	PriceFoot	DOM
Weston	610.0	2972	205.3	87
Weston	640.0	4272	149.8	113
Weston	557.5	3058	182.3	127
Weston	750.0	2852	263	61
Weston	641.5	3537	181.4	189
Weston	220.0	1482	148.5	25
Weston	587.0	1992	294.7	137
Weston	585.0	3022	193.6	87
Weston	322.2	1154	279.2	98
Weston	734.4	3085	238.1	141
Weston	375.0	1920	195.3	45
Weston	780.0	3937	198.1	77
Weston	637.5	2100	303.6	97
Weston	490.0	1988	246.5	90
Weston	740.0	3211	230.5	78
Weston	738.5	3480	212.2	52
Weston	590.0	2070	285	93
Weston	433.5	1806	240	80
Weston	540.0	3562	151.6	262
Weston	375.0	2300	163	178
Weston	311.0	1794	173.4	153
Weston	607.0	2445	248.3	360
Weston	605.0	3436	176.1	251
Weston	495.0	2464	200.9	413
Weston	325.0	2126	152.9	77
Weston	435.0	1680	258.9	66
Weston	454.9	2144	212.2	51
Weston	510.0	1673	304.8	177
Weston	590.0	2372	248.7	229
Weston	630.0	3387	186	159
Weston	699.0	2250	310.7	182
Weston	385.0	2192	175.6	77
Weston	318.5	1691	188.4	78
Weston	745.0	3624	205.6	138
Weston	494.9	1674	295.6	153
Weston	699.0	2860	244.4	199
Weston	575.5	2557	225.1	67
Weston	607.0	2471	245.7	172
Weston	250.0	4005	62.42	205
Weston	575.0	2634	218.3	127
Weston	772.5	3100	249.2	198

References

Borsboom, D. (2008, February). Latent variable theory. *Measurement*, *6*(1/2), 25-53. http://dx.doi.org/10.1080/15366360802035497

Dorr, T. G. (2015). My home's market value: An active market pricing model. *Journal of Management and Innovation, 1*. Retrieved from http://jmi.mercy.edu

Frost, J. [The Minitab Blog]. (2015, September 3). The danger of overfitting regression models [Blog post]. Retrieved from http://blog.minitab.com

Listingbook Website. (n.d.). http://www.listingbook.com/home/

National Association of Realtors. (2014). *Profile of home buyers and sellers*. Retrieved from http://www.realtor.org/

United States Census Bureau. (2010). *Housing Vacancies and Homeownership* (CPS/HVS). Washington, DC: Government Printing Office.

[i] Hedonic pricing models are based on the premise that the price of an item can be decomposed into its elements—internal or relating to the property and external relating to things such as the quality of the school or the air quality. As originally conceived hedonic modeling is based on decomposing a complex item into a subset of components that have individual market value. As such these variables could be measured and should be independent of each other (Sirmans, Macpherson, & Zietz, 2005).