

# Considerations for a Regression-Based Real Estate Valuation and Appraisal Model: A Pilot Study

Shawn L. Robey<sup>1</sup>, Mark A McKnight<sup>1</sup>, Misty R. Price<sup>2</sup> & Rachel N. Coleman<sup>1</sup>

<sup>1</sup> University of Southern Indiana, USA

<sup>2</sup> Pasco-Hernando State College, USA

Correspondence: Mark A McKnight, University of Southern Indiana, USA

Received: February 21, 2019

Accepted: March 28, 2019

Online Published: March 30, 2019

doi:10.5430/afr.v8n2p99

URL: <https://doi.org/10.5430/afr.v8n2p99>

## Abstract

This paper advocates a more scientific approach to residential real estate valuation as opposed to more traditional approaches, which are flawed for two main reasons: (1) appraiser judgements are almost exclusively used and (2) appraisers' sample sizes are too small to provide adequate estimated values. Using a specific case study approach to a single real estate market for historical data, the current research explores the impacts of different characteristics on market value. Three hundred and fifteen properties in Evansville, Indiana, were analyzed testing twelve different variables via regression analysis. This model suggests that 91.8% of the total market value variation is explained by four independent variables – total square feet of home, year constructed, property tax for most recent year and original list price. These findings provide evidence that multiple linear regression could be used to better predict a property's value, in place of more traditional market comparison models.

**Keywords:** real estate, valuation, regression, appraisal

## 1. Introduction

One of the central issues in the valuation of real estate is that market value must be estimated. Houses have been viewed as assets which require valuation (Liu, Wang, & Zha, 2011). The value of property cannot be simply observed in the marketplace. Unlike the stock market, which has many buyers and sellers actively trading throughout the day, prices are not fixed in residential real estate. Because these transactions occur many times a day, the value of the stock is almost instantaneous. Additionally, in real estate, an active negotiation is involved in price setting and location is a central element of the price, as opposed to stocks, which are location-independent. Ling (2013) explains that "since a parcel of real estate cannot be moved from its location, its value is subject to the effects of economic, social, or political developments" (p. 161). Finally, due to the infrequency inconsistencies of transactions, the data needed in establishing comparable price is often scarce. Because the real estate market is so unique, estimating market value is complex. In attempting to sort through the complexity, the appraisal market has established multiple traditional methods for estimating price. These specific approaches to real estate valuation include (1) the income approach, (2) the sales approach, and (3) the cost approach. The importance of housing markets and their role in macroeconomic fluctuations has been documented (Guo, 2017). Maintaining accurate and relevant values and valuation processes for real estate is critical in preventing financial crises such as those in 2008.

The purpose of the current project is to use a single real estate market and a case study approach to validate differences, and ultimately, advantages, of using a regression-based valuation system in place of traditional approaches of estimating market value of real estate. Case study research allows for exploration and understanding of complex issues (Zainal, 2007). Yin (1984) defined case studies as a research method "as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used" (p. 23). In this instance, the multiple traditional approaches of estimating market value are estimated to be incomplete, and a case study methodology is used to show how a regression model is superior with regarding to the inefficiencies of traditional approaches.

## 2. Literature Review and Research Question Development

### 2.1 Traditional Approaches of Estimating Market Value

There are three primary approaches commonly used by real estate appraisers to estimate the value of real estate: the sales comparison approach, the income approach, and the cost approach. Typically, the sales comparison approach can be used in the valuation of one- to four-family residential properties. Second, is the income approach, which is used to estimate the value of income-producing property. The value is determined by calculating the present value of future cash flows. This approach is not appropriate for residential real estate. The cost approach is the third conventional method used to estimate the market value of real estate. It involves estimating the cost of replacing the property net accrued depreciation. Estimating the value of accrued depreciation can be potentially very difficult and uncertain, yielding this approach as more likely to provide inaccuracies (Ling, 2013).

According to Fannie-Mae (2009), there are three types of depreciation: physical, functional, and external. Because there are various types of depreciation, estimating the decline in the property's value is very difficult. Therefore, appraisers rely on actual sales of comparable properties to estimate a subject property's value. This enables the appraiser to use the value judgement of actual market participants (Ling, 2013).

### 2.2 Sales Comparison Approach

The sales comparison approach involves comparing the property listed for sale (the subject property) with several similar properties recently sold (Pagourtzi, Assimakopoulos, Hatzichristos, & French, 2003). These are referred to as "comparable properties." If the comparable properties were perfect substitutions, then no adjustment to the sale prices would be made. Since no two real estate properties are the same, the appraiser must make adjustments. This leads to the first of three steps in the sales comparison approach, determining comparable sales. The appraiser will need to search through public records, multiple listing services (MLS), and private data services in order to obtain the necessary information on the subject and comparable properties.

Once the appraiser has identified the comparable properties, he or she can move to the second step, which is to adjust the comparable properties' sale prices into an approximation of the subject property (Ling, 2013). "Since no two properties are identical the appraiser must adjust the selling price of each comparable to account for differences between the subject and the comparable" (Pagourtzi et al., 2003, p. 386). Once differences have been taken into consideration, a value for the subject property can be inferred from the adjusted sales prices of comparables (Pagourtzi et al., 2003). Pagourtzi (2003) and others also point out that information related to comparables is "heavily dependent on the availability, accuracy, completeness, and timeliness of sale transaction data" (p. 386).

The third and final step of the sales comparison approach is to calculate the indicated value of the subject property. To do this, the appraiser takes the final adjusted sale price from step two and calculates the weighted average price. These weights are based on the appraiser's professional opinion. The resulting weighted average price is the indicated value of the subject property.

Literature points out that there are concerns with using only one of these three approaches to determine appraisal value (Albright, 1986; Ratcliff, 1975; Smith, 1986). Smith (1986) outlines nine such inconsistencies. These include: the timing of influences requiring adjustments; the application of percentage adjustments and explanatory statements; date of the appraisal and date of the report; definition of market value cited and concept actually employed; timing of data used to estimate capitalization rates and income data when using the income approach; the timing of yield rates and other components of capitalization rate models when using the income approach; theory and application of highest and best use analysis when using the income approach; penalties for curable functional obsolescence and market realities; and the measurement of incurable functional obsolescence.

Dugan (1999) developed and advocated an extension of the comparable sales model. In his hold, he used an "Ideal Point System (IPS)" to create a measure of how desirable a property was, and then used comparable sales based on that calculation. More recently, Bin et. al (2019) developed a model that focused on proximity of one property to others, relying on comparable properties that were also focused on staying a relatively close location to approximate real estate value. Manganelli, De Paola & Del Giudice (2018) developed a multi-objective analysis model in mass real estate appraisal that utilized not only a case study approach but integrated it into a regression model as well to maximize results via a spline smoothing model.

### 2.3 Using Multiple Regression to Estimate Market Value

Many subjective assumptions must be made in real estate valuation. Whether one is using the Sales Comparison Approach, Income Approach, or Cost Approach, the value of the property derives from the Appraiser's "expert"

opinion. These approaches are also known as the Market Comparison Approach (MCA).

A non-traditional approach to real estate valuation is using multiple regression. Pasymowski (2007) warns that “appraisers have little or no training in statistics and econometrics (regression analysis) and thus violate two simple rules in statistics: 1. Sample size is too small 2. Variance in the real estate market is not calculated” (p. 1). Even so, Dell (2017) argues that the ideas of regression provide the “foundation, framework, and vision for new valuation paradigms” (p. 218).

Unlike traditional approaches, the use of regression analysis provides unbiased comparable properties, and the process can be performed repeatedly with great accuracy. This is possibly because the method uses an infinitely larger sample size and the variance in the real estate market can be included. Using regression analysis in real estate provides a more scientific approach to the valuation process.

Ragsdale (2012) asserts that “regression analysis is a modeling technique for analyzing the relationship between a continuous dependent variable  $Y$  and one or more independent variables  $X_1, X_2, \dots, X_k$ ” (p. 433). Once a regression model has been created and the function has been run, the results can then be analyzed. The first number of importance is  $R^2$  value, which can range from 0 to 1. The closer  $R^2$  is to 1, the better the model “fits”. A model “fits” when it can account for the total variation in the dependent variable  $Y$ . For example, regression model A with a  $R^2$  of .86 will have a better “fit” than model B with a  $R^2$  of .8. This means that in Model A approximately 86% of the total variation in the dependent variable  $Y$  around its average has been accounted for by the independent variable(s)  $X$  in the estimated regression function. The second number of great importance is the adjusted- $R^2$ . Though not foolproof, “the adjusted- $R^2$  value can be used as a rule of thumb to help us decide if an additional independent variable enhances the predictive ability of a model or if it simply inflates the  $R^2$  statistic artificially” (Ragsdale, 2012, p. 460).

“Inflating  $R^2$ ” can happen anytime an independent variable is added to the model. This results in artificial inflation or “overfitting” if the new independent variable isn’t related to the dependent variable. For example, Model A currently has two independent variables with a  $R^2$  of .86 and an adjusted  $R^2$  of .87. If a third independent variable is added to model A and increases the  $R^2$  to .88 and the adjusted  $R^2$  decreases to .83, model A has had  $R^2$  artificially inflated.

There is much literature about the use of regression analysis, specifically multiple regression analysis in appraising real estate values. Dell (2017) outlines the importance of using data science to determine the ideal data set and the prediction that can best be made, using regression analysis. “Predictive regression directs...a probable sale price (or transaction zone) from within the competitive market segment data set” (p. 223).

Murphy (1989) outlines the seven tests one must consider in order to determine the most reliable multiple regression equation when predicting estimated sales value. These include the coefficient of correlation, coefficient of determination,  $t$ -statistic, standard error of the estimate,  $f$ -statistic, multicollinearity, and the Durban-Watson test (for time-series, not cross-sectional data). He also states that “a more reliable multiple regression equation can be achieved by securing a larger quantity of comparable sales and analysis of numerous combinations of independent variables” (p. 508). Goldberg and Mark (1988) reiterate that in order to have a reliable equation, it is necessary to have a large sample with information on sales prices and attributes that might affect value. They clarify, asserting that at a minimum, the number of observations must be larger than the number of variables.

Multiple regression analysis in the use of appraising values is not without concerns. Newsome and Zietz (1992) note problems associated with heteroscedasticity. They recommend minimizing these problems by segmenting the sale data by price. Isakson (2001) outlines two major pitfalls of using multiple regression analysis in real estate appraisal. They are model specification and the robustness of the results of the regression. To overcome these pitfalls, he has two recommendations. First, a large sample size. Second, the types of statistical tests used to determine a reliable equation. Finally, Kubus (2016) echoes the concern related to the presence of redundant or irrelevant variables by arguing that these variables can decrease the stability of the models and points out that they can even reduce prediction accuracy.

#### 2.4 Research Questions

Dell (2017) calls for “an acceptance that some customary procedures may no longer be ideal” (p. 218) and asserts that “regression analysis, used sensibly is essential” (p. 219) to a new valuation paradigm where predictive regression can direct to a probable sale price. The current research seeks to extend and apply this logic via a sample and regression analysis related to traditional elements (variables) that are thought to impact sale price of residential real estate. When using the sales comparison approach, “comparable properties” provide the basis upon which to arrive at the most appropriate appraisal value. Many characteristics of the “comparable properties” are taken into

consideration. A similar approach is used when using multiple regression to arrive at the most appropriate appraisal value.

Do and Grudnitski (1992) found that in empirical studies, a traditional valuation model for single family properties included the following predictor variables: age of the structure in years, the number of bedrooms, the number of bathrooms in increments of ¼ baths, total square footage of the house, the number of garages, the number of fireplaces, the number of stories, and the lot size measured in square feet.

In a study performed by Narula, Wellington, and Lewis (2011) using parametric programming, ten predictor variables were used. They included: taxes, number of baths, frontage (feet), lot size (square feet), living space (square feet), number of garages, number of rooms, number of bedrooms, age of home (years), and number of fireplaces. Similarly, in a study performed by Newell (1982) comparing multiple linear regression and ridge regression, ten predictor variables were used. They included: sales date, lot area, style, quality, condition, year built, number of rooms, number of bathrooms, percentage financed, and square feet.

Finally, in a study performed by Nguyen and Cripps (2001), comparing multiple regression and artificial neural networks, six predictor variables were used. They included: the square feet of living area, the number of bedrooms, the number of baths, the number of years since the property was built, the quarter the property sold, and whether the property had a garage or carport. For this study, several of the same variables will be used to determine which are significant in predicting the most appropriate appraisal value.

RQ1: To what extent are traditional variables useful as predictors of sale price for a residential real estate regression model?

RQ2: What is the relevant importance of traditional variables when applied to a regression-based real estate model?

### **3. Methodology and Design**

The three core research methods available for research are the qualitative method, the quantitative method, and the mixed method. It is important to the researcher that the methodology aligns with and supports the research questions (Yin, 2006). Properly selecting the method at the beginning of this study will help ensure alignment and that the research questions are supported. For this study, a quantitative approach was selected as the appropriate method for addressing the research questions and purpose of the study.

#### *3.1 Method*

Creswell (2014) highlights that quantitative methods can be used as a method of inquiry to measure cause and effect of numerical variables. Insight can be gained based on the results of statistical analysis (Stake, 2010). The goal of this study is, through regression analysis, to predict a probable sales price. This is in alignment with a quantitative research approach of testing theory through postpositive knowledge claims (Creswell, 2014), due to the need to identify and assess variables that influence appraisal values and estimated sales values.

#### *3.2 Design*

The study used a correlational design that examined the relationship between 12 predictor variables (number of bedrooms, number of bathrooms, total square feet of home, lot size/acreage, year of construction, covered parking square feet, interest rate at closing date, property tax for the most recent year, school ratings for elementary, middle, and high schools, and original list price) and one criterion variable (sales price). Specifically, a multiple regression design was chosen due to its fit with the purpose and the questions guiding the current inquiry: to study the relationship between a criterion (sales price) and several predictors (number of bedrooms, number of bathrooms, total square feet of home, lot size/acreage, year of construction, covered parking square feet, interest rate at closing date, property tax for the most recent year, school ratings for elementary, middle, and high schools, and original list price). Field (2013) and Gall, Gall, and Borg (2015) suggest using multiple regression when conducting a study with multiple predictor variables and a single criterion variable. This design provides both statistical significance and magnitude of the relationships between variables (Gall et al., 2007).

#### *3.3 Data Collection and Reduction*

For this model, usable sales data included homes located in Vanderburgh County, Indiana sold for the two-year period between January 1, 2016 and December 31, 2017. Initially, this resulted in a total of 11,406 sales. Next, the data was filtered to include only residential one family dwellings and residential mobile homes, on a “platted” or “un-platted” lot. This reduced the number of applicable sales to 10,512. A second filter was applied that eliminated mobile homes, which resulted in a total of 9,011 applicable sales. In order to eliminate damaged, partially

constructed or unlivable properties, or vacant land, cases with sales prices less than \$100,000 were removed. This further reduced the number of usable properties to 4,177.

A probability type of sampling strategy was used to select homes. The probability sampling method offers increased validity to the research (Vogt, 2007). Specifically, simple random sampling will be used. With a random sample, there is an equal chance of selecting each participant from the populations being studied when creating the sample (Connaway & Powell, 2010). In addition, simple random sampling reduces the potential for human bias in the selection of homes included in the sample. Therefore, a simple random sample provided a highly representative sample of the population (Vanderburgh County, Indiana).

A precise and accurate conclusion can only happen with an appropriate sample size (Nayak, 2010). Using the sample size calculator for multiple regression (Soper, 2019), the ideal sample size for each of the population was calculated. An anticipated effect size of .2, statistical power level of .95, 12 predictors, and a probability level of .05 were used to complete the calculation. To have a statistically significant sample size, the study required 141 homes from Vanderburgh County. Of the 500 cases, 315 subject properties had complete and usable data. The sample size obtained was more than twice the number necessary to have statistical significance.

From the usable sales data (4,177 sales), a sample size of 500 cases were randomly selected. For each of the 500 cases, data was gathered using a combination of Zillow and the county Assessor's office websites, yielding 16 variables for each case. The dataset was then examined for missing information; any samples with missing information were removed, which resulted in a total of 315 usable sales for analysis.

Of the 315 subject properties in the current study, the mean sale price was \$179,227.70. The average home in the study contained 3.17 bedrooms and 2.13 bathrooms and contained 2,197.59 square feet. The average lot size was .42 acres and was originally constructed in 1974. The average inside parking area was 399.94 square feet.

Interest rate data was identified using national averages at the time of purchase and indicated an overall average interest rate for 30-year fixed rate mortgages of 3.78%. The average property tax amount of the subject properties was \$1,684.69. Using the Zillow school estimate ranges, average ratings for elementary schools for the subject properties was 4.86/10. The average ratings for middle schools was 5.07/10 and the average ratings for high schools for the subject properties was 4.25/10. Finally, the average original listing price of the subject properties was \$189,475.22. This indicates an overall average gap between original list price and final sale price of \$10,247.52.

### 3.4 Variables

Sale price (Sp) served as the dependent variable for the study. The twelve independent variables tested included number of bedrooms (Be), number of bathrooms (Ba), total square feet of home (Sqft), lot size/acreage (L), year of construction (Yr), covered parking square feet (Pk), interest rate at closing date (In), property tax for the most recent year (Tx), school ratings for elementary (Er), middle (Mr), and high (Hr) schools, and original list price (Op). Nominal data collected, but not included in analysis, included zip code of property and township of property. Because subject properties were selected from a single county, these data were redundant to property tax (Tx).

## 4. Results

Using regression analysis, and sale price (Sp) as the dependent variable, with twelve independent variables, the model generated a  $R^2$  value of .922, with an adjusted  $R^2$  of .919. To test significance of the model, an ANOVA is presented below, in Table 1.

Table 1. ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	2.114E+12	12	1.762E+11	297.618	.000
Residual	1.788E+11	302	591939833.4		
<b>Total</b>	<b>2.293E+12</b>	<b>314</b>			

Table 2 indicates only five variables (one dependent, four independent) that were statistically significant at the .05 level. These included the dependent variable sale price (Sp), and the independent variables of total square feet of home (Sqft), year constructed (Yr), property tax for the most recent year (Tx) and original list price (Op). Based upon these variables, a second linear regression analysis was performed. Results confirmed the significance of these variables. Specifically, the  $R^2$  using only four independent variables was .918, with an adjusted  $R^2$  of .917.

Table 2. Coefficients

	Unstandardized B	Coefficients Std. Error	Standardized Coefficients B	t	Sig.
<b>(Constant)</b>	-355470.371	140144.070		-2.536	.012
<b>Be</b>	4426.666	2659.887	.034	1.664	.097
<b>Ba</b>	-5575.104	3154.206	-.046	-1.768	.078
<b>Sqft</b>	5.988	2.562	.066	2.338	.020
<b>L</b>	4905.641	3032.400	.032	1.618	.107
<b>Yr</b>	171.274	71.047	.056	2.411	.017
<b>Pk</b>	-6.286	6.578	-.021	-.956	.340
<b>In</b>	7216.691	5717.599	.021	1.262	.208
<b>Tx</b>	10.791	3.277	.119	3.293	.001
<b>Er</b>	1665.137	1236.066	.036	1.347	.179
<b>Mr</b>	280.669	1369.627	.004	.205	.838
<b>Hr</b>	-438.451	1252.543	-.009	-.350	.727
<b>Op</b>	.678	.032	.788	20.948	.000

Additionally, a second ANOVA (based on the four independent variables) is presented in Table 3, below.

Table 3. ANOVA (Four Dependent Variables)

	Sum of Squares	df	Mean Square	F	Sig.
<b>Regression</b>	2.105E+12	4	5.263E+11	870.179	.000
<b>Residual</b>	1.875E+11	310	604854806.7		
<b>Total</b>	2.293E+12	314			

The R<sup>2</sup> statistic in the second model is 0.918, which suggests that approximately 92% of the total variation in the market value is explained by four independent variables.

Table 4. Coefficients (Four Dependent Variables)

	Unstandardized B	Coefficients Std. Error	Standardized Coefficients B	t	Sig.
<b>(Constant)</b>	-274713.333	114966.452		-2.390	.017
<b>Sqft</b>	5.280	2.476	.058	2.133	.034
<b>Yr</b>	149.069	58.160	.049	2.563	.011
<b>Tx</b>	11.201	3.196	.124	3.505	.001
<b>Op</b>	.681	.031	.791	21.724	.000

From the regression results above, predicted sale price (P<sub>Sp</sub>) can be calculated by multiplying each unstandardized coefficient by the characteristic of the subject property, summing these products and adding the intercept term.

## 5. Discussion of Findings

Findings from this study add to research on real estate valuation by identifying key variables contributing to the value of residential one family dwellings using multiple linear regression. These findings provide further evidence of how multiple linear regression could be used to better predict a property's value and sale price in future research. The framework provided may improve our understanding of what variables truly have an impact on value. The following provides a summary of the findings as they relate to the guiding questions of this study.

In relation to research question one (RQ<sub>1</sub>), twelve traditional variables explained just over 92% of the total variation in market value. However, only 4 out of the 12 dependent variables were statistically significant at the .05 level. These findings suggest the number of bedrooms (Be), number of bathrooms (Ba), lot size/acreage (L), covered parking square feet (Pk), interest rate at closing date (In), school ratings for elementary (Er), middle (Mr), and high (Hr) schools did not have high level of impact on the sale price of a residential single-family home.

Of these variables, number of bedrooms (Be) and number of bathrooms (Ba) were the only ones with any level of significance. Levels of significance for number of bedrooms (Be) and number of bathrooms (Ba) were .097 and .078, respectively. The six remaining variables, lot size/acreage (L), covered parking square feet (Pk), interest rate at closing date (In), school ratings for elementary (Er), middle (Mr), and high (Hr) schools, had significance levels ranging from .107 to .838. Specifically, with regards to school ratings, elementary school rating (Er) did approach statistical significance when compared to middle (Mr) and high (Hr) school ratings. At a significance level of .179, elementary mattered more to the prediction of sale price than middle (Mr) and high (Hr) school.

Pertaining to research question two (RQ<sub>2</sub>), four variables explained nearly 92% of variation when using sale price (Sp) as the dependent variable. Variables with a significance at the .05 level included: total square feet of home (Sqft), year constructed (Yr), property tax for the most recent year (Tx) and original list price (Op). Of the four variables, total square feet of home (Sqft) was the least significant with a significance level of .034. Year constructed (Yr), property tax for the most recent year (Tx) and original list price (Op) had significance levels of .011, .001, and .000, respectively.

### 5.1 Regression Equation

Finally, based on these initial findings, a regression equation can be estimated for predicted sale price. That model would be represented via the following equation:

$$PSP = \text{Intercept} + Sqft_{b1} + Yr_{b2} + Tx_{b3} + Op_{b4}$$

In the current sample, the equation estimating real estate sale price would be:

$$\text{Predicted Sale Price} = -274,713.333 + 5.28_{b1} + 149.069_{b2} + 11.201_{b3} + .681_{b4}$$

Mark and Goldberg (2001) note several factors to consider when selecting the best regression model. These include, maximizing R<sup>2</sup>, maximizing adjusted R<sup>2</sup>, minimizing the standard error of the estimate, maximizing the number of coefficients with significant *t* values, and minimizing the coefficient of variation. Further, they recommended that the R<sup>2</sup> on a randomly chosen holdout sample be the basis for choosing the model that predicts the event in question most accurately. Nguyen and Cripps (2001) base the specifications in one model in their study on other studies that show that square feet, age, the number of bedrooms, and the number of bathrooms influence selling price. In the above noted recommended equation, results are similar with R<sup>2</sup> at approximately 92% and two significant variables, square feet and age, influencing predicted selling price.

## 6. Limitations

Several limitations impact the generalizability of findings for this pilot-level study. Specifically, limitations for the present research are related to the geographic location of the sample used for the study as well as a few inherent conditions related to the data used for the study.

Because of the pilot nature of the research, usable sales were only collected for a two-year period. Additionally, data were pulled from a single county. A wider span of sales data that considers a longer timeframe of sales data, as well as an expansion of data to state, regional and national level information would increase generalizability of the research findings.

Assumptions were made in the present research related to school choice of subject properties. Zillow's listing of nearby schools was used to identify school ratings (based on ratings provided by GreatSchools.org). These data however do not take into consideration actual schools of attendance for students of families who live in the related areas. For example, parents might choose private school options, home school options or otherwise enroll students in other area high schools.

Because of the logistics involved of identifying private interest rates for specific purchases of residential real estate, national interest rate data was used from FreddieMac. This data was a monthly average of national interest rates for 30-year mortgages since 1971. This data assumes 30-year mortgages were used in the subject home properties and does not account for variations in the rate based on specific home buyer characteristics, such as credit worthiness or specific financing terms.

## 7. Suggestions for Further Research

Several possibilities emerge from the current research as potential extensions and opportunities for further research. Most logically, because of the relatively localized sample used in this study, expansions and includes of larger areas should be included. Comparison samples should seek to expand the geographic area for testing the model into state,

regional and even national subject properties. Such as study would seek to test the model in different markets, but also provide interesting comparison groups as well.

One potential variable not addressed in the study is the impact that time on market has on the price of a home. Traditionally, the longer a home is listed for sale on the real estate market, the more the price declines. Further research should seek sample properties in order to track and investigate the role of time on market and its direct impacts on sale price.

One interesting (though not clearly significant) finding in the current study was the degree to which school ratings impacted sale price of a property. Elementary school did approach significance as a variable, while middle and high schools did not. Additional research, perhaps qualitative in nature, should investigate the relative importance of various school ratings. Additionally, various school ratings scales should be explored to investigate potential impacts of more impactful rating systems.

Contextual factors were not an initial consideration for the present research. Possible research should analyze the health of pertinent economies (economies for local markets) including issues such as unemployment rate in relation to the impact on sale prices of residential properties. Other related issues impacting context might include cost of living variances within state, regional or national real estate markets.

An additional issue that provides the basis for additional future research is the potential for resale value. Specifically, future research should consider intent of buyers – isolating residential real estate buyers into homeowners as compared to those who purchase residential properties as an investment strategy.

Finally, potential research could replicate the present study, and provide a comparison to other real estate valuation methods, such as ratification neural networks, parametric programming, use of ridge regression, etc. The comparison would provide insights as to the best tools in terms of predicting sales prices.

## 8. Conclusions

One of the more documented issues in the valuation of real estate is that market value must be estimated. Due to the inconsistencies of transactions in a specific property, the data needed in establishing comparable price is often scarce, if available at all. In attempting to sort through the complexity, the appraisal market has established multiple traditional methods for estimating price. However, due to the subjective assumptions made in the valuation of real estate, traditional methods are flawed.

A non-traditional approach to real estate valuation is the use of multiple regression. Unlike traditional approaches, the use of regression analysis provides unbiased comparable properties, and the process can be performed repeatedly with great accuracy. The accuracy of a model can be increased using a large sample size, statistical tests to determine reliability, and the removal of redundant or irrelevant variables.

The results of the study suggest that nearly 92% of the total variation in market value is explained by five variables - one dependent variable and four independent variables. These findings provide further evidence of how multiple linear regression could be used to better predict a property's value and sale price in future research. The framework provided may improve overall understanding of the degree to which certain variables impact value.

## References

- Albright, S. J. (1986). Appraisal guidelines created by laypersons. *Appraisal Journal*, 54(1), 39-43.
- Bin, J., Gardiner, B., Li, E., & Liu, Z. (2019). Peer-dependence valuation model for real estate appraisal. *Data-Enabled Discover and Applications*, 3(2). <https://doi.org/10.1007/s41688-018-0027-0>
- Connaway, L. S., & Powell, R. R. (2010). *Basic research methods for librarians*. Santa Barbara, CA: ABC-CLIO.
- Creswell, J. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Dell, G. (2017). Regression, critical thinking, and the valuation problem today. *Appraisal Journal*, 85(3), 217-230.
- Do, A. Q., & Grudnitski, G. (1992). A neural network approach to residential property appraisal. *The Real Estate Appraiser*, 58(3), 38-45.
- Dugan, J. W. (1999). *Real estate appraisal method and device for standardizing real property marketing analysis by using pre-adjusted appraised comparable sales*. US Patent US5857174A. Retrieved from <https://patents.google.com/patent/US5857174A/en>.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage.



- Gall, M. D., Gall, J. P., & Borg, W. R. (2015). *Applying educational research: How to read, do, and use research to solve problems of practice*. Boston, MA: Pearson/Allyn & Bacon.
- Goldberg, M. A., & Mark, J. (1988). Multiple regression analysis and mass assessment: A review of the issues. *Appraisal Journal*, 56(1), 89-109.
- Guo, Z. (2017). Housing dynamics, empirical facts and the business cycle. In *Global Conference on Business and Finance Proceedings*, 12(2), 46-53. <https://doi.org/10.2139/ssrn.3013790>
- Isakson, H. R. (2001). Using multiple regression analysis in real estate appraisal. *Appraisal Journal*, 69(4), 424-434.
- Kubus, M. (2016). Assessment of predictor importance with the example of the real estate market. *Folia Oeconomica Stetinensia*, 16(2), 29. <https://doi.org/10.1515/fofi-2016-0023>
- Ling, D. (2013). *Real estate principles: A value approach*. New York, NY: McGraw Hill.
- Liu, Z., Wang, P., & Zha, T. (2011). *Land-price dynamics and macroeconomic fluctuations*. Retrieved March 24, 2019 from <https://www.nber.org/papers/w17045>. <https://doi.org/10.3386/w17045>
- Manganelli, B., De Paola, P., & Del Giudice, V. (2018). A multi-objective analysis model in mass real estate appraisal. *International Journal of Business Intelligence and Data Mining*, 13(4), 441-455. <https://doi.org/10.1504/IJBIDM.2018.094981>
- Murphy, L. T., III. (1989). Determining the appropriate equation of multiple regression analysis. *Appraisal Journal*, 57(4), 498-517.
- Narula, S. C., Wellington, J. F., & Lewis, S. A. (2012). Valuating residential real estate using parametric programming. *European Journal of Operational Research*, 217(1), 120-128. <https://doi.org/10.1016/j.ejor.2011.08.014>
- Nayak, B. K. (2010). Understanding the relevance of sample size calculation. *Indian Journal of Ophthalmology*, 58(6), 469-470. <https://doi.org/10.4103/0301-4738.71673>
- Newell, G. J. (1982). The application of ridge regression to real estate appraisal. *Appraisal Journal*, 50(1), 116-119.
- Newsome, B. A., & Zietz, J. (1992). Adjusting comparable sales using multiple regression analysis--the need for segmentation. *Appraisal Journal*, 60(1), 129-135.
- Nguyen, N., & Cripps, A. (2001). Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, 22(3), 313-336. <https://doi.org/10.5555/rees.22.3.r47443706m17v575>
- Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. *Journal of Property Investment & Finance*, (4), 383-401. <https://doi.org/10.1108/14635780310483656>
- Pasymowski, E. (2007). How to discredit most real estate appraisals in one minute. *TriState REALTORS Commercial Alliance Newsletter Spring, 2007*.
- Ragsdale, C. (2012). *Spreadsheet Modeling and Decision Analysis: A Practical Introduction to Business Analytics*. Stamford, CT: Cengage Learning.
- Ratcliff, R. U. (1975). Appraisal is market analysis. *Appraisal Journal*, 43(4), 485-490.
- Smith, H. C. (1986). Inconsistencies in appraisal theory and practice. *Journal of Real Estate Research*, 1(1), 1-17.
- Soper, D. (2019). *A-priori sample size calculator for multiple regression*. Retrieved March 25, 2019 from <https://www.danielsoper.com/statcalc/calculator.aspx?id=1>
- Stake, R. E. (2010). *Qualitative research: Studying how things work*. New York: Guilford Press.
- Status of study concerning appraisal methods and the home valuation code of conduct*. (2010). Stonehenge International.
- Vogt, W. P. (2007). *Quantitative research methods for professionals in education and other fields*. New York, NY: Pearson
- Yin, R. K. (1984). *Case study research: Design and methods*. Beverly Hills, CA: Sage Publishing.
- Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools*, 13(1), 41 – 47.
- Zainal, Z. (2007). Case study as a research method. *UTM Jurnal Kemanusiaan*, 5(1), 1-6.