Selection of a Property Size Variable for Multi-Family Property Valuation

Note: being submitted for publication in The Appraisal Journal

Ason Okoruwa, Ph.D., MAI, AI – GRS

Jim L. Sanders, MBA

Abstract

In valuing a multi-family residential property, several property productivity characteristics are analyzed. Property productivity characteristics examined in this study are size variables: gross building area, number of units, total number of rooms, and unit mix – studio, one bedroom, etc. It is reasonable to expect correlation between any two characteristics. In predicting the expected selling price of real property, regression analysis, a technique of the sales comparison approach, may be employed. Linear and nonlinear models may be specified in terms of the form of the predicted variable (expected sale price) and predictor variables. Specifically, multiple regression models may be estimated such as using the predicted variable sale price, in the original units, or using the natural log of sale price. Similarly, predictor variables may be in their original units or log-transformed. Considering theory supported variables in an equation, the best functional form would be the form that provides the best linear unbiased estimate (BLUE).

**Introduction**

All models make assumptions for simplicity. This is true even in physics where conditions that exist, such as friction, are assumed to not exist for the basic Newton's law of motion. Real estate by its very nature is more complex because of the multiple variables; some of which have non-linear and other complex relationships with the predicted variable; some of which are difficult to quantify and also involves transactions between people. The quantum physicist Richard Feynman is reported to have said "imagine how much harder physics would be if electrons had feelings." Thus, a regression analysis model for real estate is simplified since not all variables can be added to the model due to practical and statistical reasons.

When some sample data (it is a sample of data because the characteristics and range of the data is determined by the appraiser) the parameter estimates (coefficients in the equation) and any estimate of value (from the left side of the equation) are mean values with a confidence interval for the property value. Thus, all parameter estimates (coefficients) and an estimate of value will use the confidence interval and not the prediction interval as a measure of variability. Also, since all the inputs for a specific property being appraised cannot be included into the model, the predicted means from the model are only predictions for a property with those inputs. This means that the appraiser must still use judgment applying the results of a model to provide value estimates for the property being appraised.

A numerical measure for selecting the best size variable of the four size variables (gross building area, number of units, total number of rooms), and unit mix variables (studio, one bedroom, etc.) may be the coefficient of determination (r-squared) of the estimated equations using each variable, if all the models have the predicted variable in the same functional form and the same number of predictor variables; while adjusted r-squared is the numerical measure of best fit if the number of predictor variables are different. Even when the number of predictor variables are the same and the predicted variable is in different units, using the computer-generated r-squared to choose between linear and nonlinear estimated models would be in error. Essentially, the problem would be comparing apples to oranges. To put linear and nonlinear models on the same footing, this paper presents a way to derive the r-squared for a nonlinear estimated model to ensure that one is not comparing an apple to an orange.

## Prior Research

Decker, Nielsen, and Sindt (2005) estimated three models; namely, a linear, a double-log, and a semi-log. They reported adjusted r-squared of 0.734, 0.767, and 0.771 for the linear, double-log, and semi-log models, respectively. On the basis of the highest adjusted r-squared, the semi-log model was selected as the best model.[1] In estimating depreciation for a large distribution warehouse, Ramsland and Markham (1998) estimated a straight-line, exponential, logarithmic, and power regression models. They reported r-squared of 68.3%, 87.4%, 72.1%, and 77.7% for the straight-line, exponential, logarithmic, and power functions, respectively. They concluded that the exponential specification, with the highest r-squared was the best model.[2] In these two papers, selection of the best was made based on the highest r-squared or adjusted r-squared ignoring the fact that the dependent variables were stated in different units.

## Specification of Models

In this study, our focus is on selecting the best property size variable for valuing multifamily properties. We analyze gross building area, number of apartment units, number of rooms, and unit mix. The five model specifications below are considered.

Linear Model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta p X pi + \varepsilon_i$ (1)

Log-Log (Double Log or Exponential) Model: $Ln(Y_i) = \beta_0 + \beta_1 ln(X_{1i}) + \beta_2 ln(X_{2i}) + \cdots +$

$\beta p ln(X pi) + \varepsilon_i$ (2)

[1] Christopher S. Decker, PhD, Donald A. Nielsen, PhD, and Roger P. Sindt, PhD, "Is Pollution a Homogeneous Determinant of Value?" *The Appraisal Journal* (Spring 2005) 183-197.
[2] Maxwell O. Ramsland, Jr., MAI and Daniel E. Markham, "Market-Supported Adjustments Using Multiple Regression Analysis," *The Appraisal Journal* (April 1998): 181-191.

Semi-log Model: $Y_i = \beta_0 + \beta_1 \ln(X_{1i}) + \beta_2 \ln(X_{2i}) + \cdots + \beta p \ln(Xpi) + \varepsilon_i$ (3)

Semi-log Model: $Ln(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta pXpi + \varepsilon_i$ (4)

Combination Functional Form: $Ln(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 \ln(X_{2i}) + \cdots + \beta pXpi + \varepsilon_i$ (5)

where:

$Y_i$ = the ith observation of the predicted variable; i is from 1 to n

$X_1$ - Xp = predictor variable 1 to predictor variable p

$X_{1i}$ - Xpi = the ith observation of predictor variable 1 to predictor variable p

$Ln(Y_i)$ = the natural log of ith observation of the predicted variable; i is from 1 to n

$\ln(X_1)$ - ln(Xp) = the natural log of predictor variable 1 to natural log of predictor variable p

$\ln(X_{1i})$ - ln(Xpi) = the ith observation of the natural log of predictor variable 1 to natural log of predictor variable p; i is from 1 to n

$\beta_0$ = intercept or constant term

$\beta_1 - \beta p$ = regression coefficient for predictor variable 1 to predictor variable p

$\varepsilon_i$ = the ith observation of the random error term

A linear model (Model 1) specification is appropriate when the predictor variables are linearly related to the predicted variable. The coefficients or slopes of the model are assumed to be constant. In a log-log model (Model 2), the natural logs of the predicted variable and predictor variables are used. In a Log-Log specification, the coefficients or slopes of the model are interpreted as elasticities and are constant. In a semi-log model (Model 3), the predictor variables are in natural log form and the other form of a semi-log model (Model 4), the predicted variable is in log form. In the combination functional form (Model 5) both the predicted variable and some of the predictor variables are log-transformed.

**Case Study**

Data on sixty-two multi-family transactions in Tucson City, Arizona, were gathered for this study. Summary statistics of the data set is presented below.

**Exhibit 1: Summary Statistics**

| Variable | Count | Minimum | Mean | Median | Maximum | Standard Deviation |
|---|---|---|---|---|---|---|
| Sale Price | 62 | $1,000,000 | $9,845,281 | $6,903,750 | $37,250,000 | $8,956,283 |
| Sale Date | 62 | 9/27/2017 | 10/19/2015 | 2/27/2016 | 1/18/2011 | N/A |
| Sale Age (days) | 62 | 7.00 | 715.13 | 584.50 | 2451.00 | N/A |
| Building SF | 62 | 21,378 | 128,382 | 98,092 | 401,766 | 99,377 |
| Year Built | 62 | 2013 | 1981 | 1980 | 1965 | N/A |
| Building Age (years) | 62 | 4 | 36 | 37 | 52 | 9 |
| Building Class B | 1 | | | | | |
| Building Class B | 37 | | | | | |
| Building Class C | 24 | | | | | |
| Number of Units | 62 | 50 | 162 | 142 | 368 | 94 |
| Land SF | 62 | 37,997 | 303,917 | 227,383 | 1,258,709 | 276,528 |
| Land to Building Ratio | 62 | 0.85 | 2.32 | 2.34 | 3.64 | 0.64 |
| Studios Units | 62 | 0 | 10 | 0 | 87 | 20 |
| 1-Bedroom Units | 62 | 0 | 76 | 65 | 235 | 58 |
| 2-Bedroom Units | 62 | 0 | 65 | 48 | 253 | 58 |
| 3-Bedroom Units | 62 | 0 | 8 | 0 | 64 | 14 |
| Total Number of Rooms | 62 | 60 | 392 | 316 | 1004 | 249 |
| Parking Spaces | 62 | 48 | 233 | 190 | 572 | 145 |

Information on the multifamily property to be appraised follows: gross building area: 129,336 sq. ft., year built: 1987, age: 30 years, land area: 384,199 sq. ft., land to building ratio: 2.97, building class A: 1, building class B: 37, building class C: 24, number of floors: 2, number of units: 144, studio units: 0, 1-bedroom units: 48, 2-bedroom units: 72, 3-bedroom units: 24, parking spaces: 264, and effective date of appraisal: 9/29/2017.

**Correlation Analysis**

On the following page is Exhibit 2: Correlation Report, a correlation report on the predicted variable and predictor variables. For example, a review of Exhibit 2 shows that the correlation coefficients between gross building area and number of units and total number of rooms are 0.789251 and 0.908099, respectively. The values of correlation coefficients can range from -1 to +1. Values of correlation coefficient are interpreted with reference to -1, 0, or +1. The closer to -1 or + 1, the higher the relationship. The positive values indicate that the predictor variables move in the same direction. Because this is a multivariate model, a predictor variable may not have a high correlation when compared directly with the predicted variable, but could have a high correlation within the context of the model. For example, square footage may explain much of the variance in a model. However, once square feet are controlled for, other variables can explain other differences that are useful but not highly correlated directly with the predicted variable.

The high values of the coefficient among the four predictor variables of interest indicates that they are all good candidates for predicting the expected selling price for multi-family properties. In a subsequent section, we will use r-squared to measure the magnitude of the variation in the predicted variable that is explained when each of the four candidates of possible predictor variable is included in a hedonic price formula.  Given the high correlation among the four predictor variables, often, only one should be included to avoid multicollinearity problems. Exhibit 2 presents the correlation coefficients for the four property productivity characteristics examined in this study.

| Exhibit 2: Correlation Cofficients of Selected Variables | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SalePrice | BldgSF | NoOfUnits | TotNoOfRms | Studio | 1-Bedroom | 2-Bedroom | 3-Bedroom |
| SalePrice | 1.0000 | | | | | | | |
| BldgSF | 0.8512 | 1.0000 | | | | | | |
| NoOfUnits | 0.6313 | 0.7893 | 1.0000 | | | | | |
| TotNoOfRms | 0.7900 | 0.9081 | 0.9015 | 1.0000 | | | | |
| Studio | -0.1714 | -0.1414 | 0.1747 | -0.1549 | 1.0000 | | | |
| 1-Bedroom | 0.2117 | 0.3566 | 0.7262 | 0.4948 | 0.0205 | 1.0000 | | |
| 2-Bedroom | 0.6561 | 0.7308 | 0.4406 | 0.7274 | -0.343373 | -0.1400 | 1.0000 | |
| 3-Bedroom | 0.4065 | 0.2525 | -0.0203 | 0.2460 | -0.191515 | -0.3373 | 0.2300 | 1.0000 |

Where:

SalePrice = sale price of a multifamily property

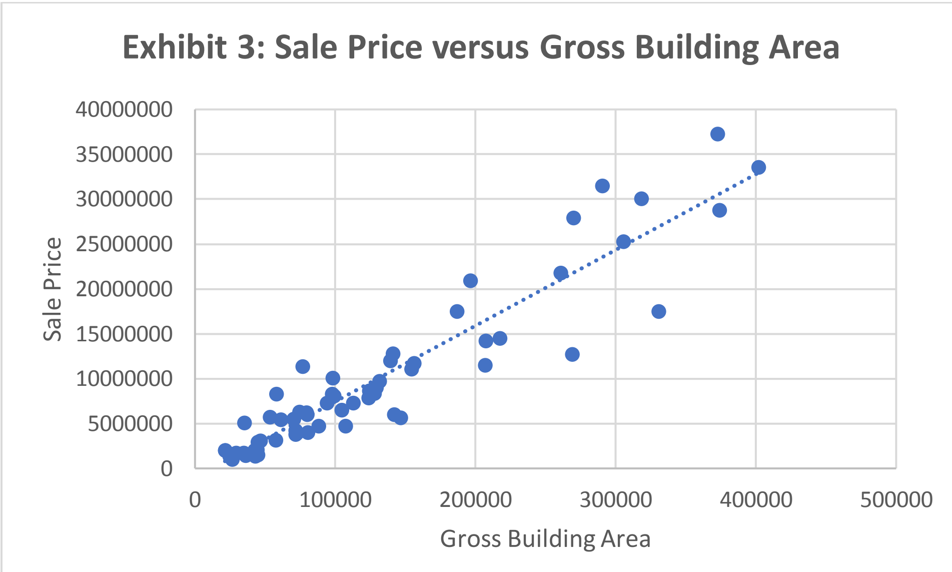
BldgSF = building(s) square footage

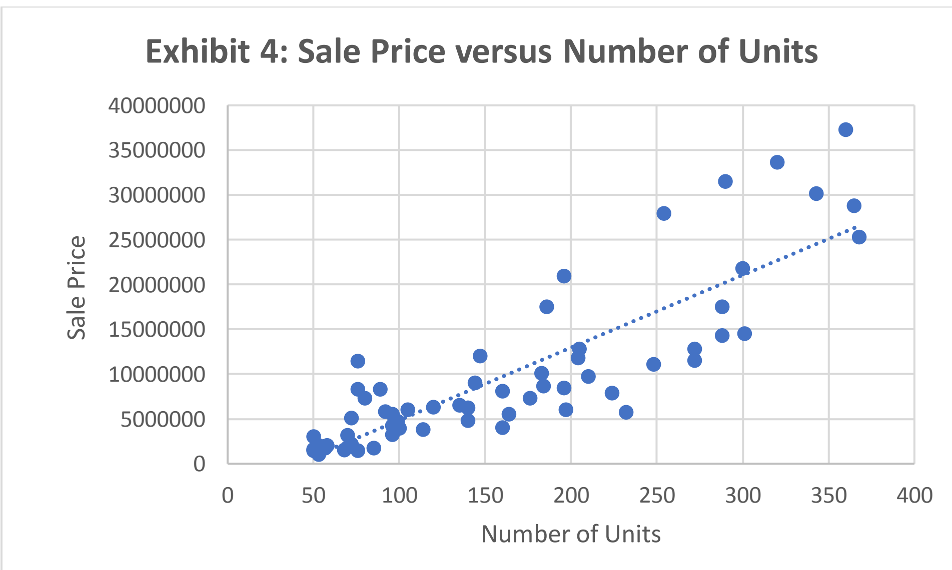
NoOfUnits = total number of units in a multifamily property

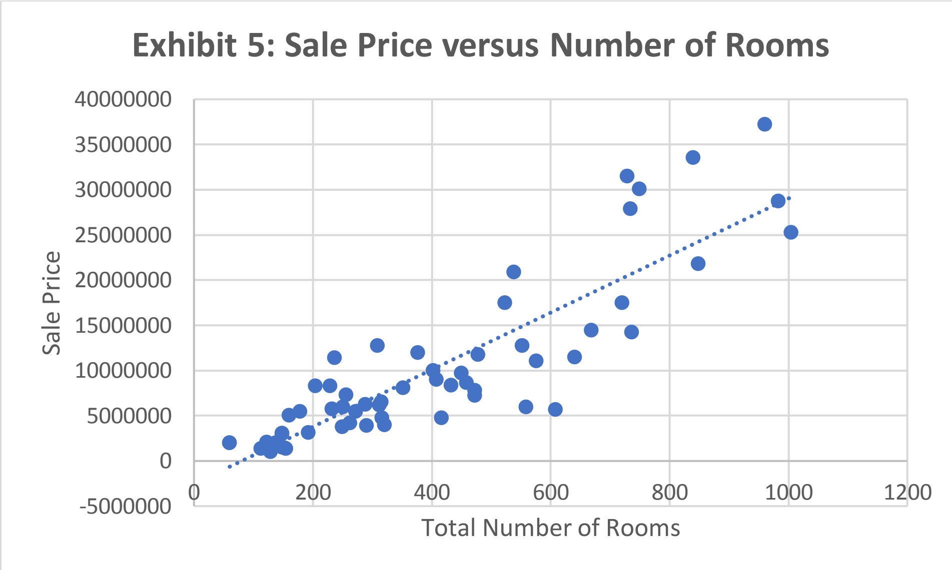
TotNoOfRms = total number of rooms in a multifamily property; studio = one room,

1-bedroom = two rooms; 2-bedroom = three rooms, 3-bedroom = four rooms

## Graphic Analysis






A review of Exhibits 3 – 5 do not indicate a discernable curvilinear relationship between sale price and gross building area, number of units, and total number of rooms. Residual plots show some heteroscedasticity (the dots get further from the regression line as size increases. This heteroscedasticity can be an indication of an omitted variable but here it is more likely a function of predictor variable scale. For example, the range of, say square footage, is from 21,378 square feet to 401,766 which is a large range. This problem can be minimized by transforming this predictor variable by transforming them. If the heteroscedasticity is not cured, the parameter estimates are not affected but standard errors are unequal such that confidence interval for any coefficient prediction changes as a function of the variable. What this means is that different square footage values inputted into the model may have different confidence intervals.

**Estimated Models**

Fourteen models are estimated using the four highly correlated predictor variables: gross building area, number of multi-family units, total number of rooms, and unit or bedroom mix. A linear model and combination functional form were estimated having as a predictor variable only one of the size correlated predictor variables. For the unit mix predictor variable, linear and semi-log models were estimated. The regression models were estimated using NCSS Statistical Software.[3]

The predicted variable is either the natural log of sale price or sale price. The coefficient of determination, r-squared, measures the magnitude of the variation in the predicted variable accounted for by the predictor variable(s) in the hedonic equation. The same predicted variable in two different units will have different sum of squares total (SST) that is used in computing the coefficient of determination. To make the coefficient of determination for a predicted variable stated in natural log unit to be the same variable in its original unit, the remedy is the conversion of the natural logged predicted variable to its original unit. This is done by finding the exponent of the natural-logged variable or its anti-log. To convert back to its original unit, simply finding its anti-log would underestimate the expected sale price. The estimated equation is adjusted by adding ½ of the square of the standard error of the estimated equation to it before finding the antilog of the modified equation as shown below.[4] In equation form, it is stated as:

$$\text{Sale Price} = \exp(\ln(\hat{y}_i) + S_e^2/2) \tag{6}$$

The column of predicted sale price for each observation is converted to sale price by applying Equation (6). Now, the correlation coefficient between the sale price and the predicted sale price is calculated. The coefficient of determination is the square of the correlation coefficient. When the predicted variable is the natural log of sale price, for comparability purposes, the coefficient of determination is derived using the procedure described above.

For example, Exhibit 6 which follows, presents the calculations for deriving the

[3] NCSS. LLC, *NCSS Statistical Software*, Ver. 10, (Kaysville, UT).
[4] Sanjiv Jaggia and Alison Kelly, *Business Statistics: Communicating with Numbers*, (New Delhi, India: McGraw Hill Education, 2013), 501.

coefficient of determination when the predictor variable is in natural log form. Model 4 of the subsequent Exhibit 7 provides the inputs for the calculations in Exhibit 6.

| **Exhibit 6: Calculation of Unlogged Predicted Sale Price and Correlation Coefficient-Derived R-Squared (Input Source: Model 4 of Exhibit 7)** | | | |
|---|---|---|---|
| Actual LnSalePrice | Predicted LnSalePrice | Actual SalePrice | Predicted Sale Price exp(Col. 2 + Se^2/2) |
| $ 14.253765 | $ 14.84688 | $ 1,550,000 | $ 3,032,916 |
| 14.209578 | 14.75816 | 1,483,000 | 2,775,427 |
| 14.508658 | 14.76905 | 2,000,000 | 2,805,816 |
| 14.569282 | 14.86037 | 2,125,000 | 3,074,107 |
| 15.607270 | 15.25896 | 6,000,000 | 4,579,568 |
| 15.150512 | 15.03093 | 3,800,000 | 3,645,795 |
| 15.256293 | 15.13611 | 4,224,000 | 4,050,153 |
| 14.508658 | 15.14399 | 2,000,000 | 4,082,194 |
| 15.656060 | 15.68676 | 6,300,000 | 7,024,504 |
| 15.373655 | 15.05746 | 4,750,000 | 3,743,813 |
| 15.644550 | 15.70725 | 6,227,900 | 7,169,920 |
| 15.607270 | 16.02960 | 6,000,000 | 9,897,127 |
| 15.567316 | 15.05482 | 5,765,000 | 3,733,942 |
| 15.803385 | 15.85455 | 7,300,000 | 8,307,798 |
| 16.012735 | 15.78959 | 9,000,000 | 7,785,278 |
| 15.907375 | 15.60978 | 8,100,000 | 6,504,047 |
| 16.217941 | 16.12900 | 11,050,000 | 10,931,457 |
| 15.973070 | 15.47334 | 8,650,000 | 5,674,512 |
| 16.091752 | 15.86818 | 9,740,000 | 8,421,808 |
| 16.249124 | 15.68923 | 11,400,000 | 7,041,876 |
| 16.123083 | 15.26063 | 10,050,000 | 4,587,222 |
| 16.279364 | 16.11138 | 11,750,000 | 10,740,531 |
| 15.929957 | 15.03553 | 8,285,000 | 3,662,605 |
| 16.489659 | 16.58688 | 14,500,000 | 17,279,565 |
| 16.361042 | 16.86341 | 12,750,000 | 22,783,909 |
| 16.364956 | 16.04116 | 12,800,000 | 10,012,202 |
| 16.300417 | 15.68898 | 12,000,000 | 7,040,115 |
| 17.046315 | 17.08287 | 25,300,000 | 28,375,173 |
| 17.175018 | 17.57090 | 28,775,000 | 46,226,099 |
| 17.330037 | 17.40549 | 33,600,000 | 39,178,752 |
| 17.433163 | 17.67435 | 37,250,000 | 51,264,298 |
| 17.265498 | 17.19825 | 31,500,000 | 31,845,452 |
| 17.219703 | 17.39881 | 30,090,000 | 38,917,910 |

Exhibit 6 (Continued)

| | | | |
|---|---|---|---|
| 17.144137 | 17.03882 | 27,900,000 | 27,152,376 |
| 16.897421 | 16.92192 | 21,800,000 | 24,156,767 |
| 16.855260 | 16.54329 | 20,900,000 | 16,542,529 |
| 16.679139 | 17.24853 | 17,525,000 | 33,487,578 |
| 16.677711 | 16.41326 | 17,500,000 | 14,525,483 |
| 16.472267 | 15.97895 | 14,250,000 | 9,408,321 |
| 16.257858 | 16.27527 | 11,500,000 | 12,653,256 |
| 15.943742 | 15.85427 | 8,400,000 | 8,305,472 |
| 15.934173 | 15.29030 | 8,320,000 | 4,725,364 |
| 15.874749 | 15.84840 | 7,840,000 | 8,256,862 |
| 15.800985 | 15.94971 | 7,282,500 | 9,137,205 |
| 15.691152 | 15.42216 | 6,525,000 | 5,391,397 |
| 15.553342 | 15.96020 | 5,685,000 | 9,233,559 |
| 15.520259 | 15.57602 | 5,500,000 | 6,288,135 |
| 15.515703 | 15.50091 | 5,475,000 | 5,833,135 |
| 15.444751 | 15.41139 | 5,100,000 | 5,333,643 |
| 15.377498 | 15.17028 | 4,768,289 | 4,190,938 |
| 15.201805 | 15.02218 | 4,000,000 | 3,614,034 |
| 15.184585 | 14.93619 | 3,931,711 | 3,316,250 |
| 14.973963 | 14.75357 | 3,185,000 | 2,762,717 |
| 14.953344 | 14.89735 | 3,120,000 | 3,189,916 |
| 14.907434 | 14.94560 | 2,980,000 | 3,347,603 |
| 14.508658 | 14.93519 | 2,000,000 | 3,312,935 |
| 14.346139 | 14.75733 | 1,700,000 | 2,773,124 |
| 14.346139 | 14.86852 | 1,700,000 | 3,099,264 |
| 14.288634 | 14.78372 | 1,605,000 | 2,847,281 |
| 14.151983 | 14.65840 | 1,400,000 | 2,511,913 |
| 14.151983 | 14.78629 | 1,400,000 | 2,854,608 |
| 13.815511 | 14.53611 | 1,000,000 | 2,222,771 |
| Correlation Coefficient between | | | |
| Actual Sale Price and Predicted Sale Price | | | 0.9380 |
| Correlation Coefficient-Drived R-squared | | | 0.8799 |
| Computer Generated R-squared | | | 0.8306 |
| | | | |
| where: | | | |
| | | | |
| Se^2 is the standard error of the estmimated model squared | | | |

Starting on the next page, Exhibit 7 to 10, present the estimated models.

**Exhibit 7: Gross Building Area Regression**

| | Predicted Variable: SalePrice | | | | Predicted Variable: LnSalePrice | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1: Predictor Variables: BldgSF & Untransformed Var. | | Model 3: Predictor Variables: LnBldgSF & Untransformed Var. | | Model 4: Predictor Variables: BldgSF & Untransformed Var. | | Model 5: Predictor Variables: LnBldgSF & Untransformed Var. | |
| Predictor Variables | Regression Coefficient | P-Value | Regression Coefficient | P-Value | Regression Coefficient | P-Value | Regression Coefficient | P-Value |
| Intercept | 3676925 | 0.0846 | -85106320 | 0.0000 | 14.92045 | 0.0000 | 4.8759890 | 0.0000 |
| BldgAge | -96591.41 | 0.0452 | -143294.6 | 0.0534 | -0.006007828 | 0.3390 | -0.006669061 | 0.1536 |
| BldgSF | 78.80862 | 1.0000 | N/A | N/A | 6.804872E-06 | 1.0000 | N/A | N/A |
| LnBldgSF | N/A | N/A | 8769475 | 0.0000 | N/A | N/A | 0.9577456 | 0.0000 |
| BldgClassB=1 | 565557.1 | 0.5293 | 392122.1 | 0.7832 | 0.4551168 | 0.0003 | 0.3140664 | 0.0009 |
| SaleAge | -1169.015 | 0.0835 | -1080.023 | 0.3004 | -0.0001804704 | 0.0441 | -0.0001317713 | 0.0494 |
| | | | | | | | | |
| Correlation Coefficient | N/A | | N/A | | 0.9380 | | 0.9485 | |
| Correlation Coefficient-Derived $R^2$ | N/A | | N/A | | 0.8799 | | 0.8996 | |
| Computer-Generated $R^2$ | 0.8956 | | 0.7495 | | 0.8306 | | 0.9061 | |
| F-Ratio | 122.308 | 0.0000 | 42.628 | 0.0000 | 69.871 | 0.0000 | 137.442 | 0.0000 |

**Exhibit 8: Number of Units Regression**

| | Predicted Variable: SalePrice | | | | Predicted Variable: LnSalePrice | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1: Predictor Variables: NoOfUnits & Untransformed Var. | | Model 3: Predictor Variables: LnNoOfUnits & Untransformed Var. | | Model 4: Predictor Variables: NoOfUnits & Untransformed Var. | | Model 5: Predictor Variables: LnNoOfUnits & Untransformed Var. | |
| Predictor Variables | Regression Coefficient | P-Value | Regression Coefficient | P-Value | Regression Coefficient | P-Value | Regression Coefficient | P-Value |
| Intercept | 6448833 | 0.0380 | -31220690 | 0.0001 | 15.07293 | 0.0000 | 10.7447700 | 0.0000 |
| BldgAge | -216452.4 | 0.0023 | -228406.9 | 0.0073 | -0.01583925 | 0.0202 | -0.015944270 | 0.0139 |
| NoOfUnits | 75248.24 | 0.0000 | N/A | N/A | 7.022385E-03 | 0.0000 | N/A | N/A |
| LnNoOfUnits | N/A | N/A | 1.022769+07 | 0.0000 | N/A | N/A | 1.1203200 | 0.0000 |
| BldgClassB=1 | -619403.2 | 0.6547 | -223576.4 | 0.8955 | 0.3072802 | 0.0264 | 0.2448357 | 0.0643 |
| SaleAge | -992.9391 | 0.3152 | -1210.807 | 0.3125 | -0.0001521465 | 0.1182 | -0.0001455153 | 0.1153 |
| | | | | | | | | |
| Correlation Coefficient | N/A | | N/A | | 0.9120 | | 0.9002 | |
| Correlation Coefficient-Derived $R^2$ | N/A | | N/A | | 0.8318 | | 0.8103 | |
| Computer-Generated $R^2$ | 0.7750 | | 0.6695 | | 0.7995 | | 0.8196 | |
| F-Ratio | 49.090 | 0.0000 | 28.871 | 0.0000 | 56.816 | 0.0000 | 64.76 | 0.0000 |

**Exhibit 9: Total Number of Rooms Regression**

| | Predicted Variable: SalePrice | | | | Predicted Variable: LnSalePrice | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1: Predictor Variables: | | Model 3: Predictor Variables: | | Model 4: Predictor Variables: | | Model 5: Predictor Variables: | |
| | ToNoOfRms & Untransformed Var. | | LnToNoOfRms & Untransformed Var. | | ToNoOfRms & Untransformed Var. | | LnToNoOfRms & Untransformed Var. | |
| Predictor Variables | Regression Coefficient | P-Value | Regression Coefficient | P-Value | Regression Coefficient | P-Value | Regression Coefficient | P-Value |
| Intercept | 6040178 | 0.0343 | -3.35E+07 | 0.0000 | 15.04727 | 0.0000 | 10.2973900 | 0.0000 |
| BldgAge | -194993.8 | 0.0028 | -200785.3 | 0.0177 | -0.0139252 | 0.0310 | -0.01257152 | 0.0333 |
| ToNoOfRms | 28957.3 | 0.0000 | N/A | N/A | 2.674294E-03 | 0.0000 | N/A | N/A |
| LnToNoOfRms | N/A | N/A | 8769764 | 0.0000 | N/A | N/A | 0.9944662 | 0.0000 |
| BldgClassB=1 | 32423.98 | 0.9791 | 1199864 | 0.4592 | 0.3738069 | 0.0040 | 0.382596 | 0.0013 |
| SaleAge | -881.6923 | 0.3316 | -982.9462 | 0.4122 | -0.0001436967 | 0.1197 | -0.0001132591 | 0.1800 |
| | | | | | | | | |
| Correlation Coefficient | N/A | | N/A | | 0.9112 | | 0.9034 | |
| Correlation Coefficient-Derived $R^2$ | N/A | | N/A | | 0.8303 | | 0.8161 | |
| Computer-Generated $R^2$ | 0.8103 | | 0.6716 | | 0.8199 | | 0.8498 | |
| F-Ratio | 60.872 | 0.0000 | 29.144 | 0.0000 | 64.886 | 0.0000 | 80.600 | 0.0000 |

**Exhibit 10: Bedroom Mix Regression**

| | Predicted Variable: SalePrice | | Predicted Variable: LnSalePrice | |
|---|---|---|---|---|
| | Model 1: Predictor Variables | | Model 4: Predictor Variables: | |
| | Unit Mix Var. &Untransformed Var. | | Unit Mix Var. & Untransformed Var. | |
| Predictor Variables | Regression Coefficient | P-Value | Regression Coefficient | P-Value |
| Intercept | 5941877 | 0.0533 | 14.98071 | 0.0000 |
| (BldgClassB=1) | 140823.9 | 0.9105 | 0.3896567 | 0.0048 |
| BldgAge | -194334.2 | 0.0045 | -1.217741E-02 | 0.0858 |
| SaleAge | -975.2624 | 0.2788 | -0.0001393951 | 0.1466 |
| Studios | 69058.88 | 0.0123 | 0.001727902 | 0.5443 |
| 1Bedroom | 42193.88 | 0.0002 | 0.005242722 | 0.0000 |
| 2Bedrooms | 100309.4 | 0.0000 | 0.007836548 | 0.0000 |
| 3Bedrooms | 118023.8 | 0.0088 | 0.01326227 | 0.0057 |
| | | | | |
| Correlation Coefficient | | | 0.9042 | |
| Correlation Coefficient-Derived $R^2$ | | | 0.8176 | |
| Computer-Generated $R^2$ | 0.8300 | | 0.8218 | 0.8218 |
| F-Ratio | 37.669 | 0.0000 | 35.573 | 0.0000 |

| | | | | |
|---|---|---|---|---|
| **Exhibit 11: Cofficients of Determination** | | | | |
| | Model 1 | Model 3 | Model 4 | Model 5 |
| Predictor Variable | Predicted Variable: SalePrice - Computer Generated R-Squared | Predicted Variable: SalePrice - Computer Generated R-Squared | Predicted Variable: LnSalePrice - Correlation Coefficient-Derived R-Squared | Predicted Variable: LnSalePrice - Correlation Coefficient-Derived R-Squared |
| BldgSF | 0.8956 | | 0.8799 | |
| LnBldgSF | | 0.7495 | | 0.8996 |
| NoOfUnits | 0.7750 | | 0.8318 | |
| LnNoOfUnits | | 0.6695 | | 0.8103 |
| ToNoOfRms | 0.8103 | | 0.8303 | |
| LnToNoOfRms | | 0.6716 | | 0.8161 |
| Unit Mix | 0.8300 | | 0.8176 | |

**Selection of Best Model**

A review of Exhibit 9 shows that Model 5 for the LnBldgSF with LnSalePrice as predicted variable, has the highest coefficient of determination. However, the coefficient of determination of the specification of Model 1 (0.8956) with SalePrice as the predicted variable and BldgSF as predictor variable, is very close to the Model 5 specification (0.8996) for LnBldgSF.

**Conclusions**

This study examined the best property size productivity feature among four productivity features to predict the expected selling price of a multi-family residential property. The four property productivity features examined are gross building area, number of units, total number of rooms, and bedroom mix. For each of the four characteristics, except unit mix, four different specifications of the regression models are examined with the predicted variable as SalePrice and the LnSalePrice with predictor variable in original unit or log-transformed. For the unit mix predictor variable, two models were estimated with the predicted variable in original unit or log-transformed. In order to compare models based on coefficient of determination, it was necessary to derive the coefficient of determination of a model specified as the LnSalePrice as the predicted variable, based on the correlation coefficient of the actual sale price and the unlogged predicted sale price. This makes sure that one is not comparing apples to oranges. In addition, the paper shows how to add a correction factor for the under prediction of expected sale price if the predicted variable was in its natural log form in the estimated model. The specification with the natural log of sale price as the predicted variable and the natural log of gross building area as the predictor variable was the best property productivity characteristic among gross building area, number of units, total number of rooms and unit room mix, based on coefficient of determination measure. Given the high collinearity between gross building area and number of units, total number of rooms, and unit mix, gross building area serves as a proxy for these variables.

**A. Ason Okoruwa, PhD, MAI, AI – GRS,** is the President of Bedrock Valuation and Consulting Services Corp., Omaha, Nebraska. He has been engaged in valuation of real property, consulting, market analysis, and litigation support, since 1996. He previously taught at North Carolina Central University, Durham, North Carolina and The University of Northern Iowa, Cedar Falls, Iowa. He has published in The Appraisal Journal, The Journal of Real Estate Research, Journal of Property Research, Journal of Urban Economics, Communications in Statistics, European Journal of Operational Research, and Journal of Information and Optimization Sciences, He received his BBA in health administration and MBA in finance from Georgia State University, Atlanta, Georgia, and PhD in business, with a concentration in real estate, from The University of Georgia, Athens, Georgia. **Contact: ason@bedrockvacs.com**

**B. Jim L. Sanders, MBA**, is the president of Real Estate Appraisal Litigation, LLC, based in Tucson, Arizona. He has been an appraiser since 1975. For past 25 years, his business focus has been expert witness work involving Eminent Domain, including issues involving high voltage transmission lines, cell phone towers, noise, stigma, etc. Mr. Sanders graduated from The University of Arizona with an MBA in Decision Science, including 15 units graduate level statistics and classes in market research. Undergraduate degree is in Chemistry and Mathematics. He has published in the Journal of Real Estate Research. Formerly, Mr. Sanders was an instructor for the Appraisal Institute, for the seminar he developed titled ***Mathematically Modeling Real Estate Data*** that focused on the statistical analysis of real estate data. **Contact: real@cox.net**