RESIDENTIAL MARKET ANALYSIS WITH STATISTICAL MODELS AND GIS

AUTHOR

Patrick M.O'Connor, ASA is president of O'Connor Consulting Incorporated, which provides valuation and location analysis consulting services in real estate valuation, CAMA and AVM modeling for residential, apartment, commercial and industrial properties.

ABSTRACT

Most computer-assisted mass appraisal (CAMA) and automated valuation model (AVM) systems are enhanced through better market analysis developed through the use of statistical software and geographic information systems (GIS). This article will demonstrate effectiveness of market analysis using statistics for data review, market models and review of model results statistical quality. GIS enhances data review and contributes to market and locational analysis.

There are several sources of statistical software and GIS. SPSS statistical software and ESRI's ArcGIS are used to perform this market analysis. This article advocates use of statistical software and GIS by appraisers to enhance their real estate market analysis abilities.

BACKGROUND

Most North American assessment jurisdictions use computer-assisted mass appraisal (CAMA) systems to provide estimates of market value. Automated valuation model (AVM) systems provide estimates of market value for use by financial institutions and other users of private-sector valuation services. While CAMA and AVM systems are designed to provide services to different clients, there are many similarities between these property valuation software systems. As with CAMA, AVM must have some appraiser/analysts, who perform market analysis to develop their valuation formulas. CAMA/AVM systems are designed to calculate estimates of valuation for multiple properties. CAMA/AVM systems support programs that use formulas or tables to rapidly calculate estimates of real estate value. CAMA/AVM systems calculate real estate property values one property at a time. CAMA/AVM systems provide some level of statistical quality.

Forty-years of industry experience with statistical models within mass appraisal profession have proven that appraisers can learn use of statistical and GIS software to enhance their market analysis skills. Most appraisers (mass or single property) perform market analysis in preparation of estimates (opinions) of market value. Appraisers know or have the skills to analyze communities that they serve. Using their existing appraisal experience and adding some statistical training, appraisers may develop accurate statistical models of their markets. Statistical software and geographic information systems (GIS) provide appraisers with additional (supplementary) analysis tools to support their professional opinions.

DATA ANALYSIS

No matter how good system edit programs are, all CAMA/AVM systems databases and any group of single property appraisal reports contain errors. Range edits will only identify some errors. Data analysis provides better understanding of data and provides comparisons of variables that are not possible from range edits. Use of statistical and GIS software for analysis data will provide an entirely different set of actual/potential data errors.

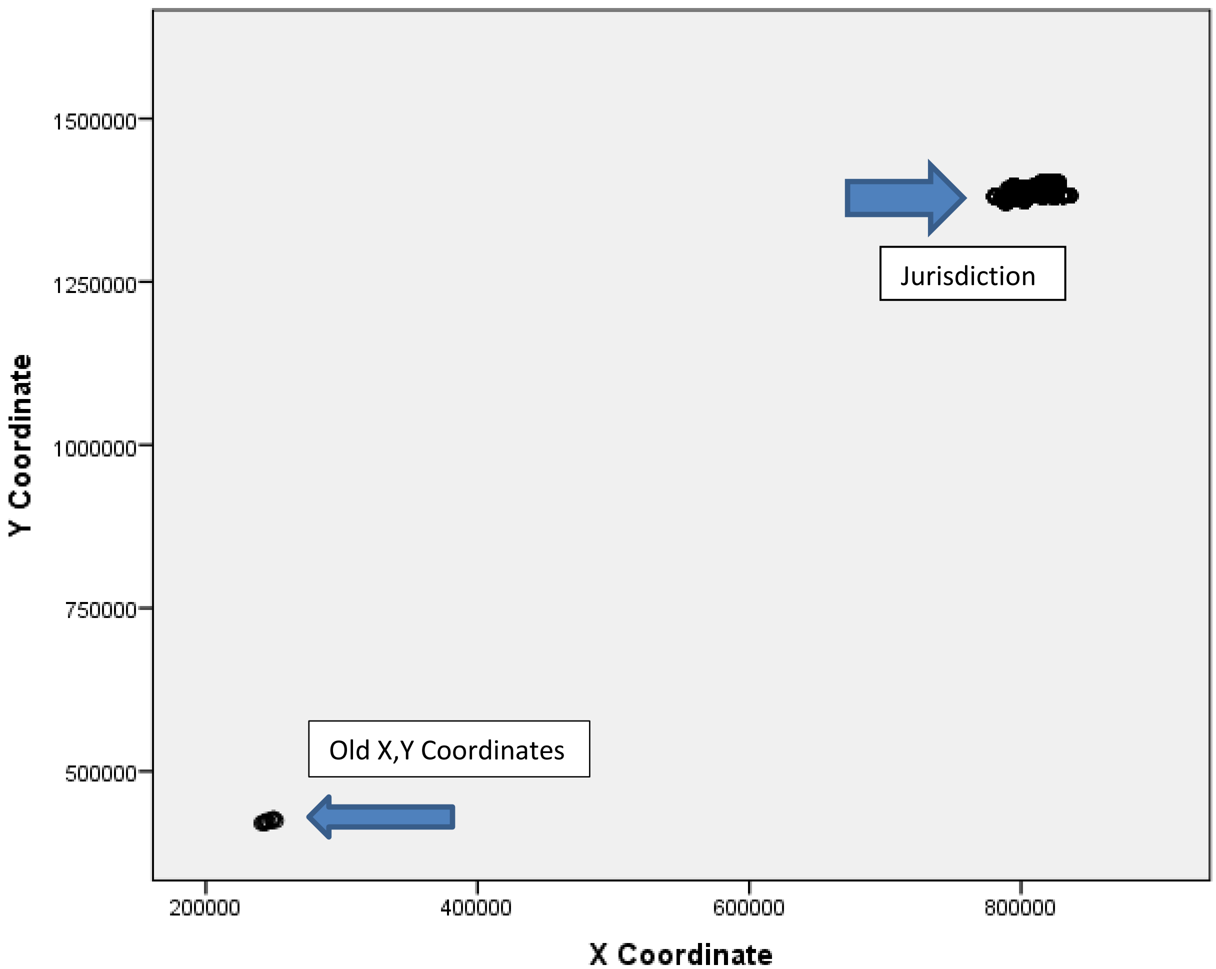
Basic data edits are variable value ranges. For example, in an improved property sales file, simple descriptive statistics of minimum and maximum range will highlight missing or data extremes, CHART 1 presents the range of adjusted sales price (ADJ_SP) starting at $1.00. This is obviously wrong information. CHART 1 also has zeros under neighborhoods numbers (NH) and year built (yrblt). This indicates one or more missing neighborhood numbers. Year built of zero is impossible and indicates one or more missing year built.

CHART 1
SET OF DESCRIPTIVE STATISTICS

Variable	Case	Minimum	Maximum
ADJ_SP	6,577	1	2,390,314
NH	6,577	0	111
GFLA	6,577	147	11,261
TOTAL_SF	6,577	402	6,709
yrblt	6,577	0	2,011

There are also data errors that come from third parties. In this community, the geographic information system (GIS) manager changed map datum (scale measurement), the assessor found two different X,Y scales within historical sales transactions file. GRAPH 1 presents sales properties by X,Y coordinates using statistical software to get measurements in graphic format. The larger blob (upper right) is current assessment jurisdiction sales transactions. But a few sales transactions are in prior scale (lower left).

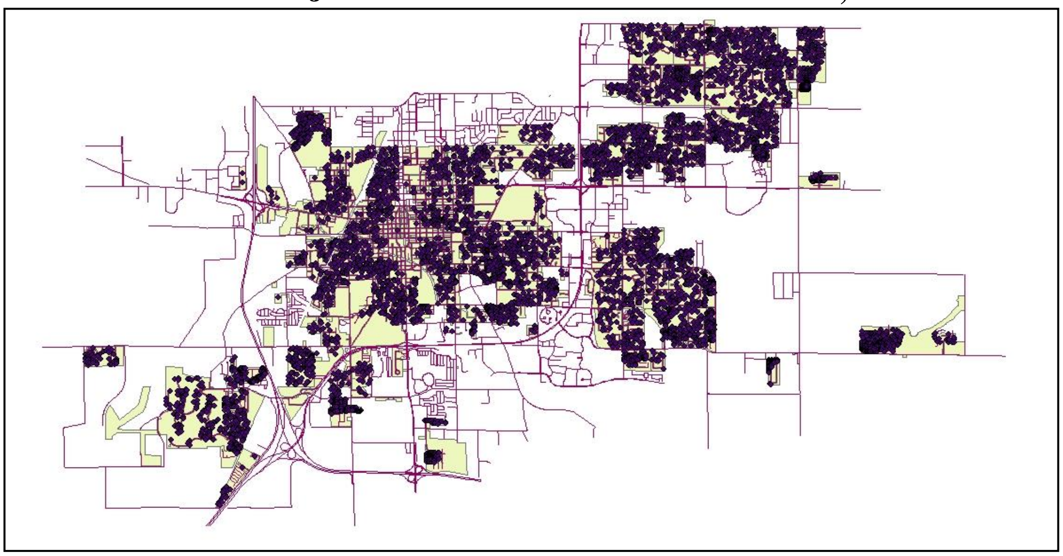
GRAPH 1
ERRORS IN X,Y COORDINATES, CURRENT AND OLD SALES TRANSACTIONS



Once X,Y coordinate errors are corrected, this assessment jurisdiction comes into view as in GIS map (GRAPH 2), which presents street centerlines, neighborhoods, and sales points (black dots).

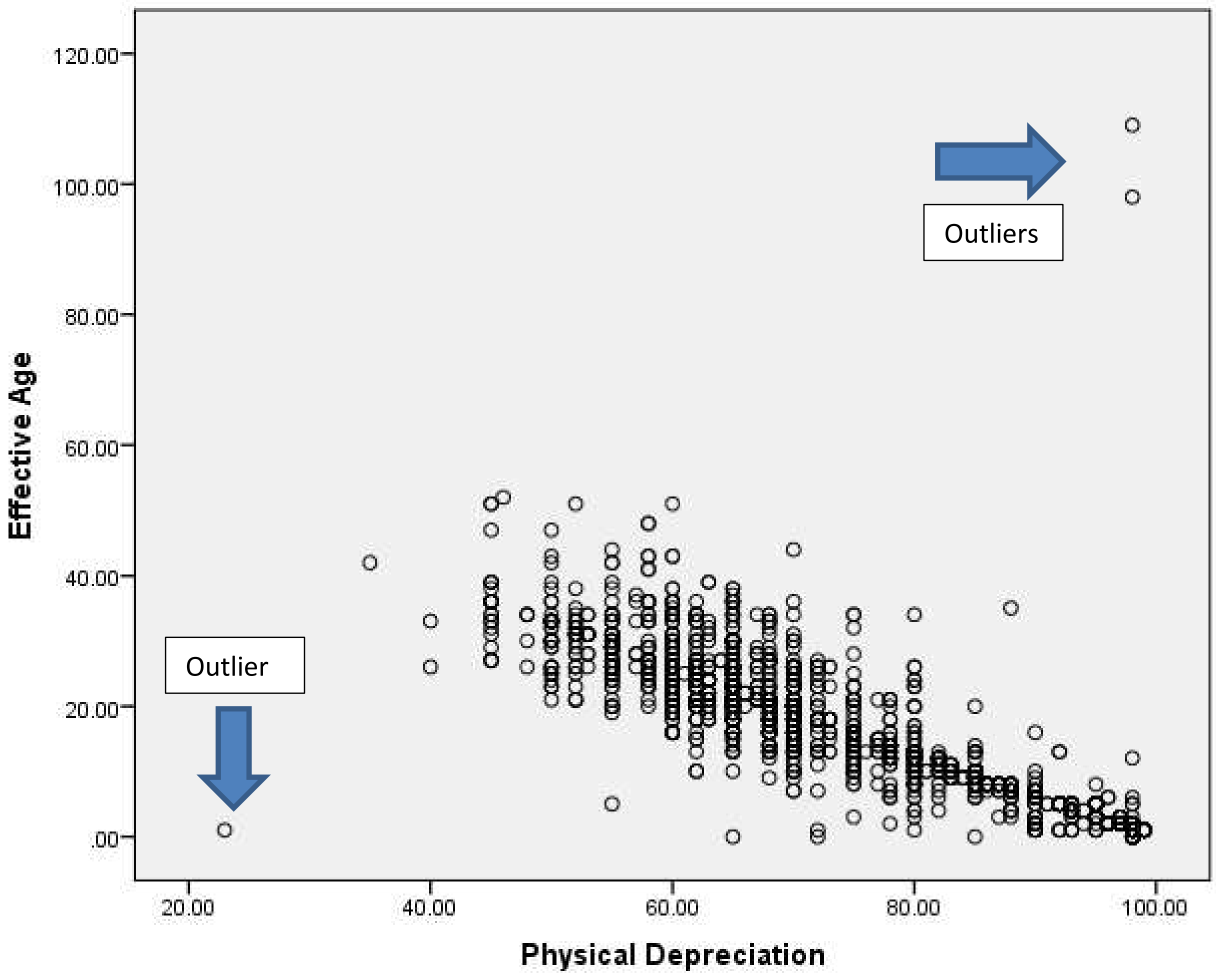
GRAPH 2

MAP OF ASSESSMENT JURISDICTION WITH CORRECT X,Y SALES POINTS



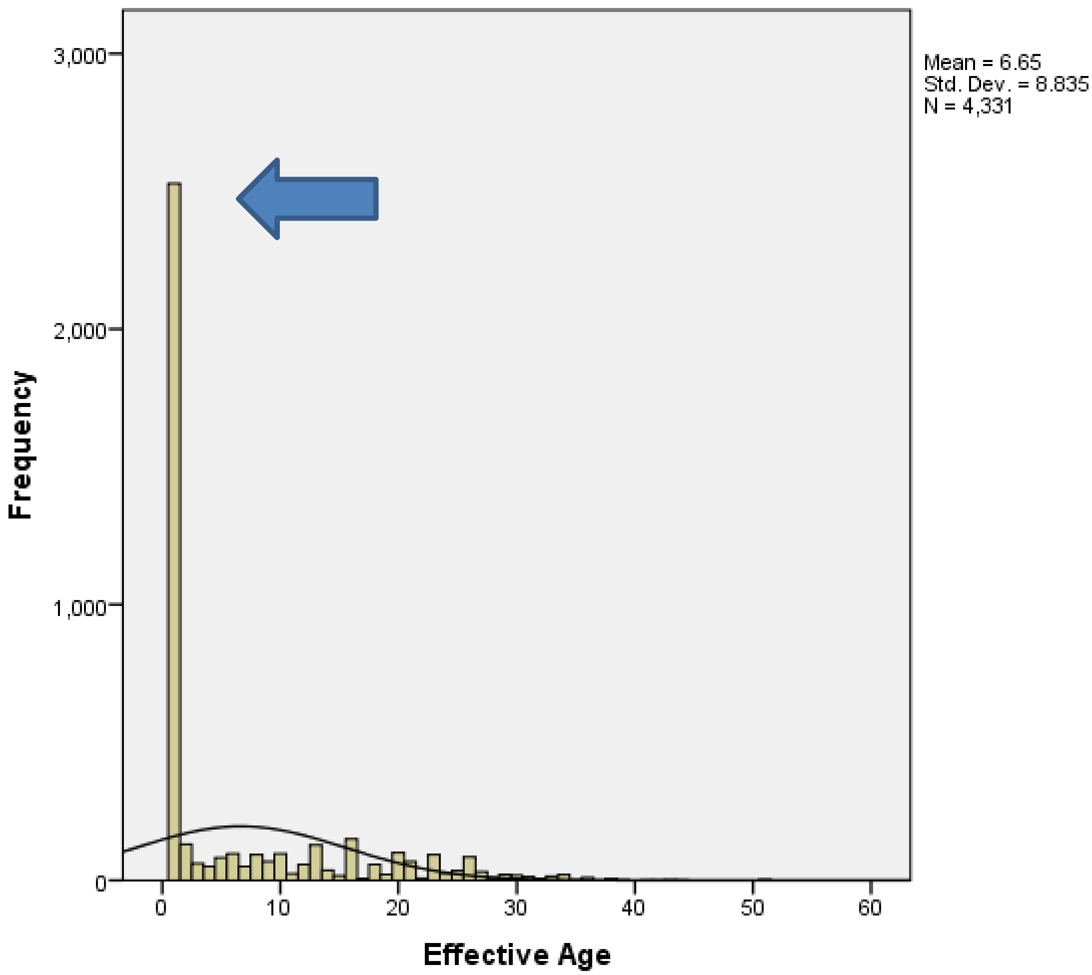
Data analysis graphs assist appraisers understand their data. GRAPH 3 presents two appraiser-assigned variables, effective age and physical depreciation. This shows a contradictory relationship between these two variables. In fact, the variable labeled as physical depreciation is actually percentage good. Along with better understanding of these two variables, data errors are identified in lower left-hand and upper right-hand corners of GRAPH 3. There is a new house with low percent good and two oldest houses with high percent good (physical depreciation). Both extremes are outliers and should not be used in modeling for market analysis.

GRAPH 3
SCATTER PLOT DIAGRAM OF EFFECTIVE AGE COMPARED TO PHYSICAL DEPRECIATION



Data analysis requires looking at data from more than one perspective. Multiple regression determines best fit (central tendency) through data. Multiple regression requires data variation to calibrate differences between independent variables (property information) and how those differences relate to differences in sales prices (dependent variable). Although effective age provides a good relationship to percent good in GRAPH 3, too many effective ages are listed as one. Most of the variation is missing from variable, effective age (GRAPH 4). Over 2,500 of 4,331 sales records have an effective age of one. There is just not enough variation or reliability in this variable's economic value (appraisers' effective age).

GRAPH 4
HISTOGRAM OF FREQUENCY OF EFFECTIVE AGE

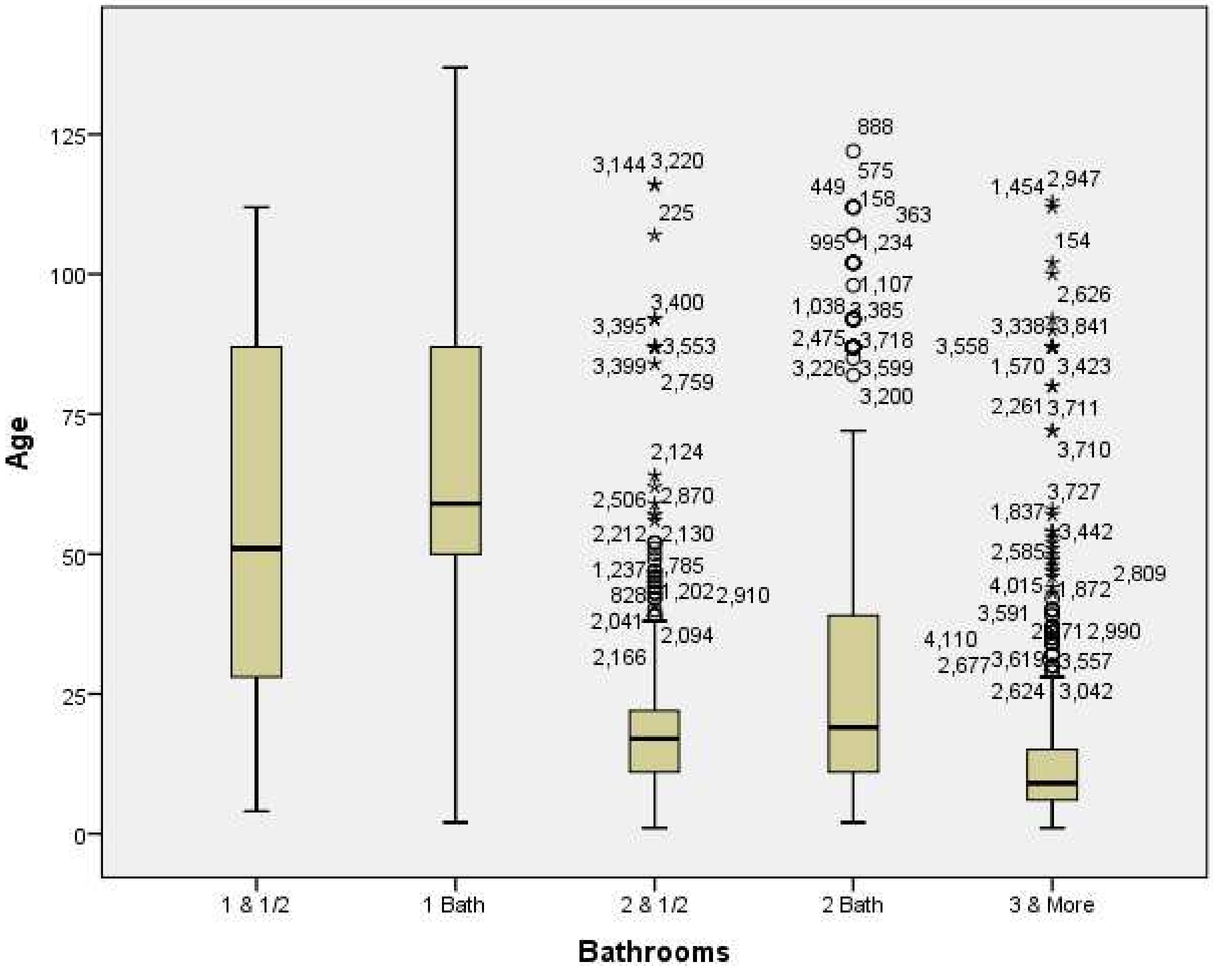


Once these problems were identified, this appraiser/analyst questioned the data and found that these variables had not been updated since about 1990 and are no longer used by this assessment jurisdiction. The variable, year built, is used to calculate a new variable, age.

Box plots provide information about data and some relationships that may not be visualized in any other chart or graphic pattern. GRAPH 5 presents five subcategories of bathroom. Five rectangular boxes represent fifty percent of all properties in each subcategory. The middle horizontal line in each box represents subcategory median values. The T formations (above and below each box) represent ranges of data points between ten percentile and ninety percentile including the mid-fifty percent within each box. Circles and asterisks (dots) above and below these T formations are extreme (remaining) twenty percent within each subcategory range.

Data analysis may provide some information about market place of improved residential property in this community. GRAPH 5 is comparison of age-to-bathrooms per building. GRAPH 5 indicates the oldest age buildings have only one bathroom. The next oldest group has one bath plus extra half baths. Two bathroom and two bathrooms with extra half bathrooms houses are newer houses than the first two categories. Finally, newest house subcategory (box) has three or more bathrooms. So there is definitely a relationship between number of bathrooms and houses age.

GRAPH 5
BOX PLOT OF AGE AND NUMBER OF BATHROOMS



There are many ways to perform data analysis that will inform appraiser/analyst about their databases. While performing data analysis, bad data or very unusual property characteristics will become evident. Therefore, outlier information (not conforming to standard patterns of each community) needs to be filtered or removed from further consideration in valuation process.

MARKET ANALYSIS, FIRST CYCLE

There are several statistical methods to select variables for use in valuation modeling. However, this article is designed to encourage real estate appraisers to use their appraisal skill sets in market analysis. Therefore, the process presented in this article will rely on appraisers' experience in an interactive process to perform market analysis. Property characteristics used in these regression programs are based on existing appraisal principles, so that appraisers will relate to this approach from their market analysis experience.

There are no sales prices in this database, only adjusted sales prices, which are this assessment jurisdiction's standard time adjustments to sales prices. Some variables are more related to adjusted sales price per square foot than to adjusted sales price. For example, CHART 2 presents comparison of exterior condition-to-adjusted sales price per square foot. In CHART 2 the largest volume of sold residential property is in subcategory, good, which has the highest subcategory values in both median and maximum columns.

In this communities' sales database, the most common physical condition is good rather than average. This may indicate condition's inflation by field appraisers or that sold properties may represent the buyers' desire for better build houses. With some subcategory exceptions, adjusted sales prices per square foot seem reasonably related to properties' exterior condition. Appraiser/analysts should question these anomalies in adjusted sales prices per square foot.

CHART 2
COMPARISON OF EXTERIOR CONDITION TO ADJUSTED SALES PRICE PER SQUARE FOOT

EXTR_COND	Cases	Minimum	Median	Maximum
1 Poor	43	15.53	**69.63**	163.80
2 Fair	152	13.65	**70.89**	161.76
3 Average	537	15.14	**84.91**	188.48
4 Good	2,740	33.93	**107.59**	269.77
5 Very Good	352	17.05	**106.13**	211.02
6 Excellent	714	44.42	**101.52**	235.77
Total	4,538	13.65	**103.08**	269.77

One way to determine variables that will be usable in valuation modeling is to test property characteristics against sales price. CHART 3 presents Pearson's correlation of adjusted sales price to ground floor living area (GFLA) and total building square foot (TOTAL_SF). As the second line in CHART 3 shows total square foot is correlated to sales prices at 0.842 as compared to adjusted sales prices correlation to ground floor living area of 0.707. Ground floor area is positively correlated (0.648) to total square footage. Comparing these two building size variables indicate high correlation but not extremely high correlation. Therefore, it is possible to test both variables in regression models.

CHART 3
PEARSON CORRELATION OF ADJUSTED SALES PRICE TO GROUND FLOOR LIVING AREA AND TOTAL SQUARE FOOT

Varaibles		ADJ_SP	GFLA	TOTAL_SF
ADJ_SP	Pearson Correlation	1	.707**	.842**
	Sig. (2-tailed)		0.000	0.000
	N	4552	4552	4552
GFLA	Pearson Correlation	.707**	1	.648**
	Sig. (2-tailed)	0.000		0.000
	N	4552	4552	4552
TOTAL_SF	Pearson Correlation	.842**	.648**	1
	Sig. (2-tailed)	0.000	0.000	
	N	4552	4552	4552

**. Correlation is significant at the 0.01 level (2-tailed).

When correctly used, all valuation approaches will calculate credible estimates of value. All direct market calibration methods (regression) perform market analysis of electronic databases. When using SPSS, non-linear regression program does not produce individual, variable significance 't' test as found in linear regression programs. Linear regression provides a method to quickly review the importance of each variable in its relationship to every other tested variable to explaining variance in dependent variable, adjusted sales price. As developed in this article, linear regression is only used as model preparation, so linear regression model specification is not fully developed. The important results of this linear regression modeling are 't' test values. If non-linear regression of statistical software provides 't' test, it is not necessary to perform 't' tests in linear regression.

CHART 4 presents ‘t’ test values. A limited set of property characteristics presents a portion of linear regression calibration output using characteristics of lot size (SFD_LOT_SF), building sizes (GFLA and Other_SF) and age. These are basic property specific characteristics that explain the majority of variation in adjusted sales prices. These four individual variables have ‘t’ tests above two (2.0), which represents lower limit for stable model variables. While the un-standardized coefficients (CHART 4, Column B) are all correctly aligned (positive and negative) to appraisers’ expectation, coefficients must be adjusted for linear model’s constant value. In this case, a negative constant value requires that independent variables’ coefficients be increased to make up for the negative constant value. At least some independent variables’ coefficients have higher unit values than necessary or market related. Explainability and model rationality suffer as a result.

CHART 4
LINEAR MODEL CORRELATION OF BASIC QUANTITATIVE VARIABLES

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-28807.208	3217.710		-8.953	.000
SFD_LOT_SF	2.175	.214	.095	10.144	.000
GFLA	143.878	2.468	.548	58.300	.000
Other_SF	81.932	1.556	.422	52.648	.000
Age	-602.428	27.966	-.173	-21.541	.000

a. Dependent Variable: ADJ_SP

Once significance is determined in linear regression model’s ‘t’ test, independent variables with ‘t’ test at two or higher are tested with nonlinear regression models. The following is the basic non-linear regression model. MODEL PROGRAM line shows starting coefficients (appraiser assigned). COMPUTE PRED line contains land size and unit value (dollar) represented by letter ‘a’. The following line shows this model’s building specification variables or formula with coefficients as letters (‘b’ to ‘e’) representing building values and percentage adjustments. The last line, NLR ADJ_SP, indicates adjusted sales price (ADJ_SP) is dependent variable.

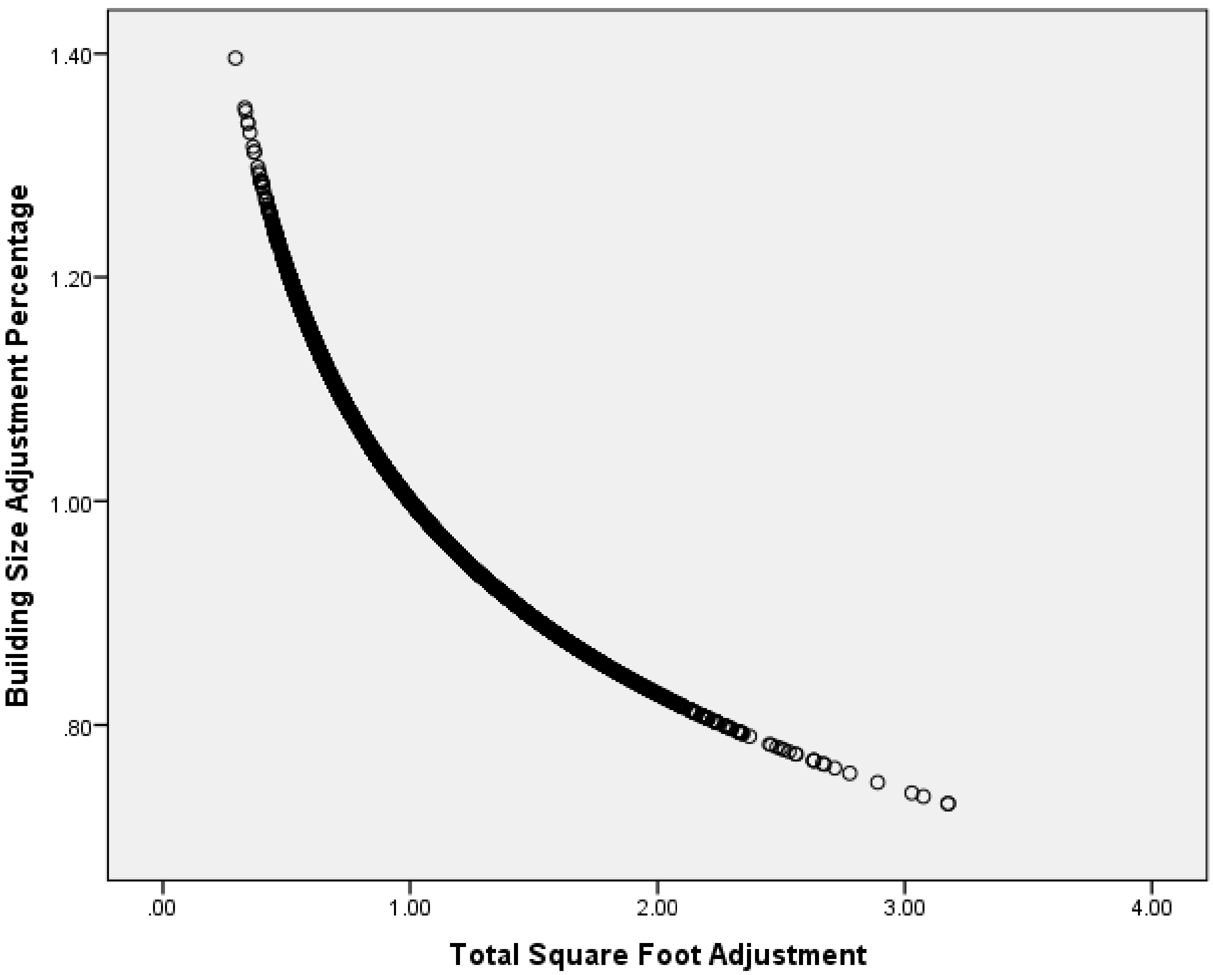
NonLinear Regression.
MODEL PROGRAM a=2 b=150 c=100 d=1 e=1.
COMPUTE PRED_=(SFD_LOT_SF*a)
+((GFLA*b)+(Other_SF*c))*(Total_SF_Size_Adjd)*(Age_Adjust**e).**
NLR ADJ_SP

SPSS non-linear model specifications use one asterisk to represent multiplication. Two asterisks represent raising variables to power curves.

The variable, Other_SF, is total square footage less ground floor living area. Building value computation is ground floor living area (GFLA) and other building square foot areas (Other_SF), which are multiplied by their coefficients (dollars). The sum of these two building values is multiplied by product of non-linear components, total size adjustment (Total_SF_Size_Adj) and age adjustment (Age_ Adjust). Originally, these variables are size and age divided by their median values (percentages). As such these original scales are completely (100 percent) correlated to original variables of total size and age. Once these percentages are raised to their power coefficients, they represent separate adjustment curvilinear factors (percentages).

A good example is building size adjustment percentage (Total_SF_Size_Adj) raised to power coefficient) that demonstrates diminishing utility for increased size for the same quality (grade) of building. Total square footage is divided by its median size. GRAPH 6 presents the straight line adjustment of Total square footage of building (horizontal scale) compare to the curvilinear percentage that has been raised to a power (vertical scale). GRAPH 6 also present this negative power curve reduction of percentage range from straight line range of zero to four, whereas the range of percentages for power curve is from below eighty percent to one hundred and forty percent.

GRAPH 6
TOTAL SQUARE FOOTAGE ADJUSTMENT TO TOTAL SQUARE FOOTAGE POWER CURVE ADJUSTMENT PERCENTAGE



Non-linear regression requires multiple passes (iterations) through the data. CHART 5 presents non-linear regression's Iteration History as coefficients adjust until residual sums of squares (CHART 5, second column) converges to stated minimum amount of change. CHART 5 presents changes in coefficients from starting coefficients (see Parameter, Columns 'a' to 'e'). With each derivative evaluation, coefficients change at a diminishing rate until arriving at stabilized values. There are fourteen model evaluations and seven derivative evaluations. In CHART 5 the final set of coefficients (line 7.1, bottom line of CHART) are model variable parameters in CHART 6.

CHART 5
NON-LINEAR REGRESSION ITERATION HISTORY

Iteration Number[a]	Residual Sum of Squares	Parameter				
		a	b	c	d	e
1.0	847986500841532.100	2.000	150.000	100.000	1.000	1.000
1.1	62545968731326.550	14.479	24.282	44.151	.958	.838
2.0	62545968731326.550	14.479	24.282	44.151	.958	.838
2.1	25094893445655.656	13.112	23.663	58.318	.757	.113
3.0	25094893445655.656	13.112	23.663	58.318	.757	.113
3.1	15725597092321.943	3.224	98.455	52.339	-.016	-.427
4.0	15725597092321.943	3.224	98.455	52.339	-.016	-.427
4.1	8832223799695.793	3.485	101.309	44.035	.291	-.217
5.0	8832223799695.793	3.485	101.309	44.035	.291	-.217
5.1	8763273747216.593	3.335	100.349	44.648	.304	-.201
6.0	8763273747216.593	3.335	100.349	44.648	.304	-.201
6.1	8763246964150.885	3.341	100.268	44.751	.303	-.201
7.0	8763246964150.885	3.341	100.268	44.751	.303	-.201
7.1	8763246891790.673	3.340	100.274	44.753	.303	-.201

Derivatives are calculated numerically.

a. Major iteration number is displayed to the left of the decimal, and minor iteration number is to the right of the decimal.

b. Run stopped after 14 model evaluations and 7 derivative evaluations because the relative reduction between successive residual sums of squares is at most SSCON = 1.000E-008.

The other important section of SPSS's non-linear regression output report is parameter estimates (CHART 6). This chart relates parameter letters from non-linear regression programming line above to final estimate of coefficients (Parameter). Model variable names are added for clarity.

Compare coefficients from CHART 4 to CHART 6. The linear relationships of building square footage coefficients (CHART 6) are lower than in CHART 4 as is the variable, Other_SF, coefficient. This reflects the lack of a constant coefficient in non-linear regression model. This makes these coefficients (CHART 6) more relevant to appraisers as they may better relate to these values from their experiences.

Non-linear relationships of size and age adjustments are power coefficients that convert straight line relationships of property characteristics (building size and age) into their curvilinear relationships with adjusted sales price. While most appraisers have been exposed to regression's raw formula format, as appraisers, they are familiar with percentages and various curved functions such as land size, building size, age/depreciation, etc. as part of the cost approach.

This non-linear regression model specification uses three living area variables: ground floor area, other area and total square foot area. To determine if a difference existed between various types of building square footage, linear and non-linear models tested these three building square footages: ground, other and total square footage. Ground floor square footage and other square foot area proved to be more explanatory than total square footage. In this case, total square footage is used for building size adjustment as it represents overall square footage and reduces the need for second size adjustments.

In non-linear regression, total size adjustment and age adjustment are percentages (variable divided by its median) then are market adjusted by raising each variable to its power coefficient. In linear regression model specification, appraiser/analysts need to predetermine the best curvilinear relationship from statistical functional curves, such as square root or log linear. In CHART 6 the age adjustment curve is age divide by median age (21 years), and then is raised to the power of negative (-0.201) to create a curve function.

One quality statistic for this model specification's calibration is R-square of 0.838. In this residential sales file, these four variables explain 83.8 percent of sales price variation.

CHART 6
NON-LINEAR REGRESSION
PARAMETER ESTIMATES

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
a, (SFD_LOT_SF)	3.340	.184	2.979	3.701
b, (GLAF)	100.274	1.412	97.506	103.042
c, (Other_SF)	44.753	1.297	42.210	47.297
d, (Total_SF_Size_Adj)	.303	.012	.280	.327
e, (Age_Adjust)	-.201	.005	-.211	-.190

*Variable names are added for clarification.

HOLDOUT FILE

Holdout files are a random sample (subset) of modeling (sales) files that are used to test final model formula (specification and calibration). Modeling and holdout files should be separated early in market analysis to preserve independence of holdout files as model formula test files. Usually, holdout files are about five or ten percent of modeling files. Holdout sales are completely separated from modeling file and not used in modeling. Another method to test market models is to use sales transaction dates after last date in sales modeling file.

Keep track of all edits but especially after this point as once modeling process is completed, filters, variable transformations, lookup variables, and final model formulas will have to be applied to holdout file to determine if modeled value quality statistics are approximately the same as derived in modeling file. The main test is median sales ratio of calculated values-to-sales prices. Eventually, variable transformations and final model formulas will have to be applied to related population (unsold) property records.

OTHER ADJUSTMENTS

Grade

Appraisers' judgmental variables may be hard to relate to economic value as appraisers tend to use judgment variables of grade and condition as manual adjustment variables. Grade represents quality of building construction. If properly listed by field appraisers, grade should have a strong rational relationship to adjusted sales prices or price per square foot. Most appraisers will assume that grade should be a continuum of values from the lowest grade with lowest percentage (below 100 percent) to highest grade with highest percentage (above 100 percent). In this community, grade's scheme is letters with plus and minus five and ten percent around each letter grade. These letter grade categorical sublevels range from a low of D to a high of A. This is standard grade scheme used in North America. Due to changes in minimum building codes in this community, the most common occurrence is C+10 rather than C.

Interestingly, as grade is added to linear model, the variable, age, loses some of its significance and falls below the 't' test value of two. This correlation between grade and age may be due to the upgrade of minimum building codes. Both grade and age will have to be tested in non-linear regression. When using age in linear regression, age's relationship is linear to adjusted sale prices. In non-linear regression, age adjustments are a percentage scale (age divided by its median) that is raised to a power curve. Up to this point in this non-linear regression model, age has been the only power curve that is negative. When grade is added to the model, negative coefficients occur for both age and size of building adjustments. These negative directions, which reduce the variance as age and size increase, are what most appraisers expect of these adjustment variables.

Lot Size Adjustments

Normally, land lot size does not enter improved property models or it may have both low 't' value and small coefficient. Lot size enters this model with a value of over two dollars per square foot. Land size adjustment has a curvilinear relationship that provides for efficiencies in size as land size increases. In many community models, land requires a size adjustment.

This community is a city with no rural areas of large land parcels with residential housing.
Land size and land size adjustment derived from market analysis corrupt quality statistic' price related differential (PRD). PRD changes from within an acceptable range to above upper range limit of 1.03, which indicates assessment regressivity. This indicates smaller lots are valued at a higher level of value than larger lots. Value per square foot also increases to over six dollars per square foot from two dollars before testing land size adjustments. Overall model quality statistics did not improve. This is another case of multicollinearity within model specification. Eventually, land size per square foot is fixed at a unit value of four dollars with lot size adjustments providing for marginal utility (size curve).

LOCATION

Location is the geographic position of each property. With each property having a unique location, adjustments are required to measure degrees of uniqueness. Location adjustments relates to variance in sales prices or economic values based on each property's geographic location. Traditionally, appraiser delineates their own sets of fixed neighborhood boundaries. Among appraisal organizations, there is little agreement on how various neighborhood boundary lines should be defined. Fixed neighborhood boundaries requires a balance between having enough properties within each neighborhood to have sufficient sales over a period of time and designing the smallest possible neighborhoods to properly reflect locational influences.

Once delineated, many jurisdictions are hesitant to change the neighborhood boundaries. Delays in updating fixed neighborhood boundaries, eventually requires more boundaries changed; and causes more difficulty when changing neighborhood boundaries. More changes make it harder to explain neighborhoods; and/or changes to taxpayers/owners. Thus, use of geographic coordinates can provide a modern substitute for fixed neighborhoods or act as supplements to fixed neighborhood boundary adjustments.

Coordinate systems can be as simple as X,Y grids lines laid over the community to give each property's location a relationship to other properties within the community. Worldwide latitude and longitude coordinates have become more prevalent with GIS use. Location adjustments are applied at specific property's locations or summarized by city blocks, neighborhoods, census tracts/numbers, zip code or any appraiser-defined geographic area. While appraisal principles indicate that each property has a unique position and location adjustment, most appraisers will choose some cluster of properties (neighborhoods) for location adjustments.

Four base location attributes to be tested are: existing neighborhood adjustment and response surface for adjusted sales price, adjusted sales price per square foot, and residual value of preliminary model. The three variables are divided by their median values to produce a range of values that's centered on one or 100 percent (CHART 7). As modeling calibration continues and sales are filtered (eliminated) as outliers, these location base variables and property specific characteristics, such as size adjustments, age adjustments have medians that may vary slightly from 100 percent during modeling cycles. Unless median variables begin to vary considerably from 100 percent, these base variable medians do not need to be recalculated for each model cycle.

CHART 7
FOUR TEST LOCATION BASE VARIABLES,
EXISTING NEIGHBORHOOD ADJUSTMENT
ADJUSTED SALES PRICE, ADJUSTED SALES PRICE PER
SQUARE FOOT AND RESIDUAL OF MODEL

Base Location	NH_07FINL_Qual	SP_Med	SP_TotSF_Med	RCPMV01
Cases	4,331	4,331	4,331	4,331
Minimum	0.67	0.09	0.13	0.18
Median	1.00	1.00	1.00	1.01
Maximum	1.87	5.78	2.62	2.30

A potential problem with existing fixed neighborhood adjustments systems is that some neighborhoods may have less than five sales necessary to provide sufficient support for neighborhood location adjustment. Where there are neighborhoods with less than five sales within a fixed neighborhood system, appraisers have to use professional judgment to determine appropriate base location percentage. Neighborhood adjustments may be used in linear and non-linear model specification.

Three location value response surface (LVRS) median variables could be summarized by existing neighborhoods (clusters of geographic location) to determine test location variables. LVRS clusters nearest sales prices (economic information) by geography to determine mean of each grid cell, which are similar to having an Excel spreadsheet laid over a geographic area. Each grid cell relates to one another in a response surface, thus the term location value response surface. Using various mathematical techniques, GIS programs identify nearest property (neighbor) or economic characteristic to each grid cell by their coordinates. This creates a response surface. Once response surface is calculated, it predicts response surface values in geographic areas (fixed neighborhood) with or without sufficient or no economic points.

The GIS system is used to calculate base (test) LVRS variables (cluster of economic variables) to be used in modeling. Property or record numbers are used to rejoin location variable files back to primary modeling file. If appraisers prefer to use some cluster of properties such as neighborhood, a separate neighborhood file with centroids of each neighborhoods is used to identify location adjustment variable(s) to be used in next model specification and calibration. These neighborhood percentages will be tested in regression for their contribution to explaining variance in adjusted sale prices.

For beginners, using ESRI's Geostatistical Analyst, most of the default choices will help to generate response surface, experience indicates that searching neighbors (property characteristic or economic points) should start with eight maximum neighbors. As this is not software designed specifically for appraisers, neighbors are variables being examined, which are normally relative economic value points. When sufficient sales transactions and sales prices are not available, benchmark appraisals, gross rents, vacancies, net income and cap rates may be tested for locational differences that contribute to explanations of economic variance. Simple Kriging is the mathematical (spatial autocorrelation) function used to provide weighted grid cell adjustments. The word, neighbor, used in this response surface software is not related to traditional appraisal term called, neighborhood.

CHART 8 presents location variables after GIS file is merged with modeling file. Cases in this modeling file are reduced due to separation of model and holdout files. Median values stay at about 100 percent. Minimums and maximums are smoothed, because these locational values are means of each grid cluster of eight (user selected) neighbor points based on geographic locations. These three variables are tested with other property specific variables, which will determine usefulness of one or more of these variables to explain variance in adjusted sales prices for the specified model.

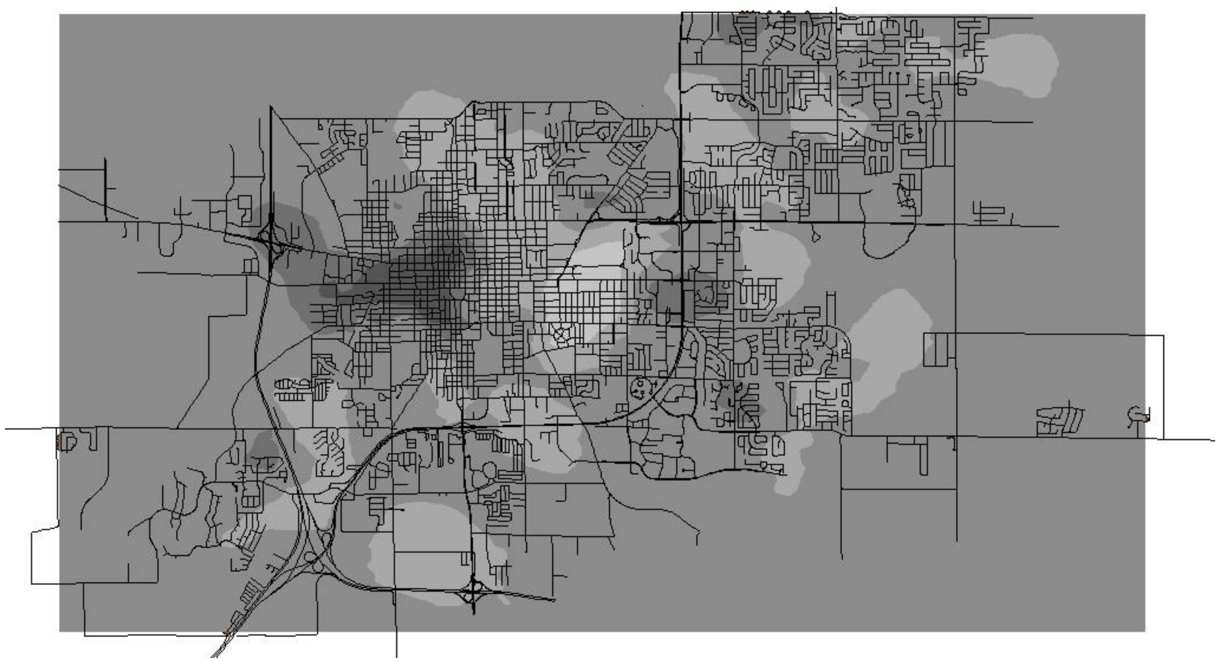
CHART 8
THREE TEST LOCATION VARIABLES BASED ON SALES PRICE (SP_Loc01) AND SALES PRICE PER SQUARE FOOT (PSF_Loc01) DIVIDED BY THEIR MEDIANS AND RESIDUAL OF PRELIMINARY MODEL (RCPMV01_Loc01)

Base Location	SP_Loc01	PSF_Loc01	RCPMV01_Loc01
Cases	4,331	4,331	4,331
Minimum	0.38	0.53	0.74
Median	1.01	1.01	0.99
Maximum	2.77	1.42	1.51

GRAPH 7 presents street centerlines. The default response surface is a rectangle surrounding the community. This provides residential adjustments for all property location within the community. However, extending the shape to community boundaries provides better visual aids. In GRAPH 7, dark shade represents lower value location percentages and light shade represents higher location percentage values.

GRAPH 7

LVRS OF RESIDENTIAL PROPERTY LOCATION PERCENTAGES



General cautions about ESRI's Geostatistical Analyst:

- Limit size of file names to eight characters as Geostatistical Analyst may not be able to manage file with longer names,
- All cases must have X,Y coordinate or Geostatistical Analyst will abruptly close while generating predictions,
- Make sure to save after each step in case errors occur and system abruptly shut down, and
- Extreme outlier ratios will impact response surfaces.

MODELING WITH LOCATION VARIABLES

Valuation modeling is a repetitive process of market analysis, which includes review of data, transformation (syntax), lookup tables, correlation, model specification, model calibration, and review of quality statistics. This process is repeated several times, until quality statistics meet IAAO Ratios Study standards and quality statistics do not materially change with each repetition of process.

CHART 9 presents a limited review of correlations for four location adjustments-to-adjusted sales prices. Correlation of sales prices-to-location adjustments indicates that sales price location 01 (SP_Loc01) is best correlated and residual location adjustment (RCPMV01_Loc01) is least correlated. While correlation is an important tool for market analysis, it represents one-to-one relationships not multiple relationships.

CHART 9
CORRELATION, ADJUSTED SALES PRICE TO SEVERAL LOCATION ADJUSTMENTS

	ADJ_SP	NH_07FINL_Qual	SP_Loc01	PSF_Loc01	RCPMV01_Loc01
Adjusted SP	1	0.401	0.854	0.639	0.042

When used in linear regression, these location variables (CHARTS 9) are considered in relationship to all other variables in their ability to explain variance in sales prices. For example, in linear regression with addition of these three location variables, age 't' test rises from minus (-14.088) to minus (-3.280). Age is negatively correlated to these location variables by over sixty percent. Location is also highly correlated to building size and may be highly correlated to other variables such as grade and condition that are still to be considered.

This assessment jurisdiction has maintained good neighborhood boundaries. They have performed market analysis necessary to identify changes in neighborhood boundaries and adjusted neighborhood valuation percentages. Existing neighborhood adjustments show strong correlations with sales prices and compete well in explaining variance in adjusted sales prices. All variables may be initially tested in linear regression to determine their interactions and relationships to adjusted sales prices. When testing location variables in linear and non-linear regression, each location variable is added one at a time and tested in combination.

After model specifications and calibrations of base model with these four location variables, three location variables contribute to explaining variance in adjusted sales prices. Location variables, sales price per square foot divided by its median (PSF_Loc01), did not add to explanation (improve R-squared) and is a negative coefficient indicating that this variable has a strong correlation with one or more other location variable. When two variables are highly correlated with one having a positive value and another negative value, this is an indication of multicollinearity. So PSF_Loc01 is dropped from future consideration in this model. By adding these location variables to the non-linear regression formula, the R-squared increased from 0.838 to 0.907.

END OF FIRST CYCLE

All steps in first cycle are first tested in linear regression to generate variable's 't' tests of significance. Variables that are significant in terms of their 't' tests are re-examined in non-linear regression. Some variables that represent property features may not be significant in linear model and may be excluded during linear model report (lot square footage and lot size adjustment are perfect correlation). When one of these variables represent size adjustments to property characteristics, both should still be tested in non-linear regression model as adjustment variables will represent curvilinear relationships of sizes to sales prices. In this non-linear model specification, both variables represent different relationships to sales prices. Lot square footage's coefficient will relate to average value per square foot. Land lot size adjustments will represent curvilinear function of diminished utility of larger residential lots over smaller residential lots.

Through this testing (model specification and calibration) process, the best model for first cycle will be established. It should represent major components of land and building (improved property) but need not represent all potential variables.

One of the potential review statistics of modeling process is removal of outliers. The International Association of Assessing Officers, IAAO, defines outliers as:

> **Outliers.** Observations that have unusual values, that is, differ markedly from a measure of central tendency. Some outliers occur naturally; others are due to data errors.[1]

[1] Standard on Ratio Studies, 2013, International Association of Assessing Officers, Page 42

While there are complex mathematical methods of identifying and removing outliers, such as described in the IAAO's Standard on Ratio Studies, Appendix B, Outlier Trimming Guidelines. A simple method to identify and remove outliers is to use sales ratios distribution of range percentiles to identify outliers based on sales ratios between estimates of market values-to-sales prices. Outliers are identified by distribution range percentiles from low to high break points of sales ratios. CHART 10 presents frequency distribution percentile ranges used to identify lowest sales ratios at distribution ranges of 2.5 and 5 percentile. Upper end of sales ratios distribution range are 95 and 97.5 percentile. This identifies by sales ratios the lowest five percent and highest five percent of sales ratios distribution. Usually, exclusion of outliers is based on more conservative ranges at 2.5 and 97.5 percentiles. When these two sales ratio percentages (.7538 and 1.4623) are applied as a filter, this will remove five percent of worst sales ratios. Outliers at the extremes are removed between modeling cycles to provide more representative regression coefficients and quality statistics.

CHART 10
SALES RATIOS DISTRIBUTION PERCENTILES USED TO IDENTIFY VALUE-RELATED OUTLIERS

Valid Cases	4329
Missing	2
Percentiles	
2.5	.7538
5	.8008
95	1.2788
97.5	1.4623

Some reasons to remove value-related outliers are:

- Quality of original data,
- Quality of sales validation, and
- Relationship of property data-to-sales prices (economic values).

Along with removal of value related outliers, check to see if each model formula calculates estimates of value for all cases. Two sales ratios are listed as missing from CHART 10. Missing estimates of market values (sales ratio) indicate that some model variables (property or economic characteristics) are missing data and/or are blank. If many cases are missing estimates of value, then responsible variables should be recollected, recalculated or removed from model specification.

MARKET ANALYSIS, SECOND PLUS CYCLES

Second and subsequent cycles follow same steps as presented in first cycle. New response surfaces are prepared. New location adjustments are set up in model specification and tested by calibration. Transformed variables may need to be reviewed and adjusted. If available, additional property specific variables are tested. Outliers are removed. This process continues until market analysis produces best quality statistics. Once this is determine, modeling process continues with two seemingly contradictory steps. First remove variables with least contributory value (lowest 't' test values) and recalibrate models until quality statistics start to decrease by more than a few thousandth of a point. This provides most stable variables' coefficients and leads to accurate, stable estimates of value (model year-to-model year).

FINAL MODEL REVIEW

One way to review statistical quality of continuous variables at smaller subgroup levels is to develop clusters of ten percentiles by variables range. Some continuous variables to review are sales price, square foot of land, total building square footage and age. These variables' categories are reviewed with other model variables to see if there is any variance in sales ratios by group or categorical sublevels. CHART 11 presents age-ten subgroups, range of ages and median sales ratios. All median sales ratios within these subcategories are within the IAAO recommended statistical quality guidelines. Most categorical variables, such as count of bathrooms or bedrooms, are reviewed without creation of subcategories.

CHART 11

AGE GROUPS WITH AGE RANGE AND MEDIAN SALES RATIOS

Group of Ages	Cases	Minimum Age	Maximum Age	Median Sales Ratio
0	518	1.00	7.00	1.01
1	401	8.00	10.00	1.00
2	370	11.00	13.00	0.99
3	587	14.00	18.00	0.99
4	284	19.00	21.00	0.97
5	410	22.00	31.00	0.99
6	422	32.00	48.00	1.00
7	406	49.00	61.00	1.01
8	460	62.00	87.00	0.98
9	254	90.00	137.00	1.00
Total	4,112	1.00	137.00	0.99

CHART 12 presents linear regression report with variables, coefficients and ‘t’ values. Unstandardized Coefficient, sub-column ‘B’ are multipliers that correspond to variable at their left. CHART 12, second column from the right is the ‘t’ values that indicate each variable’s value to explain variance in sales prices. Sale year and sale quarter adjustments are significant in this linear model. This indicates that assessment jurisdiction’s standard time adjustment for sales prices is not sufficient to explain variance in sales prices over this time period.

CHART 12
LINEAR REGRESSION OUTPUT REPORT, T-TEST VALUES

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-2039852.187	117969.998		-17.291	.000
SFD_LOT_SF	.841	.150	.036	5.619	.000
GFLA	59.290	2.060	.230	28.781	.000
Other_SF	40.203	1.275	.211	31.521	.000
Age_Adjust	-5959.659	505.972	-.080	-11.779	.000
FIN_BS_SF	23.219	1.108	.112	20.956	.000
GAR_SF	24.798	2.965	.053	8.362	.000
NH_07FINL_Qual	168173.896	6925.406	.139	24.284	.000
RCPMV05_Loc02	110770.443	7979.709	.074	13.882	.000
Total_Room_Adj	345097.717	56070.408	.027	6.155	.000
Grade_Adj_SP01	62134.579	1829.327	.323	33.966	.000
Lake_Front_Adj	473883.257	37356.076	.057	12.686	.000
AC_Adj	459754.181	84093.508	.028	5.467	.000
SaleYear_Adj	221090.323	23304.359	.042	9.487	.000
SaleQtr_Adj	116151.734	43471.442	.012	2.672	.008
Area_Adj (MLS)	105452.599	15744.679	.032	6.698	.000
Block_SP_Loc02	32037.652	2067.665	.153	15.495	.000

a. Dependent Variable: ADJ_SP

Of all variables available for testing, these fifteen variables represent best indicators to explain variance in sales prices. Area adjustment is a location variable representing MLS geographic indicators. MLS adjustments are the reciprocals of median sales ratio for MLS areas.

The second response surface variable, Block_SP_Loc02, is summarized at city block levels. It is based on sales prices divided by its median and Kriging regression technique (LVRS).

There are five location adjustments in final model: appraiser designated neighborhood/percentages; appraiser designated lake front; multiple listing system (MLS) area (neighborhoods); and two location value response surfaces. These response surfaces are calculated by simple Kriging interpolation technique measuring ratios of sales prices divided by its median and reciprocal ratios of prior model. These percentage location values are tested by final linear and adjusted by non-linear regression models. When compared to all other model variables, non-linear regression model weighs relative contribution of each of these five location variables to explain variances in adjusted sales prices.

R-squared for this model is 0.921. As this is a linear model, three building size variables (GFLA, Other_SF and FIN_BS_SF) entered model and building size adjustment did not, because it is the combination of GFLA and Other Area. In straight line relationship, overall building size variable is highly correlated to other two building size variables.

In series of non-linear regression calibrations, land lot size unit value per square foot ranged from two to six dollars. In final model specification, unit value per square foot is constrained at four dollars (not permitted to change). Based on this unit value, size adjustment for land provides for market-oriented size adjustment curve that represents marginal utility.

There are several slight adjustments to non-linear model that should be considered. First remove variables that least explains variance ('t' test) in sales prices considering other variables in this model specification. After removal of variables based on 't' tests, this statistical model contains only fourteen variables:

```
NonLinear Regression.
MODEL PROGRAM  a=1 b=100 c=50 d=25 e=1 f=1 g=1 h=1 k=1 m=1 n=1 p=1 q=1.
COMPUTE PRED_=(((SFD_LOT_SF*4)*(Lot_Size_Adj**a))
+((GFLA*b)+(Other_SF*c)+(FIN_BS_SF*d))
*(Total_SF_Size_Adj**e)*(Age_Adjust**f)*(Grade_Adj_SP01**g))
*(NH_07FINL_Qual**h)*(RCPMV05_Loc02**k)*(Lake_Front_Adj**m)*(Area_Adj**n)*
(SaleYear_Adj**p)*(Block_SP_Loc02**q).
NLR ADJ_SP
```

Garage square footage, which is significant in linear regression model, caused disruption of other model coefficients and is dropped from non-linear model specification. R-squared of this non-linear model is 0.950.

This model and next model specification show the land portion of model on COMPUTE PRED line. Building linear portions of these models are on next line followed by building non-linear portion of model. There are two lines of non-linear, general quality variables of location and time variables.

CHART 13 provides some sales ratio quality statistics. The median sales ratio is .994. The PRD is 1.010 and the COD is .088. These are all within the IAAO's recommended quality statistics for ratio studies

CHART 13
SALES RATIO QUALITY STATISTICS ESTIMATED VALUES TO SALES PRICES

Median	Minimum	Maximum	Price Related Differential	Coefficient of Dispersion
.994	.694	1.690	1.010	.088

TRADITIONAL PROPERTY VALUATION CHARACTERISTICS

A criticism of many regression programs is that variables (property characteristics) that best explain variance in dependent variable (sales prices) do not always include all traditional appraisal variables. These traditional variables may be miscoded or just do not represent variance in sales prices as do other variables. For example, Condition may not be properly specified as indicated in CHART 2. Condition did not prove useful in this final model.

To show principle of adding traditional property characteristics, one variable is added to model specifications listed below. Traditional value-related variables should be added and tested one at a time to review impact of each variable. In following model specification, air conditioning multiplier, AC_Adj01, is simply added to multiplicative building terms. These traditional value-related variables are constrained without any movable coefficients (fixed) or without a coefficient (multiplicative variable). This permits property characteristics to be added to model specification without unusual changes in market analysis or reappraisal-to-reappraisal due to low 't' values.

In prior years, appraiser/analysts might not have used these variables or added variables after final model is complete. While this removed the problems of weak 't' values, variables added after model specification/calibration are completed may cause some overall values to be too high or too low. By incorporating traditional variable constrained values into model specification/calibration, market supported coefficients for other variables will adjusted for these traditional variables and related values. Be careful when adding traditional variables to model specification/calibration, these variables may significantly impact model's generated values (coefficients) and quality statistics. If major changes occur that are undesirable, remove offending variable (traditional) from the model speciation and calibration.

Land values are required by some assessment jurisdictions that require separation of land and building values. This model uses variables, square footage (SFD_LOT_SF) for median unit value (dollar) and land size adjustment factor (coefficient 'a') provides the curvilinear percentage adjustment to median dollar amount for land size.

This non-linear regression model shows air conditioning percentage, AC_Adj01 (right end of line five), is added to best statistical models (compare to above model).

NonLinear Regression.
MODEL PROGRAM a=1 b=100 c=50 d=25 e=1 f=1 g=1 h=1 k=1 m=1 n=1 p=1 q=1.
COMPUTE PRED_=(((SFD_LOT_SF*4)*(Lot_Size_Adja))**
+((GFLA*b)+(Other_SF*c)+(FIN_BS_SF*d))
***(Total_SF_Size_Adj**e)*(Age_Adjust**f)*(Grade_Adj_SP01**g)*(AC_Adj01))**
(NH_07FINL_Qual**h)*(RCPMV05_Loc02**k)*(Lake_Front_Adj**m)*(Area_Adj**n)
(SaleYear_Adjp)*(Block_SP_Loc02**q).**
NLR ADJ_SP

CHART 14 presents this non-linear regression calibration iteration history. When land (SFD_LOT-SF) unit value is four dollars per square foot, the lot size adjustment parameter 'a' represents the adjustment for lot size. Parameter 'b' is unit value (dollars) for ground floor living area (GFLA), etc. This model calibration has eighteen model evaluations and nine derivative evaluations. Since appraiser's judgment for land values per square foot and air conditioning do not change per iteration (fixed unit value), they are not shown in iteration history.

CHART 14
NON-LINEAR REGRESSION MODEL CALIBRATION SEQUENCE

Iteration History[b]

Iteration Number[a]	Residual Sum of Squares	Parameter												
		a	b	c	d	e	f	g	h	k	m	n	p	q
1.0	7.024E+15	1.000	100.000	50.000	25.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.1	4.444E+14	.409	36.253	24.392	6.263	.828	.847	.967	.995	1.385	1.166	1.066	1.653	.834
2.0	4.444E+14	.409	36.253	24.392	6.263	.828	.847	.967	.995	1.385	1.166	1.066	1.653	.834
2.1	6.224E+13	-.146	45.273	33.267	4.428	.325	.382	.757	.970	2.375	1.059	1.193	2.825	.511
3.0	6.224E+13	-.146	45.273	33.267	4.428	.325	.382	.757	.970	2.375	1.059	1.193	2.825	.511
3.1	7.068E+12	-.563	68.562	51.945	5.172	-.135	-.010	.598	.785	2.118	.524	1.013	2.214	.131
4.0	7.068E+12	-.563	68.562	51.945	5.172	-.135	-.010	.598	.785	2.118	.524	1.013	2.214	.131
4.1	2.439E+12	-.430	78.982	58.899	11.619	-.257	-.115	.469	.511	1.275	1.195	.751	1.548	.076
5.0	2.439E+12	-.430	78.982	58.899	11.619	-.257	-.115	.469	.511	1.275	1.195	.751	1.548	.076
5.1	2.356E+12	-.397	78.966	58.617	13.991	-.269	-.124	.431	.441	1.157	1.353	.752	1.530	.108
6.0	2.356E+12	-.397	78.966	58.617	13.991	-.269	-.124	.431	.441	1.157	1.353	.752	1.530	.108
6.1	2.356E+12	-.391	78.932	58.721	14.065	-.272	-.125	.430	.437	1.155	1.364	.750	1.527	.110
7.0	2.356E+12	-.391	78.932	58.721	14.065	-.272	-.125	.430	.437	1.155	1.364	.750	1.527	.110
7.1	2.356E+12	-.390	78.928	58.735	14.071	-.273	-.125	.430	.437	1.155	1.365	.750	1.526	.110
8.0	2.356E+12	-.390	78.928	58.735	14.071	-.273	-.125	.430	.437	1.155	1.365	.750	1.526	.110
8.1	2.356E+12	-.390	78.928	58.736	14.072	-.273	-.125	.430	.437	1.155	1.365	.750	1.526	.110
9.0	2.356E+12	-.390	78.928	58.736	14.072	-.273	-.125	.430	.437	1.155	1.365	.750	1.526	.110
9.1	2.356E+12	-.390	78.928	58.736	14.072	-.273	-.125	.430	.437	1.155	1.365	.750	1.526	.110

Derivatives are calculated numerically.

a. Major iteration number is displayed to the left of the decimal, and minor iteration number is to the right of the decimal.

b. Run stopped after 18 model evaluations and 9 derivative evaluations because the relative reduction between successive residual sums of squares is at most SSCON = 1.00E-008.

CHART 15 presents final parameter estimates of value for each variable except land square footage unit value and percentage adjustment for air conditioning. Parameters 'b', 'c' and 'd' are the linear (additive) portion of this model calibration. Remaining model's variables have curvilinear (power) coefficients that modify percentage adjustments.

CHART 15
NON-LINEAR MODEL REGRESSION PARAMETERS

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
a, Lot_Size_Adj	-.390	.027	-.442	-.338
b, GFLA	78.928	.473	78.000	79.855
c, Other_SF	58.736	.775	57.216	60.256
d, FIN_BS_SF	14.072	.732	12.636	15.507
e, Total_SF_Size_Adj	-.273	.014	-.300	-.246
f, Age_Adjust	-.125	.004	-.132	-.118
g, Grade_Adj_SP01	.430	.010	.410	.450
h, NH_07FINL_Qual	.437	.022	.394	.480
k, RCPMO05_Loc02	1.155	.037	1.081	1.228
m, Lake_Front_Adj	1.365	.091	1.186	1.544
n, Area_Adj	.750	.062	.628	.872
p, SaleYear_Adj	1.526	.084	1.361	1.692
q, Block_SP_Loc02	.110	.009	.093	.127

This model's R-squared is 0.950. Addition of air conditioning variable (categorical percentage) did not adversely impact any market derived parameters and quality statistics.

CHART 16 presents this model's sales ratio quality statistics. Median sales ratio is 99.4 percent. Price related differential (PRD) is 1.01, which indicates little to no vertical bias in valuation estimates. Coefficient of dispersion (COD) is 8.8 percent, which indicates a tight set of ratios cluster around the median sales ratio. In this case, median sales ratio, price related differential and coefficient of dispersion have not changed, since best statistical model without air conditioning variable (see CHART 12). Both sets of sales ratio statistics are well within International Association of Assessing officers' recommend quality statistics. In this case adding air conditioning (AC_Adj01) variable improved the explainability but did not adversely affect the performance statistics from the previously specified model.

CHART 16
SALES RATIO QUALITY STATISTICS

Median	Minimum	Maximum	Price Related Differential	Coefficient of Dispersion
.994	.690	1.718	1.010	.088

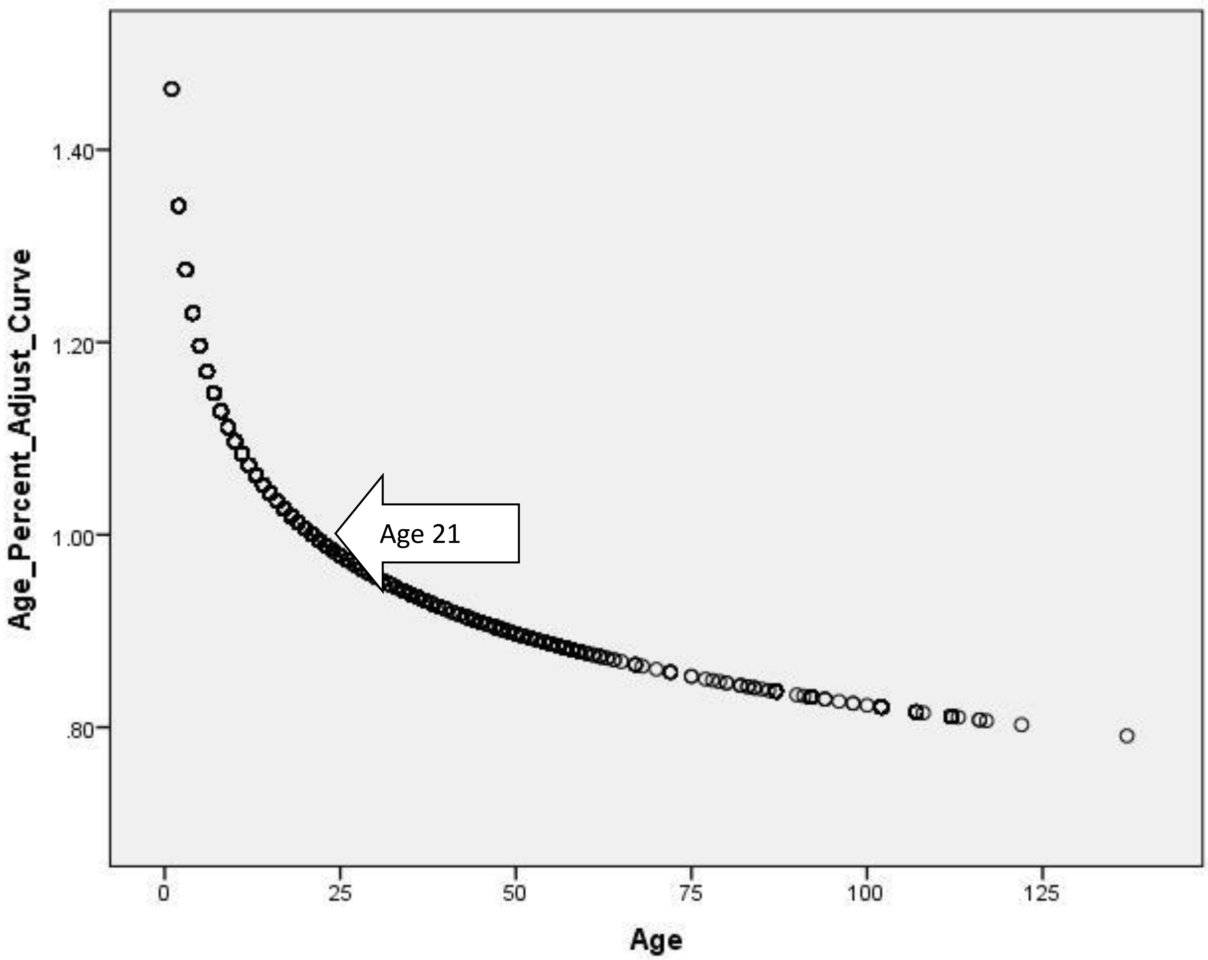
MODEL DECOMPOSITION

Some newer CAMA/AVM systems have non-linear regression programs or at least accommodate non-linear models directly within their system. Non-linear regression may be specified so that models decompose into tables and are easily fitted into any CAMA/AVM systems or developed in Excel as a valuation worksheet. This model specification provides formatting into cost approach structures and coefficients are decomposable into land and building values. To demonstrate these concepts, Graph 8 and Charts 17, 18 present several individual variables and their final coefficients.

GRAPH 8 presents curvilinear relation between age and modeled effective age curve. GRAPH 8 non-linear curve rises up to 146 percent for new buildings and down to 79 percent for the oldest buildings. This is based on a premise of regression, which provides central tendency values of all model variables. In this case, median age is 21 years old. GRAPH 8's curve is centered on age 21, which is at 100 percent point. This curve represents buyers' desire for new housing that drops off quickly as housing ages. In older houses, depreciation or percent good changes reflect smaller changes per year possibly due to improvements made by owners (sellers) and/or buyers' indifference to specific ages of older residential housing stock. All non-linear regression curve and transformation can be duplicated as cost tables to be used in CAMA/AVM systems.

GRAPH 8

COMPARISON OF AGE CURVE PERCENTAGE TO AGE



Most North American appraisers (mass or single property) are familiar with the cost approach. Traditionally, cost approaches assume that new houses represent 100 percent of percent good or zero depreciation. This does not consider premiums paid for new houses.

In non-linear regression, once this age percentage curve is created, it can be transformed to have new houses represented by percent good starting at 100 percent. If the age curve is converted to meet requirements of CAMA system or appraisers' opinions, then adjusted new curvilinear relationship must replace original age variable in model specification and be constrained (not changeable) so other variables' coefficients will compensate for the new age curve fixed values.

All categorical variables may be converted to tables. CHART 17 presents lake frontage subcategories. Lake frontage reflects a ten percent increase (1.0968 rounded to 1.10) over residential houses without lake frontage (subcategory 2). In this community, lakes may be as small as retention ponds and are scatter throughout the community. There are no large lakes in this community. Buyers are willing to pay extra for lake fronts on small ponds. Since lake front properties are scattered throughout this jurisdiction, lake front property's impacts are not completely consider in LVRS. Location value response surface adjustments (LVRS) require sufficient cases within response surface cluster areas to register changes in value. When location influences do not have sufficient cases within designated small geographic areas, these location influences need to be considered as property specific variables that are separate adjustment variables.

CHART 17
PERCENTAGE ADJUSTMENTS FOR PROPERTIES WITH LAKE FRONTAGE

Lake_Front	Cases	Percentage Adjustment
1, Lake_Front_Adj	145	1.0968
2, No Lake Frontage	3,967	1
Total	4,112	1

In this model specification ground floor living area, other living area and finished basement area are three linear variables that are modified by one curvilinear (non-linear) adjustment for total area size. This provides better quality estimates of market value but complicates building size adjustment variable. In this example:

Ground floor living area is multiplied by $78.92
Other living area is multiplied by $58.74, and
Finished basement area is multiplied by $14.07

Once these building areas are multiplied by their coefficients (dollars), values are summed. Next total values for various square footages are adjusted by total square foot of living area adjustment percentage as listed in CHART 18. Most CAMA/AVM systems have programs to interpolate between building sizes in left column of CHART 18 to arrive at correct building size adjustments. Model size adjustment percentages are continuous but shown here as table per 1000 square feet intervals.

CHART 18
BUILDING SIZE ADJUSTMENT PERCENTAGES

House Total Square Footage	House Adjustment Percentage
1,000	1.15
2,000	0.95
3,000	0.85
4,000	0.79
5,000	0.74
6,000	0.71
7,000	0.68
8,000	0.65

Non-linear regression model (following) shows model specification variables and corresponding parameters (coefficients). COMPUTE line shows variable name to be calculated, land size at four dollars and land size adjustment. Line two is linear variables (dollar adjustments) of building sizes. Line three is building curvilinear (non-linear) adjustments (multipliers). Lines four and five are overall adjustments of location and time (multipliers). One asterisk represents multiplication. Two asterisks represent raising variable percentages by power curves.

COMPUTE Calc_MV10_AC=(((SFD_LOT_SF*4)*(Lot_Size_Adj-.390))**
+((GFLA*78.92)+(Other_SF*58.74)+(FIN_BS_SF*14.07))
***(Total_SF_Size_Adj**-.273)*(Age_Adjust**.125)*(Grade_Adj_SP01**.430)*(AC_Adj01))**
***(NH_07FINL_Qual**.437)*(RCPMV05_Loc02**1.155)*(Lake_Front_Adj**1.365)**
***(Area_Adj**.750)*(SaleYear_Adj**1.526)*(Block_SP_Loc02**.110).**

Lot size adjustment (Lot_Size_Adj), total size adjustment (Total_SF_Size_Adj) and age adjustment (Age_Adjust) are continuous variables adjusted by power curves to form curvilinear relationships. Grade adjustment (Grade_Adj_SP01) is a categorical variable, whose subcategories are adjusted by variable's power coefficient. All location factors, except location value response surface variable, Reciprocal of Sale Ratio for Model Five Location 02 (RCPMV05_Loc02), are categorical variables that are adjusted by being raised to their power coefficients and then multiplied in a string that includes sales year adjustment. Location value response surface variable, RCPMV05_Loc02, is based on model five's reciprocal ratios that exclude location consideration. This location percentage is also changed by its power curve coefficient. Sum of land and building values are multiplied by product of these overall categorical variables to produce estimates of market value.

COMMENTS AND CONCLUSIONS

Market analysis of electronic real estate data is efficiently performed using statistical software and geographic information system (GIS) software.

Statistical software valuation formulas should be decomposable for explanatory use.

Appraisers should assume that commonly used software in the appraisal profession, will provide correct calculations. Appraisers only need to understand what tests need to be performed and what these test output reports represent.

Good design, maintenance, and location percentage values for neighborhoods are very important in valuation process. Generally, well delineated neighborhoods using market adjustments will assist in providing credible market value estimates, resulting in performance statistics that not only meet but often exceed industry recommend standards. When neighborhood delineations have not maintained, location value response surface (LVRS) applied at specified level for individual properties, city block centroids, census tracts, zip codes and existing neighborhoods will provide reasonable substitutes for fixed neighborhood boundaries.

One benefit of LVRS percentage location adjustments is that every geographic location (grid cells) within each surface will have supportable number of sales to generate location value. This is often not the case in fixed neighborhood location adjustments.

This valuation model uses fixed neighborhoods, MLS geographic areas, two different LVRS summarized at property level and city block geographic sizes.

This article shows that appraisers' judgment may be used to select important variables and determine starting weights or percentages for categorical variables, which establishes relative relationship between sub-categorical levels. The process used in this article is not a computer-driven selection process such as stepwise regression, where many variables are added one at a time and regression determines importance of each variable to all other variables.

Statistical models such as regression provide midpoint values (coefficients). This must be recognized in valuation process.

Since regression will perform best at midrange values (coefficients), lowest and highest estimates of market values are less reliable than typical properties market value estimates. Whenever possible, all regression market estimates should be reviewed. Estimates of values at the extreme ends of value spectrum should always be reviewed by appraisers for accuracy to the market place.

Because real estate markets are imperfect (buyers and sellers are not fully knowledgeable), for any group of appraisals/model estimates there will be variance normally between estimates of market value and sales transaction prices (see Coefficient of Dispersion, COD). Valuation models attempt to reduce this variance. When variance (COD) is too small, there may be problems with model specification, calibration or appraisers may be chasing sales prices.

Appraisers have knowledge of their markets (geographic competence). Appraisers can use this knowledge and logic to perform real estate market analysis aided by statistical and GIS software.

REFERENCES

Borst, Richard A., 2013, Spatial Temporal Methods for Mass Appraisal, International Property Tax Institute, Seminar Proceedings

Environmental Systems Research Institute, Inc., 2003, Arc GIS 9, Using the ArcView Spatial Analyst, Redland: ESRI

Figueroa, Roberto A., 1999, Modeling the Value of Location in Regina Using GIS and Spatial Autocorrelation Statistics, Assessment Journal, November/December, Page 29

Jensen, David L., 2012, Spatial Analysis via Response Surface Methodology, Journal of Property Tax Assessment and Administration, 9, 3, 5

Jensen, David L., Tricks of the CAMA Masters, 2013, Presented at 17th Annual GIS/CAMA Technologies Conference,

O'Connor, Patrick M., 2008, Automated Valuation Models by Model Building Practitioners: Testing Hybrid Model Structure and GIS Location Adjustments, Journal of Property Tax Assessment and Administration, April, 5

O'Connor, Patrick M., 2008, Automated Valuation Models of Time and Space: Best Practice, Journal of Property Tax Assessment and Administration, April, 57

O'Connor, Patrick M., 2002, Real Estate Valuation Using Statistical Modeling and GIS published Appraisal Institute's 2002 Annual Summer Conference proceeding, July 2002; reprint IAAO Fair and Equitable, 5, 2005

O'Connor, Patrick M., 2004, Basics of Non-linear Modeling published Integrating GIS & CAMA 2004 Annual Conference proceedings, 4, 2004; reprint IAAO Conference on Assessment Administration proceedings, 8, 2004

O'Connor, Patrick M., 2002, Comparison of Three Residential Regression Models: Additive, Multiplicative and Nonlinear, published Vision Beyond Tomorrow, Integrating GIS & CAMA 2002 Annual Conference proceedings, 4, 2002; reprint Assessment Journal, 7, 2002, 37

O'Connor, Patrick M., Jack Eichenbaum, 1988, Location Value Response Surfaces: The Geometry of Advanced Mass Appraisal, Property Journal, 277

Patterson Mike C., Bob Harmel, 2006, A Regression Approach to Forecasting Sale Price of Single-Family Residences in Wichita County Texas, Journal of Property Economics, Page 65

Pena, Sergio; Cesar M. Fuentes; Luis E. Cervera; and Vladimir Hernandez; 2012, Planning Support Systems: A Computer-assisted Mass Appraisal (CAMA) System for Ciudad Juarez, Mexico, Property Tax Assessment & Administration, 9, 4, 25

Ward, Richard D., Jason Guilford, Brian Jones, Debbie Pratt and Jerome German, ASA, IFAS, 2002, Piecing Together Location: Three Studies by Lucas County Research and Development Staff, Assessment Journal; September/October 9, 5, 15-48.