EXPLANATION OF MARKET MODELING
IN THE CURRENT KANSAS CAMA SYSTEM

I have been asked on numerous occasions to provide a lay man's explanation of the market modeling system of CAMA. I do not claim to be an expert on multiple regression or of the specifics of the market model approach as applied in Kansas. Some may disagree with the observations that I make from experience through the use of this system for the last seventeen years. I hope that this overview will at least give the user a better understanding of its use and how it can be adjusted and utilized.

KSCAMA Market Modeling provides a way to generate and/or display several different valuation methods for a specific property within one sheet called a **Comp Sheet**. This document is typically utilized by an appraiser to determine the final value assigned to that property. Within the current CAMA system, this option is only applied to parcels improved with a dwelling(s).

The Comp sheet is not unlike an appraisal form utilized by fee appraiser's for single family dwellings. The subject property and comparable properties are listed, adjustments are made for differences between the subject and comps and a final value is established. There are however, some important differences:
The comp sheet is designed to display up to five comparable properties (if available) whereas a fee form is typically limited to three.

Adjustments to comps are standardized within a model and applied to all properties within that model. Fee appraisals are less concerned about such standardization and adjustments are typically more arbitrary. This allows the fee appraiser much more flexibility when establishing a value and the credibility of these adjustments are rarely questioned. (The fee appraiser does not have to justify his valuation of a property and the techniques he employed to the owners of the property next door or the house down the street.)

The Comp sheet uses some additional valuation techniques and reconciles these generated values into a final recommended value to be accepted or rejected by the reviewer.

MARKET MODEL GROUNDWORK

The first phase of market modeling is the definition of valuation areas—what areas are similar or affected by similar market forces, what areas should be grouped together and what areas should be modeled together.

Step one is nbhd definition. I won't go into detail about nbhd definition other than to say they usually consist of somewhat similar properties, affected by similar market forces and are geographically related. Nbhd boundaries are often established by subdivision boundaries, land use changes, improvement or value changes.

Step two is nbhd grouping—what nbhds are similar in many if not most components (value, age, size, etc…) and would be reasonable comps for each other in the valuation process.

A nbhd group may contain one or many nbhds. Within the KSCAMA system, nbhd groups are assigned in the Categorical Name Table (XXXCNT CNT). Each nbhd is assigned to a nbhd group number.

Step three is model grouping—the appraiser must decide which nbhd groups are assigned to which models that he chooses to create. Group assignment to model is done in the base model parameter XXX430. Nbhd groups are listed and assigned to one or more models. The data from sales within these nbhds and nbhd groups is used to establish coefficients for value generation, comp adjustment and comp select criteria.

By utilizing these stages of grouping, KSCAMA provides the ability to utilize multiple models to generate comp sheets. If markets within a jurisdiction are significantly different, the use of additional models means that different adjustments and base coefficients can be used to reflect the market differences. <u>However, determining the number of models to utilize is one of the most difficult decisions in the valuation process.</u>

Most counties in Kansas are small and rural. This presents three unique problems that must be addressed when choosing the number of models to utilize:

There is not usually a large quantity of sales upon which to establish market coefficients. Models created using too few sales are not going to be as statistically reliable because all components of value may not exist in sufficient quantity for regression analysis. The appraiser has to choose between statistical reliability or the flexibility in coefficients/adjustments that an additional model might provide. Testing can usually provide a reasonable answer to this question. If the two models produce coefficients that are generally similar, it is likely that the models can be combined back to one.

NBHDs are sometimes not homogeneous. Especially in the smaller communities, improvements are very mixed with new houses often next door to older homes. Homes of all different sizes, values, styles, etc.. may exist in one nbhd. The result is that houses that would normally be assigned to different nbhds, nbhd groups and models are assigned together. Under these circumstances, it may not be reasonable to apply more than one model.

Due to limited sales, applying multiple models may limit the size of the comp population within each model. Too few comps to choose from can result in excessively large adjustments or the use of older sales that may no longer reflect value. NBHD group lines can be crossed. (If required a comp can be pulled from a different nbhd group)—model lines cannot. (*If a comp is assigned to model 1, it can never be pulled as a comp for a subject in model 2.*)

My opinion:

Some appraisers choose to use multiple models often based simply upon the fact that there are sufficient sales within the jurisdiction to allow their creation. When there are many sales and properties are very similar such as can be found in many subdivisions, this probably works very well – except for those areas where there may not be sufficient sales. Properties may be similar but in different nbhds and can no longer comp each other because models are defined too narrowly.

I am generally of the opinion that less is more. Fewer models mean more statistically significant coefficients and more choices for comps within any given model. When clear differences are present, multiple models based upon value, age, etc, are appropriate. But it should always be understood that there will be some areas that will not fit neatly into any specific model. If market influences are similar (school districts, taxes – specials, economics) then a reasonable argument can be made that a single model can be utilized.

I many small counties, the decision as to the number of models is an easy one. There are often too few sales to generate even one reliable model. Several years ago, PVD developed age models for use in smaller counties. Again sales were often too limited to utilize two models (pre-war, post-war) but an additional technique was developed that has been useful. Semi-age modeling has some of the aspects of modeling based upon improvement age while utilizing only one model. Some major coefficients like age adjustment, grade, etc, are calculated separately based upon the improvement age. I use semi-age modeling almost exclusively in smaller counties and very often in larger counties as well.

MODEL BUILDING

Once nbhd group assignments are made and the number of models is determined, the actual modeling process begins. There are several procedural steps that must take place prior to initial model runs. I won't detail those here other than to indicate that the number of years of sales must be chosen. For example, in the smallest of counties, it is often necessary to utilize as many as five years of sales to have a statistically sufficient number for modeling. In larger counties it may be as few as two years. This is an appraisal decision dictated by the number of sales available and the change in the market over the time frame considered.

COEFFICIENTS:

Coefficients are the variables defined that can be considered in the market modeling process. Depending on the county's setup, there can be as many as 200 coefficients. Many of these coefficients are purely descriptive or are a component required to define an additional coefficients. For example, year-built is a coefficient but its main purpose is as a component for the definition of the age coefficient which is then applied in the definition of other coefficients that are more specific or useful such as AGE*SFO. (AGE*SFO makes an adjustment for age based upon square footage of the house and for houses only within an older age range.)

From a practical standpoint, each model is limited to a consideration of thirty (30) coefficients. Appraisal judgment determines which coefficients are considered. Most counties utilize similar coefficients and these coefficients are typically much more detailed and numerous than would be defined in a fee appraisal.

For an initial (first) market run, the coefficients are selected with no constraints except for land and base OBY values. (These coefficients are set to pull their value directly from the cost approach and to utilize this in regression analysis and in establishing value.) The coefficients to be used in the model are listed in the model run (usually run #5) and following are listings of all the coefficients and the vital statistics associated with each (mean, high-low value, occurrences, etc..)

MULTIPLE REGRESSION

The first step in market modeling is to perform **multiple regression** using the coefficients selected. In simplest terms, multiple regression is a complicated mathematical procedure that correlates the relationship between the variables (coefficients the appraiser has chosen) and predicts values based upon this relationship within the given base of data.

A more precise definition follows:

A statistical technique used to assess the relationship between one dependent variable and two or more input variables. The purpose of multiple regression analysis is to measure the degree of association between variables or predict the dependent variable as a function of the input variables.

All of this means very little to someone unless they are statistician. Because of the sheer number of calculations and the complications of the formulas used and the need to test results using different inputs, multiple regression is only possible through the use of a computer in mass appraisal.

Before we go further into multiple regression and its pertinent terms and statistics, the reader should understand **that multiple regression is a component of the market modeling process in KSCAMA and not the sole process.** It helps the appraiser establish variables that are used for determining one of several indicated property values on the comp sheet and establishing values for adjustments made to comps. The appraiser has the option to vary from and adjust these coefficients based upon his knowledge and experience of the market. The appraiser also establishes a select criteria for comparable properties, independent of multiple regression.

Market modeling encompasses all factors that are dependent upon the appraiser's input—how he establishes valuation areas—how he groups valuation areas (assigns to models)—which variables he tests and includes in the multiple regression process—how he constrains or changes the variables to fit his knowledge of the market—and how he defines the criteria for selecting comps within any given model.

Limitations & Problems of Multiple Regression in our Application:

The general computational problem to be solved with multiple regression is **to fit the best straight line to a number of points** (different indicated values of a specific coefficients). This straight line is to minimize the squared deviations of these points from the line. In reality, a straight line does not necessarily reflect what's happening in the market place as it relates to a specific set of variables. This can be seen by looking at the results of a depreciation study in the cost approach. Improvements do not depreciate in straight equal increments over their economic lives. Depreciation schedules curve and flatten over time. One can draw a straight line from beginning to end, but any given point on that line may not be reflective of the market. This is a somewhat similar issue for multiple regression.

Sample size versus number of variables: Most statistician indicate that the number of observations (sales/samples) should be 10 to 20 times the number of variables considered. The inclusion of too many coefficients in a limited data base increases the likelihood of chance results. Given that our techniques use as many as thirty variables (coefficients) our base of samples should be at least 300. This is not always possible given county size etc… The option of reducing the number of coefficients is also not appealing as these coefficients must be used to establish adjustments for comps for items

that may be prevalent in the population but not in the sales base. In other words, eliminating these coefficients, takes away our ability to adjust for them. The easiest example to understand would be inground pools. Few exist in most parts of Kansas and they are not likely to show up as a meaningful coefficient in multiple regression analysis- but most appraisers would expect to make some adjustment for the existence of a pool regardless of whether regression analysis indicates it to be significant or not.

Things may not be as they appear: multiple regression predicts relationships between variables but can't define the cause of the relationship. For example- there would be a strong correlation between the damage caused by a fire and the number of firefighters sent to fight it, yet it would be improper to assume that the firefighters were the cause of the fire. It is more likely that we should consider another variable (fire size) rather than number of firemen when predicting damage.

The inclusion or exclusion of the right variable can change results dramatically. For example and analysis that relates height to hair length would likely show a strong negative relationship. (The shorter the person, the longer the hair). On the surface this does not seem reasonable. By adding gender as a variable in the analysis, the results change completely-the correlation is strong to gender and nonexistent for height. (Women are usually shorter than men – women generally have longer hair than men.)

Redundancy or multicollinearity can also be a problem. This is trying to predict a result using two or more similar variables.

First Run Of Market Model Building

As stated previously the first run of market modeling should be conducted with the minimum of constraints on variables considered. Through the process, it is likely many coefficients will be eliminated as not significant as predictors of value. If the variables chosen by the appraiser are the best variables to consider, this run will provide the best statistical results of R squared and standard error.

Because of what is usually a limited sales base, the subjectivity of the market, etc. some coefficients will not produce what appear to be reasonable values. An example that is common is fireplaces indicating values of $8-10k within a data base where the average sales price is $35k. However, with accurately listed data, the most important variables will produce reasonable coefficients values-variables such as grade, condition, age, etc.

It is at this point that the appraiser begins to test and adjust the coefficients. Removing coefficients that seem redundant or extremely limited in the database-trying different coefficients-and forcing in coefficients that in his opinion are important and that do exist in the database. The appraiser attempts to have a set of coefficients that utilizes the multiple regression generated coefficients for the most important of variables along with a set of variables based upon the multiple regression data but that may be adjusted or constrained to reflect reasonable values and adjustments within the marketplace.