Standard on Automated Valuation Models (AVMs)

Approved September, 2003

International Association of Assessing Officers

The assessment standards set forth herein represent a consensus in the assessing profession and have been adopted by the Executive Board of the International Association of Assessing Officers. The objective of these standards is to provide a systematic means by which concerned assessing officers can improve and standardize the operation of their offices. The standards presented here are advisory in nature and the use of, or compliance with, such standards is purely voluntary. If any portion of these standards is found to be in conflict with the *Uniform Standards of Professional Appraisal Practice* (*USPAP*) or state laws, *USPAP* and state laws shall govern.

Acknowledgments

The AVM Standard was reviewed and completed through the dedicated efforts of an Ad Hoc committee comprising Alan S. Dornfest, AAS, *Chair*, Larry J. Clark, CAE, Robert J. Gloudemans, Michael W. Ireland, CAE, Patrick M. O'Connor, and William M. Wadsworth. The Committee worked closely with Nancy C. Tomberlin, who was chair of the Technical Standards Committee at that time.

Special thanks and appreciation also go to the many individuals who served as reviewers for this standard:

Richard Almy
Richard A. Borst
Man Cho
John S. Cirincione
Robert C. Denne
Brian G. Guerin
M. Steven Kane
Josephine Lim
Mark R. Linne, CAE
Wayne D. Llewellyn, CAE
Ian W. McClung
John F. Ryan, CAE
Ronald J. Schultz
Russ Thimgan
James F. Todora, CAE
Robert Walker

Published by

International Association of Assessing Officers
130 East Randolph
Suite 850
Chicago, IL 60601-6217
312/819-6100
Fax: 312/819-6149
http://www.iaao.org

ISBN 0-88329-180-0

Printed in the United States of America.

Contents

Standard on Automated Valuation Models (AVMs)

1. SCOPE

This standard is intended to provide guidance for both public sector CAMA and private sector AVM systems. This standard provides recommendations and guidelines on the design, preparation, interpretation, and use of automated valuation models (AVMs) for the appraisal of property. The standard presents market analysis based appraisal applications and aspects of such models. The principles addressed in this standard are considered applicable to all appraisals of real property, which are designed to estimate market value.

The standard does not address appraisal of personal property, such as machinery and equipment, and AVMs are not considered applicable for appraisal of highly specialized or unique property.

As presented in this standard, the development of an AVM conforms to *USPAP* Standard 6 (Appraisal Foundation 2003, 46–56). The appraiser using AVM output should follow *USPAP* standards that relate to their assignment.

2. INTRODUCTION

2.1 Definition and Purpose of an AVM

2.1.1 Definition

An automated valuation model (AVM) is a mathematically based computer software program that produces an estimate of market value based on market analysis of location, market conditions, and real estate characteristics from information that was previously and separately collected. The distinguishing feature of an AVM is that it is an estimate of market value produced through mathematical modeling. Credibility of an AVM is dependent on the data used and the skills of the modeler producing the AVM.

2.1.2 Purpose

The purpose of an AVM is to provide a credible, reliable, and cost-effective estimate of *market value* as of a given point in time. Market value is the most probable price (in terms of money) that a property should bring in a competitive and open market under the conditions requisite to a fair sale—the buyer and seller each acting prudently and knowledgeably, and assuming the price is not affected by undue stimulus. AVM values reviewed for reliability, and generated in compliance with *USPAP* Standard 6 are considered appraisals.

AVMs are developed and used by both the public and private sector. Assessment officials use AVMs to produce estimates of value as of a common date for purposes of property assessment and taxation. Private sector appraisers and their clients use AVMs to estimate the value of a subject property at a given point in time for a wide variety of purposes.

2.1.3 Applicability

AVMs are applicable to any type of property for which adequate market information and property data are available in the relevant market area. The relevant market area is the area that would be considered by potential purchasers. For residential properties, this is typically all or a portion of a metropolitan area, one or more towns in a geographic area, or a given rural or recreational area. The market area for larger multi-family, commercial, and industrial properties can be regional or even national in scope, depending on the relevant investors and market participants.

The development of an AVM is an exercise in the application of mass appraisal principles and techniques, in which data are analyzed for a sample of properties to develop a model that can be applied to similar properties of the same type in the same market area. These may be either individual properties of interest or all properties that meet the requirements of the model.

Although the same underlying principles are applicable to all AVMs, the specific formulation and calibration techniques will vary with the purpose of the AVM, type of property, available data, and experience and preferences of the market analyst. Sections 3 and 4 discuss the general principles of model specification and calibration. Section 5 addresses residential AVMs. Section 6 focuses on commercial and industrial AVMs and section 7 focuses on AVMs developed for vacant or improved land.

2.1.4 Distinction from Traditional Valuation Applications

Although AVM development requires skilled analysis and attention to quality assurance, AVMs are characterized by the use and application of statistical and mathematical techniques. This distinguishes them from traditional appraisal methods in which an appraiser physically inspects properties and relies more on experience and judgment to analyze real estate data and develop an estimate of market value. Provided that the analysis is sound and consistent with accepted appraisal theory, an advantage to AVMs is the objectivity and efficiency of the resulting value estimates. Of course, sound judgment is required in model development and an appraiser should review the values produced by the model.

2.2 Purpose and Use of AVMs

2.2.1 General

AVMs are used to provide estimates of market value for a variety of public and private sector purposes. AVM estimates reflect a given time period and should be calibrated to produce market values as of a specific date. Although past market trends can be projected over a short time horizon, the credibility of appraisal estimates increasingly suffers as the projection is lengthened.

AVMs have the advantage of objectivity and consistency, reduced cost, and faster delivery time. It is important, however, that the AVM follow sound statistical and mathematical modeling practices and be tested for accuracy and uniformity before application. Section 8 discusses the important area of model testing and quality assurance and section 9 focuses on reporting of results.

2.2.2 Analysis of Impaired Properties

Properties subject to significant defects or that are affected by atypical circumstances impairing market value, including superadequacy or functional obsolescence, cannot be accurately modeled with an AVM. An appraiser may choose to apply the AVM to the property, but the defect or unique circumstance should be noted and a special adjustment made to compensate for the defect or special circumstance.

2.3 Steps in AVM Development and Application

The remaining portion of this section outlines the steps to take in development of an AVM. The following sections of this standard provide clarification and details concerning these steps and their application to particular property types.

2.3.1 Property Identification

The first step in any appraisal problem is to identify the property to be appraised. In developed economies, identification is normally straightforward, as maps, ownership records, property addresses, and legal descriptions will identify the property and owner. The appraisal assignment will usually require identifying physical characteristics and property rights to be valued as of the appraisal date. When applying an AVM to a particular property, improvements and renovations made before this date should be included in the appraisal; those made subsequent to the appraisal date should not.

The bundle of rights to be appraised generally includes the fee simple interest or full bundle of rights inherent in ownership of property. Nevertheless, the market analyst should make clear what rights are assumed and any limitations to full use or restrictions to transfer of the property.

2.3.2 Assumptions

The AVM supporting documentation should state all assumptions, special limiting conditions, extraordinary assumptions, and hypothetical conditions. A key assumption in many AVM applications concerns the assumed use of the property. Most real estate databases contain the actual use of property as of the inspection date. In some property tax systems, current use is stipulated as the basis for valuation. However, comparable market sales reflect the concept of highest and best (most probable) use. Market analysts and users of AVMs need to be aware of these subtleties.

Another key assumption relates to whether or not the fee simple bundle of rights is being appraised. This is generally the case for residential properties, but many commercial appraisals are made to estimate only the leased fee or leasehold interest when there is an existing lease (or leases) on the property.

Government appraisal agencies are responsible for collecting and maintaining property databases, although they often contract with private vendors for this purpose. Commercial AVM providers generally use data maintained by a government agency or third party service. In all cases, it is imperative that AVM market analysts test the reliability of the data and clearly state assumptions concerning its accuracy. If data important to value estimation are missing or the statistical process has shown the data to be inconsistent or unreliable, the AVM provider has a responsibility to not provide a potentially misleading value estimate to the intended user.

2.3.3 Data Management and Quality Analysis

The reliability of any appraisal depends on accurate data. Appraisal data fall into two general categories: property data and market data. Property data relate to location, land characteristics, and building features. Market data include sales, income, and cost information. Asking prices and independent appraisals can sometimes be used to supplement sparse sales data.

Computerized statistical tools used to develop AVMs afford the opportunity to screen data for missing or out-of-range occurrences and inconsistencies; examples include homes with more than two fireplaces or a bi-level home with no listed lower level living area.

Geographic information systems (GIS) can also help in data reviews. GIS software is used to maintain computerized maps and provide geographic representations of property attributes and features. It can be used to highlight properties with impossible, unlikely, or inconsistent data. For example, properties coded as being waterfront can be color-coded, displayed on a map, and reviewed for accuracy.

Only valid, open market sale and income data should be used in model development. (As mentioned, asking prices and independent appraisals can sometimes also be used to bolster sample sizes.)

Since the reliability of an AVM is dependent on the data from which it is generated, the integrity of the database should be monitored on a systematic and ongoing basis.

2.3.4 Model Specification

Model specification is the important process of determining the format (model structure) of the AVM. The market analyst must determine the type of model to be employed and specify the variables to be used in the model.

AVMs that employ property features, often characterized as "hedonic" models, can be categorized as additive, multiplicative, or hybrid models (see Section 3 on Specification of AVM Models). Market analysts must also determine the variables to be included in hedonic AVMs. These can represent property characteristics (e.g., square feet of living area and building age), location information, demographic data (e.g., income levels or school quality), or variables derived from property characteristics (e.g., the square root of lot size or living area multiplied by a quality index). The objective is always to include property features important in value determination and to capture actual market relationships. Skilled analysis is required to adequately specify an effective model structure.

Some models that are referred to as AVMs have only a time component; in other words, they merely track changes in property values over time. Where property characteristic information is unavailable or limited, these models can be used to trend a previous sale or value estimate to the target appraisal date.

2.3.5 Model Calibration

Calibration is the process of determining the coefficients in an AVM as well as which variables should be retained or deleted due to statistical insignificance. Several statistical tools can be used to calibrate AVM models (see Section 4 on Calibration Techniques). Proper use of these tools requires experience and training in statistical analysis and the software employed.

2.3.6 Model Testing and Quality Assurance

An AVM must be tested to ensure that it meets required accuracy standards before being deployed. This is accomplished through statistical diagnostics and a ratio study in which value estimates (e.g., estimated sale price or estimated rent) are compared to actual values (e.g., sale price or reported rent) for the same properties. GIS can be used to display color-coded ratios on maps and help spot groups of under- or over-valued properties. For more information, see Section 8 on Automated Valuation Model Testing and Quality Assurance. Before it is implemented, the AVM also should be tested on a holdout sample, which is a set of properties and their selling prices that were not used in the calibration process.

Properties with unusually large errors, termed "outliers," should be reviewed. It is likely that the sale price (or other value serving as the dependent variable in the model) is not representative, the data are partially incorrect, or the property exhibits atypical features that cannot be adequately accounted for in the model. Except where the data can be corrected, the property should be removed from the sample, and it and similar properties with similar features should not be valued by the AVM alone.

2.3.7 Model Application and Value Review

Once tested and validated, the AVM can be applied to estimate the value of other properties of the same type in the area or region where the model applies. These values should be reviewed for reasonableness and consistency with recent sales, either of the subject property itself or of similar properties in the same neighborhood or surrounding area, or where sales are not available, recent asking prices.

It is also good practice to systematically review the generated values for reasonableness and consistency with nearby properties in the same neighborhood. This affords the opportunity to ensure that the data are accurate, and to make individual adjustments to properties with unique features or that are subject to special influences, such as being located at a busy intersection or having a premium or obstructed view.

2.3.8 Stratification

Stratification is the process of grouping properties for modeling and analysis. Stratification begins with property type. Properties are delineated into generic use categories such as: single-family residential, condominium (if applicable), multi-family, commercial, and industrial. The number of property types will depend on the size and diversity of the geographic area being analyzed and the number of sales available within the proposed strata.

Residential properties in urban areas are generally stratified into "market areas." Market areas are broad, somewhat homogeneous geoeconomic areas that appeal to buyers in similar economic brackets. One AVM may be developed for each market area, or a regional model may be developed and individually calibrated for each market area. Location within the market area can be handled through neighborhood variables or other variables related to geographic location and desirability. Alternatively, a location value response surface analysis

(LVRSA) may be used to measure and adjust for location within the model formula (see Section 3.4 Location).

Commercial properties are usually modeled across a wider geographic area than residential. For example, one model may be sufficient for all properties of a given type (e.g., office, retail, or warehouse) in an entire urban county or metropolitan area.

2.3.9 Value Defense

Market analysts must be prepared to review and defend values developed through AVMs. The review process begins with checking the accuracy of the data. If no problem is found, the estimated value should be evaluated for consistency with similar properties and with any recent sales of the subject property or similar properties. The fact that a property sold for a price different from the AVM estimate does not mean that the AVM estimate is wrong. The sale date may differ significantly from the appraisal date, and the property may have sold for a relatively low or high price, depending on the peculiarities of the situation and motivations of the buyer and seller. If the estimated value appears to be unreasonable or inconsistent with market evidence, the AVM estimate is not reliable and should be discarded or adjusted (the reason for the breakdown should be investigated and corrected). If the estimate is supported by market evidence, then it should be defended.

The best support for an AVM value is recent sales of comparable properties. Current listings can also be used, although they must be given less credence than consummated sales. For income properties, it may be possible to support a value estimate derived from one AVM (say, a sales comparison model) with estimates derived from alternative methods (e.g., an income model). The consistency of the value estimate with others produced by the AVM model, as well as the overall reliability of the AVM model as evidenced by a ratio study of the holdout sample or other statistical measures can also be evaluated and used to defend the value.

AVM developers should prepare documentation that will allow clients and other appraisers to understand in non-technical terms how the model was developed and applied.

3. SPECIFICATION OF AVM MODELS

The two major components of valuation are specification and calibration. Model specification is the process of developing the proposed model structure. Model calibration relates to testing the specified model structure using data sets to generate the model variable coefficients.

In practice the specification and calibration are performed in an iterative process which includes the following steps:

1. Specify a model
2. Test the specification with calibration
3. Make adjustments to model specification
4. Test new specification with calibration
5. Continue to repeat the process until statistically significant improvement is minimized

The AVM specification and calibration iterative process makes the assumption that data are collected and verified in a consistent and professional manner.

3.1 Data Quality Assurance

The model specification process begins with an evaluation of the data availability. The availability of data will influence the specification of the model and may indicate the need for revisions in the specification and/or limit the usefulness of the resulting value estimates. Publicly available data from government sources, such as government assessors, deed recorders, registrars and census agencies, are the basis for most statistical models. Commercial sector information services may be used to supplement that data. Because more than one source will provide information toward the AVM model process, the AVM market analyst must use statistical data analysis to confirm the assumption that the quality of the data will provide reasonable support for the modeling process.

AVM models are based on a sample of the universe of data. The specification process must review the sample data used to develop the model as well as the population to which the model will be applied. The sample should be representative of the population in all key elements of value including the types of properties, market conditions, value range, land and building sizes, and building ages. Property types where market information is not available, should be excluded from both the sample and total population files as the model specification will not be representative of these properties.

Indicators of value may include sale prices, rents, expenses, and capitalization rates. Limitations in the integrity and availability of the data are important determinants of the model specification. Knowledge of key property characteristics is crucial to model specification. Models should not be specified without an understanding of the data in the sample and population.

Data field verification is common in public, but not in commercial, AVM development. Commercial AVM market analysts rely on the accuracy of the data provided to them. In cases where AVM data is not field verified, data quality can only be measured by its typical relationship to the value. When data items that appraisers would consider highly correlated to value do not prove to have such a relationship (correlation matrix or regression T or F values), this could be an indication of inconsistent data collection or scarcity of data. Data that are not consistently collected or that are mostly missing from the population should not be used in the model specification or calibration phases, as it can be insignificant and may produce misleading results.

Data may be qualitative or quantitative. Quantitative data is objective and can be counted or measured. Qualitative data is usually descriptive, subjective, and subject to judgmental decisions that require experience by the person collecting the data.

3.2 Model Specification Methods

AVM models are based upon one or more of the three approaches to value (cost, sales comparison, and income).

3.2.1 Cost Approach

Model specification for the cost approach requires the estimation of separate land and building values.

The cost approach formula converts to a model specification:

$$MV = \pi GQ * [(1-BQ_D) * RCN + LV]$$

- MV is the market value estimate.
- πGQ represents the general qualitative variables such as location and time;
- BQ_D is a building qualitative variable representing depreciation;
- RCN is the replacement/reproduction cost new;
- LV is the land value; and

(Gloudemans 1999, 124.)

If a third party provides the cost tables, it is the responsibility of the AVM market analyst to calibrate the cost tables to the local market in order to provide a valid indicator of value by the cost approach.

3.2.2 Sales Comparison Approach

The sales comparison approach can involve either a two-step process, in which comparable sales are identified and adjusted to the subject property, or the specification and calibration of a direct sales comparison model.

3.2.2.1 Comparable Sales Method

In the two-step process (also referred to as the "appraisal emulation" method), one model is developed to identify comparable sales and a second model is developed to make adjustments for differences between the subject property and the identified comparables. The first model will include data items important in determining comparability and may involve the calculation of a dissimilarity measure, such as the Minkowski or Euclidean metrics. A second model will include data items significant in directly estimating value from the market and is used to adjust the selected comparable sales to the subject. Model specification for the comparable sales method can be summarized as follows:

$$MV_S = SP_C + ADJ_C$$

- MV_S represents the market value estimate;
- SP_C represents the selling price of a comparable sale property; and
- ADJ_C represents adjustments to the comparable sale.

(Gloudemans 1999, 124.)

3.2.2.2 Direct Market Method

The direct market method involves specification and calibration of a single model to predict value directly. The model may take one of three forms: additive (also termed "linear"), multiplicative, or hybrid (also termed "nonlinear'). Basically, in an additive model, the contribution of each variable in the model is added together. In a multiplicative model, the contributions are multiplied. Hybrid models can accommodate both additive and multiplicative components. The choice of model specification usually depends on the prior experience of the market analyst and the type of property being appraised. Additive models are the most prevalent of the three, based on tradition and wide availability of software programs. Nonlinear (hybrid) models are used the least due to limited software availability, but these models more accurately reflect the combination of additive and multiplicative relationships in the real estate market.

Additive models have the form:

$$MV = B_0 + B_1*X_1 + B_2*X_2 + \ldots$$

- MV is the dependent variable;
- B_0 is a constant;
- X_i represents the independent variables in the model; and
- B_i are corresponding rates or "coefficients."

In a direct sales comparison model, "MV" is either sale price or sale price per unit. In an income model, the dependent variable is income or income per unit. Additive models are relatively easy to calibrate and understand.

In a multiplicative model the contribution of the variables is multiplied rather than added:

$$MV = B_0 * X_1^{B1} * X_2^{B2} * \ldots$$

In this example each variable is raised to a corresponding power. However, the process can also be reversed as illustrated by the third variable in the equation below:

$$MV = B_0 * X_1^{B1} * X_2^{B2} * B_3^{X3} \ldots$$

Multiplicative models consist of a base rate (B_0) and percentage adjustments. They have several advantages, including the ability to capture curvilinear relationships more effectively and the ability to make adjustments proportionate to the value of the property being appraised. Multiplicative models are usually calibrated using linear regression packages. This requires some of the variables to be converted to logarithmic format for calibration, which can complicate model development and application.

Hybrid (nonlinear) models are a combination of additive and multiplicative models. As such, they are theoretically the best alternative of the three, but software is relatively limited.

A general hybrid model specification that separates value into building, land, and "other" components (e.g., outbuildings) is:

$$\mathbf{MV = \pi GQ * [\pi\ BQ * \Sigma BA) + \pi LQ * \Sigma LA) + \Sigma OA]}$$

- MV is the estimated market value;
- πGQ is the product of general qualitative variables;
- πBQ is the product of building qualitative variables;
- ΣBA is the sum of building additive variables;
- πLQ is the product of land qualitative variables;
- ΣLA is the sum of land additive variables; and
- ΣOA is the sum of other additive variables.

(IAAO 1990, 351; Gloudemans 1999, 124.)

3.2.3 Income Approach

Income-producing real property is usually purchased for the right to receive future income. The appraiser evaluates this income for quantity, quality, direction, and duration and then converts it by means of an appropriate capitalization rate into an expression of present worth: market value. If expense data are available, the steps in this approach are:

1. Estimate gross income, expenses, and net income from market data.
2. Select the appropriate capitalization method (model specification).
3. Estimate a capitalization rate or income multiplier (model calibration).
4. Compute value by capitalization.

(IAAO 2002.)

While there are many model specifications of the income approach, the basic overall direct capitalization formula is:

$$\mathbf{MV = NOI/R}$$

- MV is the price examined in the calibration and resulting estimate of market value;
- NOI is the net operating income; and
- R is the overall capitalization rate.

Another income approach methodology uses gross income multipliers (GIMs):

$$\mathbf{MV = GI * GIM}$$

- MV is the price examined in the calibration and resulting estimate of market value;
- GI is the gross annual income; and
- GIM is the gross income multiplier.

Gross rent multipliers are the same as gross income multipliers but relate to monthly gross incomes.

3.3 Stratification

In stratification, parcels are sorted into relatively homogeneous groups based on use, physical characteristics, or location. Properties are first stratified by use such as agricultural, apartments, commercial, industrial, and residential. Additional stratification by physical characteristics or value ranges may be performed to minimize the differences within strata and maximize differences among strata. Geographic stratification is appropriate wherever the value of various property attributes varies significantly among areas and is particularly effective when housing types and styles are relatively uniform within areas (IAAO 1990, 119). Location stratification reduces the need for complex models. However, excessive stratification may provide too little variation in the data.

When the market for a given type of property is national in scope, it may be possible to create national valuation models without stratification if location adjustments are included as part of the model specification and calibration processes.

3.4 Location

Location is the numerical or other identification of a point (or object) sufficiently precise so the point can be situated. Location has a major influence upon property value. Location analysis can be used to measure the relative impact on value from the neighborhood level down to the individual property level. Location influences within a given model area can be measured by including location variables in the model, or can be established through an analysis of the residuals (errors) from a model developed without location factors.

Two specific methods to develop location adjustments are the creation or use of existing neighborhoods and LVRSA. Neighborhoods are the traditional and most common form of location analysis. In AVMs, neighborhoods may be based upon streets and natural boundaries, government assessor-designated areas, census tracts, or postal delivery codes. LVRSA techniques relate relative prices as measured, for example, by the ratio of each sale price to the median price to each property's unique location, as represented by its geographical coordinates. Software that provide the ability to perform LVRSA use a variety of smoothing techniques to compute a unique location adjustment, termed the relative location value (RLV), for each property. At a more sophisticated level, residuals from a first model developed without location variables can be plotted and analyzed to create the RLV grid. This variable is then included, along with other variables, in a multiple regression or other model to capture location influences.

4. CALIBRATION TECHNIQUES

Model Calibration is the development of the adjustments or coefficients through market analysis of the variables to be used in an AVM. The definition of an AVM used in this standard, emphasizes the use of statistical models and procedures in the development of the AVM. The majority of AVMs in use rely strictly on statistical models as the method of calibration, however *USPAP* Standard 6 (Appraisal Foundation 2003, 46–56) provides recognition of other acceptable methods.

Multiple linear regression and nonlinear regression are clearly based in statistics, while adaptive estimation procedure is based on a tracking method from the engineering sciences. Neural networks emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. Artificial neural networks are collections of mathematical models that can emulate some of the observed properties found in the real estate market.

4.1 Calibration Using Multiple Regression Analysis (MRA)

MRA is a statistically based analysis that evaluates the linear relationship between a dependent (response) variable and several independent (predictor) variables, and extracts parameter estimates for independent variables used collectively to estimate value in a mathematical model. Models produced using MRA come with a rich set of diagnostic statistics that provide evaluation tools for the market analyst to compare results between and among specified models. These goodness-of-fit statistics provide information about each variable's significance in predicting value, and how well the variables in the model work together to produce creditable results overall. Users of AVMs should be familiar with the key measures of goodness-of-fit, and review them before accepting AVM results generated by the MRA process.

4.1.1 MRA Assumptions

The accuracy and credibility of an MRA model depend on the degree to which certain assumptions are met. The most important assumptions are complete and accurate data, linearity, additivity, normal distribution of errors, constant variance of the errors, uncorrelated independent variables, and sample representativeness.

Complete and accurate data is required if MRA is to achieve predictive accuracy.

Linearity assumes the marginal contribution to value by an independent variable is constant over the entire range of the variable. When additive models are used, this assumption may not be supported in the market place, requiring a transformation of the variable. Additivity continues with the concept of marginal contribution in that any one independent variable is unaffected by the other variables in the model. In other words, linear additive models do not possess the ability to measure nonlinear effects or interactive effects of market conditions, without transforming raw variables. In such cases, one must consider using nonlinear or hybrid models.

Normal distribution of errors follows the assumption that the data are normally distributed, and therefore, any error in predictions is also normally distributed. Without the assumption of the normally distributed errors, the inferences for using the standard error of estimate and coefficient of variation (COV) as a measure for goodness of fit are meaningless. Constant variance of the error term implies that the residuals are uncorrelated with the dependent variable, which is the sale price. In other words, as the price level changes, the error term remains constant or homoscedastic; when unequal variances occur at different price ranges, it is heteroscedastic.

A term known as multicollinearity describes the condition where independent variables are correlated (measure the same thing) with each other. Depending on the method used, regression may reject one variable as insignificant or exaggerate coefficients for both variables, if multicollinearity is introduced into the model. A correlation matrix is a good tool when testing for multicollinearity.

It is assumed the sold properties data from which models are constructed are representative of the properties to which they are applied. It is important that both low and high value properties be represented in the model. Data should also be divided into training samples used to develop the model, and holdout samples (control samples) used to test model results.

Because of its robust character, minor violation of this assumption will not dramatically impact results. Poor data quality or samples not representative of the population will produce poor performing models.

The market analyst must be able to present the MRA results in an understandable and defensible format that appraisers and AVM clients can easily understand.

To avoid seriously violating assumption of linearity, additivity, and constant variance of the error term, the market analyst must consider the use of transforming variables or other calibration methods described in the standard. A multiplicative, nonlinear, or hybrid model structure is best for measuring interactive effects.

4.1.2 Diagnostic Measures of Goodness-of-Fit

Both the market analyst using regression and the user of AVM output must be aware of and understand how the various key statistical measures used in regression relate to the reliability of results. These statistics fall into two categories: overall measures that aid in the interpretation of model performance and individual variable measures that assist in the understanding of how well an individual variable performs in helping to estimate value, as well as keeping the standard error term to a minimum. Primary measures of goodness-of-fit for overall model performance are the coefficient of determination (R^2), standard error of the estimate (SEE), COV, and average percent error.

Goodness-of-fit measures for individual variables in a model are produced by most MRA software packages and include the coefficient of correlation (R), T-statistic, F-statistic, and beta coefficients. Each of these measures will provide information about an individual variable's linearity or importance of contribution toward improving predictive success, and relative importance, as variables are compared to each other.

(D'Agostino and Stephens 1986.)

When all the measures are used collectively, along with an understanding of data quality issues, those skilled in developing and using MRA can fully evaluate the credibility of the AVM estimates. Appraisers asked to review AVM results must understand the role that *goodness-of-fit* statistics play in evaluating AVM results. The application of AVM results to a single property may be better evaluated using historical market comparisons selected from a subset of data. Appraisers asked to review AVM results should review the Appraisal Standards Board's *USPAP* Standard and AO-18.

(Appraisal Foundation 2003, 46–56, 180–187; IAAO 1990; D'Agostino and Stephens 1986.)

4.1.3 MRA Software, Options and Techniques

MRA is the most widely used method for calibrating models. As such, the availability of MRA software provides users many choices. No one software package is deemed superior to another, as success using MRA is a combination of modeling skills and software familiarity. Variations of a selected MRA technique can be a decisive factor in selecting an MRA and statistical application package. Many MRA techniques have been adopted over the years to help regression take better advantage of its predictive powers. Stepwise, constrained, robust, ridge regression, and others are acceptable techniques used to improve predictive success. Many of the statistical software packages include variable selection routines that aid the market analyst in selection of significant variables.

4.1.4 MRA Strengths

1. Goodness-of-fit statistics—gives credence to the validity of results.
2. Software availability—many regression software products are available.
3. Widely-accepted calibration method.
4. Broad education network—MRA is taught at most colleges and universities around the world.
5. Credible values—in the hands of a skilled market analyst, MRA is proven to produce results that meet the test of model performance.

4.1.5 MRA Weaknesses

1. Requires a high level of statistical knowledge—market analysts must possess significant background in data analysis and statistical methods.
2. Predictive accuracy is restrained by assumptions.
3. Requires data sets that meet the test of sample size.
4. Interactive and nonlinear market trends are difficult to measure without transforming data.

4.2 Calibrating Using Adaptive Estimation Procedure (AEP)

Adaptive Estimation Procedure (AEP) is a calibration technique that was adapted to real estate value in the early 1980s. Also known as feedback, AEP is based on an engineering concept that relies on continual adjustment to coefficients as the calibration engine passes, or tracks, back and forth through the data until convergence, (minimum error is achieved) thus the feedback. For property valuation, the algorithm tracks the sale price as a moving target. It compares property characteristics as variables that measure the change in sale price, and calibrates a coefficient for each variable. The coefficients are used to estimate value that is then compared to sale price. A running tally is kept on the error term as the process continues. Figure 1 depicts the feedback loop.

AEP will make multiple passes through the sales file constantly adjusting coefficients before a final solution is reached. Success using AEP is dependent upon the market analyst's ability to properly specify a model with characteristics that measure and evaluate local market conditions. Market analysts using AEP have considerable control over the variables used in the model and the coefficient amounts. AEP uses whatever variables are introduced into the model. No variable is excluded because of insignificance. As part of the specification phase, the model can be pre-calibrated with starting, minimum, and maximum coefficients in order to help it converge sooner, and to help ensure that rational coefficients will be produced. Setting the starting, minimum, and maximum coefficients is analogous to constraining coefficients used in constrained regression.

4.2.1 AEP Model Structure

A hybrid model structure has the ability to directly deal with interactive and nonlinear effects found in the market place. The structure closely resembles a cost model; however, the calibrations give it the benefit of a direct market model. The flexibility of the hybrid model built into AEP allows the qualitative variables to be calibrated in two different ways: as multiplicatives, that is Xi^{B1} (rates), or binaries $B1^{Xi}$. Deployment of a feedback model in an AVM format allows for flexibility without the added complexity of transformations found with additive models.

4.2.2 Variable Control in AEP

Calibration of individual variables in AEP differs significantly from the fitting of a straight line or curve in linear or nonlinear regression. Controlling for extremes in the coefficient amounts is a concern when using feedback. The use of smoothing and damping factors will help provide model stability during the calibration phase. Smoothing is applied to only the quantitative variables. Using an algorithm, smoothing keeps track of each variable's exponentially smoothed mean (moving average) as a way of learning until a final solution is reached. Smoothing factors are used in conjunction with damping factors. The market analyst provides the settings for the smoothing factor. Additionally, damping factors control the amount of movement each coefficient (quantitative and qualitative) will have as each new case is introduced into the model while calibrating. Some feedback systems will dynamically adjust damping and smoothing for optimized results. Locking or constraining coefficient movement, forces residuals onto another variable. With so much control over the model, even similarly specified models may produce different final answers.

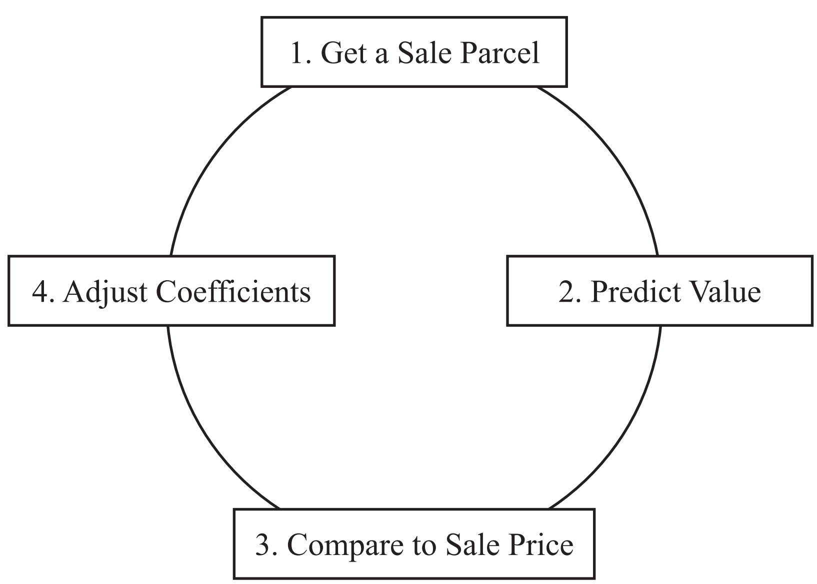
4.2.3 Results and Goodness-of-Fit Measures

Final results using AEP are measured first by the comparison of how close the estimated price comes to the actual price. Another measure, the reasonableness of coefficient amounts, is based on the skill and knowledge of the analyst in pre-defining the model prior to calibration. AEP does not care if a model uses square foot of living area at a price or the window count at a price. If either can logically predict accurate value estimates, AEP will generate a coefficient that produces the lowest error term. Feedback understands that grouped patterns of property characteristics are the determinants of price and the individual characteristics do not necessarily produce marginal contributions to price.

The AEP is not reliant on statistical measures of the model, or variable significance. Convergence occurs

FIGURE 1.

AEP–Feedback Loop



when the average absolute error does not change appreciably from one iteration to another. Some software allows other criteria to be set by the user (e.g., maximum iterations, pre-defined absolute error). There is no statistical measure that accounts for significance of a variable. Pseudo-R^2 statistics can be generated after the feedback model is complete. Output should include the accounting for calibration of each variable by giving information on the number of observations, starting, minimum and maximum ranges, and the low, high, and final coefficient of the variable.

4.2.4 AEP Advantages

1. Produces separate estimates for land and improvements.
2. Based on reducing the absolute error term, not just minimizing the squared error term.
3. Outliers' influence can be diminished during variable calibration cycle.
4. Requires fewer observations than regression.
5. Individual variable movement can be easily constrained.
6. Cost system attributes can be directly calibrated.

4.2.5 AEP Disadvantages

1. Software availability is limited, and there is no standardized algorithm.
2. Does not contain standard goodness-of-fit statistics found in regression software.
3. Requires initial model be specified carefully.

As an alternative, some market analysts have turned to using nonlinear regression software. Nonlinear regression supports the hybrid model and can calibrate interactive effects and curves simultaneously like the AEP/Feedback routine.

(Ward and Steiner 1988; Gloudemans 1999, 196; Woolery and Shea 1985; Carbone 1976.)

4.3 Artificial Neural Networks

The most recent adaptation for use in calibrating real estate valuation models is Artificial Neural Networks (ANN). The concept is borrowed from the biological sciences and functions of the human brain. The key element of the ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to neurons and are tied together with weighted connections that are analogous to synapses. As the name implies, ANN comes as close to producing artificial intelligence models as any calibration method. Like nonlinear regression and feedback, neural networks can calibrate models that consist of both linear and nonlinear terms simultaneously. The user inputs each variable with assigned weights (coefficients). The software exposes the data using an algorithm in a hidden layer where the weights are adjusted (calibrated) in a manner that reduces the squared error. This is an iterative process much like those found with feedback and nonlinear regression. The final output results in a single estimate of value with the exact formula remaining hidden from the market analyst.

4.3.1 The Artificial Neuron

The basic unit of neural networks, the artificial neurons, simulates the four basic functions of natural neurons. Those functions are represented by inputs, the processing of inputs (summation), transfer (linear, sigmoid, sine, and so on), and outputs an answer. Artificial neurons are much simpler than the biological neuron; Figure 2 shows the basics of an artificial neuron.

Inputs to the network are represented by the mathematical symbol x(n). Each of these inputs are multiplied by a connection weight that is represented by w(n). In the simplest case, these products are simply summed, fed through a transfer function to generate a result, and then output.

Even though all artificial neural networks are constructed from this basic building block, the fundamentals may vary in these building blocks.

4.3.2 Strengths of Neural Networks

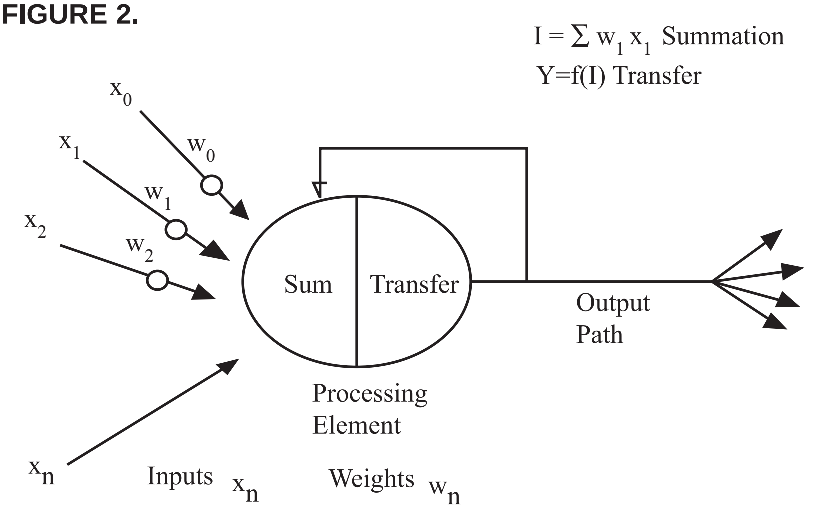
1. The ability of the neural network to "learn" as it goes and to take new information and process as it has been trained.
2. Neural networks can recognize and match complicated, vague, or incomplete patterns in data.
3. Options that provide analysts confidence about future use of neural network applications, such as helping to improve data quality.
4. Studies completed indicate that the accuracy of neural networks is comparable to other calibration methods found in the standard.

4.3.3 Weakness of Neural Networks

1. The complexity of how the process actually works in the hidden layer.
2. Lack of a definable model structure at the output stage makes explanation of value and support of the value more difficult.
3. Requires considerable background in data analysis, data structure, and mathematical concepts.
4. Limited research links pertaining to use in real property valuation.
5. Requires considerable investment in computer power and software.

(Gloudemans 1999, 329.)

FIGURE 2.



4.4 Time Series Analysis

Time series analyses are a family of techniques that can be used to measure the cyclical movements, random variations, seasonal variations, and secular trends observed over a period of time. In property valuation, these analyses can be used to develop a multiplier or index factor to update existing appraised values or to adjust sales prices for individual properties to the valuation date. Since values can change at different rates in different markets, separate factors should be tested for each property type and market area.

Four methods used to develop time trend factors in the appraisal and assessment industries are: (1) value per-unit analysis, (2) re-sales analysis, (3) sales/assessment ratio trend analysis, and (4) inclusion of time variables in sales comparison models. These methods are summarized below (for a more detailed explanation and discussion, see *Mass Appraisal of Real Property* (Gloudemans 1999, 263-270).

Value per-unit analyses track changes in sale price per unit (e.g., per square foot for residential properties or per unit for apartments) over time. The method is easily understood and lends itself well to graphical representation, as well as to statistical modeling to extract the average rate of change. A downside is that the method does not account for the myriad of other value influences, such as age and construction quality, that impact per-unit values.

Re-sales analysis uses repeat sales occurring over a given time period. Price changes between sales are converted to monthly rates and an average (or median) rate of change is extracted. As can be imagined, the larger the number of repeat sales, the more reliable the estimated rate of change. The method can overestimate rates of change if repeat sales reflect substantial improvements (or other alterations) made to the property since the first sale.

Sales/assessment ratio trend analysis involves tracking changes in the ratio of sales prices to existing assessments made as of a common base date. Increases in the ratios indicate inflation and vice versa. The ratio also provides the index factor required to convert assessed value to a full value estimate. Like value per-unit analysis, the method lends itself well to graphical and statistical analysis. An advantage of the method is that assessments account for most value determinants and thus can isolate time trends better than the value per-unit method. The method assumes that the assessments share a common basis, and its reliability depends partly on the accuracy or uniformity of the assessments.

Time variables can be included directly into AVM models to capture the rate of price change over the period of analysis. This is usually the most accurate of the various methods. However, model developers must be careful that time variables are properly specified so that coefficients developed from the model reflect the desired valuation date.

Once a time trend is established, it can be used to adjust values to any point within the sales period.

Trend factors can be extrapolated for a short period beyond the sales period, but this must be done with caution and grows increasingly unreliable as the time frame is lengthened. If more than several months are involved, the first three methods can be used to calibrate the trend (one would not ordinarily develop time adjustments through use of a modeling approach without recalibrating the entire AVM model).

(The Appraisal Institute 2002, 291.)

4.5 Tax Assessed Value Model

Tax assessed value models derive an estimate of value by examining values attributed to properties by the local taxing authorities. As a matter of local law and custom, the values reported by the taxing authorities often (but not always) vary from the current market value in some reasonably predictable manner. For example, some jurisdictions require the taxing authority to report the assessed value at 25 percent of the estimated market value. Some jurisdictions may not have reappraised in a long time, so values lag far behind the current market. Also, some jurisdictions report multiple values: assessed, appraised, and market values. By examining local laws and customs with respect to how values are determined, as well as applicable time trends (see Section 4.4) and information reported in local or state-level ratio studies (see Section 8.4), it may be possible to develop adjustment factors to apply to values reported by taxing authorities in order to approximate current market values.

The reliability of a tax assessed model will depend on the uniformity of appraisals to which the adjustment factors are applied, as well as the accuracy of the adjustment factors themselves, which can vary with how current the assessments are, and the reliability of the ratio studies or other information on which they are based. Extreme caution must be exercised when local assessment uniformity is poor, because factoring an unreliable assessed value will only result in an unreliable market value. On the other hand, local assessments that meet IAAO standards can provide a sound basis for market values estimation.

4.6 Calibration Summary

The various methods and procedures used to calibrate the AVM are the engines that drive accuracy and credibility of the estimate made. By itself, no one calibration method is better than another. Data integrity and the skill level of the analyst define the accuracy of one calibration technique as compared to others. Users of AVM products must be aware of the interdependence between skills and technologies of calibration when deciding how well the AVM will perform.

The use of MRA has been the longstanding choice for calibration and has a proven track record. Feedback, nonlinear regression, and neural networks are emerging technologies that require different levels of skill and knowledge concerning modeling real property values. Understanding calibration in relation to this standard encourages the AVM market analysts and clients to understand that AVM development is not a black box process; instead, it is based on well-defined concepts surrounding the appraisal process. Details for learning and understanding the skills and technical aspects of calibration are found in the references throughout this section of the standard.

AVM clients must understand that developers of AVM products are not limited to using a single method of calibration. Product market analysts often base their value estimates on multiple technologies. Included in these technologies are simple sales listings of automated sales comparison selections, with adjustments derived from the modeling process. Appraisers asked to use or review an AVM should read Advisory Opinion 18—published as part of *USPAP* Standard 6 by the Appraisal Foundation (2003, 46–56, 180–187).

5. RESIDENTIAL AVMS

The residential property class has the longest history of being valued by AVMs. Residential property includes detached single-family homes, condominiums, townhouses, and zero-lot-line property. Other property types included in the residential class are properties with four units or less. Traditional methods of valuing these properties are cost approach and direct sales comparison. Both methods have been automated and are considered a part of the AVM category of methods and techniques available.

5.1 Detached Single-Family

When adequate sales data is available, the direct sales comparison approach is the preferred method of valuing residential property. The approach may take two forms: direct market models and comparable sales.

Direct market models developed from sales analyses use various model structures, with coefficients derived via a mathematical calibration method. The comparable sales method is a two-part method in which comparable sales are found and then adjusted to the subject property.

Some AVMs combine the strengths of direct market models and comparable sales models, to the point where comparable sales model coefficients are derived from direct market model analysis.

Cost models, like sales comparison models, have a strong history of reliability and credibility for valuing residential property. However, the origin and accuracy of coefficients are unknown to most users and may not reflect the actual market.

5.1.1 Cost Models

The cost approach works best when applied to newer properties that do not exhibit a great deal of measurable depreciation, and where the land value can be reasonably estimated from recent land sales. Cost models are anchored in tables developed by studying local building cost data. In the AVM format, the tables are converted to a formula and applied by simply entering basic building (improvement) information. Such models are used for deriving the Replacement Cost New (RCN). The initial cost coefficients supplied with a cost model represent the supply side of the residential market. These RCN estimates need further calibration for actual property condition (depreciation), location (macro and micro), and a supportable estimate of vacant land value, in order to arrive at market value. These items represent the demand side of the market. A strength of the cost model is that it can be applied to any improvement regardless of size, quality, age, condition, or style. The accuracy and credibility of the cost model is tied to the analyst's ability to calibrate depreciation, location, and land value.

5.1.2 Comparable Sales Models

Knowing the sale price of a property with attributes similar to the subject property is a concept that consumers can easily understand. This approach provides the theoretical basis for the Sales Comparison Model using comparable sales. Sales comparison of residential property has been accepted by real estate consumers and the courts for many decades; however, this method does have limitations in the automated world. It essentially requires two models. The first one is a comparable selection model. Many AVMs rely on identification and summarization of all recent sales within a specified radius of the subject. The advantage of this model is that all recent sales with close proximity to the subject are considered. This method may work well in homogeneous areas with a high sales volume. If the comparables have significant attribute differences, the confidence of the adjustments being made also begins to suffer. For quality comparables, an AVM routine may consider using a weighted selection model (e.g., regression coefficients, Minkowski or Euclidian metrics). Another choice would be cluster analysis.

All of these methods can select comparables based on attribute comparisons that pick the comparables most similar to the subject, based on defined parameters. These methods are not limited to selecting only three sales, as has been the tradition. Once the best comparables are selected, they must be adjusted for attributes that are dissimilar to the subject. How these adjustments are developed has much to do with how accurate and reliable the sales comparison estimate will be. Mathematically, the adjustments can be derived from just two sales; one sale possesses the attribute, while the other does not. The difference in sale price measures the value of the missing attribute. Sales comparison methods that rely on direct market models that use quantitative methods for deriving the adjustments, are more stable and reliable than simple match pair analysis.

In its formatted form, the comparable sales approach should display how each attribute adjustment in the AVM contributes to the overall value estimate. Users of AVMs are cautioned that matched pairs analysis is not a statistical calibration method. Any comparable sales approach claiming to be an AVM as defined in this standard must meet the criteria of being supported by an automated market analysis process.

5.1.3 Direct Market Models

The basic premise of direct market models (also termed hedonic models) is that the price of a marketed good is related to its characteristics, or the services it provides. For example, the price of a home reflects the characteristics of that home (e.g., size, construction quality, style, location). Therefore, we can value the individual attributes of a home by looking at the prices people are willing to pay for them. Direct market models lend themselves well to the calibration methods and techniques discussed in Sections 2–4 of this standard. If a value-determining attribute can be captured in a database, then the model can calibrate a coefficient that measures its contribution to the total value estimate. Properly designed direct market models will produce AVMs capable of very accurate and credible value estimates.

All three model structures introduced in Sections 2 and 3 are well suited to the valuation of single-family residences. Additive models have been the traditional workhorse and work very well in most cases. Multiplicative models carry certain advantages discussed earlier and can also be effectively adopted. Because they accommodate dollar and percentage adjustments, hybrid models provide the most flexibility. Where an additive model will add the same lump sum amount to all property having air conditioning, multiplicative and hybrid models will attribute different amounts depending on the style, quality, and location of the property. Both model structures also lend themselves well to the valuation of spatially dispersed or highly heterogeneous residences.

5.2 Attached Residential Property (Condominiums, Townhouses, Zero-Lot-Lines)

Structures built on an individually plotted lot designed for only one family to occupy, are termed "detached single family residences" and make up the majority of residential property in most communities. Zoning and other spatial changes in a community dictate the density of residential land use. Other methods of dividing land, besides using land-based boundaries, lead to other types of residential use and ownership. Structures where multiple living units are all joined together take on different forms of ownership depending on how the title is legally conveyed in the market place. These structures are commonly referred to as "attached residential units." A ten-story building with five units per floor could be an investment property with each unit rented. Property divided into air lots is known as condominiums. Another division of ownership rights is by time, where each day, week, or month represents units of ownership. Structures where the ownership is divided vertically are known as townhouses, row houses or zero-lot-lines, depending on geographic location throughout the world. All of these uses are residential in nature.

Valuing these various residential properties is somewhat similar to valuing detached single-family structures. All of the same principles apply and all can be modeled and valued using an AVM. In fact, because these properties exhibit a high degree of homogeneity compared to the detached single-family population, sales-based AVMs can produce values that are extremely reliable and accurate. The cost approach can also work well, in some cases, if adjusted to the market, but it is not appropriate for valuating condominium units because depreciated replacement cost will not properly reflect resale values. Data requirements for attached residences will not be the same as with detached residential properties. For example, floor level can be an important value determinant for condominiums, while lot size and yard improvements are irrelevant.

5.3 Two- to Four-Family Residential Property

Part of the residential housing market consists of structures built for the purpose of housing more than one family. Improvements designed to accommodate two, three, and four families within their own separate living areas are often referred to as small income-producing properties. A common theme among these property types is that the owner of the property may reside in one of the units. This concept, however, is not a requirement for classifying these structures in the market. Two-unit properties are more likely to be owner-occupied than four-unit properties. The concept to be recognized here is how such properties are treated in the marketplace, because that impacts their price and ultimately the value generated by any AVM. The ability to model the selling price of these small-income properties is reliant on what specific data is available, relating to number of units, age, condition, location and gross income. The motivation of buyers shifts when consideration is given to other property attributes that relate to producing rental income and not just owner occupancy. Direct market models, comparable sales models, and cost models are acceptable methods for valuing these small income-producing properties. With their income-producing potential, the income approach is also a model to be considered. With an adequate sample of gross income values for comparison to sale price, a model of GI * GIM will yield credible results where GI = Gross Income and GIM = Gross Income Multiplier (sale price/gross income). Some AVMs may even be set up to predict GI and the GIM. Each of these indicators can vary with size, age, location, style, and condition of a property.

5.4 Manufactured Housing

A manufactured home is a residential structure built in a factory. Construction standards for manufactured housing are controlled and monitored by the Department of Housing and Urban Development in the United States (HUD), and by the Canada Mortgage and Housing Corporation (CMHC) in Canada. While many manufactured homes are built with the same materials as site-built homes, the factory-controlled engineering process helps control cost and quality. The house can be financed as personal or real property on leased land, in a manufactured home community, or on a privately owned site. Buyers who desire to acquire land in conjunction with the home can finance the land and home together. Market conditions and trends will indicate how the manufactured homes compete in the market place. In some communities, zoning only allows manufactured homes in certain areas, confining the market area from which comparables can be derived. Once market conditions for a manufactured home are known, it can be modeled just like any other property type. Consistency is important when using an AVM for manufactured homes. Some manufactured homes are strictly treated in the market as mobile homes (i.e., personal property). An AVM developed to value manufactured homes as real property would give a false value in the case where the home was personal property, and vice versa. AVMs developed to value manufactured personal property homes cannot be used for homes classified as real property. Some manufactured homes compete in the market place with site-built homes. Where this is the case, it is possible that an AVM designed to value detached single-family structures will produce credible results, although the model should include a variable (or variables) to capture any differences between otherwise comparable manufactured and site-built homes.

5.5 Time Series Models for Residential Property

Indexed models relate to time-series analysis (see Section 4.4 on Time Series Analysis) as described earlier. Use of these models represents a common method of delivering quick automated value estimates. These models simply measure the average change in value over time and factor the value forward from a benchmark starting-place, such as the average value in a census block or market area. The accuracy of indexed models is inconsistent and less reliable than fully specified models. These models work best in areas of homogeneity where the range of value is close to the average value.

Indexing is a common method used to update cost tables to reflect current cost. As with market models, a benchmark in time is required as a starting point. Cost coefficients are then updated, using a single index factor representing the measurable change since the original cost coefficients were generated. One current method of indexing is to use an economic indicator such as the consumer price index (CPI). In the cost approach, indexed models have no way of adjusting values at the micro level for location and other market influences that impact value. Time adjustments may be developed from the analysis of known sale prices within a geographic area, such as a neighborhood or postal code, and over a specified time reference.

Users and consumers of index models must understand how the index factor is created and how the accuracy of the original value was derived before giving a lot of credibility to an AVM using an indexed model.

5.6 Summary and Conclusions for Using Residential AVMs

AVM developers (or users) must understand the intended use of the residential property. The residential housing market is diverse. AVMs lend themselves to estimating the value of residential property. However, each class of residential property has some unique circumstances that will influence how well the AVM can perform when estimating the value. When the uniqueness is captured as part of the data used to develop the AVM, the chances of the value estimate being accurate and credible increase greatly. When unique characteristics are ignored, they are not measured in the market and the error term of the values produced will increase, destroying confidence in the AVM's ability to estimate accurate and credible values.

The overall ability of the AVM to accurately estimate value can be evaluated using the quality assurance measures found in Section 8 of this standard. If the assurance standards are being achieved, then the validity of the AVM is known, and the market analyst and users can understand what degree of confidence to expect from the ensuing value estimates.

6. COMMERCIAL AND INDUSTRIAL AVMS

Commercial and industrial properties, including apartments and multifamily residences with greater than four (4) units, are usually income-producing properties acquired for their ability to generate income. As a result, commercial and industrial properties are best valued using an Income Approach where adequate income data are available or the sales comparison approach where adequate sales are available. However, a solid Cost Approach is needed where sales and/or income data are insufficient to calibrate an appropriately structured model. Also, care must be taken in developing and applying income valuations, to appraise only the real property and not the business, and to value based on typical management, not on the present management.

Commercial and industrial properties provide their unique AVM challenges. First, in some markets there are relatively few sales of commercial and industrial properties. This creates problems with land valuation for the cost approach, development of comparable sales or statistical models, and for developing capitalization rates and multipliers for the income approach. The market analyst may have the additional problem of needing to provide separate land and building values that most income models are not designed to deliver.

Location for commercial and industrial properties can range from relatively little effect to extremely important.

Finally, special purpose properties and limited market properties, such as theme parks and casinos, are generally included with commercial and industrial properties. These properties tend to be unique and, as a result, are difficult to categorize and value.

6.1 Commercial and Industrial Model Specification

Valuation of commercial and industrial properties requires market and income or cost data. Income and market data are preferred. Cost data is needed where insufficient sales and income data are available. Commercial and industrial sales comparison models, like residential models, require data on use, location, and physical characteristics.

6.1. Property Use

The property use is extremely important as a comparison characteristic. It is necessary to determine the general category of property use. The property use does not need to distinguish detailed specific uses, such as shop versus liquor store or gift shop. Use of broad categories will increase the number of properties for which information can be captured, analyzed, and compared.

6.1.2 Location

As with residential properties, location can be included either through the use of neighborhoods or market areas with binary (dummy) variables or categorical variables with percentage adjustments, LVRSA, or as a distance variable to Value Influence Centers (VICs), such as the central business district. For commercial and industrial properties, location analysis relates largely to identifying zones or groups of properties subject to similar influences. Proximity to VICs is important in commercial and industrial valuation, but a lack of commercial and industrial sales may make location of the VICs, as well as measuring their effect, difficult.

The importance of a location adjustment will also vary considerably with the property use. For example, while the value of a service station generally depends on location on a major street, such a variable may not be needed if all service stations throughout the jurisdiction, or area, enjoy such locations. Other uses, such as hotels, may be highly dependent on location, such as being on a beach area or near a convention center. Finally, what is considered a nuisance for residential properties, such as a railroad track or heavy traffic pattern, could be an important amenity for commercial and industrial properties.

The most common method of modeling location is through the delineation of economic areas or neighborhoods. Central Business Districts, cities/towns, and other areas of significant deviation from the norm, are used to identify economic areas or neighborhoods. When using economic areas or neighborhoods for location adjustment, care must be taken not to create too many as this may result in too few sales or insufficient income data for analysis and modeling.

LVRSA using GIS or manual grids is also used for developing location adjustments where sufficient data is available.

6.1.3 Physical Characteristics and Site Influences

Commercial and industrial properties require a number of physical characteristics for comparison and modeling. These may be quantitative or qualitative variables.

The most significant quantitative characteristic is building area. Different building areas may be used for cost, sales comparison, and income approaches. Areas are differentiated by type, such as basement, ground floor, and upper floors, for cost valuation; whereas income models generally use net rentable areas, differentiated by use, such as retail, office, etc. Sales comparison models also benefit from use differentiations, although either gross or rentable areas can be used. Some sales comparison and income AVMs utilize other units of comparison, such as units for apartment buildings, rooms for hotels, and spaces for parking garages. Other key quantitative variables are the year built and effective age or condition, which are used to capture accrued depreciation and Remaining Economic Life (REL). Effective age (EA) or REL is a critical factor of comparison for cost, market, and income modeling. The EA or REL is also a key variable in determining the relationship between income and value because it establishes the time remaining for the income stream.

Other significant quantitative and qualitative variables are similar to those used for residential AVMs. Such examples include building quality and lot size. While others, such as traffic patterns or ceiling height, may be important to specific property uses or occupancies.

6.1.4 Income Data

An income value is essentially a calculation of the present worth of the future benefits to an income stream. It is used to estimate the market value of a property based on what an investor would pay for the property. Income data includes revenues, expenses, net income, and capitalization rates or income multipliers, which are then used to develop a projection of an income stream to estimate the market value.

The income value is generally estimated by either capitalizing the Net Operating Income (NOI) or developing a multiplier for the potential or effective gross income. The capitalization rate can be developed as an Overall Capitalization Rate (OCR) from the market place by comparing the estimated NOI against sales prices, where available. Sales and gross income data can be used to develop a GIM from the market place. GIMs can be accurate, require less data, and eliminate the need for expense analysis. They can be developed from potential or effective gross income as long as the data is collected on the same basis. While GIMs may be easier to develop, overall rates and NOIs may more accurately and directly reflect the value of the income stream critical to investors.

Due to the sensitivity of income data, the widely varying manner in which it is kept, and the differences in information maintained for differing property types, income data is difficult to ascertain. Creating different reporting forms for different property types makes the forms easier to use and understand, thereby increasing the likelihood that more forms will be completed and returned. Breaking income and expenses into generic categories also facilitates reporting. However, creating too many categories may only complicate the form and minimize the number of completed returns while not necessarily contributing to a more accurate net income calculation. Minimizing the detail collected, including avoiding tracking information about individual tenants, serves to make the data more likely to be completed and easier to maintain.

The pool of sales and income data can be expanded by using multiple years of data and making any indicated time adjustments. However, if the income and sales data are from the same time period, neither needs to be adjusted for time for the purposes of developing capitalization rates and income multipliers. In addition, trade publications and local banks may serve as sources of information to build capitalization rates and multipliers.

6.2 Development of the Model(s)

Commercial and industrial properties can be valued by sales comparison, income, and cost AVMs. Because there are fewer commercial and industrial sales, it is often difficult to develop comparable sales and statistical market models for commercial and industrial properties. However, a number of income-approach models may be developed using sales to develop capitalization rates, and GIMs using gross incomes and expenses derived from the local market or industry-specific publications when local data are insufficient. The cost approach, while generally the least desirable, is still necessary for property types that have insignificant sales and insufficient revenue or expense information.

Income models may be developed using stratification or global methods. Stratification requires grouping commercial and industrial sales by factors that affect the relationship between income and value. This is accomplished by groupings based on use or occupancy, age or condition, and location. As with any valuation approach, the more strata you create, the fewer data in each strata are available for analysis. The use of global methods, such as MRA, can be used to overcome the limited data in many strata by combining selected property types (such as all retail-related properties) into a single model and using binary variables to differentiate the specific uses or occupancies (such as general retail, restaurant, or convenience mart).

Industrial properties may be modeled in the same manner as commercial properties, but there are even fewer industrial sales than commercial sales. Often warehouses and light industrial properties can be combined into a single model to increase sample sizes.

6.2.1 Cost Models

While commercial and industrial cost models are similar to residential cost models, they typically comprise different structural components. The commercial and industrial cost model requires a number of extra features or miscellaneous items. Cost models are most appropriate for commercial, industrial, and special purpose properties where there is insufficient sales and income information.

6.2.2 Sales Comparison Models

It is often difficult to get sufficient qualified sales to develop commercial and industrial comparable sales and statistical models. However, where sufficient sales can be found, direct market models can be developed using variables for location, size, construction quality, age or condition, land size or frontage, and relevant amenities or nuisances. Additive, multiplicative, and hybrid models can all be used; yet proper model specification is critical.

6.2.3 Income Models

The income approach can be used to develop commercial and industrial AVMs. Because these properties are frequently sold based on their income streams, the income approach can be the most desirable. The two most popular approaches are direct capitalization and GIMs. Discounted cash flow (DCF) analysis can also be used; however, the data requirements for developing yield capitalization estimates from DCF analysis make the method more challenging than direct capitalization. Also, a number of the assumptions required for DCF analysis, including anticipated yield, holding period, and value at the end of that period, can be difficult to derive from the market and, therefore, may be subjective.

6.2.3.1 Modeling Gross Income

Gross incomes may be analyzed from local market surveys or questionnaires or they may be obtained from industry publications. Typically the gross rent per unit (e.g., square feet/square meters, rental unit, or room rate) is the dependent variable in the model. Gross income models are ordinarily easier to develop than net income models because the data is easier to obtain and less subject to manipulation. Gross income models can be developed for either potential or effective gross incomes. The independent variables are those that affect the expected gross income, including: location, age or condition, amenities and nuisances, etc. Where data is limited, to develop separate models for each use or occupancy, a single model may be developed by determining a reference use or occupancy group and using binary or categorical variables for the other use or occupancy groups.

6.2.3.2 Vacancy and Collection Losses

Vacancy and collection losses are deducted from the Potential Gross Income (PGI) to account for typical losses due to vacancy and bad debts based on local market conditions. The vacancy and collection loss usually varies by property use and is expressed as a percentage of the annual PGI. The percentage may be determined by a market analysis of PGIs compared with actual income, or from information supplied by local lenders and industry trade publications.

6.2.3.3 Modeling Expenses

Expense data may be obtained from the same sources as the revenue data. Expense ratios can be developed by either stratification or a modeling approach. The expense ratio is the dependent variable, and the independent variables are similar to (but typically fewer than) those used to determine the gross income per unit. Like gross income models, a single expense ratio model may be developed, where insufficient data are available for multiple models, by determining a reference use or occupancy group and creating binary variables for the other use or occupancy groups.

6.2.3.4 Direct Capitalization

Direct capitalization involves developing an overall rate (OAR) directly from the market place. The OAR is then used with the estimated net income to estimate the value by income capitalization. Like expense ratios, capitalization rates can be developed using either stratification or a modeling approach. The advantage of OARs is that they use the NOI that includes both gross incomes and expenses, and thus may specifically reflect a typical investor analysis of commercial properties. The dependent variable in developing a direct capitalization rate is the indicated OAR (estimated net income divided by the sale price). In developing an OAR model, a single model can be developed by determining a reference use or occupancy group, and creating binary or categorical variables for the other uses or occupancy groups. As in the revenue and expense models, this permits more data to be used in the model. In addition to variables for location, age or condition, and amenities and nuisances, the OAR model should include an adjustment for attributes that affect the recapture portion of the OAR (such as land/building value ratios, REL estimates, and expense ratios).

6.2.3.5 Gross Income Multiplier

GIM models involve developing multipliers directly from the market place for either the potential gross income or the effective gross income, depending on the data collected. Effective Gross Income Multiplier models are generally more stable. GIMs have the advantage of not requiring expense data that may be missing, unreliable, difficult to interpret, or incomplete. The GIM is the dependent variable while the independent variables are typically the same as those previously described for an OAR model. However, it is important to ensure that variables related to differences in expense ratios are included because gross incomes are unadjusted for expenses.

6.2.3.6 Property Taxes

Care should be taken to treat property taxes consistently in the development and application of AVMs. Property taxes may be included as an expense or as a component of the OAR.

6.3 Quality Assurance

Commercial and industrial quality assurance is particularly critical due to the limited amount of sales and income data available for analysis and modeling. Commercial and industrial quality assurance is accomplished in much the same manner as with any other type of property. Valuation research and appraisal procedures are subject to review, and the values tested and statistically analyzed for accuracy and consistency.

In addition, quality assurance must be extended to the income data collected. Estimated gross incomes, expense ratios, OARs, and GIMs should all be reviewed for consistency. Gross income and expense data should be compared with like properties to identify outliers that may need to be removed from the modeling process unless the data can be corrected.

7. LAND MODELS

If ample sales are available, vacant land is generally best valued using a sales comparison approach. The most significant exceptions to this are leased land and rural/agricultural land that are usually valued using the income approach.

Land provides a set of unique problems for AVMs. Land is highly speculative and there frequently are relatively few sales for analysis and modeling (see Section 2.3.3 on Data Management and Quality Analysis).

Land values are highly affected by location. This is also one of the reasons why land values appear to be more speculative. Other factors affecting land values include Federal, state, and local regulations affecting development and what stage the neighborhood is in its life cycle. Developed land will command a significant premium over underdeveloped land, especially when there is no guarantee that the purchaser or potential developer will be successful. In addition, neighborhoods evolve from growth to stability, to decline, and potentially to being a land-driven market where the improvements have no value.

GIS is extremely valuable as an aid in establishing the effect of location on land. Where a GIS is not available, neighborhoods can be developed based on appraisal judgment, or grids can be developed and x, y coordinates manually derived from the grids to better handle location.

Land that is significantly distant from urban areas may be best valued based on its income potential.

7.1 Land Valuation Model Specification

Market land valuation modeling requires data on use, location, and physical characteristics. Land models, like improved models, require qualitative and quantitative variables, as well as data transformations.

7.1.1 Property Use

The analyst must estimate the property use of a parcel of land for any AVM. This will serve to determine how it should be appraised as well as provide a key variable for comparison and to determine what sales are best suited for building the model. Although many states and provinces provide use codes for reporting, currently there are no generally accepted standards for classifying land uses. The American Planning Association (APA) has recently provided an update on their Web site of the 1965 Standard Land Use Coding Manual (APA 2003). However, the APA is not an appraisal organization and its solution contains multiple dimensions—whereas appraisers generally focus on the current use and the highest and best use.

7.1.2 Location

Location and parcel size are arguably the most important pieces of land data. The most common method of modeling location is through the delineation of economic (or submarket) areas or neighborhoods. More recently, variations of LVRSA have been developed to determine location adjustments both with and without delineating economic areas.

Appraisers, using maps and their judgment (based on knowledge of market conditions), generally decide neighborhood or submarket boundaries. All parcels in a neighborhood or submarket receive the same location adjustment. There are two factors to be aware of when using this approach. First, boundaries may be drawn to coincide with major streets, natural barriers, and/or political subdivision boundaries. And, secondly, the market analyst should be aware that location adjustments can change abruptly from one submarket or neighborhood to another.

LVRSA is another method of developing location adjustments. LVRSA uses a geographic grid to display value residuals or sales ratios based on values derived from a model lacking a location variable, to develop factors that quantify the relative locational advantage or disadvantage of the property. This process may include identifying positive and negative VICs. Distance variables from all of the VICs are computed for each parcel. If VICs are used, the distance variables are then included in the model to calculate the location adjustment for each parcel. The location adjustments determined in this manner may be developed for cost, sales comparison, and income AVM models. Due to the method LVRSA uses to develop the location adjustment, it will include anything that is not accounted for elsewhere in its estimate of the location adjustment.

Geographic grids for LVRSAs are best obtained from a GIS. However, where one does not exist, a geographic grid can be manually developed by using maps and arbitrary grids, such as every 100 feet. The x, y coordinates can then be determined for each parcel and entered into the database. Although not as accurate and effective as a GIS, this approach can be used where one does not exist or is not yet available to the market analyst.

When using neighborhoods or submarkets for location adjustment, care must be taken not to create too many neighborhoods or submarkets; because this may result in too few sales for effective analysis and modeling. Central Business Districts, cities/towns, natural features, and major streets can be used to define neighborhood boundaries. Because the single property appraiser generally values only a single, or few properties, and the AVM market analyst must value many parcels and sometimes deal with adjacent parcel review by the public, the AVM market analyst might prefer to use blocks, subdivisions, or neighborhoods for location adjustments so that adjacent and nearby parcels receive the same adjustment.

7.1.3 Physical Characteristics and Site Influences

In addition to land use and location variables, AVMs require a number of physical characteristics and site influences for comparison and modeling. These may be quantitative or qualitative variables.

The most significant quantitative characteristic is land size. Land size is determined by the number of land units by type such as lot, site, front feet or meter, square feet or meter, acre, hectare, etc. Therefore, it is usually necessary to develop some form of land size adjustments to reflect the changing rate per unit based on the total parcel size.

Most of the other important characteristics and influences are qualitative. These include topography, site amenities (such as government services), property access, water and sewer, proximity to negative influences (like railroads or treatment plants), and proximity to positive influences (like view, golf courses, water frontage, or recreational areas). However, keep in mind that a negative influence under one condition might be considered positive in another situation. An example might be high traffic volume that could be positive for commercial properties and negative for residential properties.

7.2 Land Data Collection

Sales and income data for land are collected, verified, and maintained in the same manner as improved parcels. Maps and aerial photographs are used to supplement field reviews to effectively collect, maintain, and review land data.

Land use/soil productivity data for income modeling of agricultural property may be obtained from Federal and state/province agricultural agencies, universities, and agricultural cooperatives and associations. When insufficient arms length sales are available, data to develop capitalization rates for agricultural properties may be obtained from farm lenders such as the Federal Land Bank and Farm Credit Bank, as well as local lenders.

7.3 Development of the Model(s)

Sales comparison is the primary approach for estimating the market value of land. The valuation of land by sales comparison shares many of the same analyses and modeling processes with improved valuation models. The dependent variable in a sales comparison model should be sales price or sales price per unit. For example, if land sales in an area are based on square feet of land area, then the dependent variable should be sale price per square foot. Typical independent variables include property use, zoning, size, or location; site characteristics including physical characteristics; amenities (positive influences); and negative influences.

For leased land, and agricultural and rural properties, where insufficient sales are available, a capitalized income stream is commonly used to estimate the market value. Income land appraisal relies on capitalized income analysis.

7.3.1 Land Valuation Modeling by Sales Comparison

Land values may be modeled separately from improved values, or vacant and improved property may be modeled in a single combined valuation model (Guerin 2000). The primary benefit of a combined model is that both vacant and improved sales are used, which significantly increases the sales sample size for analysis and modeling. When developing a combined model, a binary variable should be used to separate vacant and improved sales. In addition, separate time and size adjustments should be tested for vacant and improved sales.

7.3.2 Land Valuation Modeling by Income

Income data can be used to value rented or leased land. Income capitalization for land follows the same general principles as commercial and industrial properties.

8. AUTOMATED VALUATION MODEL TESTING AND QUALITY ASSURANCE

AVM testing and quality assurance is necessary to determine the applicability of the model and/or the need for further specification. The process of developing and deploying an automated valuation model must include safeguards to insure the accuracy of data used and the integrity of results produced. Those safeguards are similar in kind and effect to those employed in evaluating the performance of any mass appraisal project.

8.1 Data Quality Assurance

All data used in model specification and calibration must pass the following screening tests:

1. Data must be sufficient to produce reasonable predictive models with regard to the property characteristics utilized in model calibration and implementation. As a general rule, the number of sales should be at least five times (fifteen times is desirable) the number of independent variables (Gloudemans 1999, 127).
2. Sales data must reflect, to the maximum extent possible, the conditions requisite to market value transactions.
3. Subjective data must be consistent across the population of properties to be valued using the model. Examples would include quality, physical condition, and effective age.
4. Accurate property characteristic data is essential to model quality. If the data were to be verified through a field audit, it should be found to be correct 95 percent of the time.

Data quality assurance should measure the quality and quantity of data, as well as provide a means of evaluating the application of the developed AVM formula to a specific population of properties. The product of that evaluation may include the acceptable ranges of specific property characteristics and ranges of estimated market values to which the model can be applied.

In addition to the quality assurance statistics discussed below, it is good practice to provide the user with a measure or index of the relative confidence that can be placed in individual value estimates, especially at the extremes of the data ranges. Using stratified ratio studies to examine the extreme low and high ends of various property characteristics in the modeling and holdout data sets, the market analyst will be able to determine the applicability of the model at these extremes.

The market analyst should decline to provide an estimate at those points where the value estimates become unreliable due to the data falling outside of acceptable parameters (see Section 8.4 on Sales Ratio Analysis).

8.2 Data Representativeness

Because AVMs use a relatively small sampling of properties from which inferences about the total population of properties are drawn, care must be taken to ensure that the sample adequately represents the total population of properties to be valued. In many kinds of statistical studies, samples are selected randomly from the population to ensure representativeness. Because sales do not represent true random samples, extra care must be taken to ensure representativeness. A sample is considered representative when the distribution of values of properties in the sample reflects the distribution of values in the population. Because the distribution of values in the population cannot be directly ascertained and appraisal accuracy may vary from property to property (depending on property type and characteristics), representativeness can be achieved by selecting a sample that adequately reflects salient value-related property characteristics. A property should be included in a sample based on characteristics of the property and not actions or characteristics of the owner.

This same degree of care should be taken in selecting sales samples used to test the quality of the AVM once it is developed

(IAAO 1999, 12.)

8.3 Model Diagnostics

The specific diagnostic tools available to market analysts and users of automated valuation models will vary with the model methodology employed. Multiple regression analysis provides the market analyst and user with a wide range of diagnostic statistics that may not be available with other calibration methodologies. In any event, the market analyst must make effective use of the diagnostic tools available during model calibration and be prepared to explain their use and significance to end users.

Standards do not exist for goodness-of-fit statistics (such as the coefficient of determination) or measures of individual variable significance (such as the T-statistic). Nonetheless, the market analyst should be able to explain how those statistics were used and how they relate to the predictive quality of a specific model in relation to the sales data available for calibration.

8.4 Sales Ratio Analysis

Sales ratio analysis is a type of statistical study based on comparisons between an estimated value and market value as indicated by sales prices. For AVM use, the numerator would be the estimated value generated from the model, while the denominator would be the sale price. The ratios thus calculated are subjected to statistical analysis to determine central tendency (level), and vertical (value related) and horizontal uniformity or variation. Central tendency statistics provide information about the overall or typical level in relation to market value that would be achieved given the results of the model. Variability statistics provide information about the degree to which model-determined values for individual properties are similar with respect to market value.

Sales based ratio studies are among the most objective methods for testing the performance and quality of any mass appraisal system. Much of the information in this section has been reprinted from the *Standard on Ratio Studies* (IAAO 1999).

8.4.1 Measures of Appraisal Level

Statistically, measures of central tendency provide an indication of the overall level of appraisal for any group of properties represented by a particular sales sample. Point estimates of these measures are calculated as shown in table 1. Reliability statistics should also be calculated around each of these measures (see Section 8.4.3 on Measures of Reliability). Common measures of appraisal level include the mean, sales weighted mean, and median ratios.

8.4.2 Measures of Variability

Several statistical tests are available and should be used to determine the degree of variability (uniformity) in the products of any AVM model. Common measures of appraisal variability include the coefficient of dispersion (COD) and coefficient of variation (COV).

8.4.2.1 Coefficient of Dispersion

The most useful measure of variability is the COD, which measures the average percentage deviation of the ratios from the median ratio and is calculated by (1) subtracting the median from each ratio, (2) taking the absolute value of the calculated differences, (3) summing the absolute differences, (4) dividing by the number of ratios to obtain the "average absolute deviation," (5) dividing by the median, and (6) multiplying by 100. For the data in table 1:

Average Absolute Deviation =

9.271 ÷ 36 = 0.2575;

COD = (0.2575 ÷ 0.864) * 100 = 29.8.

The COD has the desirable feature that its interpretation does *not* depend on the assumption that the ratios are normally distributed. Standards for interpreting CODs are contained in Section 14.2 of the *Standard on Ratio Studies* (IAAO 1999). Note that the COD represents the mean (not the median) percent deviation from the median. In general, more than half the ratios will fall within one COD of the median.

The COD should not be calculated about the mean because the mean is more affected by extreme ratios than the median, and because of the inherent (upward) bias of the mean of a set of ratios. The COD also should never be calculated about the weighted mean, which implicitly weights each ratio based on its sale price.

(IAAO 1999, 24.)

8.4.2.2 Coefficient of Variation

The COV can be another important measure of appraisal variability. The COV for a sample is calculated by (1) subtracting the mean from each ratio, (2) squaring the calculated differences, (3) summing the squared differences, (4) dividing by the number of ratios less one to obtain the "variance," (5) taking the square root to obtain the "standard deviation," (6) dividing by the mean, and (7) multiplying by 100. Note that the COV is calculated only about the mean—not the median or weighted mean (although other methods permit calculation about the weighted mean). For the data in table 2:

Variance = 3.0808 ÷ 35 = 0.0880;

Standard Deviation = sqrt 0.0880 = 0.2966;

COV = (0.2966 ÷ 0.900) * 100 = 33.0.

The interpretation of the standard deviation and COV rests on the assumption that the ratios are normally distributed. When this is the case, approximately 68 percent of the predicted ratios in the population will lie within one standard deviation of the mean, and approximately 95 percent will lie within two standard deviations of the mean. When the ratios do not approximate a normal distribution, these relationships no longer hold (although there always will be at least 75 percent of the ratios in any population within two and at least 89 percent of the ratios within three standard deviations of the mean). Hence, one should determine whether ratios are approximately normally distributed before using the COV. When the normality assumption is met, the COV provides the most precise measure of variability.

Because the deviations between each ratio and the mean ratio are squared in determining the COV, ratios that differ greatly from the mean influence the COV more than they do the COD, in which the deviation of each observation from the median is equally weighted.

(IAAO 1999, 25.)

8.4.3 Measures of Reliability

Reliability, in a statistical sense, concerns the degree of confidence one can place in a calculated statistic for a

Table 1. Example of Ratio Study Statistical Analysis

Data analyzed

Rank of ratio of observation	Appraised value (AV in $)	Market value (MV in $)	Ratio (AV/MV)
1	48,000	138,000	0.348
2	28,800	59,250	0.486
3	78,400	157,500	0.498
4	39,840	74,400	0.535
5	68,160	114,900	0.593
6	94,400	159,000	0.594
7	67,200	111,900	0.601
8	56,960	93,000	0.612
9	87,200	138,720	0.629
10	38,240	59,700	0.641
11	96,320	146,400	0.658
12	67,680	99,000	0.684
13	32,960	47,400	0.695
14	50,560	70,500	0.717
15	61,360	78,000	0.787
16	47,360	60,000	0.789
17	58,080	69,000	0.842
18	47,040	55,500	0.848
19	136,000	154,500	0.880
20	103,200	109,500	0.942
21	59,040	60,000	0.984
22	168,000	168,000	1.000
23	128,000	124,500	1.028
24	132,000	127,500	1.035
25	160,000	150,000	1.067
26	160,000	141,000	1.135
27	200,000	171,900	1.163
28	184,000	157,500	1.168
29	160,000	129,600	1.235
30	157,200	126,000	1.248
31	99,200	77,700	1.277
32	200,000	153,000	1.307
33	64,000	48,750	1.313
34	192,000	144,000	1.333
35	190,400	141,000	1.350
36	65,440	48,000	1.363

Note: Due to rounding, totals may not add to match those on following table, which reports results of statistical analysis of above data.

Results of statistical analysis

Statistic	Result calculated on preceding data
Number of observations in sample	36
Total appraised value	$3,627,040
Total market value	$3,964,620
Average appraised value	$100,751
Average market value	$110,128
Mean ratio	0.900
Median ratio	0.864
Geometric mean ratio	0.849
Weighted mean ratio	0.915
Price-related differential (PRD)	0.98
Coefficient of dispersion (COD)	29.8%
Standard deviation	0.297
Coefficient of variation (COV)	33.0%
Probability that population mean ratio is between 90% and 110%	49.7%
95% mean two-tailed confidence interval	0.799–1.000
95% median two-tailed confidence interval	0.684–1.067
95% weighted mean two-tailed confidence interval	0.806–1.024
Shape of distribution of ratios	Normal (based on binomial distribution)
Date of analysis	9/99/9999
Category or class being analyzed	Residential

sample (for example, how accurately does the sample median ratio approximate the true [population] median appraisal ratio?). There are two related measures of reliability: confidence intervals and standard errors. A confidence interval consists of two numbers that bracket a calculated measure of central tendency for the sample; one can have a specified degree of confidence that the true measure of central tendency for the population falls between the two numbers. Standard errors relate to the distance one must add to and subtract from certain measures of central tendency to compute the confidence interval.

For the data in table 1, the 95 percent confidence interval for the median is 0.684 to 1.067 (calculations not shown)—from the sample data, one can be ninety-five percent confident that the median level of appraisal for the population is in this range. Although most commonly calculated around the mean, confidence intervals can be calculated about various measures of appraisal level and variability, or about a resulting property value estimate; standard errors can be properly calculated about the mean and weighted mean, or about an estimate of value for the population. (See IAAO [1990, 515–546] and Gloudemans [1999, 257–339] for information on performing these calculations.) The article, "Confidence Intervals for the COD: Limitations and Solutions" (Gloudemans 2001), provides criteria for evaluating whether CODs can be deemed to have exceeded standards.

Measures of reliability explicitly take into account the errors inherent in a sampling process. In general, these measures will be tighter (better) when samples are relatively large and the uniformity of ratios is relatively good. Although the mathematics of calculating these measures is comparatively straightforward, their correct interpretation is critical and requires someone well grounded in the underlying statistical principles.

Users must give careful consideration to reliability measures in evaluating AVM output.

(IAAO 1999, 25.)

8.4.4 Vertical Inequities

The COD and COV relate to "horizontal," or random, dispersion among the ratios in a stratum, regardless of the value of individual parcels. Another form of inequity may be systematic differences in the appraisal of low-value and high-value properties, termed "vertical" inequities. When low-value properties are appraised at greater percentages of market value than high-value properties, appraisal regressivity is indicated. When low-value properties are appraised at smaller percentages of market value than high-value properties, appraisal progressivity is the result. Appraisals should be neither regressive nor progressive.

An index statistic for measuring vertical equity is the PRD (Price-Related Differential), which is calculated by dividing the mean by the weighted mean:

Mean/Weighted Mean =

Price-Related Differential

This statistic should be close to 1.00. Measures significantly above 1.00 tend to indicate appraisal regressivity; measures below 1.00 suggest appraisal progressivity. For the data in table 1, the PRD is 0.983, suggesting slight progressivity. When samples are small or the weighted mean is heavily influenced by several extreme sales prices, however, the PRD may not be a reliable measure of vertical inequities. If not representative, extreme sales prices may be excluded in calculation of the PRD. Similarly, when samples are very large, the PRD may be too insensitive to show small pockets of properties in the population where there is significant vertical inequity.

(IAAO 1999, 26.)

8.4.5 Guidelines for Evaluation of Quality

Because the development and utilization of automated valuation models are ongoing, without definitive beginning or end dates, sales ratio studies should be performed on a scheduled, periodic basis to establish the current performance status of the model. Such ratio studies should be conducted utilizing holdout samples accumulated according to Section 8.7. Model accuracy should be measured against the *Standard on Ratio Studies* (IAAO 1999) for the particular property type valued by the model. The *Standard on Ratio Studies* (IAAO 1999) suggests that the level of AVM estimate-to-sale price in each stratum (group of like properties) should be within 5 percent of the overall estimate-to-sale ratio for all strata; and the overall estimate-to-sale level should be within 10 percent of the desired level of 100 percent. For residential properties, variability, as measured by the coefficient of dispersion (average percent of error about the median estimate-to-sale price ratio), should be 15 percent or less in older, heterogeneous areas and 10 percent or less in areas of newer and fairly similar residences. Variability within strata composed of income-producing properties requires a coefficient of dispersion of 15 percent or less in larger, urban areas, and 20 percent or less in small or rural areas. Within all other types of property strata, the coefficient of dispersion should be 20 percent or less.

Table 2 is taken from the IAAO *Standard on Ratio Studies* (IAAO 1999) and provides guidelines for evaluating the quality of appraisal level and variability based on statistical measures previously discussed.

8.4.6 Importance of Sample Size

There is a general relationship, between statistical precision and the number of observations in a sample,

Table 2. Ratio Study Performance Standards

Type of property	*Measure of central tendency*	*COD*	*PRD**
Single-family residential			
Newer, more homogenous areas	*0.90–1.10*	*10.0 or less*	*0.98–1.03*
Older, heterogeneous areas	*0.90–1.10*	*15.0 or less*	*0.98–1.03*
Rural residential and seasonal	*0.90–1.10*	*20.0 or less*	*0.98–1.03*
Income-producing properties	*0.90–1.10*		
Larger, urban jurisdictions	*0.90–1.10*	*15.0 or less*	*0.98–1.03*
Smaller, rural jurisdictions	*0.90–1.10*	*20.0 or less*	*0.98–1.03*
Vacant land	*0.90–1.10*	*20.0 or less*	*0.98–1.03*
Other real and personal property	*0.90–1.10*	*Varies with local conditions*	*0.98–1.03*

****The standards for the PRD are not absolute when samples are small or when wide variations in prices exist. In such cases, appropriate tests are more useful (see table 5 of the Standard on Ratio Studies [IAAO 1999, 27]).***

drawn from a given population: the larger the sample, the greater the precision. The required sample size for any given degree of precision depends primarily on acceptable sampling error and the variability in the population. When there are insufficient sales to achieve target levels of precision, all valid sales should be used unless this results in nonrepresentativeness. If an abundance of sales is available, it is permissible to randomly include sufficient sales to obtain uniform or reasonably small margins of error.

Table 3 demonstrates the relationship between sample size requirements and variability as measured by the COV with the values in the table indicating margins of error that must be added to and subtracted from the sample mean to determine the confidence intervals. For example, a sample consisting of ten sales with a COV of 20 percent would produce a 95 percent confidence interval with a width of ±14.3 percent around the mean. Given the same COV with a sample size of 100 sales, the 95 percent confidence interval width would be reduced to ±3.9 percent around the mean, thus providing greater precision.

8.5 Property Identification

AVM developers must accurately identify property in order to produce an accurate valuation estimate for that property. The common property identification for the commercial AVM industry has become the property address. However, third party data providers use different variations of addresses. Many assessment jurisdictions have not fully standardized their addresses. Some condominium complexes have the same street address for all units. Condominium unit numbers assigned by the assessment jurisdiction may be the postal number or the lot number of the subdivision. These are just some of the variations in addresses that causes errors or misidentification of the properties requested by AVM users.

AVM developers attempt to minimize property identification errors by using address standardization software for all data to be used in the AVM system. All electronic real estate property systems should move to standardized addressing systems such as the Coding Accuracy Support System (CASS) certified by the United States Postal Service. While CASS is a way to standardize addresses across the U.S., it is primarily intended to ensure the accuracy of addresses for mail delivery purposes. This is a slightly different goal than the identification of the physical location of the property, especially in rural areas.

One precondition of address standardization is parsing the address into separate fields for the number, directional, street name/number, prefixes, and suffixes. Once this is accomplished, the correction or standardization of the address can begin. For example, Florida may be represented by the word Florida or abbreviations such as “Fla.” or “FL.”

Geographic information systems can be used to match AVM system property addresses to addresses in the U.S. Census Tiger files (or enhanced Tiger files provided by third parties). These GIS files have identified/located addresses by latitude and longitude at the street address segment level for most of the United States. Other countries have similar methods to geocode addresses to locational reference systems.

AVM users also have a responsibility to provide accurate addresses when requesting an AVM report. They should review the returned AVM report to confirm that the value estimate is for the property in question. Valid AVM reports are important for measuring the quality of the AVM system. This is called the hit rate, which is a measure of the number of usable AVM valuation reports compared to the total number of valuation reports requested. The hit rate will vary by several factors such as address mismatch; missing data within the property record that prevents the estimation of value; type of property is outside the scope of the AVM model; and the size or valuation of the subject property is outside the range of acceptable quality as determined by the quality assurance review of the model.

(Collateral Risk Management Consortium (CRC) 2003, 6.)

8.6 Outliers

The term "outliers" is defined in the *Glossary for Property Appraisal and Assessment* (IAAO 1997) as observations that have unusual values; that is, they differ markedly from a measure of central tendency. Some outliers occur naturally; others are due to data errors. In valuation models, outliers may include parcels with unusual characteristics as well as those with extreme estimated values per unit. Large, difficult to explain differences with respect to previous or control model runs may also identify outliers. Failure to understand and address outlier influences may result in unstable models that produce unpredictable changes in value over time. Documentation accompanying the automated valuation model must describe the methodology used to identify outliers and the procedures/trimming criteria followed once outliers are identified.

In ratio studies, outlier ratios are very low or high ratios as compared with other ratios in the sample. When the sample is small, outlier ratios may distort calculated ratio study statistics. Some statistical measures, such as the median ratio, are resistant to the influence of outliers. However, the COD and mean are sensitive to extreme ratios.

Outliers in AVM models can result from any of the following:

1. an erroneous sale price
2. a nonmarket sale
3. unusual market variability
4. a mismatch between the property sold and the property appraised
5. an error in the appraisal of an individual parcel
6. an error in the appraisals of a subgroup

(IAAO 1999, 19–20.)

One extreme outlier can have controlling influence over some statistical measures. Particular care must be taken to identify outliers if point estimates are used to make inferences about population level or variability. If, after proper verification, screening, and editing, an outlier with a nonrepresentative ratio remains in a study, statistical results will not reflect population level and variability. The potential distortion is greater when sample size is small. If outliers can be identified, trimming procedures are acceptable methods for creating a more representative sample. One outlier identification method is based on the interquartile range; however, because of the skewed distribution of ratios, this procedure may locate only extremely high ratios. If one or two high outlier ratios are trimmed from a small sample, the statistical measures of level may be shifted significantly lower. (See Tomberlin [1997] and Hoaglin, Mosteller, and Tukey [1983] on trimming small samples.)

(IAAO 1999, 20.)

Table 3. Confidence Intervals and Sample Size: 95 Percent Confidence Interval

Sample size	*COV = 10.0*	*COV = 20.0*	*COV = 30.0*
5	*±12.4*	*±24.8*	*±37.2*
10	*±7.2*	*±14.3*	*±21.5*
50	*±2.8*	*±5.5*	*±8.3*
100	*±2.0*	*±3.9*	*±5.9*
300	*±1.1*	*±2.3*	*±3.4*

8.7 Holdout Samples

Holdout samples represent groups of valid sales selected in a manner that guarantees their group characteristics match those of the population of properties covered by the automated valuation model. Such samples should be accumulated at the same time sales are collected for model calibration, but used for testing the calibrated model. Inherent in the definition of holdout samples is the premise that the sales not be used in developing the original model. Sales that occur after model calibration can also be used in testing and validating the model, and this method may be preferable when few sales are available.

8.8 Value Reconciliation

When a model is designed to produce more than one value estimate for a subject property, model documentation must contain a thorough explanation of the procedures followed to reconcile those candidate estimates into a final estimate of value. Those procedures must include analysis of the relative strengths and weaknesses of the candidate estimates, and specification of how that analysis results in a final value estimate.

In those instances in which all candidate estimates are presented to the user for their reconciliation, the system must report the quantity and quality of data supporting each of the candidate estimates. If the product of an automated valuation model is a set of value estimates derived from more than one of each of the three approaches, that product must also include sufficient information to allow the user to weigh the validity of those estimates, based on the quality and quantity of data available to support them.

When the model is designed to produce estimates of value for individual properties, those estimates must be accumulated and compared to their actual selling prices using ratio studies conducted at regular intervals. In addition, confidence intervals can be calculated around value estimates developed for individual parcels. Narrow intervals indicate greater likelihood that the estimate reflects market value. Additionally, z scores can be calculated and show the number of standard deviations by which an AVM estimated value misses actual sales price. Properties with value misses outside of a ±3 standard deviation range should be reviewed for systematic model error.

8.9 Appraiser Assisted AVMs

When an appraiser reviews or changes an AVM report prepared by a separate AVM provider, the results are called appraiser-assisted AVMs (AAVMs). The appraiser can provide an additional opinion of the estimated value and usually will sign the report and confirm the value. All AVM reports can have their estimates of value overridden by an appraiser's opinion of value. In most cases, appraisers are limited in their ability to change an AVM report. AVM reports based on the traditional formats of the cost, sales comparison and income approaches are the easiest for appraisers to change.

8.10 Frequency of Updates

AVM estimates of value are based on formulas derived from market analysis of a specific geoeconomic area during a specified time frame. Because AVM value estimates represent trends in time as applied to a specific property with known characteristics (physical and/or economic), AVM providers must update their formulas, estimates of value, characteristics, and economic databases regularly. Movements in the market and the availability of market information should dictate the frequency of this process.

9. AVM REPORTS

There are three general types of reports that are considered part of the AVM reporting process. They are the detailed documentation report, the restricted use report, and the appraiser-assisted report. In all cases, the reports should be in compliance with the respective portions of *USPAP*.

9.1 Types of Reports

There are several report formats associated with the development of an AVM and the reporting of an individual property's estimate of value. Documentation reports, restricted use reports, and CAMA/AAVM reports each provide different reporting levels of appraisal analysis within the report.

9.1.1 Documentation Report

There are several report formats associated with the development of an AVM and the reporting of an individual property's estimate of value. The development of an AVM formula involves the analysis of the historical market place (real estate) information in order to create value estimates at a particular point in time. This market analysis should comply with *USPAP* Standard 6: Mass Appraisal, Development and Reporting (Appraisal Foundation 2003, 46–56). There should be a detailed report to document and support the market analysis process and the final valuation formula. This includes the sections of *USPAP* Standard Rule 6–7 (report format) and 6–8 (certification) (Appraisal Foundation 2003, 53–55).

9.1.2 Restricted Use Report

When requesting an AVM, the client is normally not interested in complete narrative reports as described in Standard 2: Real Property Appraisal Reporting, or Standard 6: Mass Appraisal, Development and Reporting (Appraisal Foundation 2003, 21–31, 46–56). AVM clients want quick standardized indicators of value, that may be retrieved from the AVM systems by support personnel without professional real estate training or knowledge. This includes the general public, which may be interested in an indication of value for properties that they already know, such as property owners who request an AVM to check the current market value before making various economic decisions. This requires a restricted use report that is limited to the immediate intended user (client) of the AVM report. These restricted use reports are typically limited to generally acceptable property identification such as street address, indication of value, some basic property descriptive characteristics, known additional indicators of value (such as last sale price/date and property tax assessment), and report date. There may be additional qualification and limiting conditions information as described in *USPAP* Standard 6–7 (mass appraisal report) (Appraisal Foundation 2003, 53–55) that is not of general interest to the intended user. These restricted reports are generally one to a few pages in length. These are the reports referred to in *USPAP* AO–18, Use of an Automated Valuation Model (AVM) (Appraisal Foundation 2003, 180-187), which states that the output of an AVM is not, by itself, an appraisal. These restricted use reports are simply the application of an AVM model formula to an individual property and do not contain the supporting documentation of the appraisal process performed to create the formula, which should be in the documentation report.

9.1.3 CAMA or AAVM Report

A third type of report is the combination of AVM formulas with appraisers' review and verification of valuations. These are sometimes called CAMA in government tax assessment and appraiser-assisted AVM reports, in the commercial AVM field. This type of report combines the most desirable parts of the AVM (unbiased market analysis and consistently applied model formulas) with the most desirable parts of the field appraiser (property inspection, local knowledge and experience). The AVM provider sends the AVM report to the appraiser in electronic format. The appraiser performs a desktop review or one of various levels of inspection, as desired by the client, and corrects/confirms the AVM report and value estimate before delivery of the appraiser's final opinion of value to the client.

While all AVM reports can have their estimates of value overridden by an appraiser's opinion of value, in most cases, appraisers are limited in their ability to change the comparable selections, calculations, and variable adjustments within an AVM report. AVM reports based on the traditional formats of the cost, sales comparison, and income approaches, are the easiest for appraisers to change or adjust at the individual variable level.

9.2 Uses of AVM

AVM reports may have many uses. This standard will only list some of the typical uses.

9.2.1 Real Estate Lenders

- Reduce time to approve real estate loan applications
- Provide unbiased estimate of value for loan underwriting
- Provide real estate value/scores to compliment borrower's credit scoring
- Standard estimates for annual review of individual appraiser's performance
- Quality assurance for selling pooled loans
- Review of loan portfolios
- Support for lending decisions and geographic distribution required by the Community Reinvestment Act
- Statistical support for litigation
- Updates current valuation of portfolio properties
- Support in purchase of loan portfolios or lending institutions
- Portfolio valuation reviews by secondary mortgage markets and bond rating firms
- Systematic review of mortgage loan transaction to assist in the discovery of potential fraud

9.2.2 Real Estate Professional

- Support in setting listing price
- Support in negotiation between sellers and buyers
- Central database for appraisers
- Support for appraiser's opinions of value
- Support for appraiser's review and desktop appraisal assignments
- Support for appraisal consulting assignments that involve large numbers of properties
- Statistical support for litigation

9.2.3 Government

- Planning and land use decisions
- Development of value estimates for review by assessment staff appraisers
- Standardized estimates of value to annually review field appraisers' performances
- Valuation substitutes for appraisals in ratio study reports
- Screening of sale prices for valid market sales transactions
- Audits of lenders by state and federal regulators
- Assist states with standardized values to review property assessments in school funding formulas
- Fraud identification and prevention by enforcement, taxation, customs, and oversight agencies (such as GSE, HUD, IRS, Canada Mortgage and Housing Corporation, Statistics Canada, and state and national bank regulators)
- Fraud prosecution by comparing transactions to standardized values
- Assist in valuation for right-of-way and property condemnation cases

9.2.4 General Public

- Support for various business development and economic decisions
- Assistance in determining best listing price
- Assistance in determining best offering price
- Review of local government tax assessments
- Estate estimates of real estate value by attorneys and estate administrators

AVM reports may be sufficient as stand-alone products, or they may lead to a request for a more detailed appraisal report based on the needs and usage of the intended user. This listing is only a portion of the potential uses of AVMs. When clients request AVMs for a limited and specific use, the AVM report will provide quality information to the intended user quickly and inexpensively.

10. GLOSSARY

Algorithm—Computer-oriented, precisely defined set of steps that, if followed exactly, will produce a prespecified result (for example, the solution to a problem).

Additive Model—A model in which the dependent variable is estimated by multiplying each independent variable by its coefficient and adding each product to the constant.

Appraisal Emulation Model—The appraisal emulation model (see Section 3.2.2.1 Comparable Sales Method) follows the steps that an appraiser might follow in forming a value estimate (although not with the same insight or flexibility that a qualified appraiser brings to the assignment). The model selects "comparable sales" using some standard criteria. It then rates those comparable sales by suitability, based on the physical and sales characteristics of each comparable sale, by adjusting the varying elements (much as is done on an appraisal form); the model then calculates an estimate of value.

Automated Valuation Model—An automated valuation model (AVM) is a mathematically based computer software program that produces an estimate of market value based on market analysis of location, market conditions, and real estate characteristics from information that was previously and separately collected. The distinguishing feature of an AVM is that it is a market appraisal produced through mathematical modeling. Credibility of an AVM is dependent on the data used and the skills of the modeler producing the AVM.

Binary (Dummy) Variable—(1) Binary variables are qualitative data items that have only two possibilities—yes or no (for example, corner location). (2) A variable for which only two values are possible, such as results from a yes-or-no question; for example, does this building have any fireplaces? Used in some models to separate the influence of categorical variables. Also called a dichotomous variable or dummy variable.

Blended Model—A blended model (see Section 8.8: Value Reconciliation) is one where more than one modeling technique is used in deriving the estimate of value. Typically, the technique involves running a hedonic model and a repeat sales index. The results are then compared and evaluated. Based on each result, the blended model reports a final estimate of value. In addition to the hedonic model and repeat sales index, many blended models also include the results of a tax-assessed value model.

Calibration—The process of estimating the coefficients in a mass appraisal model.

Coefficient—(1) In a mathematical expression, a number or letter preceding and multiplying another quantity. For example, in the expression "5X", 5 is the coefficient of X, and in the expression "aY", a is the coefficient of Y. (2) A dimensional statistic, useful as a measure of change or relationship.

Cluster Analysis—A statistical technique for grouping cases (for example, properties) based on specified variables such as size, age, and construction quality. The objective of cluster analysis is to generate groupings that are internally homogeneous and highly different from one another. Various cluster algorithms can be employed.

Cost Approach—(1) One of the three approaches to value, the cost approach is based on the principle of substitution–that a rational, informed purchaser would pay no more for a property than the cost of building an acceptable substitute with like utility. The cost approach seeks to determine the replacement cost new of an improvement minus depreciation plus land value. (2) The method of estimating the value of property by: (a) estimating the cost of construction based on replacement or reproduction cost new, or trended historical cost (often adjusted by a local multiplier); (b) subtracting depreciation; and (c) adding the estimated land value. The land value is most frequently determined by the sales comparison approach.

Data Management—The human (and sometimes computer) procedures employed to ensure that no information is lost through negligent handling of records from a file, all information is properly supplemented and up-to-date, and all information is easily accessible.

Direct Market Method/Analysis—One of two formats of the sales comparison approach to value (the other being the Comparable Sales Method). In the direct market method, the market analyst specifies and calibrates a single model used to estimate market value directly using multiple regression analysis or another statistical algorithm.

Economic Area—A geographic area, typically encompassing a group of neighborhoods, defined on the basis that the properties within its boundaries are more or less equally subject to a set of one or more economic forces that largely determine the value of the properties in question.

Euclidean Distance Metric—A measure of distance between two points "as the crow flies." In property valuation, it is used to find the nearest neighbor or similar property based on an index of dissimilarity between property location or attributes. When using multivariate selection, the squared difference is divided by the standard deviation of the variable so as to normalize the differences. (Also see *Minkowski Metric*.)

Hedonic Model—Hedonic pricing attempts to take observations of the overall goods or services and obtain implicit prices for the goods and services. Prices are measured in terms of quantity and quality. When valuing real property, the spatial attributes and property-specific attributes are valued in a single model. Calibration of the attribute components is performed statistically by regressing the overall price onto the characteristics.

Heteroscedasticity—Nonconstant variance; specifically, in regression analysis, a tendency for the absolute errors to increase (fan out) as the dependent variable increases.

Holdout Sample—Part of a set of data set aside for testing the results of analysis.

Homogeneous—Possessing the quality of being alike in nature and therefore comparable with respect to the parts or elements; said of data if two or more sets of data seem drawn from the same population; also said of data if the data are of the same type (that is, if counts, ranks, and measures are not all mixed together).

Hybrid Model—Model that incorporates both additive and multiplicative components. (See also *Additive Model*, *Hedonic Model*, and *Multiplicative Model*.)

Income Approach—One of the three approaches to value, based on the concept that current value is the present worth of future benefits to be derived through income production by an asset over the remainder of its economic life. The income approach uses capitalization to convert the anticipated benefits of the ownership of property into an estimate of present value.

Geographic Information System (GIS)—(1) A database management system used to store, retrieve, manipulate, analyze, and display spatial information. (2) One type of computerized mapping system capable of integrating spatial data (land information) and attribute data among different layers on a base map.

Goodness-of-Fit—A statistical estimate of the amount, and hence the importance, of errors or residuals for all the predicted and actual values of a variable. In regression analysis, for example, goodness-of-fit indicates how much of the variation between independent variables (property characteristics) and the dependent variable (sales prices) is explained by the independent variables chosen for the AVM.

Location Value Response Surface Analysis—A mass appraisal technique that involves creating value influence centers, computing variables to represent distances (or transformations thereof) from such points and using the variables in a multiple regression or other model to capture location influences. Implementation of the technique is enhanced by the use of a geographic information system. Some geographic information systems permit the value influence centers to be displayed and measured as a three-dimensional grid surface, the results of which can be likewise used in calibration techniques to arrive at the contribution of location based on the model specification.

Location Variable—A variable that seeks to measure the contribution of locational factors to the total property value, such as the distance to the nearest commercial district or the traffic count on an adjoining street.

Market—(1) The topical area of common interest in which buyers and sellers interact. (2) The collective body of buyers and sellers for a particular product.

Market Analysis—A study of real estate market conditions for a specific type of property.

Market Analyst—An appraiser who studies real estate market conditions and develops mathematical formulas that represent those market conditions.

Market Area—(See *Economic Area*.)

Market Value—Market value is the major focus of most real property appraisal assignments. Both economic and legal definitions of market value have been developed and refined. A current economic definition agreed upon by agencies that regulate federal financial institutions in the United States is:

The most probable price (in terms of money) which a property should bring in a competitive and open market under all conditions requisite to a fair sale, the buyer and seller each acting prudently and knowledgeably, and assuming the price is not affected by undue stimulus. Implicit in this definition is the consummation of a sale as of a specified date and the passing title from seller to buyer under conditions whereby:

- The buyer and seller are typically motivated;
- Both parties are well informed or well advised, and acting in what they consider to be their best interests;
- A reasonable time is allowed for exposure in the open market;
- Payment is made in terms of cash in United States Dollars or in terms of financial arrangements comparable thereto.

The price represents the normal consideration for the property sold unaffected by special or creative financing or sales concessions granted by anyone associated with the sale.

Mean—A measure of central tendency. The result of adding all the values of a variable and dividing by the number of values. For example, the mean of three, five, and ten, is their sum (eighteen) divided by three, which is six.

Median—A measure of central tendency. The value of the middle item of an uneven number of items arranged or arrayed according to size; the arithmetic average of the two central items in an even number of items similarly arranged.

Minkowski Metric—Any of a family of possible ways of measuring distance. Euclidean distance, a member of this family, computes straight-line distances (as the crow flies) by squaring differences in like coordinates, summing them, and taking the square root of the sum. In mass appraisal model building, Minkowski metric usually refers to the sum of absolute differences (not squared) in each dimension, and resembles a “taxicab” or city block pattern. Other alternatives are possible, including the distance as calculated only for the dimension of greatest difference, but the city block distance is most common.

Model—(1) A representation of how something works. (2) For purposes of appraisal, a representation (in words or an equation) that explains the relationship between value or estimated sale price and variables representing factors of supply and demand.

Model Specification—The formal development of a model in a statement or equation, based on data analysis and appraisal theory.

Model Calibration—The development of the adjustments or coefficients from market analysis of the variables to be used in an automated valuation model.

Multicollinearity—Correlation among two or more variables. In regression analysis, high multicollinearity among the independent variables complicates modeling and will compromise the reliability of the resulting coefficients.

If the multicollinearity is perfect, the multiple regression algorithms simply will not work and either an error message may result or the software may purge one or more of the problem variables.

Multiplicative Model—A mathematical model in which the coefficients of independent variables serve as powers (exponents) to which the independent variables are raised, or in which independent variables themselves serve as exponents; the results are then multiplied to estimate the value of the dependent variable.

Multiple Regression Analysis (MRA)—A particular statistical technique, similar to correlation, used to analyze data in order to predict the value of one variable (the dependent variable), such as market value, from the known values of other variables (called "independent variables"), such as lot size, number of rooms, and so on. If only one independent variable is used, the procedure is called simple regression analysis and differs from correlation analysis only in that correlation measures the strength of the relationship, whereas regression predicts the value of one variable from the value of the other. When two or more variables are used, the procedure is called multiple regression analysis.

Neighborhood—(1) The environment of a subject property that has a direct and immediate effect on value. (2) A geographic area (in which there are typically fewer than several thousand properties) defined for some useful purpose, such as to ensure for later multiple regression modeling that the properties are homogeneous and share important locational characteristics.

Neighborhood Analysis—A study of the relevant forces that influence property values within the boundaries of a homogeneous area.

Neural Network—An artificial neural network (ANN) is a collection of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. An artificial neural network has several key elements: input, processing (calibration), and output. Other names associated with neural networks include: connect-ionism, parallel distributed processing, neuro-computing, natural intelligent systems, and machine learning algorithms.

Outlier—An observation that has unusual values, that is, it differs markedly from a measure of central tendency. Some outliers occur naturally; others are due to data errors.

Ratio Study—A study of the relationship between appraised or assessed values and market values. Indicators of market values may either be sales (sales ratio study) or independent "expert" appraisals (appraisal ratio study). Of common interest in ratio studies are the level and uniformity of the appraisals and assessments.

Repeat Sales Analysis Model—Repeat sales analysis (see Section 4.4: Time Series Analysis) aggregates changes in value and statistical means for properties sold more than once during a specified period of time in a given geographic area. For example, in a zip or postal code area, estimate market-level housing price changes. If an individual property has not been substantially changed since its last sale, this analysis matches each pair of sales transactions (thus the name "repeat sales"). The amount of appreciation (or depreciation) is calculated from the time of the first sale to the second and so on, providing an estimate of the overall appreciation of that local housing market during that time period.

The larger the number of available sales pairs, the more statistically reliable the estimate of overall housing price trends will be. Because this analysis is based on identifying properties where more than one sale has occurred, the challenge is to identify enough observations to provide a meaningful index of housing values, while keeping to as small a geographic area as possible.

A repeat sales index may also overestimate market appreciation if the data contains pairs of sales in which the second sales price reflects substantial improvements (or other alterations) made to the property after the first sale. On the other hand, repeat sales indices can and do provide very useful valuation estimates in jurisdictions where the data is insufficient to support hedonic models. In addition, they may prove to be more accurate in tracking housing values for the houses that a hedonic model may struggle with (especially those subject to extreme positive or negative influences) when a prior sale is known on the property.

Sales Comparison—One of the three approaches to value, the sales comparison approach estimates a property's value (or some other characteristic, such as its depreciation) by reference to comparable sales.

Stepwise Regression—A kind of multiple regression analysis in which the independent variables enter the model, and leave it, if appropriate, one by one according to their ability to improve the equation's power to predict the value of the dependent variable.

Software—Anything that is stored electronically on a computer is software. The storage device is hardware. There are two general categories of software: (a) operating systems and the utilities that allow the computer to function, and (b) applications which are programs that allow users to work with the computer (e.g., word processing, spreadsheets, databases, AVMs).

Stratification—The division of a sample of observations into two or more subsets according to some criterion or set of criteria. Such a division may be made to analyze disparate property types, locations, or characteristics, for example.

Tax Assessed Value Model—Tax assessed value models derive an estimate of value by examining market values attributed to properties by the local taxing authorities (see Section 4.5 Tax Assessed Value Model). As a matter of local law and custom, the values reported by the taxing authorities often (but not always) vary from the current market value in some reasonably predictable manner. For example, some jurisdictions require the taxing authority to report the value at 25 percent of estimated market value. In others, values are reassessed only on an infrequent basis. Some jurisdictions report multiple values—assessed, appraised and market values. By examining local laws and customs with respect to how that value is derived, it is often possible to provide a general adjustment to values reported by taxing authorities to better approximate current market value.

***Time Series Analysis*—**A family of techniques that can be used to measure the cyclical movements, random variations, seasonal variations, and secular trends observed over a period of time.

Weighted Mean—An average in which each value is adjusted by a factor reflecting its relative importance in the whole, before the values are summed and divided by their number.

Variable—An item of observation that can assume various values, such as square feet, sales prices, or sales ratios. Variables are commonly described using measures of central tendency and dispersion.

References

American Planning Association. 2003. "Land Based Classification Standards." Available from www.planning.org/LBCS/GeneralInfo.

Appraisal Foundation. 2003. *Uniform standards of professional appraisal practice* (*USPAP*). Washington, D.C.: Appraisal Foundation.

Appraisal Institute. 2002. *The dictionary of real estate appraisal.* 4th ed. Chicago: Appraisal Institute.

Carbone, R. 1976. The design of an automated mass appraisal system using feedback. PhD diss., Carnegie-Mellon University.

Collateral Risk Management Consortium (CRC). 2003. The CRC guide to automated valuation model (AVM) performance testing. Paper presented at Fidelity National Information Solutions (FNIS) Valuation Innovation and Leadership Summit, CRC, May 28, in Laguna Beach, CA.

D'Agostino, R.B., and Stephens, M.A. 1986. *Goodness-of-fit techniques.* New York: Marcel Dekker.

Gloudemans, R. J. 1999. *Mass appraisal of real property*. Chicago: IAAO.

Gloudemans, R.J. 2001. Confidence intervals for the coefficient of dispersion: Limitations and solutions. *Assessment Journal* 8 (6):23-27.

Guerin, B.G. 2000. MRA model development using vacant land and improved property in a single valuation model. *Assessment Journal* 7 (4):27-34.

Hoaglin, D.C., Mosteller, F., and Tukey, J.W. 1983. *Understanding robust and exploratory data analysis.* New York: John Wiley & Sons.

IAAO. 1990. *Property appraisal and assessment administration*. Chicago: IAAO.

IAAO. 1997. *Glossary for property appraisal and assessment.* Chicago: IAAO.

IAAO. 1999. *Standard on ratio studies.* Chicago: IAAO.

IAAO. 2002. Automated valuation modeling: A primer. Typescript.

Mendenhall, W., and Sincich, T. 1996. *A second course in statistics: Regression analysis.* 5th ed. Upper Saddle River, NJ: Prentice Hall.

Tomberlin, N. 1997. "Trimming outlier ratios in small samples." Paper presented at IAAO International Conference on Assessment Administration, September 14-17, at Toronto, ON, Canada.

Waller, B.D. 1999. The impact of AVMs on the appraisal industry. *The Appraisal Journal* 67 (3):287-292.

Ward, R.D., and Steiner, L.C. 1988. A comparison of feedback and multivariate nonlinear regression analysis in computer-assisted mass appraisal. *Property Tax Journal* 7 (1):43-67.

Wollery, A., and Shea, S. 1985. *Introduction to computer assisted valuation.* Boston, MA: Oelgeschlager, Gunn & Hain, Publishers, Inc.

Additional Suggested Readings

Gloudemans, R.J. 2002. Comparison of three residential regression models: Additive, multiplicative, and nonlinear. *Assessment Journal* 9 (4):25-36.

O'Connor, P.M. 2002. Comparison of three residential regression models: Additive, multiplicative, and nonlinear. *Assessment Journal* 9 (4):37-44.

Gloudemans, R.J, and O'Connor, P.M. 2002. "Comparison of three residential regression models: Additive, multiplicative, and nonlinear." Papers presented at 6th Annual Integrating GIS/CAMA Conference. Urban and Regional Information Systems Association and IAAO, April 7-10, at Reno, NV. Available on CD-ROM.

Maloney, J., Ripperger, R., and O'Connor, P.M. 2001. "The first application of modern location adjustments to cost approach and its impact." Linking Our Horizons, 67th International Conference on Assessment Administration. IAAO, September 9-12, at Miami Beach, FL.

Assessment Standards of the International Association of Assessing Officers