THE COMMON THREAD IN MARKET DATA SYSTEMS

Richard A. Borst
President
Cole-Layer-Trumble Company

GENERAL TRACK
Session G 1
Monday - 1:30 P.M.

PAPER PRESENTED AT THE WORLD CONGRESS ON COMPUTER-ASSISTED VALUATION, SPONSORED BY THE LINCOLN INSTITUTE OF LAND POLICY, AUGUST 1 - 6, 1982.

THE COMMON THREAD IN MARKET DATA SYSTEMS

R. A. BORST

BACKGROUND

Multiple regression analysis (MRA) is a tool used by many assessment jurisdictions, state agencies, and private firms in the valuation of real property. Briefly stated, it is used to establish the coefficients B_i in an equation of the form

$$ESP = B_0 + B_1 X_1 + B_2 X_2 + -- + B_n X_n$$

where ESP represents the estimated selling price of a property and the X_i are property characteristics and/or functional combinations of property characteristics. The MRA technique is applied to a set of sales with known property characteristics to obtain the B_i, which in turn are used to estimate the selling price of a designated set of parcels of property.

This valuation technique can be used as the sole method of estimating a property's value or it can be used in conjunction with other valuation techniques such as comparable sales analysis or the cost approach.

In describing this technique to potential users, a number of questions are frequently asked. They include:

1. What data elements are required to make MRA work?
2. What data elements would be needed in my jurisdiction?
3. How many terms are there in a model?
4. How many models are needed?
5. How many sales are needed?

These are pragmatic questions. It is the purpose of this paper to offer answers to these questions based on the Cole-Layer-Trumble Company's experience in developing MRA models that were used to value on the order of 725,000 residential parcels of property on projects in New Hampshire, Massachusetts, New York, New Jersey, North Carolina, South Carolina, and Texas. In all, some 86 models developed by several individuals in valuing eleven jurisdictions were reviewed in preparing the information provided herein.

THE STARTING POINT

Later in the paper, results will be given in the form of property characteristics utilized to develop regression based value estimates. Recognition must be given to a simple fact. The factors so described can only be a subset of those initially considered. The Company utilizes several data collection instruments, but for the

most part they contain similar information. A typical set of characteristics collected is shown in Exhibit 1. Exhibit 2 is a sample property record card showing the coding structure and layout of the characteristics of Exhibit 1. Most of the items are self-explanatory; however, it should be mentioned that CDU is an overall rating relating to the condition, desirability, and usefulness of a residence.

It is from this or similar starting point that the observations of this paper are made.

THE MODELING PROCESS

To develop MRA models, the Company either uses backward stepwise regression analysis or constrained regression analysis, a variation on the backward elimination technique. In constrained regression analysis, some subset of the regression coefficients are constrained to fall within certain limits for pragmatic valuation reasons. One such reason is that there may not be sufficient sales information to develop a regression model that takes into account the contribution a detached garage has on the estimated value of a residence in a rational manner. This and similar problems with other property characteristics (pools, patios, decks, carports) are frequently encountered in regression modeling. The Company's approach to this problem is to specify that certain data items must be in the model and that their coefficient must fall within certain supportable bounds[1].

Assuming that a computer-readable file of sales and property description data has been created, the modeling process involves several steps, including:

1. Variable transformations.

2. Specification of variables to consider for regression.

3. Specification of coefficient constraints.

Exhibit 3 will serve to illustrate some of these points. It shows the variables considered for regression and shows the constraints placed on certain coefficients. Exhibit 4 further describes the variables of Exhibit 3. The first variable listed is TOTLND which stands for total land value which has been estimated by an independent appraisal method. Note that the constraints on the coefficient for TOTLND force the coefficient to be brought in with a value of 1.0. This, in effect, means that the independent land value estimate will be used in developing the total value estimate for the parcel. The variable AGE*SF stands for age (in years) X square feet of living area, where age is a derived variable calculated by subtracting year built from the current year. The corresponding model is shown in Exhibit 5.

It can be seen that of the original 27 factors considered, 24 were brought into or forced into the model. Where reporting results in this paper, if a factor is in a model, constrained or otherwise, it is considered as a valid value predictor.

(1) A Companion World Congress paper entitled "Use of Constrained Regression Analysis by Non-Statisticians" by the author gives more detail on this technique.

RESULTS

As was previously stated, 86 models used in eleven different jurisdictions were reviewed in compiling the results to be reported. Exhibit 6 provides a summary of general information regarding the number of parcels valued, the number of models used, the number of terms in a model, and the number of sales used in developing the models. From this exhibit, several useful observations can be made:

1. In smaller jurisdictions, two to five models tend to be used. The two largest jurisdictions had 22 and 28 models.

2. The number of sales in a model ranges widely. As few as 118 were used. Many were in the 200 - 400 range, and only one model (author's review) utilized over 1,000 sales.

   The number of sales in a model is somewhat smaller than might be expected compared to models developed in the 1970's. This is due to the fact that in many jurisdictions, the number of sales available in 1980, 1981, and 1982 is a much lower percentage of the parcels to be valued.

3. The number of terms in a model ranged from five to 36, with 20 being a typical number.

   The model with only five terms in jurisdiction 2 was developed using RCNLD as one of the factors. Thus, the figure of five terms is artificially low. The model was developed for large, high quality homes where direct modeling of the usual property characteristics proved less satisfactory. The model with seven terms in jurisdiction 8 was developed for a group of parcels where the parcel was being acquired on speculation (the value was in the land), and the building description had less influence than typical on property value.

4. The quality of a model can be judged by several measures. The statistic used most frequently by Company model developers is the ratio of standard error of the estimate to the mean selling price expressed as a percent. The author reviewed this statistic for jurisdiction 9 in Exhibit 6. The ratio ranged from a low of 5.5% to a high of 22.5%. The median was 9.7%. Examining the extremes, the low ratio was formed from a model with a $2,095 standard error, and a $37,873 mean selling price. This model was developed on newer, modestly priced homes. As expected, the model should be accurate in this housing category because the market is well understood by both buyer and seller. The high figure was the result of a standard error of $3,659 and a mean selling price of $16,268. This model was developed for older (1929 average year built), low price housing. In absolute terms, the standard error is reasonable. It is the older, low cost less homogenous housing that traditionally has been the more difficult to value accurately by any means. A $3,000 variation from selling price is a high percent if the selling price is in the $10,000 - 20,000 range.

   The model shown in Exhibit 5 is typical of the models developed in jurisdiction 9. Although not shown, the mean selling price for this model is $45,928. Using the standard error $3,792 as shown, a figure of 8.3% results. The point is that with models having 16 - 26 terms, a 10% ratio of standard error to mean selling price is a reasonable expectation. Better performance can be expected for models developed for the more homogenous properties. Higher ratios can be expected on the difficult to value less homogenous properties.

To describe the common thread in market data, the frequency of use of a descriptor in the eleven jurisdictions of the sample group will be used. The results are compiled in Exhibit 7.

If a variable was used in any model in a jurisdiction, it was counted. Thus, if a variable was used in all jurisdictions, the count will be eleven. Also, when a variable such as the number of fireplaces is listed, it could have been in the model directly or a derivative term such as the square root of the number of fireplaces may have been used. No attempt to make a distinction was made because the major idea of the paper is to identify data that needs to be considered for collection prior to the modeling process. Later on, certain frequently encountered data transformations will be identified as helpful to prospective model developers.

Comments on Exhibit 7 include the following:

1. The most frequently used descriptors for valuation purposes are really quite logical - size, number of baths, age, etc.

2. The less frequently used descriptors are utilized in exception or unusual areas. They should not be ignored, however. For example, view and waterfront factors appeared only once, but when they were in the model, their coefficients had a large influence on value.

3. Neighborhood variables do not show up very frequently. The Company's procedure is to geographically delineate neighborhoods as a part of the overall valuation process. It is a manual procedure that is performed prior to modeling by experienced appraisers. Each parcel is, therefore, identified as belonging to a specific neighborhood.

   As part of the modeling process, neighborhoods of similar characteristics (e.g., age, size, selling price, style) are combined into "valuation areas". It is on sales from a valuation area that the regression models are developed. Thus, neighborhood would tend not to be a variable in the process.

4. With reference to Exhibit 5, it can be seen that the structure of the model is quite simple. Most variables are straightforward such as the number of full-baths and pool area. Certain factors are usually weighted by (multiplied by) square feet of living area.

   In the model shown - grade factor, age, air conditioning, no heat, date of sale, and CDU rating are brought into the model in this fashion. An intuitive notion of why this is done can be seen in the air conditioning example. By forming the product AC*SF, where AC is 0 or 1, the influence of air conditioned area is being brought into the model rather than its existence or nonexistence. Experience has shown that this is a better value predictor than the unweighted term.

In conclusion, there are great similarities in the model structures encountered. With the wealth of modeling experience, the Company has been able to make the process more routine and has begun the transfer of the modeling process to appraisal oriented personnel.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.