IMPROVING DATA COLLECTION TECHNIQUES IN COMPUTER ASSISTED MASS APPRAISALS

Joseph C. Small and Gail Poe Inderfurth, Mathematica, Inc.*

1. DESCRIPTION OF THE STUDY

The research reported in this paper involves two major innovations in real estate appraisal. First, the use of lay persons, similar in attributes to survey research interviewers, to record in a systematic manner using standardized forms objective characteristics about the properties to be appraised. Secondly, the use of these data items to estimate through multiple regression techniques the appropriate weights to assign to various household characteristics in order to predict the market value or selling price of the parcel. The objective of the research was to see if it was feasible to use lay persons who were much less expensive than real estate appraisors to collect these data and if, in turn, the data collected were reliable enough and sufficient to determine the market value of the properties. The ultimate goal, of course, is to achieve more efficient and less expensive methods of annual reassessments and hence more equitable real property tax rolls.

The project was sponsored by the New York State Board of Equalization and Assessment. Mathematica, Inc. of Princeton, New Jersey under contract with the State assisted in the development of the new techniques and with the analysis of the findings.

2. CURRENT ASSESSMENT PRACTICE

Most county and municipal governments in this country rely on the property tax as their major source of revenue. The property tax is levied on all owners of nonexempted** property. This tax is calculated in a very simple manner. First, all taxable property is assessed; most States require that this assessment be at full or market value. Second, the sum of all assessed values is calculated. Third, that amount of revenue that the jurisdiction wishes to raise from the property tax is divided by the sum of all assessed values and the resulting quotient is the tax rate.

$$\frac{\text{Revenues Required}}{\text{Sum of all assessed values}} = \text{Tax Rate}$$

Lastly, each individual's assessment is multiplied by the tax rate to determine his tax for the year.

The key elements in these equations are the individual assessments. To the extent that they approach market value or are a constant percent of market value, the assessments are equitable and the tax is imposed equitably. To the extent that assessments are not at a uniform rate of full value, taxes are inequitable and the system does not function as it was designed to.

Because of the rapid changes in real estate prices over the last few years the value of almost every property has gone up--but not in proportion to each other. Unless the assessments are changed each year, an equitable assessment roll becomes full of inequities after several years.

Though many States now require annual reassessment, the requirement is honored more in the breach than in the observance. The cost of reassessment under current practices using real estate appraisors is extremely high and most jurisdictions cannot afford the cost of an annual reassessment.

Three Methods of Assessment

There are three basic methods of property assessment: 1) the income method, 2) the replacement cost method, and 3) the market value method.

1. Income

   The income method of assessment is based on the amount of income that can be earned from the property divided by the interest rate. The income method is very sensitive to the interest rate. It is useful for such properties as office buildings, rented factories, and stores where their real value is the income they produce.

2. Replacement Cost

   The replacement cost method assumes that a property is worth what it would cost to build at the time of assessment minus depreciation. The assessor has a series of physical factors: square feet, materials, bathroom fixtures, etc., that he multiplies by standard cost factors. He adds the sum of these products and comes up with a figure known as replacement cost new. He reduces this figure (or depreciates it) according to what part of its usable life remains. The final figure is replacement cost new less depreciation.

*The research reported in this paper was sponsored by the New York State Board of Equalization and Assessment under the direction of Michael O'Shea, Deputy Executive Director, and Gerow Carlson, formerly Head of the Mass Appraisals Section. The authors of this paper are indebted to their colleagues at Mathematica, George Carcagno, Cheri Marshall, and Judith Glotzer, for their contributions. The contents of this paper reflect the interpretations of the Mathematica personnel and not necessarily those of the State Board

**Government, churches, schools, and other, generally charitable, non-profit organizations are usually exempted from the property tax.

There are several problems with this method: (a) It is time consuming to collect the data and do all of the numerical manipulations. (b) Its accuracy rests on two sets of figures, the cost factors and the depreciation rate. Cost figures are as quickly out of date as other prices and failure to update them results in bad predictions of replacement cost. Further, building materials change and some are no longer available so that putting a price on them is either guesswork or unfair. As for useful life left, it is very much a guess and no assessor ever feels comfortable with this item.

In practice what has happened is that an assessor calculates replacement cost new. Then from his knowledge of the real estate market, he estimates what the market value is. He then selects a depreciation figure that will bring his calculated replacement cost down to his predicted market value. Though in theory different, the replacement cost is in actuality an estimate of the market value.

3. Market

The market value of a property is based on what someone will pay for it. Experienced assessors know the market and make their predictions accordingly. They make calculations in their heads such as $1,000 for a bathroom; $20 per square foot; and $1,200 for a fireplace.

In addition to the problem of high costs in traditional market value assessments, there is the problem that no matter how detailed and standardized their manuals and training, each assessor will have some (often subconscious) individual techniques of valuation.

Since the largest number of parcels in the country are single family houses, initial efforts have been made at improving the methods used to assess them. There are many of them, they sell relatively frequently, and much effort has been put into improving the market value assessment approach.

3. A MULTIPLE REGRESSION ANALYSIS

The multiple regression techniques for real property valuation used in our investigation are based on the following assumptions:

- All property can be described by certain characteristics (e.g., size, style, location, building material, and condition)
- Each of these characteristics can be objectively described
- All property has value
- The value is attributable to the characteristics of the property
- The relationship between the characteristics and the value can be determined (i.e., the value contribution of each characteristic or sets and subsets of characteristics is determinable)

Thus if one can determine value of the characteristics, and one can describe the characteristics of a given property, then one can value each of the properties.

The assumptions are similar to those used in calculating replacement cost except that:

- The property values are not fixed, but are determined from the actual selling prices in a given jurisdiction, and
- The total value is not simply the sum of the value of each characteristic; a more complex interrelationship among several characteristics and value is possible.

Multiple regression valuation techniques are applied in the following manner. First, all of the houses in the jurisdiction are described. Second, two files are created, a file (sales) of those properties which have sold recently (and for which there is a reliable, arms-length selling price available), and a file of all other properties. Both files have complete descriptions of the properties which are contained in them. The first file has a price, the dependent variable, associated with each description.

The sales file is then used to estimate the regression coefficients of each of the independent variables. These coefficients can then be applied to the description of the non-sales property file so that a value is predicted for each non-selling house.

This technique of revaluation has two distinct advantages over current methods:

- The valuation of all properties can be updated annually without physical inspection of all properties.
- Valuation is more consistent and equitable because it is arrived at using a single computational program applied identically to all properties.

Determination of Necessary Data Elements

One of our first steps in our project was to determine which data elements are important for three specific purposes:

1. Value Calculation

   Which physical characteristics of a house are most important in predicting value? Similarly, which items traditionally collected are of little or no importance in predicting market price?

2. Property Identification

   What items are necessary and sufficient to enable a data collector to locate a property accurately and swiftly, (i.e. street address, frontage, type of house, name of owner)? And what items would enable an auditor of the data collector's work to quickly determine whether he had accurately described the required property or not?

3. Consumer Satisfaction

   What items are necessary to collect in order to demonstrate to the taxpayer that indeed it was his property that had been inspected? What data items will the taxpayer feel are important? (The panelled family room in the basement may be the result of many Saturday's efforts but contribute little or nothing to the value of the house. It might be wise to collect this piece of information and display it in the records despite its lack of importance in predicting value.)

Next we wanted to discover which pieces of data are easy and which are difficult to collect. The two basic criteria used were consistency and cost. As the study progressed it became apparent that there was not much purpose in determining the cost (time) of collecting each specific data item. Any savings experienced by eliminating a few items from the form would be relatively insignificant given the total amount of time required to locate and describe each property. Furthermore, it would have made the data collector's task virtually impossible had we required him to record the time after he noted each data item. Thus we concentrated our efforts on determining which items were collected consistently by having a subset of the properties visited twice.

Finally, we hoped to improve the actual forms used in data collection and the operating procedures -- from training through data processing -- in order to improve quality and reduce overall cost.

The Sample

A total of 656 properties were observed in the study which took place in Colonie, New York. Out of these, 296 properties were observed twice by two different data collectors using two different data collection forms. One form was basically a longer version of the other form; the longer form contained additional items to make possible a calculation of replacement cost as well.

In our analysis, in addition to evaluating the predictive power of the two instruments, we were able to evaluate the consistency with which the data items were collected by the independent duplicate observations.

4. THE RESULTS

The variables which are needed for predicting market value include a combined set of variables from the short form and the long form which resulted from the regression analysis. Because the regression program was run separately for each type of data collection instrument two separate sets of variables were produced which are not identical. Several items are exactly the same, and most of the others bear reasonable similarity to each other. Table A shows those variables which were arrived at from the regression analysis on the short form. In the second column, R, the correlation coefficient, is shown. In the third column, R^2, the coefficient of determination, is shown; and in the fourth column the increase in R^2 is printed. The regression coefficient is in column 6 and its standard error is in the last column. Therefore, 85 percent of the variation in the market sale price can be explained by the regression of the short form variables on the actual sale price. Table B contains similar information for the long form results. As can be seen, 86 percent of the variation in sale price can be explained by the regression of the long form variables on the sale price. In arriving at a combined set of variables, it is not feasible in this instance to use the regression analysis to directly estimate R^2 for the combined variables because the sample size of properties for which we have both a long and a short form in addition to a sale price with which to compare predicted value is too small. Therefore the combined set of "predictor" variables had to be arrived at through using not only the regression results but other information on the variables such as measures of their consistency of collection. After the next large data collection effort, new regression coefficients may be computed, using all properties with recent sale prices and then these coefficients may be applied to the rest of the properties.

To assess the consistency of the collection of the data items, indexes of inconsistency, the portion of total variability contributed by random response error, were computed on identical data items which were collected independently. For items with high indexes of inconsistency but which still are included in the prediction model we did not recommend deleting these items but rather revising or modifying them in various ways to improve their accuracy. We suggested deleting those items which were highly inconsistent and time consuming or difficult to collect.

5. SUMMARY

Overall we were quite pleased with the success of the methodology and the results of the value calculations from the data collected in the Colonie study. Given the variability one might expect for the sale price on the same house given different buyers, different sellers, etc. an explanation of 85 percent of the variability in sale price seems quite good.

The results of the study indicate that it is feasible to use lay persons trained in survey research interviewing techniques to collect the data items necessary to predict market value. It should be pointed out, however, that the data collectors were not able to collect two very important variables. These variables were neighborhood and traffic flow. The neighborhood in which a property is located is obviously essential; however, defining neighborhoods is a complex task requiring thorough knowledge of the locality. As for street frontage traffic, the data collectors' impressions of traffic as opposed to a more objective measure such as traffic per hour counts seemed unreliable.

The process of designing a computer assisted property appraisal system is a dynamic and fluid procedure. Efforts must continue to improve the data variables which are collected, the method of their collection, and the training of data collectors on these variables. The next collection effort will itself provide a learning opportunity from which further improvements and refinements may be made. Since the efforts reported in this paper have been centered on one community at one point in time, it may not be appropriate to generalize results for other communities or other points in time. However, these findings will certainly aid in conducting future data collection efforts.

TABLE A

SHORT FORM - FINAL VARIABLES FROM REGRESSION

Variable	R	R^2	Increase in R^2	Coefficient	SE of Coefficient
Total square feet	.7947	.6316	.6316	10.73	.96
Attached garage	.8379	.7020	.0704	1070.37	656.03
Level 1 fireplace	.8673	.7522	.0502	6953.53	1008.53
Basement level total rooms	.8870	.7868	.0347	1108.40	400.90
Lot front footage	.8990	.8081	.0213	57.98	10.09
Full baths	.9067	.8221	.0140	2895.94	939.61
Full basement-floor other than earth	.9112	.8303	.0082	2537.36	807.96
Year built	.9147	.8367	.0064	71.98	23.87
Basement fireplaces	.9181	.8429	.0063	5937.57	1979.61
Overall appearance	.9212	.8487	.0057	1723.15	731.09
Sale date	.9225	.8511	.0024	89.28	44.56
Tile floor (in kitchen)	.9235	.8528	.0017	1415.36	854.67
Landscaping	.9238	.8534	.0006	1015.55	1125.62
Presence of swimming pool	.9241	.8540	.0006	1637.77	1823.96

TABLE B

LONG FORM - FINAL VARIABLES FROM REGRESSION

Variable	R	R^2	Increase in R^2	Regression Coefficient	SE of Regression Coefficient
Lot front footage	.5470	.2992	.2992	53.90	10.17
Sale date	.5505	.3030	.0038	123.58	45.60
Attic square footage	.8754	.7664	.4633	10.26	1.03
Total fireplaces	.8939	.7990	.0327	3665.75	848.86
Total rooms	.9049	.8189	.0199	753.45	491.05
Total "other" rooms	.9118	.8314	.0125	1155.69	597.26
Attached garage	.9130	.8336	.0023	327.17	244.24
Full baths	.9140	.8354	.0018	2849.66	1290.52
Half baths	.9145	.8364	.0010	1679.91	1163.97
Year built	.9148	.8369	.0005	5.75	5.28
House quality	.9238	.8534	.0165	3895.25	772.35
Pool square feet	.9261	.8577	.0043	0.014	.005
Roof condition	.9280	.8612	.0035	1595.05	688.49