VOLUME 10 • NUMBER 2 JUNE 1991

PROPERTY TAX JOURNAL

A QUARTERLY JOURNAL
DEVOTED TO STUDIES IN PROPERTY TAXATION
AND ASSESSMENT ADMINISTRATION

Contents

The International Association of Assessing Officers is an association of public officials engaged in property appraisal and assessment administration. The association's purposes are to improve the standards of assessment practice; to educate the taxpaying public on the true nature and importance of the work performed by assessing officers; to engage in research and publish the results of studies; to elevate the standards of personnel requirements in assessment offices; to cooperate with other public and private agencies interested in improving tax administration; and in every proper way to promote justice and equity in the distribution of the tax burden. In sponsoring the *Property Tax Journal*, the International Association of Assessing Officers seeks to provide a forum for the examination of significant issues in property taxation and assessment administration.

INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS
1313 EAST SIXTIETH STREET, CHICAGO, ILLINOIS 60637-2892

A Calibration Technique for Estimating the Effect of Location on the Values of Residential Properties

Lawrence C. Marsh and Anthony B. Sindone*

Lawrence C. Marsh is director of graduate studies and Anthony B. Sindone is graduate research assistant, Department of Economics, University of Notre Dame, Notre Dame, Indiana.

This paper uses a broken response surface method to estimate location value through the use of three-dimensional polynomial splines. Stepwise regression is chosen as the means of selecting the most statistically significant adjustment points from any prespecified set of adjustment points because it provides enough flexibility for the location of the final selection of adjustment points to be determined simultaneously with the estimation of the regression coefficients.

This paper explains and demonstrates a method for determining the effect of location on the values of residential properties. Because identical houses situated in different parts of a city could be worth different amounts depending on where they are located, it is important to be able to isolate the influence of location on property values. This can be done by using a broken response surface method in multiple regression analysis (MRA).

Conceptually, to isolate the effect of location on the values of residential properties the relatively heterogeneous set of recently sold homes must first be statistically transformed into a relatively homogeneous set. This task is accomplished by controlling for the essential characteristics of a home by including all of the relevant variables affecting sale value, other than location, in a multiple regression equation.

The initial step in carrying out the broken response surface method in

*We are grateful to Charles C. Cook for his helpful comments in his review of this work. We take full responsibility for any remaining inadequacies.

MRA is to designate a point in the city as the origin, or 0,0 position. For convenience, this point could be taken to be the southwesternmost point in the city. On the city map, the east-west and north-south coordinates for each home sold are measured as the centimeters in the west-to-east direction and the centimeters in the south-to-north direction, respectively, from the point that was designated as the 0,0 position on the map. These numbers are then used to create sets of potential linear, quadratic, and cubic adjustments for the three-dimensional spline fit that characterizes the broken response surface method. The broken response surface is fit using a stepwise multiple regression algorithm. The most statistically significant adjustment is allowed into the regression first, the second most statistically significant adjustment is allowed in next, and so forth, until the overall F-statistic for the model reaches a maximum and begins to decline. This gives us a set of regression coefficient values as well as spline adjustments for the fitted broken response surface. If there are many observations, and the assessor does not wish to consider all of the observed sale locations as possible adjustment points, then a hybrid additive-multiplicative model may be fit first, and the observations with the largest absolute residuals may be considered as the proposed set of potential adjustment points. Alternatively, a prespecified grid of such points could be designated based on the assessor's knowledge of school district boundaries and other such considerations. Instead, for simplicity we will use a prespecified grid corresponding to streets and avenues, although this is by no means necessary or even recommended in practice.

The main point of this paper is that enough flexibility must be incorporated into the estimation procedure so that the location of the final selection of adjustment points can be determined simultaneously with the estimation of the regression coefficients. This can be done by using stepwise regression to select the most statistically significant adjustment points from whatever prespecified set of potential adjustment points the assessor cares to use.

Using the points with the largest residuals from the initial regression as the adjustment points is a good first step, but will only achieve a local minimum of the sum of squared residuals that is conditional on the first-round regression coefficient estimates. In this paper we propose an iterative search, in a stepwise regression procedure, to determine a set of adjustment points and coefficient estimates that only together can find the overall global minimum. The assessor must be sure to prespecify a set of potential adjustments that will adequately represent the population whenever the sale locations in a given sample are not well distributed over the city.

An added feature of this method is its ability to identify a property with a value substantially out of line with the values of neighboring properties. This robust feature allows an assessor to locate statistically significant outliers that would otherwise throw off the pattern of sales. These sales may be due to some unique characteristic of the property that the assessor has been unable to control for in the regression or may not be arm's-length sales.

The regression results from this model may be used to adjust for location, along with a host of other variables with the hybrid additive-multiplicative model structure, to get substantially improved estimates of the values of houses with different characteristics in different locations throughout the city. A standard sample data set from the International Association of Assessing Officers (IAAO) is used for this demonstration. The calibration methods proposed herein can be applied using any standard statistical package.

Introduction

Essentially, location effect may be thought of as a residual effect in the sense that it is extracted from the data after all of the other relevant characteristics of the parcel are taken into account. Variables must be available to represent basic land quantity, land quality adjustments, basic building quantity, building quality adjustments, values of other improvements, and overall quality adjustments. The choice of variables to represent these characteristics will, of course, depend on climate, region of the country, and the specific nature of the city being analyzed. Once appropriate variables have been selected, they must be incorporated into the MRA model. Although both basic quantity variables and quality variable adjustments affect sale price, they do so in somewhat different ways. The basic quantity variables tend to have a direct, additive effect on sale price, so they enter the MRA equation in the usual linear manner. Quality adjustment variables are more appropriately seen as percentage adjustments to sale price instead of linear additions. In particular, although many land improvements, land area, and land frontage may contribute to sale value additively, other variables, such as quality of view, extent of landscaping, and building's style, contribute to sale value multiplicatively using percentage adjustments (see Jensen 1987). This hybrid approach allows for compounding effects where appropriate.

Estimating a Hybrid Additive-Multiplicative Model

Jensen (1987) has combined these effects to create hybrid additive-multiplicative models that are not intrinsically linear in their parameters (that is, they are statistically nonlinear). Models whose parameters enter linearly are called statistically linear models. For example,

$$\log \hat{y} = a_0 + a_1 \log X \qquad (1)$$

and

$$\hat{y} = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_1X_2 \tag{2}$$

are statistically linear models even though the X_1X_2 interaction term in the second model and the log terms in the first model make them both mathematically nonlinear (that is, they are nonlinear in their variables). As originally formulated, Jensen's hybrid models such as

$$\hat{y} = b^{a_0} b^{a_1 X_1} X_2^{a_2} + a_3 X_3 \tag{3}$$

are both statistically and mathematically nonlinear.

A simple way to estimate this nonlinear hybrid additive-multiplicative model when a nonlinear regression program is not available is to first reexpress the model as

$$\hat{y} = a_3X_3 + X_2^{a_2} b^{a_0 + a_1 X_1}. \tag{4}$$

Next, specify some relatively arbitrary initial set of values for the regression parameters, such as

$$a_{0(i)} = 1, a_{1(i)} = 1, a_{2(i)} = 1, a_{3(i)} = 1, \text{ and } b_{(i)} = 2,$$

and then repeat the following four steps until convergence.

Step 1. Define

$$W_{(i)} \equiv X_2^{a_{2(i)}} b_{(i)}^{a_{0(i)} + a_{1(i)} X_1}$$

and then

$$Z_{0(i)} = [\log b_{(i)}] W_{(i)}, \qquad Z_{1(i)} \equiv X_1 [\log b_{(i)}] W_{(i)},$$

$$Z_{2(i)} \equiv [\log X_2] W_{(i)}, \qquad Z_{3(i)} \equiv X_3,$$

and

$$Z_{4(i)} \equiv [a_{0(i)} + a_{1(i)} X_1] b_{(i)}^{-1} W_{(i)}$$

where $Z_{0(i)}$, $Z_{1(i)}$, $Z_{2(i)}$, $Z_{3(i)}$, and $Z_{4(i)}$ are the derivatives of $\hat{y}$ in equation 4 with respect to a_0, a_1, a_2, a_3, and b, respectively, evaluated at the ith iteration, where the first iteration begins with the relatively arbitrary set of initial values given above.

Step 2. Define

$$y^*_{(i)} \equiv y - [W_{(i)} + a_{3(i)} X_3] + a_{0(i)} Z_{0(i)} + a_{1(i)} Z_{(i)}$$
$$+ a_{2(i)} Z_{2(i)} + a_{3(i)} Z_{3(i)} + b_{(i)} Z_{4(i)}.$$

Step 3. Now run an ordinary least-squares regression without an intercept term, *using any standard linear regression computer package*, with $y^*_{(i)}$ as the dependent variable and $Z_{0(i)}$; $Z_{1(i)}$; $Z_{2(i)}$; $Z_{3(i)}$; and $Z_{4(i)}$ as the set of explanatory (independent) variables. With the least-squares estimated regression coefficients defined as a_0, a_1, a_2, a_3, and b, the resulting fitted regression may be expressed as

$$\hat{y}^{*}_{(t)} = a_0 + a_1 Z_{1(t)} + a_2 Z_{2(t)} + a_3 Z_{3(t)} + b Z_{4(t)}.$$

Step 4. Reset the initial parameter values (that is, regression coefficients) to be equal to the estimated least-squares regression coefficients from step 3:

$$a_{0(t)} = a_0, \quad a_{1(t)} = a_1, \quad a_{2(t)} = a_2, \quad a_{3(t)} = a_3, \text{ and } \quad b_{(t)} = b.$$

This means recalculating $W_{(t)}$, $Z_{0(t)}$, $Z_{1(t)}$, $Z_{2(t)}$, $Z_{3(t)}$, $Z_{4(t)}$, and $y^{*}_{(t)}$ at each iteration.

Repeat these four steps until convergence. Convergence may be defined as the point where

$$|a_{j(t+1)} - a_{j(t)}| < 0.0001 \text{ for } j = 1,\ldots,3 \quad \text{and} \quad |b_{(t+1)} - b_{(t)}| < 0.0001$$

for approximately four decimal places of accuracy. If convergence can be obtained without much difficulty, this is an easy procedure for fitting hybrid additive-multiplicative models. The above procedure is the Gauss-Newton nonlinear method, which simply runs a linear, least-squares regression at each iteration as in step 3 above. Understanding this procedure means that an assessor can estimate a nonlinear regression model even when only a linear regression computer program is available. Moreover, the coefficient estimates from the linear regression will be exactly the same at each iteration as those produced by the nonlinear computer program at that iteration when the original Gauss-Newton algorithm is used. These results will not necessarily be the same as those produced by a nonlinear computer program using the Marquardt (modified Gauss) or steepest-descent algorithms.

Alternatively, Jensen (1987) has shown that his hybrid models can be roughly approximated by models that are at least linear in their parameters (statistically linear) even though they are still nonlinear in their variables (mathematically nonlinear) due to interaction effects. Moreover, statistically linear models can be estimated using linear MRA even when they contain mathematically nonlinear terms (such as $X_i X_j$ interaction terms). This simplifies the estimation of such hybrid models. However, there may be times when it might be worth the additional time and effort to obtain the greater accuracy that might be possible by using nonlinear estimation methods to estimate these hybrid models in their original statistically nonlinear forms.

The Broken Response Surface Method

Once all the requisite control variables from official records are included in the MRA hybrid additive-multiplicative model, the effect of location on sale value using the broken response surface method can be considered. The theoretical background for understanding this method may be found in the work on splines by Smith (1979). An example and the SAS computer programming for applying this method was provided by Marsh (1983; 1986).

Consider a situation where there are many sales of houses widely scattered throughout the city. Lay out a map of the city and pinpoint the location of each house. Next, represent the sales values as vertical distances above the map. This is not necessary for carrying out the method but provides a conceptual understanding of what the method is doing. The broken response surface method fits a surface to these sales values that is capable of representing both abrupt changes in value up or down in some locations in the city and rather gradual changes in other locations depending on the statistical significance of the potential adjustments to value. Crossing railroad tracks or a busy highway may or may not result in an abrupt change in sales values, but any visible or invisible line separating houses of substantially different locational value can be identified statistically by this method. Gradual changes in locational value can also be identified. Flat land that becomes rolling hills that becomes mountainous terrain may or may not bring about gradual changes in locational value, but any statistically significant changes in locational value, regardless of the reason, will be picked up using the broken response surface method. No symmetry is required for this method, so there need not be anything symmetrical about the locational values throughout the city.

The broken response surface method of estimating location values can be applied to either the original nonlinear hybrid model or to its reformulated statistically linear version. Essentially, the broken response surface method adds statistically linear (but occasionally mathematically nonlinear) terms to either the original or reformulated hybrid models. The adjustments to value provided by the method may be either abrupt (using binary variables) or gradual (using linear, quadratic, or cubic terms for spline smoothing). The broken response surface method fits a three-dimensional smoothing polynomial to the sale price data (z-coordinate) using west-to-east distances (x-coordinate) and south-to-north distances (y-coordinate) measured from the southwesternmost point in the tax district. The z-dimension representing sale price initially generates a rugged surface. These sharp differences in value may greatly overstate the effect of location until the assessor has controlled for other characteristics of the properties by entering appropriate variables into the regression model. By controlling for all of the appropriate characteristics other than location, the subject properties are, in effect, transformed into an otherwise homogeneous set of properties. The remaining variation in sale price is then seen as largely due to the effect of location. Thus, once appropriate land and building value variables and their adjustments, as well as values for other improvements and global adjustments, have been fully incorporated into the model, the residual effect of location will appear greatly subdued.

A properly estimated general representation of the hybrid additive-multiplicative model extended to include quadratic response surface adjustments for location may be written as:

$$\hat{z} = HAM + a_0 + b_0 X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y, \quad (5)$$

where $\hat{z}$ is the predicted or fitted sale price, *HAM* is the hybrid additive-multiplicative part of the model, and X and y represent the x-coordinate and y-coordinate distance measures described above, respectively. Equation 5 represents location as an additive adjustment to value but could easily be made multiplicative by replacing each variable by the logarithm of itself (that is, $X = \log(X)$, $X^2 = \log(X^2)$, $HAM = \log(HAM)$, and so on).

The choice of additive versus multiplicative location effects is really an empirical issue. Location effects are residual effects that may implicitly represent a wide variety of location factors. Such hidden factors might include distance from school, shopping or job locations, or the physical presence or absence of such things as an auto junkyard, golf course, busy highway, or spectacular view. If luxuries like spectacular view dominate, then location effect may be essentially multiplicative, providing a percentage adjustment. This is because luxuries may be more important to people who can afford to buy larger, better-quality houses and pay for some added luxuries, including those associated with location. People buying smaller, less desirable houses may be less influenced by location effects considered luxuries. On the other hand, equal dollar adjustments may be appropriate for location effects emanating from necessity, such as proximity to school, shopping, or work sites for those with transportation constraints. Ultimately, the additive versus multiplicative nature of location must be sorted out empirically and could involve both types of effects (such additive-multiplicative models are called transcendental equations in production analysis).

The problem with equation 5 as currently written is that it generated a very smooth, symmetric representation of the effect of location on sale price. This is, of course, unrealistic. There is no reason that the effect of location on sale price has to be symmetric. Consequently, a sufficient number of adjustment terms must be allowed into the model to increase or decrease value abruptly or gradually at various locations in a completely asymmetric manner. Some abrupt adjustments may or may not be desirable depending on the location effects in the particular city of interest. Because more gradual linear, quadratic, or cubic adjustments logically follow more complicated transformations of abrupt adjustments, the abrupt adjustments will be discussed in detail first, and either augmenting them or replacing them with more gradual adjustments will easily follow. A true spline regression model does not use the abrupt adjustments at all, but only adjustments that maintain the continuity of the original function (see Marsh 1983 or Marsh 1986).

First define a set of binary (0,1) variables (also known as indicator or dummy variables) along the x-axis: D_{Xi}, where $i = 1,\ldots, I$ such that

$$D_{Xi} = 0 \text{ if } X < X_i^* \text{ and } D_{Xi} = 1 \text{ if } X \geq X_i^*, \tag{6}$$

where the asterisk, *, is used to mean a value of X that is fixed at some ith location. Similarly, along the y-axis define D_{yj}, where $j = 1,\ldots, J$ such that

$$D_{yj} = 0 \text{ if } y < y_j^* \text{ and } D_{yj} = 1 \text{ if } y \geq y_j^*. \tag{7}$$

Note that this method is not adjusting one cell at a time; rather, each location variable splits the data points into two groups differentiating all points to the west from all points to the east, or differentiating all points to the south from all points to the north, or some combination when interaction variables are used. Thus, one adjustment variable with just one coefficient can, for 10,000 sale values, differentiate 4,000 of the values from the other 6,000 values. Again, this is not a one-cell-at-a-time adjustment process. Using an equal number of binary variables for X and y will separate the observations into 4 groups with 2 variables, 121 groups with 20 variables, 676 groups with 50 variables, and 10,000 groups using 198 variables. To understand how to generate these variables in practice, consider the CAMA2 data set used as a training data set by the IAAO. The data set includes an x-coordinate (*XCOORD*) that ranges from 0 to 1,208 and a y-coordinate (*YCOORD*) that ranges from 365 to 1,064.

The original spline adjustment regression search method proposed by Marsh (1983) used only the observed, sample points (such as the observed sale values) as potential adjustment points. This works fine if the sample points are well-distributed over the region of interest but may be unrepresentative and give distorted results if sales tend to clump together more in some areas than in others for a given sample in a given year. One way of dealing with this is to allow potential adjustments at a prespecified, evenly distributed set of points. In this way important adjustment locations may be identified that may or may not correspond exactly to the points of sale, especially in areas with sparse sales. Moreover, this approach allows the location points to be selected simultaneously while at the same time determining the magnitudes of the location coefficients along with the magnitudes of the other parameters of the hybrid model. Of course, one could create potential adjustments at actual sale points augmented by additional points where sales are thin or where assessors have specific reasons to believe adjustments may be needed. A well-designed procedure may be able to allow for both types of points.

To keep things simple for purposes of exposition, the following discussion uses a hypothetical set of potential adjustment points defined by a neatly ordered grid of streets and avenues. In reality, a more appropriate set of potential adjustment points particular to the city of interest must be specified.

The particular town is not specified, but we have used these data to estimate an assessment model adjusted for location effects using our proposed method. However, we have chosen to focus on explaining our procedure rather than discussing our MRA results. Imagine a small town with just twenty streets that run north and south so that street numbers are read in the west-to-east direction (X direction) and five avenues that run east and west so that avenue numbers are read in the south-to-north direction (y direction). Although there is no particular need for the adjustment points to correspond to particular streets and avenues, this is convenient for purposes of discussion. Thus, we may select the end of each block in the west-to-east direction as a potential adjustment point, with the exception of the end of the last block because it is at the boundary. Consequently, with twenty streets there are eighteen potential interior adjustment locations in the west-to-east direction, and with five avenues there are three potential interior adjustment locations in the south-to-north direction. For those using the *NCSS* computer program as used in the IAAO training classes, the eighteen binary variables in the X dimension (ranging from 0 to 1,208) may be created with the transformation:

$$[\#1:\#18] = (XCOORD >= (([2:19] - 1) * (1208/19))).$$

This represents the eighteen interior streets from the original twenty streets separated by nineteen blocks. Similarly, binary variables for the three interior avenues from the original five avenues separated by four blocks are generated by:

$$[19:\#21] = (YCOORD >= (([2:4]) - 1) * (699/4) + 365)),$$

where the y variable (*YCOORD*) ranges from 365 to 1,064 (note that 699 + 365 = 1,064).

Now to allow for interaction (asymmetry), define

$$D_{Xiyj} = 0 \text{ if } X < X_i^* \text{ or } y < y_j^* \text{ and } D_{Xiyj} = 1 \text{ if } X \geq X_i^* \text{ and } y \geq y_j^*. \quad (8)$$

These binary variables allow for abrupt changes in property values at the ith point (ith street) in the west-to-east (that is, X) direction and at the jth point (jth avenue) in the south-to-north (that is, y) direction. Consequently, with eighteen west-to-east adjustment locations and three south-to-north adjustment locations, we have fifty-four (18 × 3) potential binary variable interaction adjustment terms. In *NCSS* the interaction terms may be generated by the transformation statements

$$[\#22:\#39] = \#19 * [\#1:\#18], [\#40:\#57] = \#20 * [\#1:\#18],$$

$$\text{and } [\#58:\#75] = \#21 * [\#1:\#18].$$

Of course, an experienced assessor who knows the tax district well may be able to hand-select some of the potential adjustment points and then check their statistical significance instead of starting with a purely generic grid of potential adjustment points.

Once the seventy-five binary adjustment variables have been created ($18 + 3 + 54 = 75$), each is considered according to its potential statistical significance for inclusion in a stepwise regression model. With the stepwise add approach, all the potential adjustment variables enter the regression model one at a time according to their statistically significant contributions, as recalculated at each step in the process, to explain the variation in sale price. The advantage of this stepwise add approach is that it starts with a relatively simple model and complicates it one step at a time. The disadvantage is that an incomplete initial model is likely to be biased and inconsistent because initially it is likely to be missing key adjustment variables that may be important in explaining sale price.

An alternative approach would be to start with all seventy-five of these variables in the initial regression model along with the hybrid additive-multiplicative variables and then delete those variables that were contributing the least to the model. This alternative approach is the stepwise delete method. This method starts by including in the regression a large number (seventy-five in the above example) of possible value adjustment locations (known as knot locations or join points) and continues deleting the least significant locations one at a time until all the remaining ones make statistically significant contributions in explaining sale price at some prespecified level of statistical significance. To the extent that the initial model was otherwise correct, having more adjustment terms than needed would make the estimated regression coefficients inefficient, but not biased or inconsistent.

One could argue that there is more to lose by being too stingy with the variables than vice versa. Nonetheless, for the purpose of clearer exposition, we proceed with the analysis in a stepwise add manner. In practice, both the automatic and manual stepwise options in *NCSS* tend to get bogged down when a very large number (close to the maximum of 250) of highly intercorrelated variables is used. Consequently, the assessor may have to run the regular *NCSS* regression routine repeatedly and hand-select or hand-delete variables until they are reduced to a manageable set.

Continuing with the example using stepwise add, if an abrupt change in property values occurs at 7th Street, the binary variable D_{X7} may be added to equation 5 to get equation 9:

$$\hat{z} = HAM + a_0 + b_0 X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y + a_7 D_{X7}. \quad (9)$$

An equation like equation 9 really represents two equations: one equation for which the binary variable is equal to zero and a second equation for which the binary variable is set equal to one.

Define X_7^* as the value of the X variable ($XCOORD$) at 7th Street, where X is measured from west to east (the streets in this example are assumed to run north and south). When X is less than X_7^* (as it is everywhere to the west of 7th Street) then D_{X7} will be equal to zero and equation 9 will revert to equation 5:

$$\hat{z} = HAM + a_0 + b_0 X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y. \quad (10)$$

On the other hand, at 7th Street and for all points to the east of 7th Street, X will be equal or greater than X_7^*, so by definition D_{X7} will be equal to one so that the term $a_7 D_{X7}$ becomes simply a_7 and the second equation will be

$$\hat{z} = HAM + (a_0 + a_7) + b_0 X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y. \quad (11)$$

The $(a_0 + a_7)$ term in this equation reveals that property values have jumped by a_7 dollars in going from west of 7th Street to east of 7th Street. Other such adjustment variables may enter the equation to account for property value changes at various streets in the west-to-east direction. Similar adjustments can be made at points (avenues) along the y (south-to-north) direction based on the variable $YCOORD$.

The interaction binary variables

$$D_{XiYj} = D_{Xi} * D_{Yj}$$

provide a simple means of making adjustments in value that are more accurately defined by the intersection of the ith street and the jth avenue. For example, if the area to the northeast of the intersection of 10th Street and 3rd Avenue exhibits significantly higher property values, then the D_{X10Y3} binary variable should reveal a statistically significant contribution to value as one moves to the northeast of that intersection. Equation 11 must then be extended to incorporate this adjustment into the model as follows:

$$\hat{z} = HAM + a_0 + b_0 X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y + a_7 D_{X7} + a_{10,3} D_{X10,Y3}. \quad (12)$$

Because of the two binary variables in equation 12, the equation actually represents 4 (2 * 2) equations as follows: for west of 7th Street, equation 5; between 7th Street and 10th Street, equation 11; east of 10th Street but south of 3rd Avenue, equation 11; and east of 10th Street and north of 3rd Avenue,

$$\hat{z} = HAM + (a_0 + a_7 + a_{10,3}) + b_0 X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y. \quad (13)$$

Depending on the signs of the a_7 and $a_{10,3}$ regression coefficients, property values can rise or fall at these adjustment locations.

Thus, for a town with $I + 2$ streets and $J + 2$ avenues such that $i = 1,\ldots, I$ and $j = 1,\ldots, J$, there are three sets of adjustment location variables: one set of I binary variables for west-to-east adjustments, $[D_{X1},\ldots,D_{XI}]$, a second set of J binary variables for south-to-north adjustments, $[D_{Y1},\ldots,D_{YJ}]$, and a third set of $I * J$ interaction binary variables for adjustments at intersections: $[D_{X1Y1},\ldots,D_{XIYJ}]$. As demonstrated above for a town of twenty streets and five avenues, this process generates up to seventy-five possible adjustment variables capable of providing an abrupt increase or decrease in value if included in the model through a stepwise add or stepwise delete selection process.

One interesting phenomenon that often arises in applying the broken response surface method is the isolation of observations that are out of line with similar properties, or at least with properties that have been made relatively homogeneous by controlling for relevant differences. These outliers might otherwise mislead an assessor by distorting the regression coefficient estimates, especially with respect to location because location is essentially being viewed here as a residual effect after controlling for all other relevant characteristics.

For example, say an adjustment term allowing for a large increase in value is observed at 15th Street at the same time that an equally large decrease in value is observed at 16th Street. Meanwhile, a substantial increase in value is detected at 3rd Avenue, followed by a similar-sized decrease in value at 4th Avenue. This suggests an unusual property located in the block between 15th and 16th Streets and between 3rd and 4th Avenues. If these four adjustments are all statistically significant when simultaneously present but much weaker and statistically insignificant when one or more is left out of the regression, then an abnormal property whose characteristics have not been adequately controlled for by the set of independent variables has been found. Either this property has some unique physical characteristic that considerably increased its value, or something is unusual about the sales transaction itself. In either case, the broken response surface method will tend to identify and isolate such properties so that they do not bias the coefficient estimates of the free market regression model representing arm's-length transaction values.

More Gradual Adjustments to Value

The binary variables defined above provide a means of allowing for abrupt, discrete jumps in property values where such sudden changes in value might be appropriate. However, some changes in value might occur gradually. This means that value may move up a slope at some particular rate of increase instead of jumping suddenly at some point. A gradual increase in value as a result of location may occur in a linear (straight-line) manner or may be better represented by a quadratic or even cubic expression in a more subtle and flexible way. Sine and cosine terms could also be introduced into this stepwise process as a simple extension of the method described in this paper.

First, consider how to introduce linear adjustments in property valuation. A gradual linear adjustment beginning at the ith street along the west-to-east direction may be expressed in general as

$$(X - X_i^*)\, D_{Xi}.$$

More specifically, continuing the above example, if this gradual linear adjustment were to begin at 12th Street, then

$$(X - X_{12}^*)\, D_{X12}$$

could be added to equation 12 to get:

$$\begin{aligned} \hat{z} = HAM &+ a_0 + b_0 X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y \\ &+ a_7 D_{X7} + a_{10,3} D_{X10,Y3} + b_{12} (X - X_{12}^*) D_{X12}. \end{aligned} \quad (14)$$

West of 12th Street, $D_{X12} = 0$ so that

$$b_{12} (X - X_{12}^*) D_{X12}$$

will be zero. East of 12th Street, $D_{X12} = 1$ so that

$$b_{12} (X - X_{12}^*) D_{X12}$$

enters the equation.

However, the value of X at 12th Street is X_{12}^* so that

$$(X - X_{12}^*) = 0$$

at that point.

This means that no abrupt jump in value occurs at 12th Street. Instead, the rate at which X has been raising or lowering value, b_0, is altered so that the new rate east of 12th Street becomes $(b_0 + b_{12})$. Thus, the value of the function did not change at this point, but its rate of increase did. Another way of looking at this is to say that $b_{12}X$ and $-b_{12} X_{12}^*$ both enter the equation when D_{X12} changes from zero to one. Ordinarily the $-b_{12} X_{12}^*$ term would cause an abrupt jump in the function when added to the intercept term. However, it is exactly countered at the point where X is equal to X_{12}^* by the $b_{12}X$ term, so no jump in the function occurs. On the other hand, there is an abrupt jump in the first derivative (that is, slope) of the function from b_0 to $(b_o + b_{12})$. West of 12th Street the regression model remains the same as expressed by equation 12. However, east of 12th Street and south of 3rd Avenue, the new regression model becomes

$$\hat{z} = HAM + (a_0 + a_7 - b_{12} X^*_{12}) + (b_0 + b_{12}) X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y. \quad (15)$$

East of 12th Street and north of 3rd Avenue, the new equation may be expressed as

$$\hat{z} = HAM + (a_0 + a_7 - b_{12} X^*_{12} + a_{10,3}) + (b_0 + b_{12}) X + c_0 y + d_0 X^2 + e_0 y^2 + f_0 X y. \quad (16)$$

Of course, many such adjustments could take place along the west-to-east dimension represented by the X axis. Similar adjustments could occur at the jth avenue represented by the Y axis by introducing the variables

$$(y - y^*_j) D_{yj},$$

where $j = 1, ..., J$.

Furthermore, interaction terms of the form

$$(X - X^*_i)(y - Y^*_j)D_{XiYj}$$

can be created to allow for adjustments at the intersection of the ith street and jth avenue.

More flexible transitions in property values can be accommodated by introducing quadratic and cubic variables such as

$$(X - X^*_i)^2 D_{Xi} \text{ and } (X - X^*_i)^3 D_{Xi}$$

in the west-to-east direction, and

$$(y - y^*_j)^2 D_{yj} \text{ and } (y - y^*_j)^3 D_{yj}$$

in the south-to-north direction. Note that for second derivative adjustments such as

$$d_0 (X - X^*_i)^2 D_{Xi}$$

both the original function and its first derivative with respect to X are unchanged at the point where X is equal to X^*_i and D_{Xi} becomes equal to 1, because both

$$d_0 (X - X^*_i)^2 \text{ and } 2\, d_0 (X - X^*_i)$$

are equal to zero at that point. However, the second derivative with respect to X, $2\, d_0$, is not equal to zero unless d_0 itself is zero. Furthermore, because this is a quadratic term, all third-order or higher derivatives will also be equal to zero. Consequently, at the point $X = X^*_i$ only the second derivative or slope of the function is altered, with no change to the function itself or any of its other derivatives. In a similar manner, the cubic adjustment term

$$g_0 (X - X_i^*)^3 D_{Xi}$$

will alter only the third derivative and not the function or any other derivative. Again, this is because at this point where $X = X_i^*$, the term $(X - X_i^*)$ will be zero so that the function adjustment

$$g_0 (X - X_i^*)^3$$

will be zero, the first derivative

$$3 g_0 (X - X_i^*)^2$$

will be zero, the second derivative

$$6 g_0 (X - X_i^*)$$

will be zero, but the third derivative $6 g_0$ will not be zero.

Furthermore, corresponding quadratic and cubic interaction terms such as

$$(X - X_i^*)^2 (y - y_j^*)^2 D_{XiYj} \text{ and } (X - X_i^*)^3 (y - y_j^*)^3 D_{XiYj}$$

can be made available for intersection adjustments. These adjustments work in a manner similar to the intersection function adjustments, D_{XiYj}, and the intersection linear adjustments,

$$(X - X_i^*)(y - y_j^*) D_{XiYj}.$$

Summary and Conclusions

This paper has focused on the problem of estimating the location values of properties. Location value may be thought of as a residual value after all other characteristics of a property have been taken into account by including them appropriately in a multiple regression equation. Consequently, specification of the hybrid additive-multiplicative model developed by Jensen (1987) has been discussed as a first step in formulating a model to estimate the location values of properties. Methods of estimating such a model using linear and nonlinear approaches have been explored.

This paper emphasizes the need to incorporate into the estimation procedure enough flexibility so that the location of the final selection of adjustment points can be determined simultaneously with the estimation of the regression coefficients. The use of stepwise regression is proposed to select the most statistically significant adjustment points from whatever prespecified set of potential adjustment points the assessor decides to use.

The broken response surface method provides an extension of the hybrid additive-multiplicative model that allows for the estimation of location value through the use of three-dimensional polynomial splines. Details have been provided on how to create appropriate variables to represent various binary variable and spline adjustments to determine the effect of location value on sale price.

Special features of the broken response surface method have been examined, including the choice between stepwise add and stepwise delete algorithms, the robust nature of the method in identifying unusual properties that don't fit the prevailing sales pattern, and the difference between abrupt adjustments and gradual adjustments in value.

Readers wishing to apply these methods to data using their own favorite regression analysis computer program should feel free to contact the authors for advice and assistance.

References

Jensen, David L. 1987. Alternative modeling techniques in computer-assisted mass appraisal. *Property Tax Journal* 6(3):193–237.

Marsh, Lawrence C. 1983. On estimating spline regressions. *Proceedings of the SAS Users' Group International* 8:723–28.

Marsh, Lawrence C. 1986. Estimating the number and location of knots in spline regressions. *Journal of Applied Business Research* 3:60–70.

Smith, Patricia L. 1979. Splines as a useful and convenient statistical tool. *The American Statistician* 33:57–62.