

Home Value - Location, Location, Location: An Application in Real Property Global Response Surface Technology

Lawrence C. Marsh, University of Notre Dame

Anthony B. Sindone, University of Notre Dame

Introduction

Current interest in the field of estimating property value has focused on the use of multiple regression analyses and other econometric tools. Of particular interest is the effect of location on property values. This effect is generally found by using non-linear hybrid models to control for all effects other than location. This would result in the relatively heterogeneous set of sold homes being transformed into a relatively homogeneous set.

Examples of this usage can be found in the work of Jack Eichenbaum from the Appraisal Research Division, New York City Department of Finance. He used a hybrid "Half-Gaussian" function to arrive at his location value response surface (LVRS) method of land value estimation.

This and other models like this use Gauss-Newton and Newton-Raphson non-linear least squares estimation techniques on previously determined and stationary location points. However forcing these points to remain fixed adds bias to the estimates. Also, traditional methods, such as that mentioned above, use the largest residual as the adjustment point for the non-linear regression. That is a good first step. However, that only determines a local minimum. Our algorithm uses an iterative procedure that produces an overall or global minimum. Additionally, this algorithm would allow more flexibility in the estimating procedure so that the location of the final selection of adjustment points can be determined simultaneously with the estimation of the other regression coefficients. We have designed this generalized Global Response Surface Method using SAS* and applied in to the location value problem in real property assessment.

Estimating the Hybrid Model

The hybrid model is one that incorporates both quantitative and qualitative variables in the estimate. This is a direct market model that assigns additive value to quantitative variables and multiplicative values to qualitative variables.

Examples of quantitative variables in property valuation estimates would be: "square feet of living space", "number of bedrooms", and "size of the lot" etc. Examples of qualitative variables would be: "shape of the lot", "proximity to schools", and "view".

Mathematically, a hybrid model could be formulated as follows:

$$y = b^{a_0} b^{a_1 x_1} x_2^{a_2} + a_3 x_3 \quad (1)$$

This expression is both mathematically and statistically non-linear. That is, both the variables and parameters are specified as non-linear. It also illustrates the additive and multiplicative properties of the variables and parameters. This estimate can be performed by using the Gauss-Newton method. In a simple way this would be performed as follows:

Reexpress the model as:

$$y = a_3 x_3 + x_2^{a_2} b^{a_0} + a_1 x_1. \quad (2)$$

Then set arbitrary values for the regression parameters:

$$a_0(i) = 1, a_1(i) = 1, a_2(i) = 1, a_3(i) = 1, b(i) = 2.$$

Define: $W(i) = x_2^{a_2(i)} b(i)^{a_0(i)} + a_1(i) x_1$

and $Z_0(i) = [\log b(i)] W(i), Z_1(i) = x_1 [\log b(i)] W(i),$

$$Z_2(i) = [\log x_2] W(i), Z_3(i) = x_3,$$

$$Z_4(i) = [a_0(i) + a_1(i) x_1] b(i)^{-1} W(i)$$

where the Z's are the derivatives of y in equation 2 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 arbitrary set of initial values given above.

Next define: $y^*(i) = Y - [W(i) + a_3(i) x_3] + a_0(i) Z_0(i) + a_1(i) Z_1(i)$

$$+ a_2(i)Z_2(i) + a_3(i) + b(i)Z_4(i).$$

Next run ordinary least squares with y^* as the dependent variable and the Z 's as the independent variables.

Reset the parameter values to be equal to the estimated least-squares coefficients from the above steps. In other words, recalculate each W and Z until convergence. Convergence can be defined as the point where:

$$[a_j(i+1) - a_j(i)] < 0.0001 \text{ for } j = 1, \dots, 3 \text{ and } [b(i+1) - b(i)] < 0.0001.$$

This is essentially what the NLIN procedure would be doing if it were given the above specified model along with the defined derivatives. It would run ordinary least squares estimating y^* and while redefining parameter values until convergence.

The Hybrid Model and Non-Linear Estimation of Property Value

Real property valuation, especially location effects, requires that other relevant characteristics of the parcel of housing be taken into account. The resulting residual would be identified as the location effect on that parcel. The problem then is deciding which variables are most significant in determining sale or market value. Then those variables would be placed into a specified model so that the location effects could be estimated.

Variables are chosen, in our method, by performing ordinary least squares on all of the relevant actual or composite variables as defined by the International Association of Assessing Officers (IAAO) on a data set containing properties.

The variables we used were defined as:

DATE, NBHTYLIN, LOTSHLIN, LOTSIZE, STORYTLA, STORYLIN, SFLAOTH, ROOMS, BATHSLIN, GRADELIN, CONDLIN, GARAGE, LOTSHLOT, LOTADJ, SQRTLOT, TLAADJ, SQLASQLT, AGECOND, XCOORD, AND YCOORD. Some of these variables are self-explanatory. Others are not so apparent. These are defined by the IAAO training course materials.

Next we defined a grid of locations based on a simulated city designating a point on the grid as the origin or (0,0). On a city map, location coordinates can then be expressed in centimeters in east-west and north-south directions. The (0,0) point could be the most southwestern point for convenience. Each property would then be located in terms of these coordinates. The east-west coordinate is designated X-COORD and the north-south coordinate is designated Y-COORD. These location variables can then be incorporated into the model.

After PROC REG is run on these variables, the property with the highest residual would be the location of the first adjustment point which would be used to find the local minimum. For efficiency, the property with the highest residual can be found by running PROC SORT by descending residuals. This would put property with the highest residual in the first position. Working one direction at a time, we designate the X-COORD variable as KA1 for the sorted observations. We then merge the original input data with the sorted data. This provides us with the ability to define our adjustment points using splines. The expression: IF X-COORD LT KA1 THEN DA1 = 0; ELSE DA1 = 1; allows us to define variable XCRDKA1 = (XCOORD - KA1) * DA1;. This creates set of observations defining abrupt changes in property values by location. OLS is run on the original set of variables plus the variables X-COORD and XCRDKA1.

The most significant variables would then be used to specify our hybrid additive and multiplicative (HAM) model and the dummy variable submodel (DUM). In PROC NLIN:

The fully specified model used is:

$$\begin{aligned} \text{HAM} &= (\text{DATE}^{B8}) * (((\text{LOTADJ}^{B13}) * \text{LOTSIZE}^{B4}) + ((\text{GRADELIN}^{B9}) \\ &\quad * \text{STORYTLA}^{B15} + \text{BATHSLIN}^{B8}) + \text{GARAGE}^{B11}); \\ \text{DUM} &= ((\text{AX} * \text{XCOORD}) + (\text{A1X} * \text{XCRDKA1} * \text{DA1}) - (\text{A1X} * \text{KA1} * \text{DA1})); \end{aligned}$$

MODEL SALE = HAM * DUM;

All qualitative variables are raised to the power of their parameters, ie. DATE^{B8} . All quantitative variables are simply multiplied by their parameters. AX is the coefficient of XCOORD. A1X is the coefficient of XCRDKA1.

No attempt is made here to present the entire model. It is quite lengthy and complex. However, the general idea is shown above. We would iterate this procedure in both directions until convergence. The important point, though, is that the

residual points would not necessarily be fixed. The adjustment points would generally be floating or changing with each iteration. Therefore, the specified parameter values would be determined within the model. Making this choice less arbitrary can only reduce bias in estimating the effect of location on the final sale value of real property.

Conclusions

This Global Response Surface Method allows the selection of adjustment points to be determined simultaneously with the estimation of the regression coefficients. Allowing adjustment points to float enhances flexibility within the estimating procedure, thus reducing bias.

* SAS is a registered trademark of SAS Institute Inc. , Cary, NC, USA.

Any questions or comments may be addressed to the authors at the following address:

Department of Economics
University Of Notre Dame
Notre Dame, IN 46556