

Chapter 487

Hybrid Appraisal Models (Old Version)

Introduction

This procedure analyzes a special prediction model often used in mass appraisal and assessment. The model is referred to as the *hybrid* model. Although **NCSS** has a nonlinear regression module for solving this model, this module was added for several reasons:

1. It provides several methods of optimization for calibrating the model parameters. Some of these methods have the same goal as the 'feedback algorithm'.
2. It streamlines model setup, including the automatic generation of binary variables.
3. It provides many new reports and statistics that aid in the calibration of the parameter values.

The use of multiple and nonlinear regression in property appraisal and assessment administration has been encouraged by such organizations as the International Association of Assessing Officers (IAAO). They publish a book by Eckert (1990) and teach courses which use these regression procedures. This program was developed to automate the hybrid model that they propose.

The Hybrid Appraisal Model

The hybrid model is a combination of both additive and multiplicative models. It relates the sales price of a property to various characteristics such as size (in square feet), lot size, construction quality, location, number of bathrooms, etc. This model computes a market value for each 'structure' of the parcel, where a structure refers to an object like the land or a building. For example, the sales price may be based on the sum of the market values of the building, the lot, and a garage. Because of variance that may occur because of location, time, etc., this sum may be adjusted by one or more overall variables. These overall variables act as percentage adjustments.

The general form of the model is:

$$\text{Sales Price} = \text{Overall} (\text{Building} + \text{Land} + \text{Garage} + \dots)$$

Each of these factors are modeled by one or more variables from the database. These factor models are made up of three types of variables: rate, binary, and amount. The amount variable usually represents the 'size' of the structure. Examples are acreage and square footage. The rate and binary variables are variables that adjust the size variables up or down, such as a quality index or an age adjustment.

Hybrid Appraisal Models (Old Version)

Using this construction, the individual factors are modeled as follows.

$$\text{Overall} = R_1^{B_1} R_2^{B_2} B_3^{I_3} B_4^{I_4}$$

$$\text{Building} = R_5^{B_5} R_6^{B_6} B_7^{I_7} B_8^{I_8} (B_9 A_9 + B_{10} A_{10})$$

$$\text{Land} = R_{11}^{B_{11}} R_{12}^{B_{12}} B_{13}^{I_{13}} B_{14}^{I_{14}} (B_{15} A_{15} + B_{16} A_{16})$$

$$\text{Garage} = R_{17}^{B_{17}} R_{18}^{B_{18}} B_{19}^{I_{19}} B_{20}^{I_{20}} (B_{21} A_{21} + B_{22} A_{22})$$

In this model, R_i represent rates or multipliers (usually centered near one) which modify the whole factor. Examples of rate variables include sales date multiplier and depreciation. These are analogous to percentage adjustment variables.

The I 's represent indicator (binary) variables. These are variables that have only two values: zero and one. Examples of these are neighborhood indicators and special feature indicators (such presence of a swimming pool). Usually, qualitative variables are broken down into individual binary variables. For example, suppose that the properties in your study come from three neighborhoods. You would designate one as the standard neighborhood, then create two binary variables, one for each of the other neighborhoods. The model will then adjust for differences among the three neighborhoods.

The A 's represent amount variables like lot size or square footage of living area. These variables are entered in the normal linear fashion.

The B 's represent the coefficients that are estimated from the data. Often, these coefficients are constrained to lie within specified limits.

Note that the number of variables in each group varies. Quite often you may have ten or fifteen binary variables and only one rate variable.

Estimation Method

The model coefficients may be estimated using various methods. Several methods are available.

Minimize Squared Errors (Nonlinear Regression)

This is the classical approach often used by statisticians because it gives reasonable estimates relatively quickly. This method tends to emphasize relatively expensive properties as compared to less expensive properties.

Minimize the Average |Percent Error|

Using the genetic search algorithm called *differential evolution*, this method finds estimates that minimize the average of the absolute percent errors. These percent errors are the difference between the actual and predicted sales prices divided by the actual price. This method treats all properties equally, regardless of price.

Hybrid Appraisal Models (Old Version)

The function minimized is

$$MAPE = \frac{\sum_{properties} \left| \frac{100(actual - predicted)}{actual} \right|}{N}$$

where $|X|$ represents the absolute (positive) value of X and $\sum_{properties} X$ is interpreted as the sum of the values of all properties. Note that this is the quantity minimized by the *feedback* algorithm.

If you want to use a solution that minimizes the percent errors, this is the solution that we recommend.

Minimize the Maximum |Percent Error|

Using the genetic search algorithm called *differential evolution*, this method finds estimates that minimize the maximum of the absolute percent errors. These percent errors are the difference between the actual and predicted sales prices divided by the actual price.

Minimize the Median |Percent Error|

Using the genetic search algorithm called *differential evolution*, this method finds estimates that minimize the median of the absolute percent errors. These percent errors are the difference between the actual and predicted sales prices divided by the actual price.

Minimize the Percentile |Percent Error|

Using the genetic search algorithm called *differential evolution*, this method finds estimates that minimize a designated percentile of the absolute percent errors. These percent errors are the difference between the actual and predicted sales prices divided by the actual price. The percentile is specified in the Min Percentile box.

Minimize the Average |Error|

Using the genetic search algorithm called *differential evolution*, this method finds estimates that minimize the average of the absolute errors. These absolute errors are the difference between the actual and predicted sales prices. These percent errors are the difference between the actual and predicted sales prices divided by the actual price. The percentile is specified in the Min Percentile box.

Minimize the Median |Error|

Using the genetic search algorithm called *differential evolution*, this method finds estimates that minimize the median of the absolute errors. These errors are the difference between the actual and predicted sales prices.

Minimize the Percentile |Error|

Using the genetic search algorithm called *differential evolution*, this method finds estimates that minimize a designated percentile of the absolute errors. These errors are the difference between the actual and predicted sales prices. The percentile is specified in the Min Percentile box.

Differential Evolution

Introduction

One of the steps in using the hybrid model in mass appraisal is to calibrate (estimate) the unknown coefficient values, the B 's. Often, the method of least squares (MRA) is used to estimate the coefficients. Least squares finds the set of estimates that minimize the sum of the squared errors. That is, the objective of the method of least squares is to minimum the sum of squared errors. The sum of squared errors is called the *objective function*, and the problem is to minimize it. Least squares is one method of minimizing it. Another possibility is simple trial-and-error (of course, trial-and-error may be very time consuming).

Because of the distortion that a few anomalies in the data can cause when least squares is used, other methods have been proposed. One alternative is the *feedback* algorithm. This algorithm seeks to minimize a different objective function: the *average absolute percent error*. This objective function quantifies the percentage accuracy of the model. Both expensive and inexpensive parcels are modeled to the same percentage accuracy. Least squares, on the other hand, concentrates on fitting the most expensive properties. The feedback algorithm is used to minimize a specific objective function: the *average absolute percent error*. Recently, mathematicians have found other algorithms with the same goal of minimizing an objective function. One such algorithm is *differential evolution*.

Differential evolution is one of a group of *genetic algorithms* (see for example, the recent book by Haupt (1998)). By studying how generations respond over time to their environment, mathematicians have discovered new, more robust, algorithms for minimizing an objective function. Differential evolution is one of these algorithms. It can be outlined as follows. A *population* of about 20 individuals has certain traits (values of the unknown coefficients being estimated). The well-being of an individual is measured by his/her value of the objective function. Each individual gives birth to a new individual and then dies, thus forming a new generation. Each new individual *inherits* traits from their parents. The well-being of each child is computed, and if it is better than the parent, it takes the parent's place. Otherwise, the parent's traits pass directly to the child. Finally, occasionally, a mutation of a trait will occur.

The main point to remember is that the big goal is to minimize the objective function. Whether we use the feedback algorithm, or the differential evolution algorithm makes little difference, as long as you find the minimum!

The Algorithm

A population consists of a small group (about 20) of individuals whose characteristics are the values of the unknown coefficients. Using these coefficients, the value of the objective function is computed for each individual. This value is an inverse measure of the wellbeing of an individual. The lower the value, the higher the well-being.

To begin the algorithm, a small group of individuals must be formed. This is done by assigning the nonlinear regression coefficients to one individual and then randomly assigning the other individuals to a grid of values around this first individual. This is the initial population.

The next step is the evolution of the population. The population progresses through a series of *generations*. At each change in generation, depending on a member's wellbeing, each population member may move on to the next generation or be replaced by a better member. For each member, a trial replacement is constructed as follows. The best member of the population is found. The attributes of each replacement member are computed as a weighted average of those of the member and the best member. The amount

Hybrid Appraisal Models (Old Version)

of weight of the best member is controlled by the *inheritance factor*. This is a value between 0 and 1. The closer this value is to 1, the more the replacement member resembles the best member. The closer this value is to 0, the more the replacement member resembles their parent. The value of 0.85 seems to work in many cases.

As in real populations, *mutations* occur at a given rate. When a mutation occurs, a particular trait is changed randomly. This tends to maintain diversity in the population. A mutation rate of about 30% (0.30) seems to work well.

The algorithm proceeds from generation to generation until the population seems to converge to a single individual. The number of generations is arbitrary. Usually, about 100 generations are needed for the algorithm to converge.

Assumptions

The main assumption needed is that the data are well represented by the model.

Data Structure

The data are entered as one dependent variable and one or more independent variables. Each row of data represents a single parcel.

Missing Values

Rows with missing values in the variables being analyzed are ignored in the calculations. When only the value of the dependent variable is missing, predicted values are generated.

Example 1 – Hybrid Appraisal Model

This section presents an example of how to estimate a prediction equation from the sales price data stored in the Assess dataset. In this example, a hybrid model of the form

$$\text{Overall(Land+Building+Garage)}$$

will be estimated.

Setup

To run this example, complete the following steps:

1 Open the Assess example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Assess** and click **OK**.

2 Specify the Hybrid Appraisal Models (Old Version) procedure options

- Find and open the **Hybrid Appraisal Models (Old Version)** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

```

Estimation Method .....Minimize the AVERAGE |PERCENT Error|
Y - Dependent Variable (Sales Price) .....Sale
Factor 1 .....Ov
  Variable .....Date
  Type.....R
Factor 2 .....Ov
  Variable .....Neighborhood(5)
  Type.....D
Factor 3 .....B1
  Variable .....GradeLinear
  Type.....R
Factor 4 .....B1
  Variable .....Sqft1stFlr-Baths
  Type.....A
Factor 5 .....Ln
  Variable .....LotAdjusted
  Type.....R
Factor 6 .....Ln
  Variable .....LotSize
  Type.....A
    
```

Hybrid Appraisal Models (Old Version)

```

Factor 7 ..... B2
Variable ..... GarageSqft
Type..... A

Options Tab
-----
Max Generations ..... 50
Min Percent..... 10
Seed ..... 12346 (for reproducibility)

Reports Tab
-----
Run Summary Report ..... Checked
Nonlinear Regression Iteration Report..... Checked
Differential Evolution Iteration Report ..... Checked
Model Report ..... Checked
Estimated Model Report ..... Checked
Assessment Ratio Report ..... Checked
Predicted Values and Residuals ..... Checked
Report
Predicted Values for New Rows ..... Checked
Exception Report ..... Checked
Exception Percentage..... 25
Ov ..... Overall
Ln..... Land
B1 ..... Building
B2 ..... Garage

Report Options (in the Toolbar)
-----
Variable Labels ..... Column Names
    
```

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Run Summary Report

Run Summary Report

Item	Value
Model	Sale = Overall(Land+Building+Garage)
Estimation Method	Average Percent Error (AAPE)
Final Value of AAPE	11.77
R-Squared (from NonLinReg)	0.8394111
Random Number Seed	-12346
Number of Variables Used	10
Number of Parameters in Model	10
Number of Rows Used	76
Number of N.R. Iterations	8
Number of D.E. Iterations	50

This report displays summary information about the analysis such as the model that was fit, the number of rows and variables, the number of iterations, and the random number seed.

Model

This shows the model that was estimated. It gives you a quick overview.

Estimation Method

This is the estimation method that was used.

Final Value of AAPE

This shows the final (minimum) value of the objective function.

R-Squared (from NonLinReg)

This is the R-Squared that was achieved by the nonlinear regression routine. There is no direct R-Squared defined for nonlinear regression. This is a pseudo R-Squared constructed to approximate the usual R-Squared value used in multiple regression. We use the following generalization of the usual R-Squared formula:

$$R\text{-Squared} = (ModelSS - MeanSS)/(TotalSS - MeanSS)$$

where *MeanSS* is the sum of squares due to the mean, *ModelSS* is the sum of squares due to the model, and *TotalSS* is the total (uncorrected) sum of squares of Y (the dependent variable).

This version of R-Squared tells you how well the model performs after removing the influence of the mean of Y. Since many nonlinear models do not explicitly include a parameter for the mean of Y, this R-Squared may be negative (in which case we set it to zero) or difficult to interpret. However, if you think of it as a direct extension of the R-Squared that you use in multiple regression, it will serve well for comparative purposes.

Random Number Seed

The random number seed is shown so that if you want to duplicate these results, you can be entering this random number seed.

Number of Variables Used

This is the number of variables from the database that were used in the analysis.

Number of Parameters in Model

This is the number of parameters that were estimated by the model.

Number of Rows Used

This is the number of rows from the database that were used in the analysis.

Number of N.R. Iterations

This is the number of iterations used by the nonlinear regression procedure.

Number of D.E. Iterations

This is the number of iterations used by the differential evolution algorithm.

Nonlinear Regression Iteration Section

Nonlinear Regression Iteration Section

Itn No.	Sum of Squared Errors	Message
0	87346.32	
1	20104.99	
2	5566.038	
3	5076.756	
4	5044.927	
5	5044.661	
6	5044.651	
7	5044.647	
8	5044.647	

Convergence criterion met.

This report displays the sum of squared errors which is the objective function of the nonlinear regression routine. It allows you to observe the algorithm's progress toward the solution. If you do not see the message 'Convergence criterion met' at the bottom, it means that the algorithm did not terminate normally, and you should take corrective action—which usually means that you should increase the maximum number of iterations or provide different starting values.

Differential Evolution Iteration Section

Differential Evolution Iteration Section

Iteration Number	Objective Function (AAPE)
0	12.08115
1	12.08115
2	12.08115
3	12.08115
4	12.08115
5	12.08115
10	12.08115
15	12.08115
20	11.93821
25	11.92409
30	11.80285
35	11.78862
40	11.77882
45	11.77437
50	11.76717

This report displays the value of the objective function that is being minimum by the differential evolution algorithm. In this example, it is the average absolute percent error between the actual and predicted sales price. This allows you to see the progress of the algorithm towards a solution. If it looks like the objective function is still shrinking, you may want to rerun the procedure with a larger value for the Max Generations parameter.

Model Specification and Estimation Section

Model Specification and Estimation Section

Component Estimates

Parameter Name	Variable Type	Variable Name	Parameter Estimate	Starting Value	Parameter Bounds
B1	Overall Rate	Date	0.04662152	0	-5 to 5
B2	Overall Binary	Neighborhood=4	0.9451685	1	0 to 5
B3	Overall Binary	Neighborhood=6	1.006669	1	0 to 5
B4	Building Rate	GradeLinear	1.776111	0	-5 to 5
B5	Building Amount	SqFt1stFlr	2.225528	1	0.001 to B
B6	Building Amount	SqFtOthFlr	1.634559	1	0.001 to B
B7	Building Amount	Baths	5.391271	1	0.001 to B
B8	Land Rate	LotAdjusted	3.064366	0	-5 to 5
B9	Land Amount	LotSize	0.4990667	1	0.001 to B
B10	Garage Amount	GarageSqFt	2.325238	1	0.001 to B

This report displays the details of the model that was fit.

Parm Name

The name of the parameter is shown on this line.

Variable Type

This shows the type of the variable as well is the factor that it belongs to.

Variable Name

The name of the variable on the database. Note that for discrete variables that were expanded into a set of binary indicator variables, a separate line is given for each generated variable. The value indicated by this binary variable is shown after an equals sign. For example, a binary variable was generated that is one when the Neighborhood value is '4' and zero otherwise. This variable is called 'Neighborhood=4'.

Parameter Estimate

This is the estimated value of the parameter in the hybrid model. Note that these values should not be analyzed separately, but together as a group. If you change the model in any way (such as including other variables), these values will change—perhaps substantially!

If you are going to use these values to predict sales prices, you should use the double-precision version of these numbers. These are obtained by setting the Precision value to 'Double' in the Report tab.

Starting Value

These are the values used by the nonlinear regression algorithm in the first iteration. Since the differential evolution algorithm uses the results of a nonlinear regression as its starting values, these values have little influence on the results of the differential evolution algorithm.

Parameter Bounds

These are the limits that were provided for the parameter estimates. If you notice an estimate that is equal to one of its bounds, you should analyze the situation carefully to determine if bound should be relaxed to allow the parameter a wider range.

Model

Model	
Factor	Detail
Overall	$Date^{(B1)} * (B2)^{(Neighborhood=4)} * (B3)^{(Neighborhood=6)}$
Land	$LotAdjusted^{(B8)} * ((B9) * LotSize)$
Building	$GradeLinear^{(B4)} * ((B5) * SqFt1stFlr + (B6) * SqFtOthFlr + (B7) * Baths)$
Garage	$(B10) * GarageSqFt$

This report displays the variables that make up each factor. The parameters that will be estimated are B1, B2, ...

Factor

The name of the factor in the hybrid equation.

Detail

The details of the construction of the factor.

Estimated Model

Estimated Model

```
Date^(0.0466215216532935) *(0.945168518786134)^(Neighborhood=4)
*(1.00666919375534)^(Neighborhood=6)*((LotAdjusted^(3.06436632259012) *((0.499066748801789)
*LotSize)))+(GradeLinear^(1.77611087858286) *((2.22552851585426) *SqFt1stFlr +(1.63455938516476)
*SqFtOthFlr +(5.39127079611118) *Baths))+((2.3252376629361) *GarageSqFt))
```

This is the model with the parameter names replaced with the parameter estimates. This expression may be copied onto the Clipboard and pasted into the transformation section of the database to allow you to predict other observations. Note that this expression is always provided in double precision.

Assessment Ratio Section

Assessment Ratio Section

Statistic Name	Actual Sale	Predicted Sale	Ratio	Percent Error
Number of Cases	76	76	76	76
Mean	55.9	55.1	1.01	11.77
Minimum	22.5	29.3	0.77	0.00
Lower Quartile	41.7	42.0	0.90	2.87
Median	51.9	49.4	0.99	9.43
Upper Quartile	64.3	62.6	1.05	16.77
Maximum	117.5	130.4	1.58	57.73
Range	95.0	101.1	0.80	57.72
I. Q. Range	22.6	20.6	0.15	13.90
Variance	418.8	379.8	0.03	150.84
Std. Deviation	20.5	19.5	0.17	12.28
Ave Dev. from Median	14.7	13.5	0.12	8.39
Coef. of Variation x 100	36.62	35.37	16.92	104.37
Coef. of Dispersion x 100	28.38	27.31	11.87	88.96
Weighted Mean			0.99	
Price Related Differential			1.02	

This report provides information that assessors have found useful in analyzing the performance of the estimated model. The four columns of the report represent the actual sales price, the predicted sales, and the ratio of the two (predicted over actual), and the percent error in the predicted. Most of the statistics are defined in the Descriptive Statistics procedure. Uncommon terms are defined next.

Ave |Dev - Median|

The average of the absolute values of the deviations of the variable from its median.

Coef. of Dispersion (COD)

This is 100 times the average absolute deviation about the median divided by the median.

IAOO standards recommend that for single-family residences, COD's of the ratios should be 15.0 or less.

Coef. of Variation (COV)

This is 100 times the standard deviation divided by the mean.

Weighted Mean

The weight ratio mean is the mean of the predicted values divided by the mean of the actual values.

Price Related Differential (PRD)

The price related differential is the mean ratio divided by the weighted mean ratio. It provides a measure of assessment regressivity or progressivity. A PRD greater than 1.0 indicates that the more expensive properties are under appraised. A PRD less than one indicates that the more expensive properties are over appraised. Experience indicates that this value is normal when it is in the range 0.98 to 1.03.

Predicted Values of New Rows Section

Predicted Values of New Rows Section

Row No.	Predicted Sale
77	86.3
78	44.6
79	82.4

The section shows the predicted sales price for rows in which values for all variables except the sales price are given.

Using the Model to Predict for New Parcels

You can use your model to predict sales for new values of the model variables. Here is how. Add new rows to the bottom of your database containing the values of the independent variables that you want to create predictions from. Leave the sales price variable blank. When the program analyzes your data, it will skip these rows during the estimation phase, but it will generate predicted values for all rows with a complete set of independent variables, regardless of whether the sales price variable is available.

Exception Report Section

Exception Report Section

Row No.	Actual Sale	Predicted Sale	Actual-Predicted (Residual)	Predicted/Actual (Ratio)	Percent Error
3	22.5	29.6	-7.1	1.32	31.62
23	26.0	38.7	-12.7	1.49	48.89
26	78.5	106.2	-27.7	1.35	35.29
47	33.0	52.1	-19.1	1.58	57.73
53	48.0	67.8	-19.8	1.41	41.35
64	36.0	50.7	-14.7	1.41	40.81
75	30.0	44.6	-14.6	1.49	48.60

This report shows those rows that had a large percentage prediction error. These are the parcels that were poorly predicted by the model. You should analyze them to determine if there is some explanation as to why they were not fit well. You may find that the explanation is as simple as an error in data entry. It may be worthwhile to rerun the analysis without these rows, especially if there is a reasonable explanation as to why they did not fit the pattern shown by most of the data.

Note that the actual cutoff value for inclusion on this report is set in the Exception Percentage box under the Reports tab.

Predicted Values and Residuals Section

Predicted Values and Residuals Section

Row No.	Actual Sale	Predicted Sale	Actual-Predicted (Residual)	Predicted/Actual (Ratio)	Percent Error
1	26.0	30.5	-4.5	1.17	17.38
2	53.0	43.8	9.2	0.83	17.31
3	22.5	29.6	-7.1	1.32	31.62
4	85.0	74.3	10.7	0.87	12.57
5	48.0	44.2	3.8	0.92	7.87
6	37.5	36.5	1.0	0.97	2.53
7	40.0	44.9	-4.9	1.12	12.27
8	42.0	42.9	-0.9	1.02	2.20
9	43.0	43.2	-0.2	1.00	0.42
10	57.0	57.0	0.0	1.00	0.02
.
.
.
75	30.0	44.6	-14.6	1.49	48.60
76	82.0	82.4	-0.4	1.00	0.44
77		86.3			
78		44.6			
79		82.4			

This report shows the actual and predicted sales as well as various measures of their disagreement. Assessors commonly study the Ratio and the Percent Error for individual predictions to determine the goodness of a mass appraisal.