Chapter 315

Nonlinear Regression

Introduction

Multiple regression deals with models that are linear in the parameters. That is, the multiple regression model may be thought of as a weighted average of the independent variables. A *linear* model is usually a good first approximation, but occasionally, you will require the ability to use more complex, nonlinear, models. Nonlinear regression models are those that are not linear in the parameters. Examples of nonlinear equations are:

$$Y = A + B \, EXP(-CX)$$

$$Y = (A + BX)/(1 + CX)$$

$$Y = A + B/(C + X)$$

This program estimates the parameters in nonlinear models using the Levenberg-Marquardt nonlinear least-squares algorithm as presented in Nash (1987). We have implemented Nash's MRT algorithm with numerical derivatives. This has been a popular algorithm for solving nonlinear least squares problems, since the use of numerical derivatives means you <u>do</u> <u>not</u> have to supply program code for the derivatives.

Starting Values

Many people become frustrated with the complexity of nonlinear regression after dealing with the simplicity of multiple linear regression. Perhaps the biggest nuisance with the algorithm used in this program is the need to supply bounds and starting values. The convergence of the algorithm depends heavily upon supplying appropriate starting values.

Sometimes you will be able to use zeros or ones as starting values, but often you will have to come up with better values. One accepted method for obtaining a good set of starting values is to estimate them from the data. We will show you how this is done with the example that we will be using throughout this chapter.

Suppose you have 44 observations on X and Y (the data are shown below). Suppose further that you want to fit the specific nonlinear model:

$$Y = A + (0.49 - A) \, Exp \, (-B(X - 8)).$$

Since there are two unknown parameters, A and B, we select two observations. To make the estimates as representative as possible, we select observations from each end of the range of X values. The two observations we select are (10, 0.48) and (42, 0.39). Putting these two observations into our model yields two equations with two unknowns:

(1) $0.48 = A + (0.49 - A) \, Exp \, (-B(10 - 8))$

(2) $0.39 = A + (0.49 - A) \, Exp \, (-B(42 - 8))$.

Solving (1) for B yields

(3) $B = \{log((0.48 - A)/(0.49 - A)\}/(-2)$.

Putting this result into the second equation yields

(4) $\{(0.39 - A)/(0.49 - A)\} = \{(0.48 - A)/(0.49 - A)\}^{17}$.

These equations appear difficult, but since we are only after starting values, we can analyze them for possible values of A and B. From (3), we see that A must be less than 0.48 and greater than 0. Suppose we pick a number in this range, say 0.1. Next, using (3), we calculate B as 0.013. These are our starting values. As a review:

1.   Select one data value for each parameter.

2.   Plug the selected data values into the model and solve for the parameters. If the model is too difficult, analyze the resulting equations for possible ranges of each parameter.

3.   Try these starting values in the program. These values need not be too accurate, just in the ball park.

Assumptions and Limitations

Usually, nonlinear regression is used to estimate the parameters in a nonlinear model without performing hypothesis tests. In this case, the usual assumption about the normality of the residuals is not needed. Instead, the main assumption needed is that the data may be well represented by the model.

Data Structure

The data are entered in one dependent variable and one or more independent variables. An example of data appropriate for this procedure, taken from page 476 of Draper and Smith (1981), is shown below. These data are contained in the DS746 dataset. In this example, the dependent variable (Y) is the proportion of available chlorine in a certain quantity of chlorine solution and the independent variable (X) is the length of time in weeks since the product was produced. When the product is produced, the proportion of chlorine is 0.50. During the 8 weeks that it takes to reach the consumer, the proportion declines to 0.49. The hypothesized model for predicting Y from X is

$$Y = A + (0.49 - A)\ EXP(-\ B(X-8)) + e.$$

Here, A and B are the parameters and e is the error or residual.

DS476 dataset

Row	X	Y	Row	X	Y
1	8	0.49	23	22	0.41
2	8	0.49	24	22	0.40
3	10	0.48	25	24	0.42
4	10	0.47	26	24	0.40
5	10	0.48	27	24	0.40
6	10	0.47	28	26	0.41
7	12	0.46	29	26	0.40
8	12	0.46	30	26	0.41
9	12	0.45	31	28	0.41
10	12	0.43	32	28	0.40
11	14	0.45	33	30	0.40
12	14	0.43	34	30	0.40
13	14	0.43	35	30	0.38
14	16	0.44	36	32	0.41
15	16	0.43	37	32	0.40
16	16	0.43	38	34	0.40
17	18	0.46	39	36	0.41
18	18	0.45	40	36	0.38
19	20	0.42	41	38	0.40
20	20	0.42	42	38	0.40
21	20	0.43	43	40	0.39
22	22	0.41	44	42	0.39

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.

Procedure Options

This section describes the options available in this procedure.

Model Tab and Parameters - Cont Tab

These panels specify the model and variables used in the analysis.

Dependent Variable

Y: Dependent Variable

Specifies a single dependent (Y) variable from the current database.

Options

Alpha Level

The value of alpha for the asymptotic confidence limits of the parameter estimates. Usually, this number will range from 0.1 to 0.001. A common choice for alpha is 0.05, but this value is a legacy from the age before computers when only printed tables were available. You should determine a value appropriate for your particular study.

Model

Model

This box contains the nonlinear equation.

This expression is made up of

1. Symbols: +, -, *, /, ^, <, >, =, (, and).

2. Functions:

   ABS(X)   Absolute value of X

   ASN(X)   Arc sine of X

   ATN (X)   Arc tangent of X

   COS(X)   Cosine of X

   EXP(X)   Exponential of X

   INT(X)   Integer part of X

   LN(X)   Log base e of X

   LOG(X)   Log base 10 of X

   SGN(X)   Sign of X

   SIN(X)   Sine of X

315-3

SQR(X)   Square root of X

TAN(X)   Tangent of X

3. One or more variables referenced by name.

4. Parameters, which are defined below.

5. Constants.

The syntax of the model expression follows that of the variable transformations, so we will not go into syntax here, but refer you to the Variable Transformations chapter. Note that only a subset of the functions available as transformations are also available here.

Model Parameters

Parameter

The model may contain up to thirty parameters. Each parameter used in the model must be defined in this section by entering a name, bounds, and starting value.

The parameter name is any combination of letters and numbers, except that the name must begin with a letter. You should not use symbols in the parameter name. All letters are converted to upper case internally, so it does not matter whether you use upper or lower case. The name cannot be one of the internal mathematical functions like SIN or TAN, as this will confuse the function parser. Also, the parameter name should not be the same as a variable name.

The name may be as long as you want, but, for readability, you should keep it short.

The model may contain up to thirty parameters.

Min Start Max

The minimum, starting value, and maximum are entered in this box. The three values must be separated with blanks or commas.

- **Minimum**

  This is the smallest value that the parameter can take on. The algorithm searches for a value between this and the Maximum Value. If you want to search in an unlimited range, enter a large negative number such as **-1E9**, which is -1000000000.

  Since this is a search algorithm, the narrower the range that you search in, the quicker the process will converge.

  Care should be taken to specify minima and maxima that keep calculations in range. Suppose, for example, that your equation includes the expression LOG(B*X) and that values of X are positive. Since you cannot take the logarithm of zero or a negative number, you should set the minimum of B as a positive number. This will insure that the estimation procedure will not fail because of impossible calculations.

- **Starting Value**

  This is the beginning value of the parameter. The algorithm searches for a value between the Minimum Value and the Maximum Value, beginning with this number. The closer this value is to the final value, the quicker the algorithm will converge.

  Although specific instructions for finding starting values are given at the beginning of this chapter, we would like to make the following suggestions here.

  1. Make sure that the starting values you supply are legitimate. For example, if the model you were estimating included the phrase 1/B, you would not want to start with B=0.

2. Before you go to a lot of effort, make a few trial runs using starting values of 0.0, 0.5, and 1.0. Often, one of these values will converge.

3. If you have a large number of observations, take a small sample of observations from your original database and work with this subset database. When you find a set of starting values that converges on this subset database, use the resulting parameter estimates as starting values with the complete database. Since nonlinear regression is iterative and each iteration must pass through the complete database, this can save a considerable amount of time while you are searching for starting values.

- **Maximum**

  This is the largest value that the parameter can take on. The algorithm searches for a value between the Minimum Value and this value, beginning at the Starting Value. If you want to search in an unlimited range, enter a large positive number such as **1E9**, which is 1000000000.

  Since this is a search algorithm, the narrower the range that you search in, the quicker the process will converge.

  Care should be taken to specify minima and maxima that keep calculations in check. Suppose, for example, that your equation includes the expression LOG(B*X) and that values of X are negative. Since you cannot take the logarithm of zero or a negative number, you should set the maximum of B as a negative number near zero. This will insure that the estimation procedure will not fail because of impossible calculations.

Options Tab

Options

Lambda

This is the starting value of the lambda parameter as defined in Marquardt's procedure. We recommend that you do not change this value unless you are very familiar with both your model and the Marquardt nonlinear regression procedure. Changing this value will influence the speed at which the algorithm converges.

Nash Phi

Nash supplies a factor he calls *phi* for modifying lambda. When the residual sum of squares is large, increasing this value may speed convergence.

Lambda Inc.

This is a factor used for increasing lambda when necessary. It influences the rate at which the algorithm converges.

Lambda Dec.

This is a factor used for decreasing lambda when necessary. It also influences the rate at which the algorithm converges.

Max Iterations

This sets the maximum number of iterations before the program aborts. If the starting values you have supplied are not appropriate or the model does not fit the data, the algorithm may diverge. Setting this value to an appropriate number (say 50) causes the algorithm to abort after this many iterations.

Zero

This is the value used as zero by the least squares algorithm. To remove the effects of rounding error, values lower than this value are reset to zero. If unexpected results are obtained, try using a smaller value, such as 1E-16. Note that 1E-5 is an abbreviation for the number 0.00001.

Reports Tab

The following options control which reports and plots are displayed.

Select Additional Reports

Iteration Report ... Predicted Value and Residual Report

Each of these options specifies whether the indicated report is displayed.

Report Options

Precision

Specify the precision of numbers in the report. Single precision will display seven-place accuracy, while the double precision will display thirteen-place accuracy. Note that all reports are formatted for single precision.

Variable Names

This option lets you select whether to display variable names, variable labels, or both.

Plots Tab

Various plots may be displayed to help you validate the assumptions of your regression analysis as well as investigate the fit of your estimated equation. The actual uses of these plots will be described later. Each of these plots includes the following options.

Select Plots

Probability Plot ... Residuals vs X Plot

Specify whether the indicated plot is displayed. Click the plot format button to change the plot settings.

Storage Tab

The predicted values and residuals may be stored on the current database for further analysis. This group of options lets you designate which statistics (if any) should be stored and which variables should receive these statistics. The selected statistics are automatically stored to the current database while the program is executing.

Note that the variables you specify must already have been named on the current database.

Note that existing data is replaced. Be careful that you do not specify variables that contain important data.

Data Storage Variables

Predicted Values

The predicted (Yhat) values.

Lower Prediction Limit

The lower confidence limit of the predicted value.

Upper Prediction Limit

The upper confidence limit of the predicted value.

Residuals

The residuals (Y-Yhat).

Example 1 – Nonlinear Regression Analysis

This section presents an example of how to run a nonlinear regression analysis of the data that was presented above in the Data Structure section. In this example, we will fit the model

$$Y = A + (0.49 - A)\ EXP(-\ B(X-8))$$

to the data contained in the variables Y and X on the database DS476.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Nonlinear Regression window.

1   Open the DS476 dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **DS476.NCSS**.
- Click **Open**.

2   Open the Nonlinear Regression window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Nonlinear Regression** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- On the Nonlinear Regression window, select the **Model tab**.
- Double-click in the **Y: Dependent Variable** box. This will bring up the variable selection window.
- Select **Y** from the list of variables and then click **Ok**. "Y" will appear in the Y: Dependent Variable box.
- In the **Model** box, enter **A+(0.49-A)*EXP(-B*(X-8))**. Note that A and B are parameters to be defined below, X is a variable of the dataset, and EXP is the name of a function.
- Enter **A** in the **first Parameter** box.
- Enter **0 0.1 1** in the first **Min Start Max** box.
- Enter **B** in the **second Parameter** box.
- Enter **0 0.013 1** in the **second Min Start Max** box.

4   Specify the reports.
- Select the **Reports tab**.
- Check all reports and plots.

5   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Minimization Phase Section

Minimization Phase Section

Itn No.	Error Sum of Squares	Lambda	A	B
0	0.01643321	4E-05	0.1	0.013
Stepsize reduced to 0.6032159 by bounds.				
Stepsize reduced to 0.8185954 by bounds.				
1	0.0147339	0.016	0.1464944	1.375224E-02
2	1.461316E-02	0.0064	0.2331648	1.792601E-02
3	1.278996E-02	0.0256	0.2486083	2.104608E-02
4	1.218322E-02	0.01024	0.3052482	2.783621E-02
5	0.0102341	0.04096	0.3141184	3.271746E-02
6	9.077431E-03	0.016384	0.3472015	0.0425624
7	7.713023E-03	0.065536	0.3519653	4.887892E-02
8	6.631856E-03	0.0262144	0.3685163	6.024325E-02
9	5.852748E-03	1.048576E-02	0.386985	8.150943E-02
10	5.045347E-03	4.194304E-03	0.3904939	9.880718E-02
11	5.001693E-03	1.677722E-03	0.390121	0.1015279
12	5.00168E-03	6.710886E-04	0.39014	0.101633
Convergence criterion met.				

This report displays the error (residual) sum of squares, lambda, and parameter estimates for each iteration. It allows you to observe the algorithm's progress toward the solution.

Model Estimation Section

Model Estimation Section

Parameter Name	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
A	0.39014	5.033759E-03	0.3799815	0.4002985
B	0.101633	1.336168E-02	7.466801E-02	0.1285979

Model	$Y = A+(0.49-A)*EXP(-B*(X-8))$
R-Squared	0.873375
Iterations	12
Estimated Model	
$(.39014)+(0.49-(.39014))*EXP(-(.101633)*((X)-8))$	

This section reports the parameter estimates.

Parameter Name

The name of the parameter whose results are shown on this line.

Parameter Estimate

The estimated value of this parameter.

Asymptotic Standard Error

An estimate of the standard error of the parameter based on asymptotic (large sample) results.

Lower 95% C.L.

The lower value of a 95% confidence limit for this parameter. This is a large sample (at least 25 observations for each parameter) confidence limit.

Upper 95% C.L.

The upper value of a 95% confidence limit for this parameter. This is a large sample (at least 25 observations for each parameter) confidence limit.

Model

The model that was estimated. Use this to double check that the model estimated was what you wanted.

R-Squared

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:

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

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.

Iterations

The number of iterations that were completed before the nonlinear algorithm terminated. If the number of iterations is equal to the Maximum Iterations that you set, the algorithm did not converge, but was aborted.

Estimated Model

This expression displays the estimated nonlinear-regression model. It is displayed in this format so that it may be copied to the clipboard and used elsewhere. For example, you could copy this expression here and paste it as a Variable Transformation.

Analysis of Variance Table

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square
Mean	1	7.9475	7.9475
Model	2	7.981998	3.990999
Model (Adjusted)	1	3.449832E-02	3.449832E-02
Error	42	5.00168E-03	1.190876E-04
Total (Adjusted)	43	0.0395	
Total	44	7.987	

The section presents an analysis of variance table.

Source

The labels of the various sources of variation.

DF

The degrees of freedom.

Sum of Squares

The sum of squares associated with this term. Note that these sums of squares are based on Y, the dependent variable. Individual terms are defined as follows:

Mean	The sum of squares associated with the mean of Y. This may or may not be a part of the model. It is presented since it is the amount used to adjust the other sums of squares.
Model	The sum of squares associated with the model.

Model (Adjusted)	The model sum of squares minus the mean sum of squares.
Error	The sum of the squared residuals. This is often called the sum of squares error or just SSE.
Total	The sum of the squared Y values.
Total (Adjusted)	The sum of the squared Y values minus the mean sum of squares.

Mean Square

The sum of squares divided by the degrees of freedom. The Mean Square for Error is an estimate of the underlying variation in the data.

Asymptotic Correlation Matrix of Parameters

Asymptotic Correlation Matrix of Parameters

	A	B
A	1.000000	0.887330
B	0.887330	1.000000

This report displays the asymptotic correlations of the parameter estimates. When these correlations are high (absolute value greater than 0.95), the precision of the parameter estimates is suspect.

Predicted Values and Residuals Section

Predicted Values and Residuals Section

Row No.	Actual Y	Predicted Y	Lower 95.0% C.L.	Upper 95.0% C.L.	Residual
1	0.49	0.49	0.4679772	0.5120228	0
2	0.49	0.49	0.4679772	0.5120228	0
3	0.48	0.4716319	0.4494232	0.4938406	0.00836813
4	0.47	0.4716319	0.4494232	0.4938406	-0.00163187
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	

This section shows the values of the residuals and predicted values. If you have observations in which the independent variables were given, but the dependent (Y) variable is missing, a predicted value will be generated and displayed in this report.
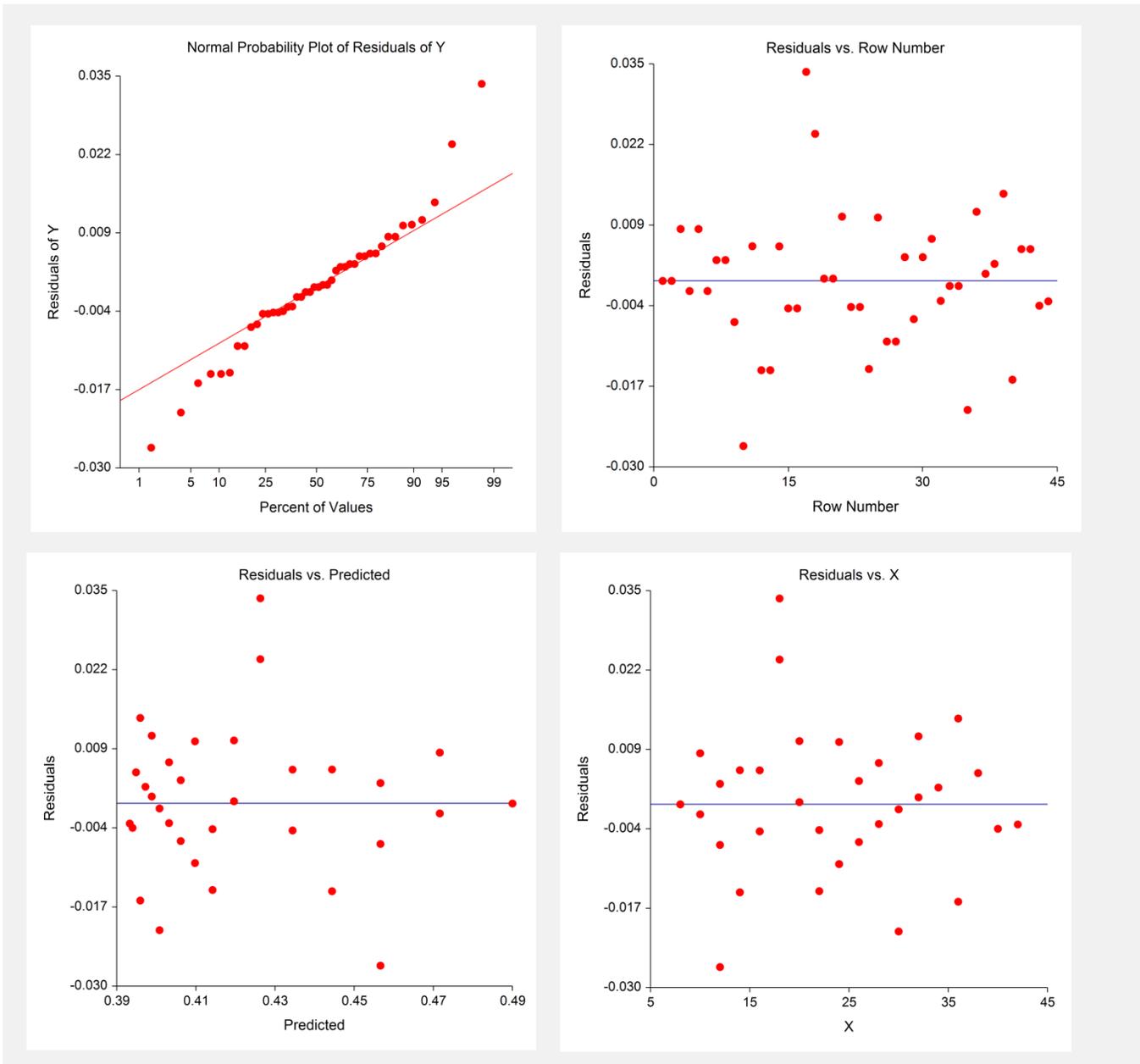
Residual Plots

Normal Probability Plot

If the residuals are normally distributed, the data points of the normal probability plot will fall along a straight line. Major deviations from this ideal picture reflect departures from normality. Stragglers at either end of the normal probability plot indicate outliers, curvature at both ends of the plot indicates long or short distributional tails, convex or concave curvature indicates a lack of symmetry, and gaps, plateaus, or segmentation in the normal probability plot may require a closer examination of the data or model. We do not recommend that you use this diagnostic with small sample sizes.

Residual versus Predicted Plot

This plot should always be examined. The preferred pattern to look for is a point cloud or a horizontal band. A wedge or bowtie pattern is an indicator of nonconstant variance. A sloping or curved band signifies inadequate specification of the model. A sloping band with increasing or decreasing variability could suggest nonconstant variance and inadequate specification of the model.

Residual versus Independent Variable(s) Plot

This is a scatter plot of the residuals versus each independent variable. Again, the preferred pattern is a rectangular shape or point cloud. Any nonrandom pattern may require a redefining of the model.

Predicting for New Values

You can use your model to predict Y for new values of the independent variables. Here is how. Add new rows to the bottom of your database containing the values of the independent variable(s) that you want to create predictions from. Leave the dependent 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 Y variable is available.