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:

$$Y = A + Be^{-CX}$$

$$Y = \frac{(A + BX)}{(1 + CX)}$$

$$Y = A + \frac{B}{(C + X)}$$

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 do not 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:

$$Y = A + (0.49 - A)e^{-B(X-8)}.$$

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. $0.48 = A + (0.49 - A)e^{-B(10-8)}$

2. $0.39 = A + (0.49 - A)e^{-B(42-8)}$

Solving the first equation for B yields

3. $B = \frac{\ln\left(\frac{0.48-A}{0.49-A}\right)}{-2}$

Putting this result into the second equation yields

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

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 ballpark.

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

$$Y_i = A + (0.49 - A)e^{-B(X_i - 8)} + e_i$$

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

DS476 Dataset

Row	X	Y
1	8	0.49
2	8	0.49
3	10	0.48
4	10	0.47
5	10	0.48
6	10	0.47
7	12	0.46
8	12	0.46
9	12	0.45
10	12	0.43
11	14	0.45
12	14	0.43
13	14	0.43
14	16	0.44
15	16	0.43
16	16	0.43
17	18	0.46
18	18	0.45
19	20	0.42
20	20	0.42
21	20	0.43
22	22	0.41

Row	X	Y
23	22	0.41
24	22	0.40
25	24	0.42
26	24	0.40
27	24	0.40
28	26	0.41
29	26	0.40
30	26	0.41
31	28	0.41
32	28	0.40
33	30	0.40
34	30	0.40
35	30	0.38
36	32	0.41
37	32	0.40
38	34	0.40
39	36	0.41
40	36	0.38
41	38	0.40
42	38	0.40
43	40	0.39
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.

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

$$Y = A + (0.49 - A)e^{-B(X-8)}.$$

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

Setup

To run this example, complete the following steps:

1 Open the DS476 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **DS476** and click **OK**.

2 Specify the Nonlinear Regression procedure options

- Find and open the **Nonlinear Regression** 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.

Model Tab

Y Dependent Variable....................................**Y**
Model..**A+(0.49-A)*EXP(-B*(X-8))** (Note that A and B are parameters to be defined below, X is a variable in the dataset, and EXP is the name of a function.)
Parameter 1...**A**
Min Start Max...**0 0.1 1**
Parameter 2...**B**
Min Start Max...**0 0.013 1**

Reports Tab

All Reports and Plots ....**Checked**

3 Run the procedure

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

Iterations (Minimization Phase)

Iterations (Minimization Phase)

Iteration Number	Error Sum Lambda	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	0.01375224
2	0.01461316	0.0064	0.2331648	0.01792601
3	0.01278996	0.0256	0.2486083	0.02104608
4	0.01218322	0.01024	0.3052482	0.02783621
5	0.0102341	0.04096	0.3141184	0.03271746
6	0.009077431	0.016384	0.3472015	0.0425624
7	0.007713023	0.065536	0.3519653	0.04887892
8	0.006631856	0.0262144	0.3685163	0.06024325
9	0.005852748	0.01048576	0.386985	0.08150943
10	0.005045347	0.004194304	0.3904939	0.09880718
11	0.005001693	0.001677722	0.390121	0.1015279
12	0.00500168	0.0006710886	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

Model Estimation

Model: Y = A+(0.49-A)*EXP(-B*(X-8))

			95% Confidence Interval Limits	
Parameter Name	Estimate	Asymptotic Standard Error	Lower	Upper
A	0.39014	0.005033759	0.3799815	0.4002985
B	0.101633	0.01336166	0.07466805	0.1285979

Model Estimation Information

R-Squared	0.873375
Iterations	12

Estimated Model

(0.39014)+(0.49-(0.39014))*EXP(-(0.101633)*((X)-8))

This section reports the parameter estimates.

Model

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

Parameter Name

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

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 and Upper 95% Confidence Interval Limits

The lower and upper limits of a 95% confidence interval for this parameter. This is a large sample (at least 25 observations for each parameter) confidence interval.

R-Squared

There is no direct R^2 defined for nonlinear regression. This is a pseudo R^2 constructed to approximate the usual R^2 value used in multiple regression. We use the following generalization of the usual R^2 formula:

$$R^2 = \frac{ModelSS - MeanSS}{TotalSS - MeanSS}$$

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^2 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^2 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^2 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	7.984499
Model (Adjusted)	1	0.03449832	0.03449832
Error	42	0.00500168	0.0001190876
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

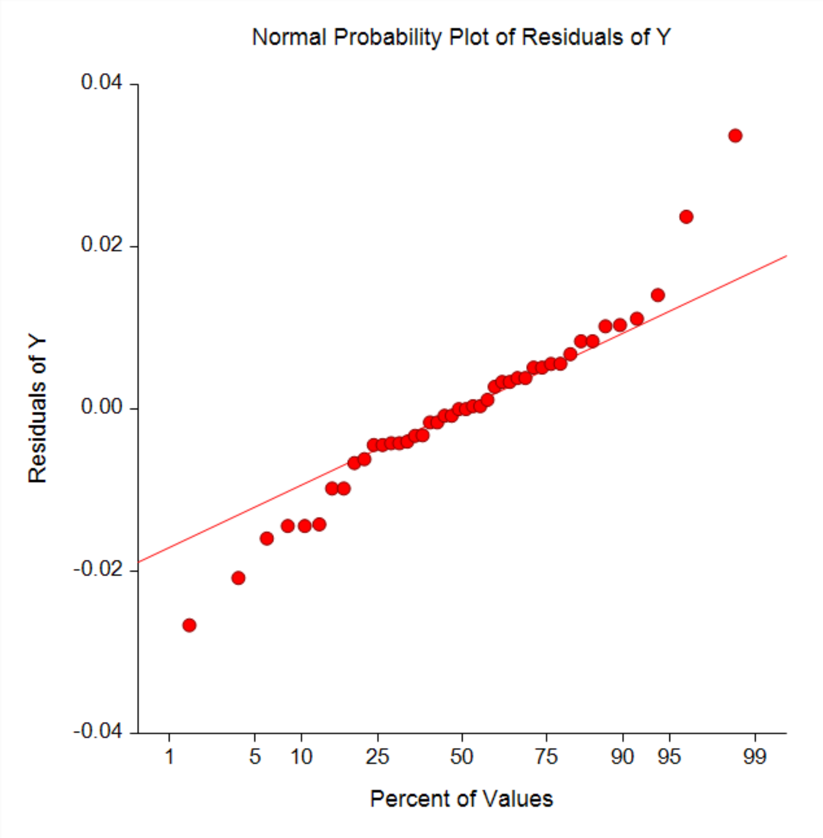
Predicted Values and Residuals

	Y		95% Confidence Interval Limits		
Row Number	Actual	Predicted	Lower	Upper	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
5	0.48	0.4716319	0.4494232	0.4938406	0.00836813
6	0.47	0.4716319	0.4494232	0.4938406	-0.00163187
7	0.46	0.4566424	0.4341762	0.4791085	0.003357648
8	0.46	0.4566424	0.4341762	0.4791085	0.003357648
9	0.45	0.4566424	0.4341762	0.4791085	-0.006642352
10	0.43	0.4566424	0.4341762	0.4791085	-0.02664235
11	0.45	0.44441	0.4217992	0.4670208	0.005590011
12	0.43	0.44441	0.4217992	0.4670208	-0.01440999
13	0.43	0.44441	0.4217992	0.4670208	-0.01440999
14	0.44	0.4344276	0.4117945	0.4570608	0.005572368
15	0.43	0.4344276	0.4117945	0.4570608	-0.004427632
16	0.43	0.4344276	0.4117945	0.4570608	-0.004427632
17	0.46	0.4262814	0.4037038	0.448859	0.03371858
18	0.45	0.4262814	0.4037038	0.448859	0.02371858
19	0.42	0.4196336	0.3971402	0.442127	0.0003663911
20	0.42	0.4196336	0.3971402	0.442127	0.0003663911
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

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.

Probability Plot of Residuals

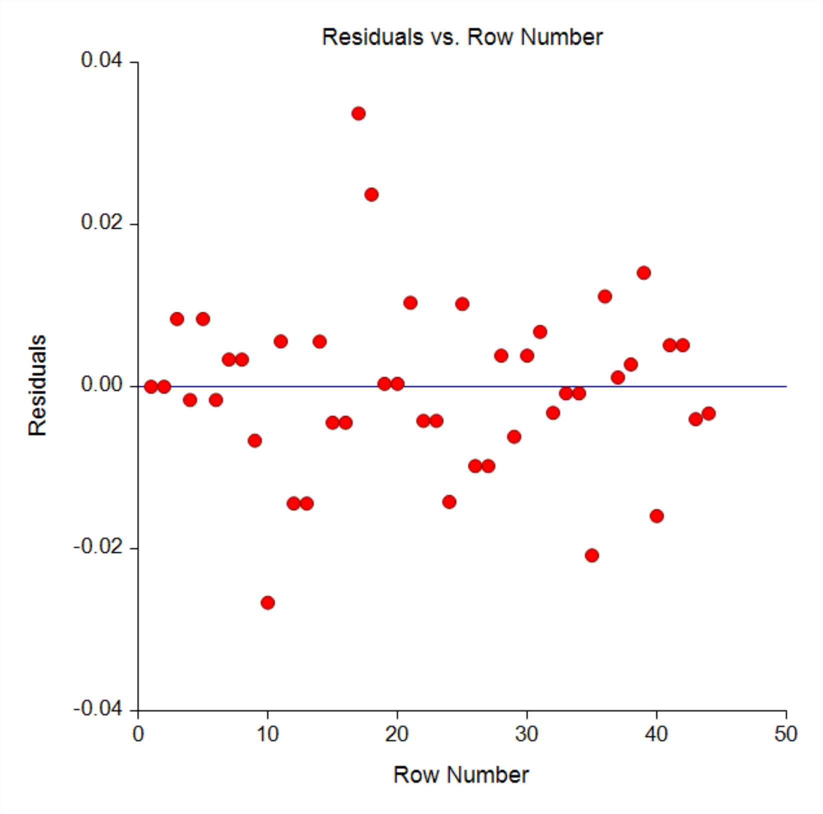
Probability Plot of Residuals



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.

Sequence Plot: Residuals vs Row Number

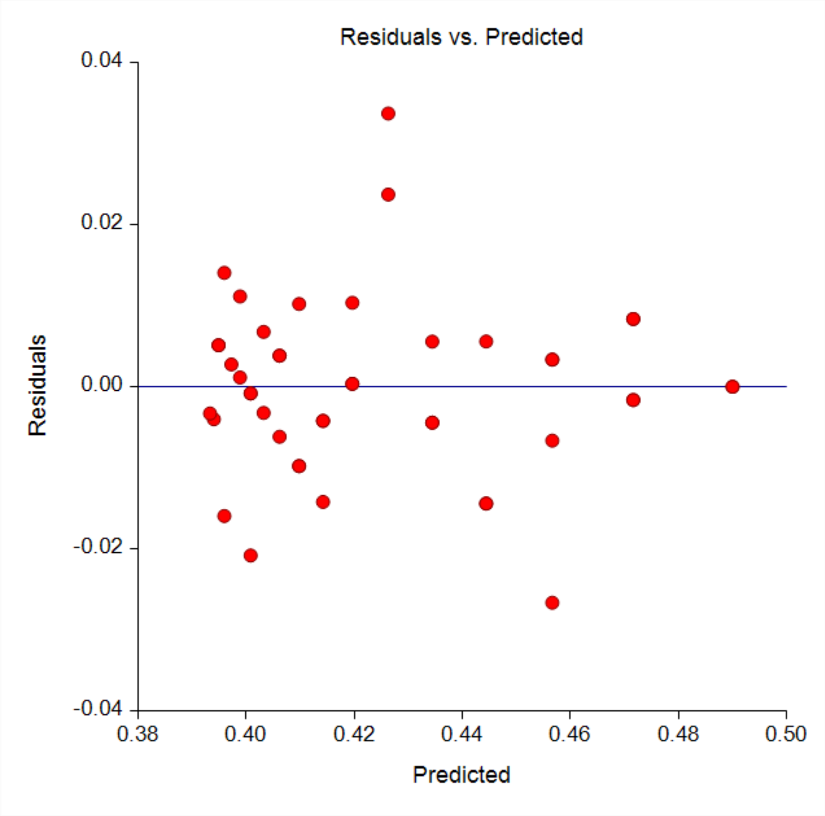
Sequence Plot: Residuals vs Row Number



When the row number can be equated to time period, this plot lets you see if there is a pattern across time.

Residuals vs Predicted Plot

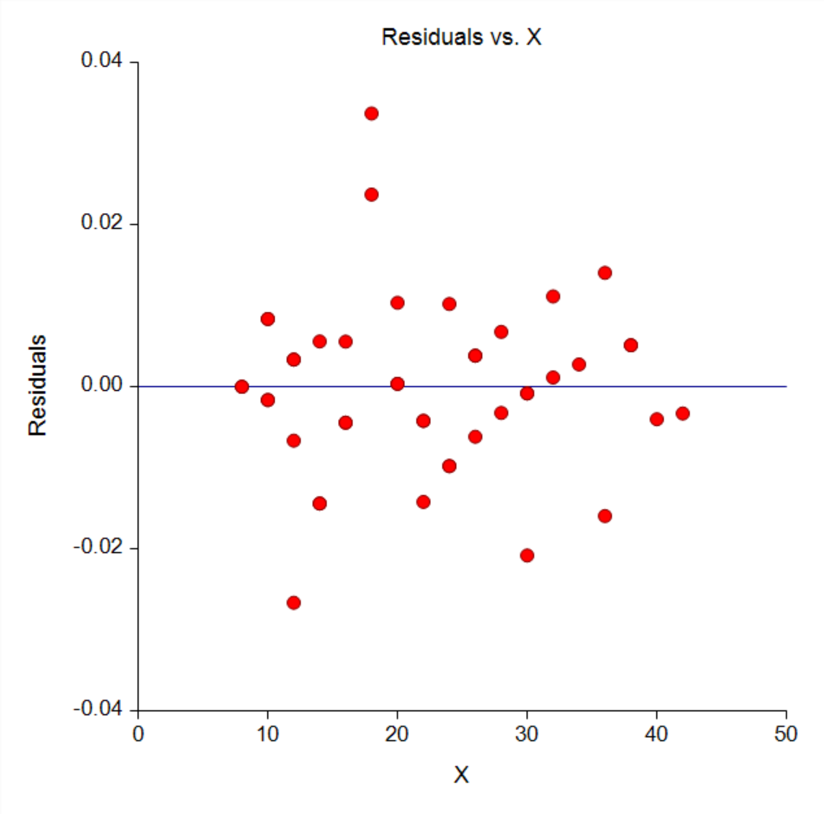
Residuals vs Predicted Plot



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.

Residuals vs X's Plot(s)

Residuals vs X's Plot(s)



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.