Chapter 380

Sum of Functions (of X) Model Fit – Y vs One X

Introduction

This program fits regression models of two types: models that are linear expressions of one variable and models that are the ratio of two linear expressions. They offer a much wider variety of curves than the usual polynomial models.

These models approximate many different curves. Since these are approximating curves and have no physical interpretation, care must be taken outside the range of the data. You must study the resulting model graphically to determine that the model behaves properly between data points.

The general form of the first type of model is:

$$g(Y) = A0 + A1f_1(X) + A2f_2(X) + A3f_3(X) + A4f_4(X) + A5f_5(X) + e$$

The general form of the second type of model is:

$$g(Y) = \frac{A0 + A1f_1(X) + A2f_2(X) + A3f_3(X) + A4f_4(X) + A5f_5(X)}{1 + B1h_1(X) + B2h_2(X) + B3h_3(X) + B4h_4(X) + B5h_5(X)} + e$$

where $f_i(X)$, $g(Y)$, and $h_i(X)$ are standard functions such as sin(X), ln(X+1), sqrt(X/2), etc. The A0, A1, …, B5 are constants (parameters) to be estimated from the data.

Starting Values

Starting values are determined by the program from the data. You do not have to supply starting values.

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 two variables: one dependent variable and one independent variable.

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.

Model – Numerator and Denominator Terms

The user may specify up to five terms for use as the numerator and/or denominator of the model. You do not have to have a denominator.

Function

There are eighteen possible transformations.

$f(z) = 1/(z^2)$	$f(z) = 1/z$	$f(z) = 1/\sqrt{z}$
$f(z) = \ln(z)$	$f(z) = \sqrt{z}$	$f(z) = z$ (none)
$f(z) = z^2 = z*z$	$f(z) = z^3$	$f(z) = z^4$
$f(z) = z^5$	$f(z) = \exp(z)$	$f(z) = \exp(-z)$
$f(z) = \sin(z)$	$f(z) = \cos(z)$	$f(z) = \tan(z)$
$f(z) = \sinh(z)$	$f(z) = \cosh(z)$	$f(z) = \tanh(z)$

where

$z = MX+A$; M and A are specified constants.

Example 1 – Fitting a Sum of Functions Model

This section presents an example of how to fit a sum of functions model. In this example, we will fit the model

$$Y = A0 + A1/(X + 0.5) + sin(X/2) + A3tanh(X)$$

to the variables Y and X of the FnReg1 dataset.

Setup

To run this example, complete the following steps:

1 Open the FnReg1 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **FnReg1** and click **OK**.

2 Specify the Sum of Functions (of X) Model Fit – Y vs One X procedure options

- Find and open the **Sum of Functions (of X) Model Fit – Y vs One X** 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

Y (Dependent) Variable ....**Y**
X (Independent) Variable................................**X**
Function 1 ....**1/X**
  Rescale X ....**Checked**
    Multiply by M...**1**
    Add A...**0.5**
Function 2...**sin(X)**
  Rescale X ....**Checked**
    Multiply by M...**0.5**
    Add A...**0**
Function 3...**tanh(X)**
  Rescale X ....**Unchecked**

3 Run the procedure

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

Model Estimation

Model Estimation

Rows Used: 100 of 100

Model: $Y = A0 + A11/(X+0.5) + A2\sin(0.5X) + A3\tanh(X)$
  Y: Y
  X: X

R^2:   0.956784

Coefficient			Asymptotic Standard Error	95% Confidence Interval Limits for the Coefficient	
Name	**Model Term**	**Estimate**		**Lower**	**Upper**
A0	Intercept	2.070158	1.075332	-0.06435942	4.204676
A1	1/(X+0.5)	4.885924	0.5628729	3.768631	6.003218
A2	sin(0.5X)	1.059697	0.06202012	0.9365882	1.182806
A3	tanh(X)	7.084624	0.9798195	5.139698	9.029551

Estimated Model (Double Precision)

Y =
(2.07015847235604 + (4.88592427829876)*1/(X+0.5) + (1.05969721661408)*sin((0.5*X)) +
(7.08462387494981)*tanh(X))

Coefficient Name

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

Coefficient Model Term

The model term that is multiplied by the regression coefficient.

Coefficient 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 for the Coefficient

The lower and upper limits of a 95% confidence interval for this coefficient. 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.

R²

When the model does not contain a denominator component then this is the usual R^2. When the model includes a denominator, there is no direct R^2 defined for nonlinear regression. In this case, 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:

$$R^2 = (ModelSS - MeanSS)/(TotalSS-MeanSS)$$

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.

Estimated Model

The model that was estimated with the parameters replaced by their estimated values. This expression may be copied and pasted as a variable transformation in the spreadsheet. This will allow you to predict Y for additional values of X.

Analysis of Variance

Analysis of Variance

Source	DF	Sum of Squares	Mean Square
Mean	1	10559.48	10559.48
Model	4	10690.45	10694.89
Model (Adjusted)	3	130.9726	43.65752
Error	96	5.915745	0.06162234
Total (Adjusted)	99	136.8883	
Total	100	10696.37	

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

	A0	A1	A2	A3
A0	1.000000	-0.972405	0.786167	-0.999209
A1	-0.972405	1.000000	-0.831952	0.964719
A2	0.786167	-0.831952	1.000000	-0.779440
A3	-0.999209	0.964719	-0.779440	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

Row Number	X	Y	Predicted Value	95% Prediction Interval Limits for the Predicted Value		Residual
				Lower	Upper	
1	0.5	10.16989	10.49218	9.92255	11.06181	-0.3222909
2	0.5959596	10.83415	10.62378	10.07584	11.17172	0.2103729
3	0.6919192	10.93412	10.77391	10.24289	11.30493	0.1602088
4	0.7878788	10.71519	10.92673	10.40745	11.44602	-0.2115456
5	0.8838384	11.25534	11.07237	10.56021	11.58453	0.182972
6	0.979798	11.5098	11.2051	10.69653	11.71368	0.3046989
7	1.075758	11.10627	11.32207	10.81472	11.82942	-0.2157996
8	1.171717	11.64065	11.42233	10.91487	11.9298	0.2183182
9	1.267677	11.22304	11.50616	10.99801	12.01431	-0.2831205
10	1.363636	11.52402	11.57457	11.06568	12.08347	-0.05055245

The section shows the values of the residuals and predicted values. If you have observations in which the independent variable is given, but the dependent (Y) variable was left blank, a predicted value and prediction limits will be generated and displayed in this report.

Function Plot(s)

Function Plot(s)

Model: A0 + A11/(X+0.5) + A2sin(0.5X) + A3tanh(X)
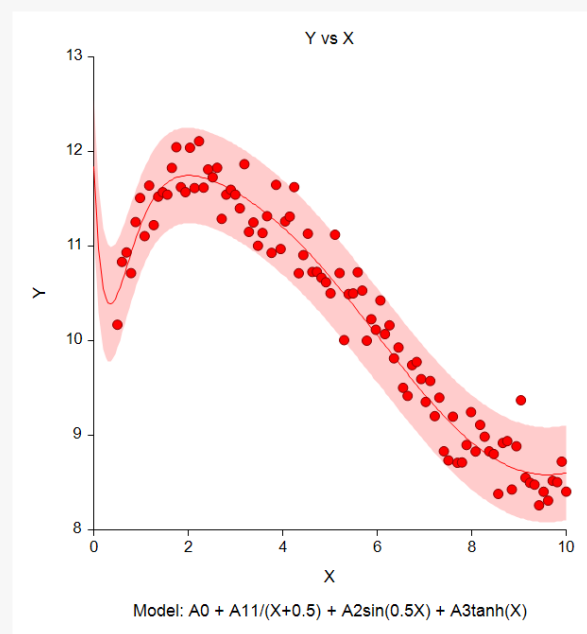
This plot displays the data along with the estimated function and prediction limits. It is useful in deciding if the fit is adequate and if the prediction limits are appropriate. In poorly fit models, we have found that it is often necessary to disable the prediction limits so that the data will show up. In these cases, the prediction limits may be so wide that the scale of the plot does not allow the data values to be separated.
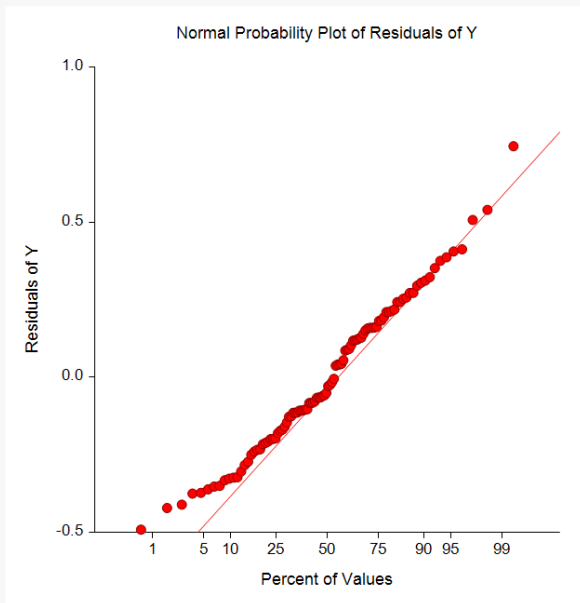
Residual Plot(s)

Residual Plot(s)
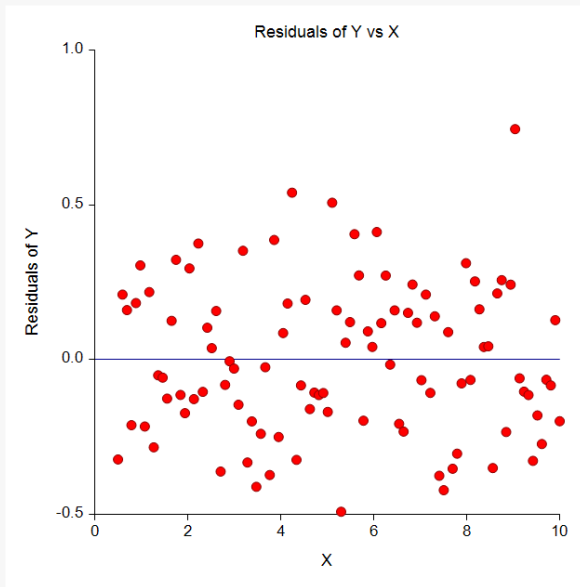
This is a scatter plot of the residuals versus the independent variable, X. The preferred pattern is a rectangular shape or point cloud. Any nonrandom pattern may require a redefining of the model.

Probability Plot(s)

Probability Plot(s)

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 or 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.

Predicting for New Values

You can use your model to predict Y for new values of X. Here's how. Add new rows to the bottom of your database containing the values of the independent variable that you want to create predictions for. 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, regardless of whether the Y variable is missing or not.