Chapter 380

Sum of Functions Models

Introduction

This program fits models that are the ratio of two linear expressions. The general form of a model is:

$$g(Y) = \frac{A0 + A1\,f_1(X) + A2\,f_2(X) + A3\,f_3(X) + A4\,f_4(X) + A5\,f_5(X)}{1 + B1h_1(X) + B2h_2(X) + B3h_3(X) + B4h_4(X) + B5h_5(X)} + e$$

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.

These models approximate many different curves. They offer a much wider variety of curves than the usual polynomial models.

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.

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.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

This panel specifies the variables used in the analysis.

Variables

Y (Dependent) Variable

Specifies a single dependent (*Y*) variable.

Y Transformation

Specifies a power transformation of the dependent variable. Available transformations are

Y'=1/(Y*Y), ***Y'=1/Y***, ***Y'=1/SQRT(Y)***, ***Y'=LN(Y)***, ***Y'=SQRT(Y)***, ***Y'=Y (none)***, and ***Y'=Y*Y***

X (Independent) Variable

Specifies a single independent (*X*) variable.

Model

Bias Correction

This option controls whether a bias-correction factor is applied when the dependent variable has been transformed. Check it to correct the predicted values for the transformation bias. Uncheck it to leave the predicted values unchanged. See the Introduction to Curve Fitting chapter for a discussion of the amount of bias and the bias correction procedures used.

Model – Numerator and Denominator Terms

These options 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

Select one of the eighteen possible transformations for this term.

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 constants that are supplied in the two options below.

Add (A)

X may be scaled using the equation z=MX+A. This option sets the value of A. If you want to ignore this option, set A=0.

Multiply (M)

X may be scaled using the equation z=MX+A. This option sets the value of M. If you want to ignore this option, set M=1.

Options Tab

The following options control the nonlinear regression algorithm.

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 nonlinear algorithm. Because of rounding error, values lower than this value are reset to zero. If unexpected results are obtained, you might 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 Reports

Iteration Report ... Residual Report

These options specify which reports are displayed.

Report Options

Alpha Level

The value of alpha for the asymptotic confidence limits of the parameter estimates and predicted values. 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 needs.

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 only.

Variable Names

Specify whether to use variable names or (the longer) variable labels in report headings.

Plots Tab

This section controls the plot(s) showing the data with the fitted function line overlain on top and the residual plots.

Select Plots

Function Plot with Actual Y ... Probability Plot with Transformed Y

These options specify which plots are 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 dataset while the program is executing.

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

Storage Columns

Store Predicted Values, Residuals, Lower Prediction Limit, and Upper Prediction Limit

The predicted (Yhat) values, residuals (Y-Yhat), lower 100(1-alpha) prediction limits, and upper 100(1-alpha) prediction limits may be stored in the columns specified here.

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 database.

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 Sum of Functions Models window.

1 Open the FnReg1 dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **FnReg1.NCSS**.
- Click **Open**.

2 Open the Sum of Functions Models window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Sum of Functions Models** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- On the Sum of Functions Models window, select the **Variables tab**.
- Double-click in the **Y Variable** box. This will bring up the variable selection window.
- Select **Y** from the list of variables and then click **Ok**.
- Double-click in the **X Variable** box. This will bring up the variable selection window.
- Select **X** from the list of variables and then click **Ok**.
- Under the Numerator Terms heading, select **1/z** in the **first Function** box.
- Under the Numerator Terms heading, enter **0.5** in the **first Add (A)** box.
- Under the Numerator Terms heading, select **SIN(z)** in the **second Function** box.
- Under the Numerator Terms heading, enter **0.5** in the **second Mult (M)** box.
- Under the Numerator Terms heading, select **TANH(z)** in the **third Function** box.

4 Specify the reports.

- On the Sum of Functions Models window, select the **Reports tab**.
- Check the **Residual Report** box. Leave all other reports and plots checked.

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 Lambda	Lambda	A0	A1	A2	A3
0	5.915745	0.00004	2.070158	4.885924	1.059697	7.084624

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. Since no denominator terms were selected, the model was solved on the first iteration using standard multiple linear regression.

Model Estimation Section

Model Estimation Section

Parameter Name	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
A0	2.070158	1.075332	-6.435943E-02	4.204676
A1	4.885924	0.5628729	3.768631	6.003218
A2	1.059697	6.202012E-02	0.9365882	1.182806
A3	7.084624	0.9798195	5.139698	9.029551

Dependent	Y
Independent	X
Model	Y=(A0+A1*(1/(X+.5))+A2*(SIN((.5*X)))+A3*(TANH(X))) / (1)
R-Squared	0.956784
Iterations	0

Estimated Model
(2.070158+(4.885924)*1/(X+.5)+(1.059697)*SIN((.5*X))+(7.084624)*TANH(X))

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. Note that the “/(1)” at the end emphasizes that there was no denominator specified.

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:

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

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

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

Analysis of Variance Table

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square
Mean	1	10559.48	10559.48
Model	4	10690.45	2672.613
Model (Adjusted)	3	130.9726	43.65752
Error	96	5.915745	6.162234E-02
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.

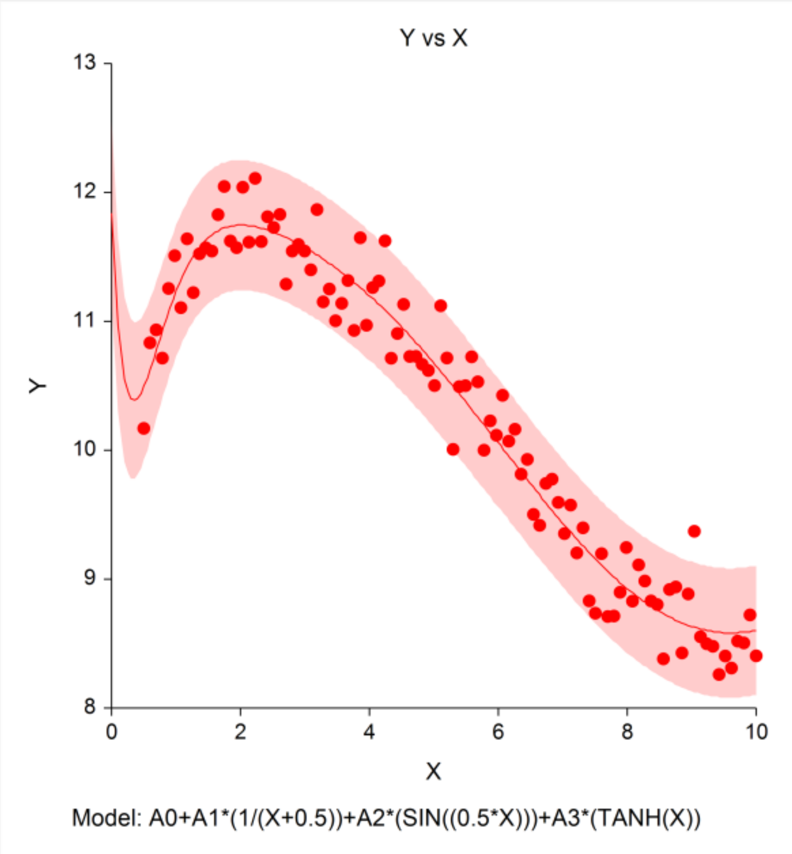
Predicted Values and Residuals Section

Predicted Values and Residuals Section

Row No.	X	Y	Predicted Value	Lower 95.0% Value	Upper 95.0% Value	Residual
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
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

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.

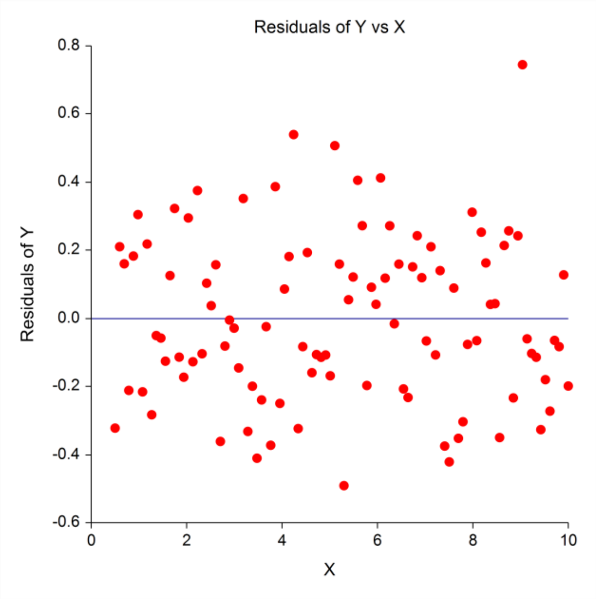
Function Plot(s)



This plot displays the data along with the estimated function and prediction limits. It is useful in deciding if the fit is adequate and the prediction limits are appropriate.

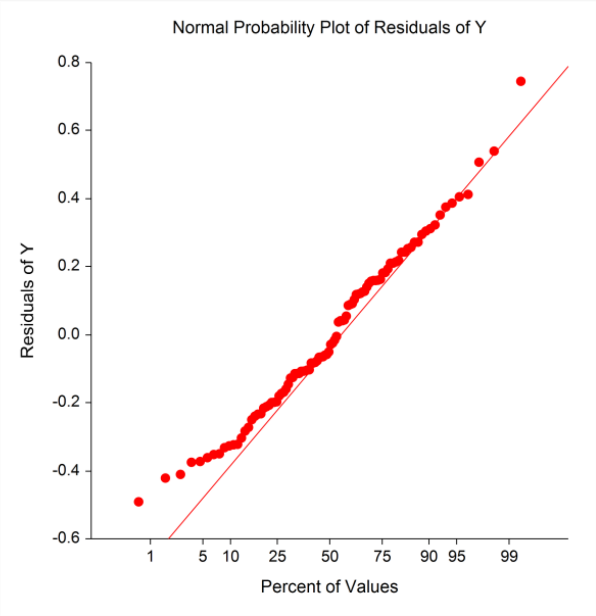
In poorly fit models, we have found that it is often necessary to disable the prediction limits so that the data will show up. In these cases, the prediction limits may be so wide that the scale of the plot does not allow the data values to be separated.

Residual Plot(s)



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)



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.