Chapter 376

Ratio of Polynomials Fit – Many Variables

Introduction

This program fits a model that is the ratio of two polynomials of up to fifth order. Instead of a single independent variable, these polynomials may involve up to four independent variables (U, V, W, and X). An example of this type of model is:

$$Y = \frac{B0 + B1\,X + B2\,X^2 + B3U + B4U^2}{1 + B5\,X + B6\,X^2 + B7U + B8U^2}$$

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.

Usually you would use the Multivariate Ratio of Polynomials Search procedure first to find an appropriate model and then fit that model with this program.

Starting Values

Starting values are determined by the program. 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 or more variables: one dependent variable and up to four independent variables.

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

Specify the variables on which to run 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*

U, V, W, X (Independent) Variable

Each option specifies an independent variable. At least one independent variable must be designated.

U, V, W, X Transformation

Specifies a power transformation of this independent variable. When the variable is referenced in the model, it refers to the transformed variable. Available transformations are

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

Model

These options specify the polynomials to be fit.

Numerator Terms

These options specify a list of terms that become the numerator polynomial of the model.

The syntax of these lists of terms follows these rules:

1. Individual terms may be listed as UiVj. If i or j is one, it may be omitted. For example, UV2X3 means (U)(V*V)(X*X*X) and U2 means U^2 which means U*U. A list of individual terms is formed by separating such terms with commas.

2. The **Oi** notation includes all terms of a particular order. The order is the sum of the exponents of the variables in a term. For example, the order of the term U2VW3 is six. If you had selected three variables and included "O2" in the list of terms, you would include the terms U3, V3, W3, U2V, U2W, V2W, UV2, VW2, and UVW in your model.

3. The **Si** notation includes all single variables to the power i. For example, if you had selected three variables and included "S2" in the list of terms, you would include the terms U, V, W, U2, V2, and W2 in your model.

4. The **Ei** notation includes all combinations of variables with at least one variable to the power i and none of the other variables to a power greater than i. For example, if you had selected three variables and included "E2" in the list of terms, you would include the terms U2, V2, W2, U2V, U2W, U2V2, U2W2, UV2, UW2, VW2, V2W, and V2W2 in your model.

5. The **Hi** notation includes all terms in the hierarchical model of order i. For example, if you had selected two variables and included "H2" in the list of terms, you would include the terms U, V, U2, V2, and UV in your model.

6. The **P** notation includes all simple paired terms. For example, if you had selected three variables and included "P" in the list of terms, you would include the terms UV, UW, and VW in your model.

7. The **T** notation includes all triplet terms. For example, if you had selected four variables and included "T" in the list of terms, you would include the terms UVW, UVX, UWX, and UWX in your model.

You can combine these notations however you like. If a term is specified twice, it will be included in the model only once. The order in which you specify terms is arbitrary. Examples are:

E2

U,V,E2,O1

O1,U2V2

Denominator Terms

These options specify a list of terms that become the denominator polynomial of the model. The syntax of these options follow the same rules as those given for Numerator Terms above.

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.

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 residual and probability plots.

Select Plots

Residual 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 to the current dataset 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 columns 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 Multivariate Ratio of Polynomials Model

This section presents an example of how to fit a multivariate ratio of polynomials model. In this example, we will fit a custom model to the variables Y, U, and X of the FnReg4 dataset. The numerator will include the terms SQRT(X), SQRT(U), and UX. The denominator will include the terms SQRT(UX), XU, and U.

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 Ratio of Polynomials Fit – Many Variables window.

1   Open the FnReg4 dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **FnReg4.NCSS**.
- Click **Open**.

2   Open the Ratio of Polynomials Fit – Many Variables window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Ratio of Polynomials Fit – Many Variables** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- On the Ratio of Polynomials Fit – Many Variables 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 **U Variable** box. This will bring up the variable selection window.
- Select **U** from the list of variables and then click **Ok**.
- Select **SQRT(z)** in the U Transformation box.
- 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**.
- Select **SQRT(z)** in the X Transformation box.
- Enter **U, X, U2X2** for **Numerator Terms** under **Model**.
- Enter **U2, UX, U2X2** for **Denominator Terms** under **Model**.

4   Specify the reports.
- On the Ratio of Polynomials Fit – Many Variables 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	B0	B1	B2	B3
0	0.0256219	0.00004	2.050828	0.7681349	-1.027562	1.102767
1	2.246299E-02	0.000016	2.002186	0.960939	-1.019122	1.408448
2	2.238311E-02	0.0000064	1.995856	0.9876834	-1.019637	1.488163
3	2.238232E-02	2.56E-06	1.99552	0.989055	-1.019808	1.495897
4	2.238232E-02	1.024E-06	1.99551	0.9890978	-1.019816	1.496198

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.

Model Estimation Section

Model Estimation Section

Parameter Name	Term	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
B0	Intercept	1.99551	0.0128966	1.970092	2.020928
B1	U	0.9890978	5.595055E-02	0.8788245	1.099371
B2	X	-1.019816	1.188279E-02	-1.043235	-0.9963957
B3	U2X2	1.496198	0.1342063	1.23169	1.760706
B4	u2	1.009521	2.531414E-02	0.9596294	1.059413
B5	ux	-1.081743	2.945009E-02	-1.139787	-1.0237
B6	u2x2	1.36935	0.1003615	1.171546	1.567153

R-Squared   0.993757
Iterations   4

Symbolic Model
Y = P1(U,X) / P2(U,X)
P1(U,X) = B0+B1*U+B2*X+B3*U2X2
P2(U,X) = 1+B4*U2+B5*UX+B6*U2X2
where
Y = Y
U = SQRT(U)
X = SQRT(X)

Estimated Model
((1.99551)+(.9890978)*(SQRT(U))-(1.019816)*(SQRT(X))+(1.496198)*(SQRT(U))^2
(SQRT(X))^2) / (1+(1.009521)(SQRT(U))^2-(1.081743)*(SQRT(U))*(SQRT(X))
 +(1.36935)*(SQRT(U))^2*(SQRT(X))^2)

This section reports the parameter estimates.

Parameter Name

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

Term

The name of the term in the model. Note that upper case letters are used for numerator terms and lower case letters are used for denominator terms.

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.

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:

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

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.

Symbolic Model

The expanded model that was fit. Any of the shortcut terms like O1 and E2 are replaced by the individual terms that they represent. Note that one list is presented for the numerator and one for the denominator. Any transformations that were applied are also listed.

Estimated Model

This is a copy of the symbolic model in which the parameter names have been replaced by their estimates. This expression may be used as a variable transformation by copying it and pasting it into the Variable Info section of the database.

Analysis of Variance Table

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square
Mean	1	675.2869	675.2869
Model	7	678.8495	96.97849
Model (Adjusted)	6	3.562525	0.5937542
Error	218	2.238232E-02	1.026712E-04
Total (Adjusted)	224	3.584907	
Total	225	678.8718	

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

|	B0	B1	B2	B3	B4	B5
B0	1.000000	-0.915182	-0.635169	-0.014330	-0.850024	-0.465655
B1	-0.915182	1.000000	0.505466	-0.065987	0.969403	0.577334
B2	-0.635169	0.505466	1.000000	-0.520124	0.301986	0.802626
B3	-0.014330	-0.065987	-0.520124	1.000000	0.093673	-0.751426
B4	-0.850024	0.969403	0.301986	0.093673	1.000000	0.405675
B5	-0.465655	0.577334	0.802626	-0.751426	0.405675	1.000000
B6	0.041458	-0.150335	-0.553984	0.991271	0.008270	-0.819380

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.	Y	Predicted Value	Lower 95.0% Value	Upper 95.0% Value	Residual
1	1.981996	1.992756	1.971384	2.014127	-1.075969E-02
2	2.028455	2.026659	2.006006	2.047311	1.796112E-03
3	2.027451	2.014322	1.993759	2.034884	1.312937E-02
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	

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

Probability Plot(s)
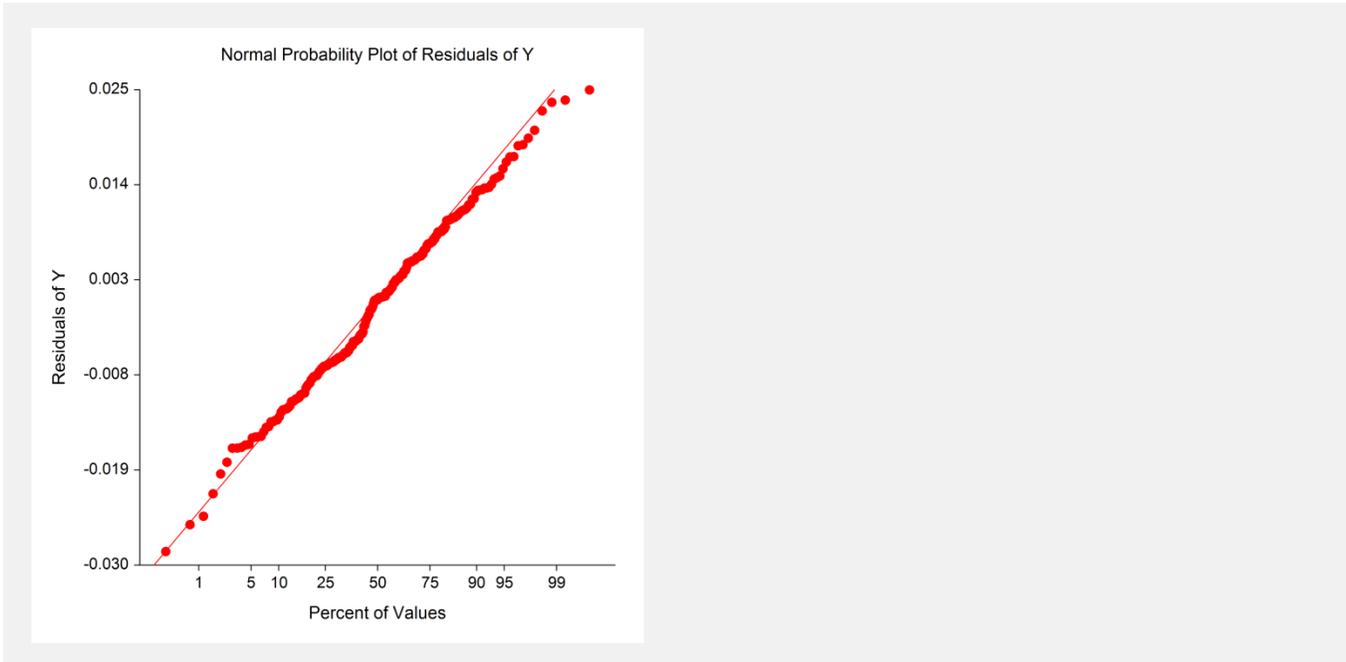
If the residuals are normally distributed, the data points of the normal probability plot will fall along a straight line. Major deviations from this ideal picture reflect departures from normality. Stragglers at either end of the normal probability plot indicate outliers, curvature at both ends of the 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.

Residual Plot(s)
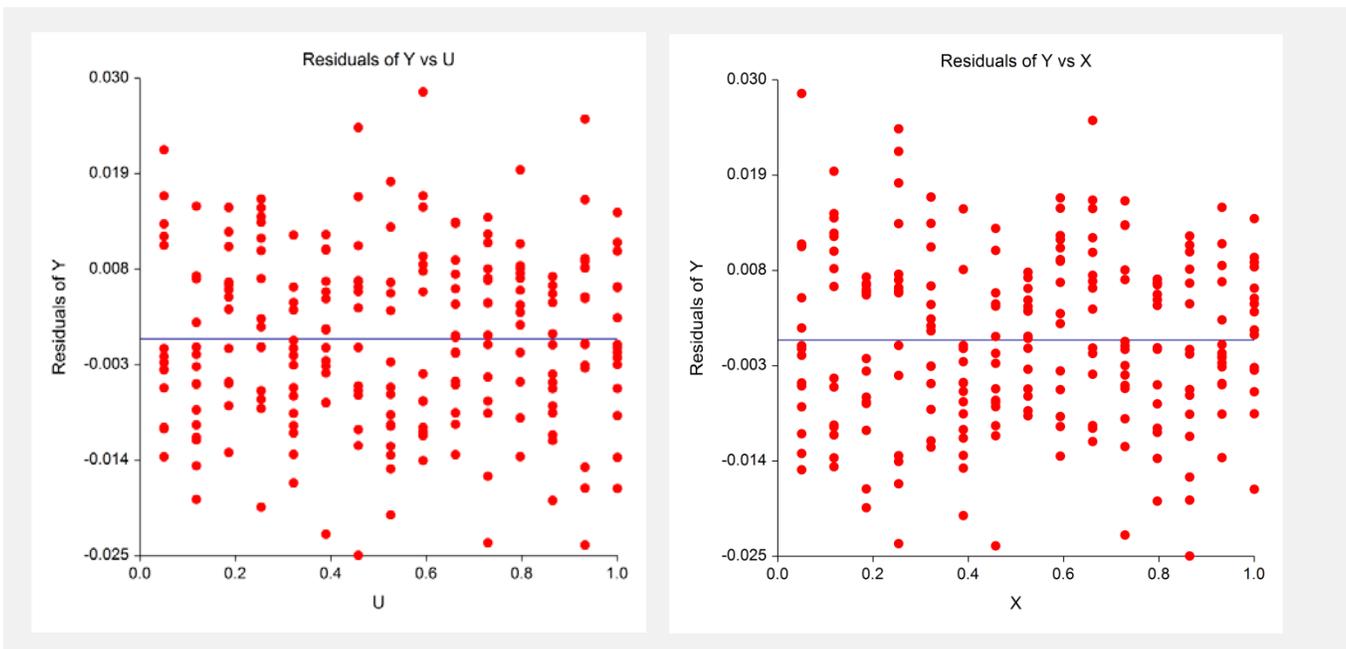
These are scatter plots of the residuals versus each of the independent variables. The preferred pattern is a rectangular shape or point cloud. Any nonrandom pattern may require a redefining of the model.

Predicting for New Values

You can use your model to predict Y for new values of the independent variables. Here's how. Add new rows to the bottom of your database containing the values of the independent variables 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.