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:

$$Y = \frac{B0 + B1X + B2X^2 + B3U + B4U^2}{1 + B5X + B6X^2 + B7U + B8U^2}$$

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.

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.

Setup

To run this example, complete the following steps:

1 Open the FnReg4 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **FnReg4** and click **OK**.

2 Specify the Ratio of Polynomials Fit – Many Variables procedure options

- Find and open the **Ratio of Polynomials Fit – Many Variables** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings template. To load this template, click **Open Example Template** in the Help Center or File menu.

Option	Value
Variables Tab	
Y Variable	**Y**
U Variable	**U**
Transformation	**SQRT(z)**
X Variable	**X**
Transformation	**SQRT(z)**
Numerator Terms	**U, X, U2X2**
Denominator Terms	**U2, UX, U2X2**
Reports Tab	
Residual Report	**Checked**
All Other Reports	**Checked**
Plots Tab	
All Plots	**Checked**

3 Run the procedure

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

Minimization Phase Section

Minimization Phase Section

Itn No.	Error Sum Lambda	Lambda	B0	B1	B2	B3
0	0.0256219	4E-05	2.050828	0.7681349	-1.027562	1.102767
1	0.02246299	1.6E-05	2.002186	0.960939	-1.019122	1.408448
2	0.02238311	6.4E-06	1.995856	0.9876834	-1.019637	1.488163
3	0.02238232	2.56E-06	1.99552	0.989055	-1.019808	1.495897
4	0.02238232	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	0.05595055	0.8788245	1.099371
B2	X	-1.019816	0.01188279	-1.043235	-0.9963957
B3	U2X2	1.496198	0.1342063	1.23169	1.760706
B4	u2	1.009521	0.02531414	0.9596294	1.059413
B5	ux	-1.081743	0.02945009	-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)+(0.9890978)*(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	0.02238232	0.0001026712
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	B6
B0	1.000000	-0.915182	-0.635169	-0.014330	-0.850024	-0.465655	0.041458
B1	-0.915182	1.000000	0.505466	-0.065987	0.969403	0.577334	-0.150335
B2	-0.635169	0.505466	1.000000	-0.520124	0.301986	0.802626	-0.553984
B3	-0.014330	-0.065987	-0.520124	1.000000	0.093673	-0.751426	0.991271
B4	-0.850024	0.969403	0.301986	0.093673	1.000000	0.405675	0.008270
B5	-0.465655	0.577334	0.802626	-0.751426	0.405675	1.000000	-0.819380
B6	0.041458	-0.150335	-0.553984	0.991271	0.008270	-0.819380	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

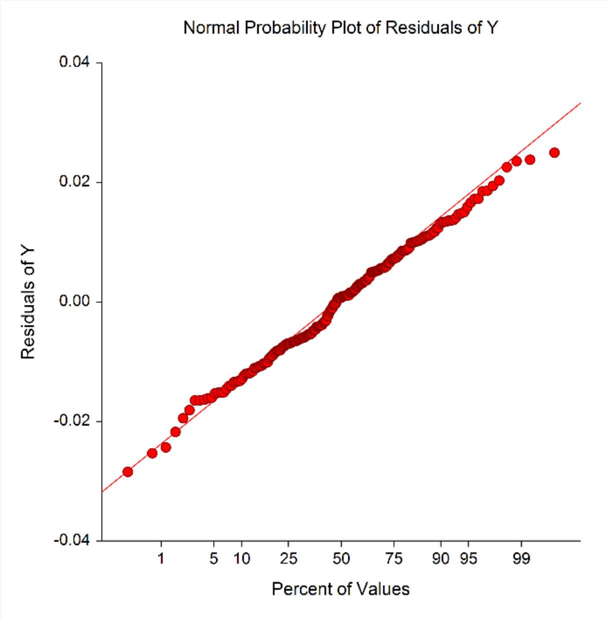
Predicted Values and Residuals Section

Row No.	Y	Predicted Value	Lower 95.0% Value	Upper 95.0% Value	Residual
1	1.981996	1.992756	1.971265	2.014246	-0.01075969
2	2.028455	2.026659	2.005891	2.047427	0.001796112
3	2.027451	2.014322	1.993644	2.034999	0.01312937
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

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)

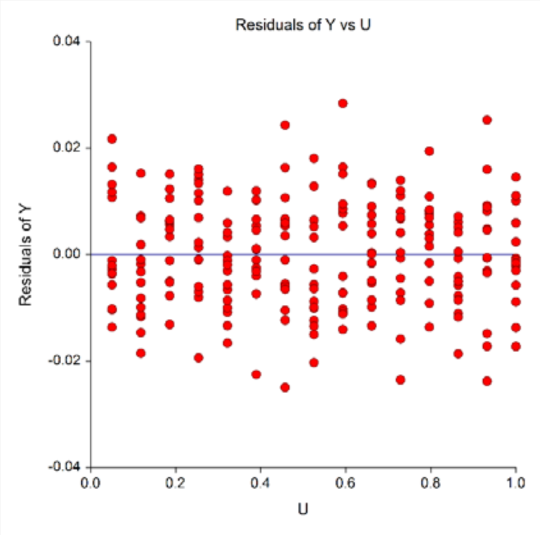
Probability Plot(s)

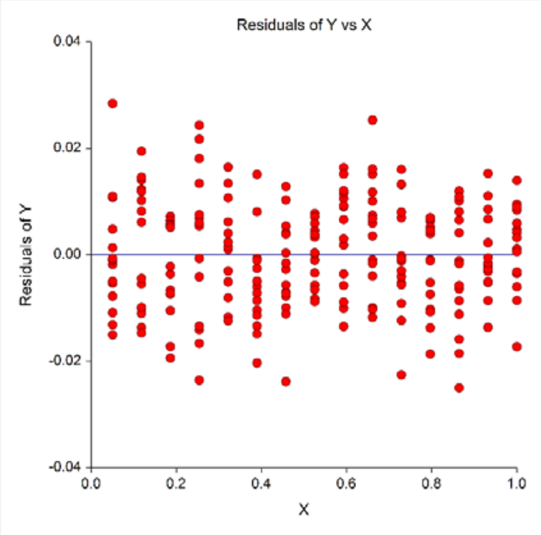


If the residuals are normally distributed, the data points of the normal probability plot will fall along a straight line. Major deviations from this ideal picture reflect departures from normality. Stragglers at either end of the normal probability plot indicate outliers, curvature at both ends of the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

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.