Chapter 375

Ratio of Polynomials Fit – One Variable

Introduction

This program fits a model that is the ratio of two polynomials of up to fifth order. Examples of this type of model are:

$$Y = \frac{A0 + A1X + A2X^2}{1 + B1X + B2X^2}$$

and

$$Y = \frac{A0 + A1X + A2X^2 + A3X^3 + A4X^4 + A5X^5}{1 + B1X + B2X^2 + B3X^3 + B4X^4 + B5X^5}$$

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 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 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.

Example 1 – Fitting a Ratio of Polynomials Model

This section presents an example of how to fit a ratio of polynomials model. In this example, we will fit a third order polynomial in the numerator and a fourth order polynomial in the denominator to the variables Y and X of the FnReg3 database.

Setup

To run this example, complete the following steps:

1 Open the FnReg3 example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **FnReg3** and click **OK**.

2 Specify the Ratio of Polynomials Fit – One Variable procedure options

- Find and open the **Ratio of Polynomials Fit – One Variable** 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 (Dependent) Variable	**Y**
X (Independent) Variable	**X**
A1 X^1	**Checked**
A2 X^2	**Checked**
A3 X^3	**Checked**
B1 X^1	**Checked**
B2 X^2	**Checked**
B3 X^3	**Checked**
B4 X^4	**Checked**
Options Tab	
Max Iterations	**10**
Reports Tab	
Residual Report	**Checked**
All Other Reports and 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	A0	A1	A2	A3
0	70.81162	4E-05	11.78766	-0.2798519	0.005809555	-4.172031E-05
1	70.73225	0.16	11.78725	-0.2798166	0.005810308	-4.172031E-05
2	70.68134	0.064	11.78944	-0.2797599	0.005810308	-4.172031E-05
3	70.61962	0.256	11.78874	-0.2797441	0.005810308	-4.172031E-05
4	70.58868	0.1024	11.78977	-0.2797041	0.005810308	-4.172031E-05
5	70.58413	0.4096	11.78925	-0.2796919	0.005810308	-4.172031E-05
6	70.5548	0.16384	11.78946	-0.2796662	0.005810308	-4.172031E-05
7	70.54813	0.065536	11.79198	-0.2796135	0.005810308	-4.172031E-05
8	70.51613	0.262144	11.7916	-0.2795997	0.005810308	-4.172031E-05
9	70.50236	0.1048576	11.79268	-0.2795659	0.005810308	-4.172031E-05
10	70.49983	0.4194304	11.79228	-0.279557	0.005810308	-4.172031E-05

Maximum iterations before convergence.

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	Parameter Estimate	Asymptotic Standard Error	Lower 95% C.L.	Upper 95% C.L.
A0	11.79254	1.012715	9.750206	13.83488
A1	-0.2795364	0.09166186	-0.4643901	-0.09468261
A2	0.005810308	0.003516727	-0.001281847	0.01290246
A3	-4.172031E-05	2.863182E-05	-9.946187E-05	1.602125E-05
B1	-0.0783304	0.002020337	-0.08240479	-0.074256
B2	0.002391857	4.689093E-05	0.002297292	0.002486421
B3	-2.86472E-05	4.332001E-06	-3.738351E-05	-1.991088E-05
B4	1.171313E-07	1.025966E-06	-1.951927E-06	2.186189E-06

Dependent	Y
Independent	X
Model	Y=(A0+A1X^1+A2X^2+A3X^3) / (1+B1X^1+B2X^2+B3X^3+B4X^4)
R-Squared	0.990159
Iterations	10

Estimated Model
((11.79254-(0.2795364)*(X)+(0.005810308)*(X)^2-(4.172031E-05)*(X)^3)/(1-(0.0783304)*(X)+(0.002391857)*(X)^2-(2.86472E-05)*(X)^3+(1.171313E-07)*(X)^4))

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.

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.

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	76141.53	76141.53
Model	8	83234.82	10404.35
Model (Adjusted)	7	7093.291	1013.327
Error	43	70.49983	1.639531
Total (Adjusted)	50	7163.791	
Total	51	83305.32	

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	B1	B2	B3
A0	1.000000	-0.668342	0.251252	-0.126446	0.045937	0.033491	-0.444426
A1	-0.668342	1.000000	-0.826039	0.725516	0.586512	-0.664699	-0.127771
A2	0.251252	-0.826039	1.000000	-0.986996	-0.937223	0.967391	0.655601
A3	-0.126446	0.725516	-0.986996	1.000000	0.978499	-0.991658	-0.765633
B1	0.045937	0.586512	-0.937223	0.978499	1.000000	-0.989097	-0.867018
B2	0.033491	-0.664699	0.967391	-0.991658	-0.989097	1.000000	0.801153
B3	-0.444426	-0.127771	0.655601	-0.765633	-0.867018	0.801153	1.000000
B4	0.769239	-0.364811	0.217490	-0.176658	-0.095811	0.123138	-0.139159

Asymptotic Correlation Matrix of Parameters (Continued)

	B4
A0	0.769239
A1	-0.364811
A2	0.217490
A3	-0.176658
B1	-0.095811
B2	0.123138
B3	-0.139159
B4	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

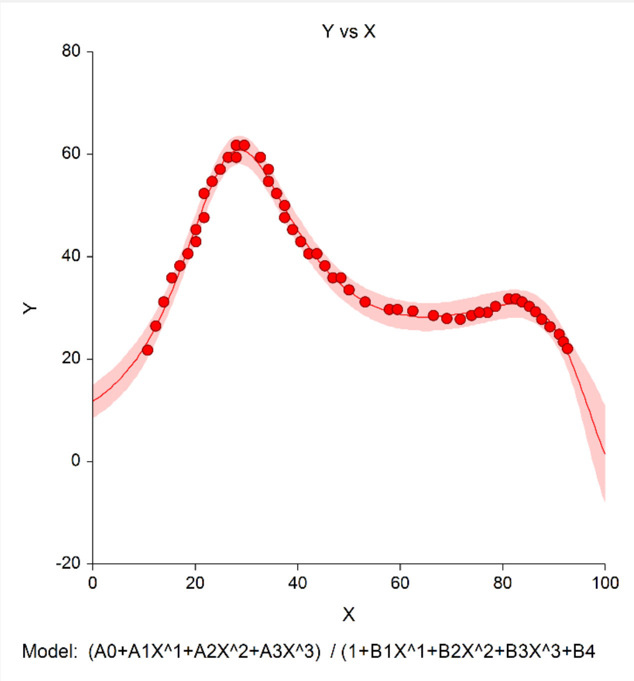
Predicted Values and Residuals Section

Row No.	X	Y	Predicted Value	Lower 95.0% Value	Upper 95.0% Value	Residual
1	10.69182	21.76471	23.39933	20.2282	26.57045	-1.634617
2	12.26415	26.47059	26.25687	23.30506	29.20868	0.2137172
3	13.83648	31.17647	29.50871	26.70471	32.31271	1.667763
4	15.4088	35.88235	33.16126	30.38701	35.93551	2.721093
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

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)

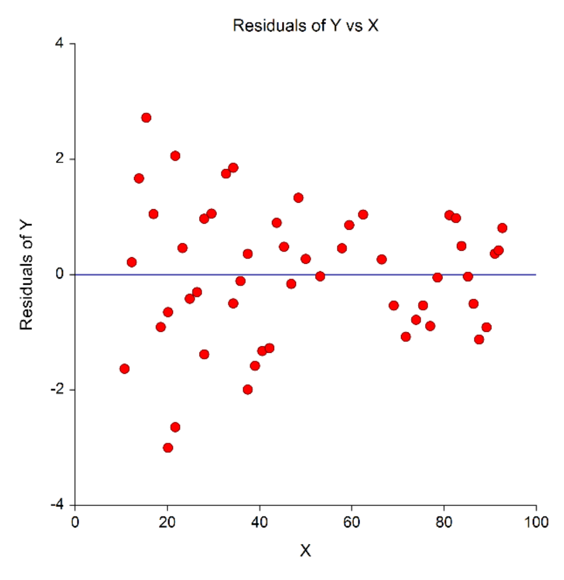


This plot displays the data along with the estimated function and prediction limits. It is useful in deciding if the fit is adequate and the prediction limits are appropriate.

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)

Residual Plot(s)

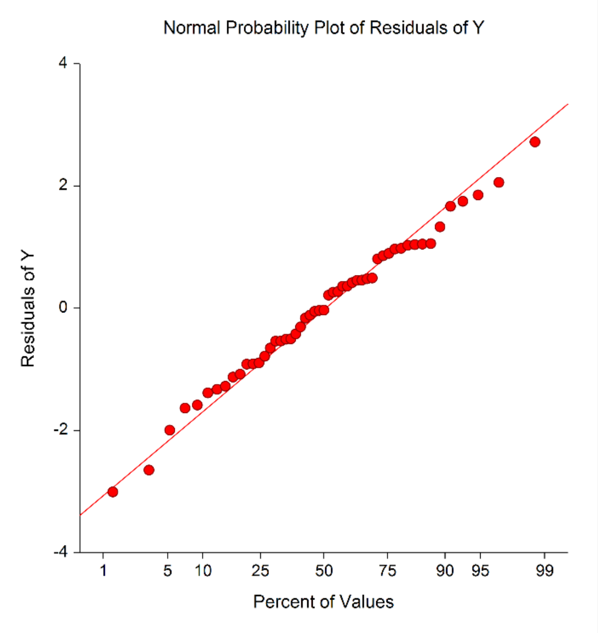


This is a scatter plot of the residuals versus the independent variable, X. The preferred pattern is a rectangular shape or point cloud. Any nonrandom pattern may require a redefining of the model.

In this example, it appears that the variance of the residuals decreases as X increases—suggesting that the equal variance assumption is violated. If this pattern was very drastic, we might want to try a variance stabilizing transformation of Y or using weighted nonlinear regression. However, the pattern does not appear severe in this case, so we probably would not take further action.

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.