

Chapter 371

Ratio of Polynomials Search – Many Variables

Introduction

This procedure searches through hundreds of potential curves looking for the model that fits your data the best. The procedure is heuristic in nature but seems to do well with the data we have tried. This procedure is more general than the Ratio of Polynomials Search procedure because it models up to four independent variables.

A general class of models called the ratio of polynomials (see Multivariate Ratio of Polynomials Fit chapter) provides a wide variety of curves to search from. Normally, fitting these models is a slow, iterative process. However, using a shortcut, an approximate solution may be found very quickly so that a large number of models may be searched in a short period of time. After the best fitting model is found, use the procedure discussed in the Multivariate Ratio of Polynomials Fit chapter to obtain a detailed analysis.

Parsimony

One of the main principles in model building is that you never use three parameters when two parameters will do. Hence, one of our tasks will be to find a model with the fewest number of parameters. A second principle in dealing with the ratio-of-polynomials model is that you should not fit a model with a numerator of higher polynomial order than that of the denominator. The models tried by default by this program follow these rules. A third rule is that all terms in a polynomial up to the desired order must be included. Hence, you would not use $Y=A+CX^2$. Instead, you would fit $Y=A+BX+CX^2$.

Goodness-of-Fit

Measuring how well a given model fits the data so that the various models can be compared is an important part of the search. This is tough since the goodness-of-fit statistics you are familiar with (like R-Squared) do not have the same meaning in this setting. However, because of the lack of other general, goodness-of-fit indices, we have chosen to base our selection on the value of R-Squared.

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 search for a model relating the dependent variable Y to the independent variables U and X of the FnReg4 dataset.

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 Search – Many Variables procedure options

- Find and open the **Ratio of Polynomials Search – 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 file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab	
Y Variable	Y
Ln(Y)	Checked
U Variable	X
Ln(U)	Checked
V Variable	U
Ln(V)	Checked
All Other Options	Default Settings

3 Run the procedure

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

Multivariate Search Report

Variables Y = Y, U = X, V = U
 Models Fit 192

Results of Multivariate Search

R-Squared	Percent of Best	Transformations					No. of X's	Model Numerator / Denominator
		Y	U	V	W	X		
0.991356	100.00	N	L	L			10	(H2) / (H2)
0.991349	100.00	L	L	L			10	(H2) / (H2)
0.982629	99.12	L	N	L			10	(H2) / (H2)
0.982347	99.09	N	N	L			10	(H2) / (H2)
0.980970	98.95	N	N	L			8	(E1,E2) / 1
0.979765	98.83	N	L	L			8	(E1,E2) / 1
0.978064	98.66	L	N	L			8	(E1,E2) / 1
0.977188	98.57	N	L	L			5	(H2) / 1
0.976036	98.45	L	L	L			8	(E1,E2) / 1
0.975872	98.44	N	N	L			5	(H2) / 1
0.972253	98.07	L	N	L			5	(H2) / 1
0.971677	98.01	L	L	L			5	(H2) / 1
0.899874	90.77	L	N	N			8	(E1,E2) / 1
0.898429	90.63	N	L	N			10	(H2) / (H2)
0.897845	90.57	L	L	N			8	(E1,E2) / 1
0.893617	90.14	N	N	N			8	(E1,E2) / 1
0.892422	90.02	N	L	N			8	(E1,E2) / 1
0.887341	89.51	L	N	L			4	(U1,V1,U2,V2) / 1
0.884633	89.23	L	L	L			4	(U1,V1,U2,V2) / 1
0.884020	89.17	N	N	L			4	(U1,V1,U2,V2) / 1

Transformation key:
 2 = 1/X^2, I = 1/X, Q = 1/SQRT(X), L = LN(X), R = SQRT(X), N = X, and S = X^2

This report displays the best models (in terms of R-Squared) found. Each row describes the results for a single model.

Models Fit

This value is the total number of models that were evaluated.

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.

Ratio of Polynomials Search – Many Variables

Percent of Best

This is the percent that the R-Squared value of this model is of the best (top) model.

Transformations (Y U V W X)

The letters correspond to the transformations that were used for each variable. The transformation key is listed at the bottom of the report. For example, the entry for the first row is N L L. This means that Y was regressed on $\ln(X)$ and $\ln(U)$.

No. of X's

This is the number of parameters fit in this model. Of course, we want models with only a few parameters and a large value of R-Squared.

Model Numerator/Denominator

This gives the model using the shorthand notation described in Models - Custom section above. You can apply this shorthand notation directly in Multivariate Ratio of Polynomial Fit procedure to obtain detailed results for a particular model.

Note that the numeral one (1) is used when no polynomial is specified.