When a Linear Model Just Won't Do: Fitting Nonlinear Models Using JMP®

JMP Discovery Conference 2013
Susan Walsh – SAS Institute

Abstract

This paper will explore the different approaches to fitting nonlinear models in JMP. With the addition of the Fit Curve personality to the nonlinear platform, fitting nonlinear models in JMP has never been easier. The first option addressed will be the use of this personality to examine a series of curves by a grouping variable. Specific attention will be given to comparing the parameter estimates, using analysis of means or the equivalence test. A second set of data will be used to fit a sequence of growth models, comparing different models to determine which model might be the best.

Although the Fit Curve personality includes a collection of popular models, it is always possible that a user will want to fit a model not included in Fit Curve. The paper will use additional data tables to demonstrate the use of the previously existing model library. The final approach addressed will be an example of creating a nonlinear model as a formula in the data table and then fitting it in the nonlinear platform. It will include locking the values of parameters in the formula to compare different models and determine the best fit.

Introduction

Before fitting a nonlinear model, it is important to understand what is meant by a linear versus a nonlinear model. A linear regression model is linear in the parameters. That is, there is only one parameter in each term of the model and each parameter is a multiplicative constant on the independent variable(s) of that term. Examples of linear models are shown in Figure 1.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$
$$Y = \beta_0 + \beta_1 \ln(X) + \varepsilon$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

Figure 1: Linear Models

In contrast, a nonlinear model is nonlinear in the parameters. The SAS/STAT documentation states "A model is nonlinear in the parameters if the derivative of the model with respect to a parameter depends on this or other parameters [1]." Examples of nonlinear models are shown in Figure 2.

$$Y = \beta_1 e^{-\beta_2 X} + \varepsilon$$
$$Y = \beta_1 + (\beta_2 - \beta_1)e^{-\beta_3 X} + \varepsilon$$
$$Y = \frac{(\beta_1 X)}{(\beta_2 + X)} + \varepsilon$$

Figure 2: Nonlinear Models

When fitting a linear model, the general form of the model is known. It is only necessary to identify the dependent and independent variables and the terms to include in the model. For a nonlinear model, the form of the model must be specified, the parameters to be estimated identified, and starting values for those parameters must be provided.

When obtaining estimates of the parameters in a linear model, specific equations are used to determine the single unique solution that results in the smallest error sum of squares possible for the model and the data table. In the case of a nonlinear model, iterative techniques are generally used to converge on a solution that provides the best fit model. It is important for the analyst to check to be sure that convergence has been attained. There are certain conditions which can prevent convergence. These conditions may include:

- incorrect specification of the model
- poor initial starting values for the parameters
- overdefined model
- insufficient iterations
- insufficient data.

Incorrect model specification is when the data do not exhibit the same relationship as the one given in the model. Poor initial starting values can cause lack of convergence or convergence to a local minimum. A model that is over defined is one that has more parameters than are necessary for the relationship. If two or more parameters are very close in value, they could, and should, be represented by a single parameter. In some cases, the estimation method is slowly approaching convergence, but convergence is not obtained before the maximum number of iterations is reached. Finally, there might be insufficient data to generate a model. In this case, additional data must be collected [2].

Fitting Nonlinear Models Using JMP

In JMP, there are three different ways to approach fitting a nonlinear model: Fit Curve, Model Library, and column formula.

The Fit Curve personality was added to the nonlinear platform in JMP 10. This personality allows the user to specify the X and Y variables for the model and choose the general form of the model. JMP then fits the chosen model automatically. The user does not need to specify the starting values of the parameters. Fit Curve includes sigmoid curves, exponential growth and decay models, peak models, pharmacokenetic models, and the Michaelis Menten model. The platform also includes the ability to fit separate models for the levels of a grouping variable and then compare parameter estimates and test for parallelism. Fit statistics are provided which allow for comparison of different models that are fit to the data.

The Model Library includes a larger variety of models than the Fit Curve personality. This library provides graphs for many of the models included as well as the capability to use the model graph with the data to set starting values for the parameters. In some cases, the Model Library will calculate suggested starting values. The Model Library is used to create a model formula column for use with the platform.

Finally, creating a column formula allows you complete freedom to specify any equation for which you want to fit a model along with the starting values for the parameter estimates.

The Fit Curve Personality

As was mentioned earlier, the Fit Curve personality includes pharmacokinetic models, including the following models:

- one compartment oral dose
- two compartment IV Bolus dose
- biexponential, 4 parameter

Fit Curve: Theophylline Example

To illustrate the one compartment oral dose model, examine data collected when studying the drug theophylline. Theophylline is used as a long term medication to control lung disease. In order to regulate the dose of the medication, researchers measure the concentration of theophylline in the bloodstream [3]. They want to determine the relationship between the concentration and the time after ingesting the drug. They administered the drug to 12 patients, measuring the theophylline level eleven times for each patient. The subject identifier, time, and concentration of theophylline were recorded.

In the nonlinear platform, you generally specify a numeric response variable. The X, Predictor Formula variable is either a column that has a formula that includes the parameters for the model to be fit or is a numeric independent variable to be used with one of the built in models. Use the Fit Curve personality of the nonlinear platform to fit a one compartment oral dose model. Select Analyze → Modeling → Nonlinear and assign the variables as shown in Figure 3.
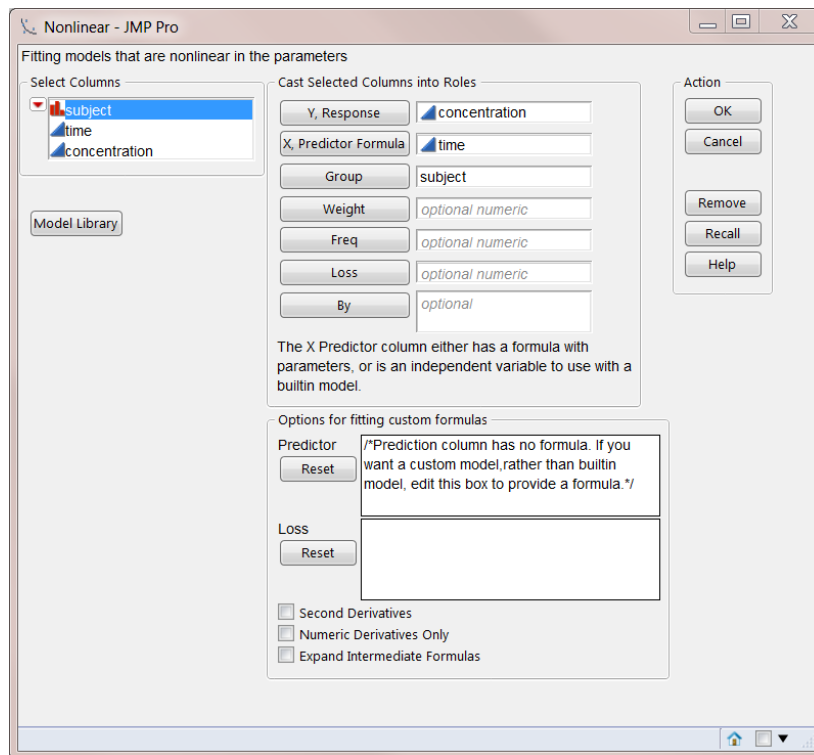

Figure 3: Nonlinear Dialog

In this case, use subject in the Group role to fit separate models for each of the subjects in this experiment. This will allow comparison of the model parameters between subjects. As the column used as the X variable does not have a formula, the Fit Curve personality is invoked and graphs of the data are displayed.
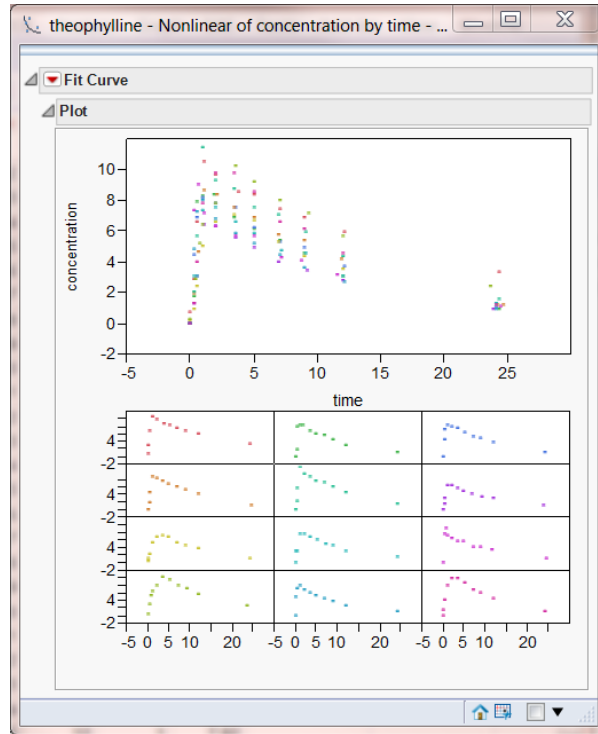
Figure 4, Fit Curve Initial Graphs

To fit the model of interest, click on the red triangle at the top of the report window. Select Pharmacokenetic Models → Fit One Compartment Oral Dose. The results include the Prediction Model, Summary of Fit, Parameter Estimates, and Plot.

The Prediction Model displays the model formula, including the parameters. When appropriate, physical interpretation of the parameters is also shown. In the case of the one compartment oral dose model, the parameters represent the area under the curve, the elimination rate, and the absorption rate.
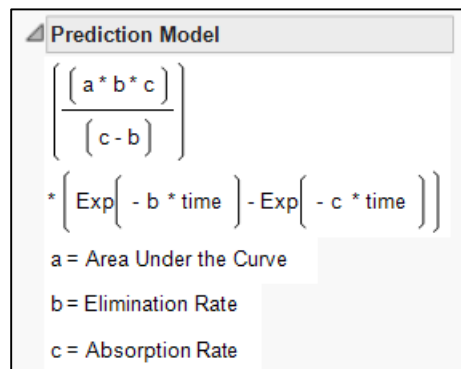
Figure 5: Prediction Model with Parameter Definitions

The summary of fit statistics are provided both within the model results and in the Model Comparison section at the top of the report. In cases where multiple models are fit to the same data table, the summary of fit statistics for all models fit in the window will be shown in the Model Comparison report.

For this model, there are three parameters estimated for each of the subjects. Analysis of means can be used to compare the parameters across subjects [4]. To generate the analysis of means decision charts, click on the red triangle next to the model name and select Compare Parameter Estimates.
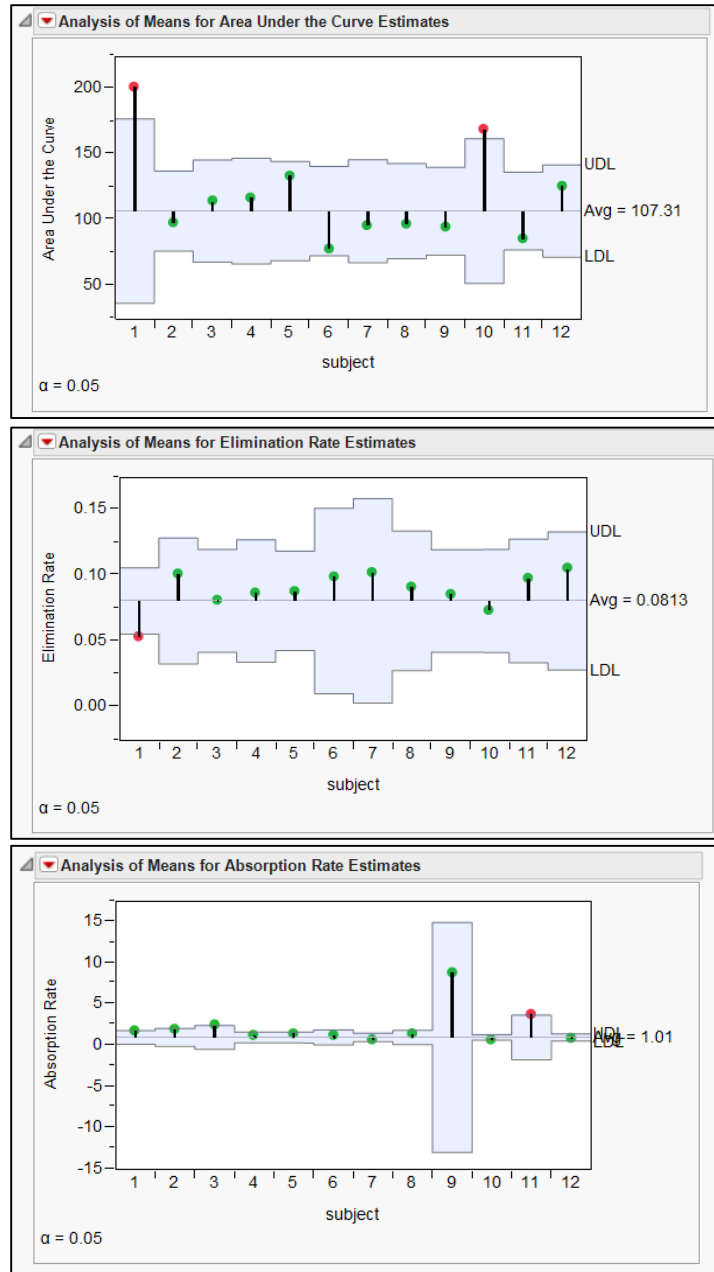
Figure 6: Analysis of Means for Parameter Estimates

For each category of parameter estimates there is at least one subject for whom the parameter estimate is found to be statistically different than the others at the 0.05 level of significance. This is an indication that a single model may not be appropriate for this data. It may also prompt additional examination of patients 1, 10, and 11 to see if there is a particular demographic or medical history reason that they are different from the other subjects.

A series of plots are also provided to illustrate the models that are fit. There is a single plot that overlays all of the subjects, and a matrix of plots that shows each patient individually.
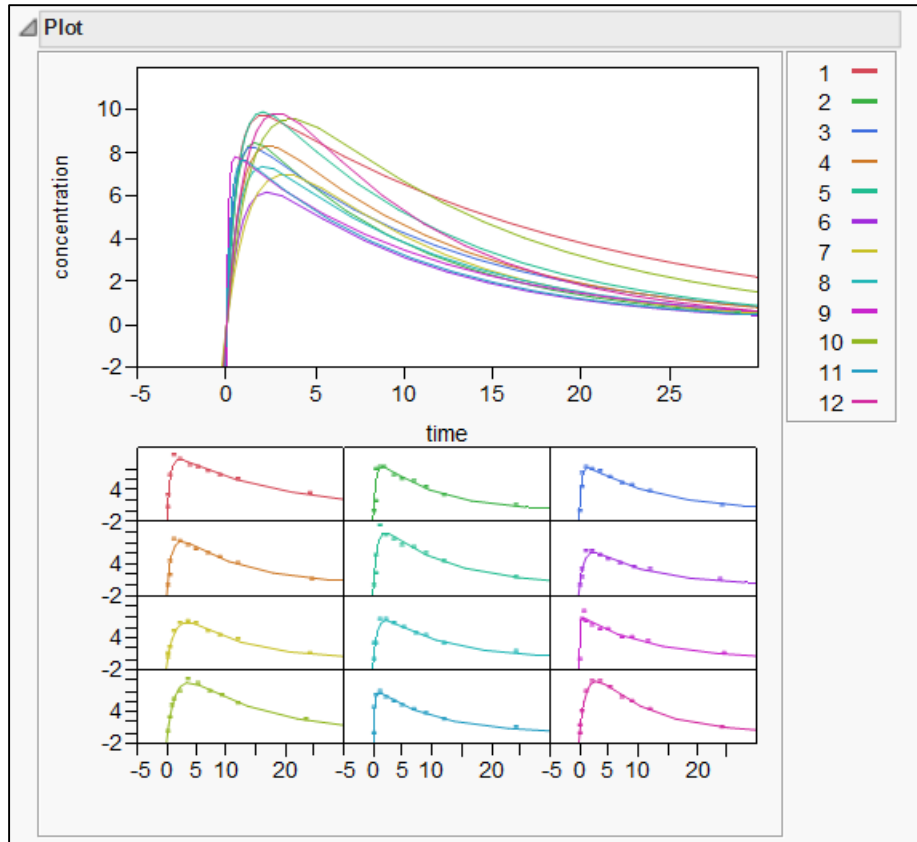
5

Figure 7: Model Plots

Fit Curve: Enzyme Example

Another type of model that can be fit with the Fit Curve personality is a growth model. Consider the relationship between concentration and velocity for a particular enzyme/substrate pair. The reaction rates (velocity) are observed at different substrate concentrations [5]. The data table is shown in Figure 8.
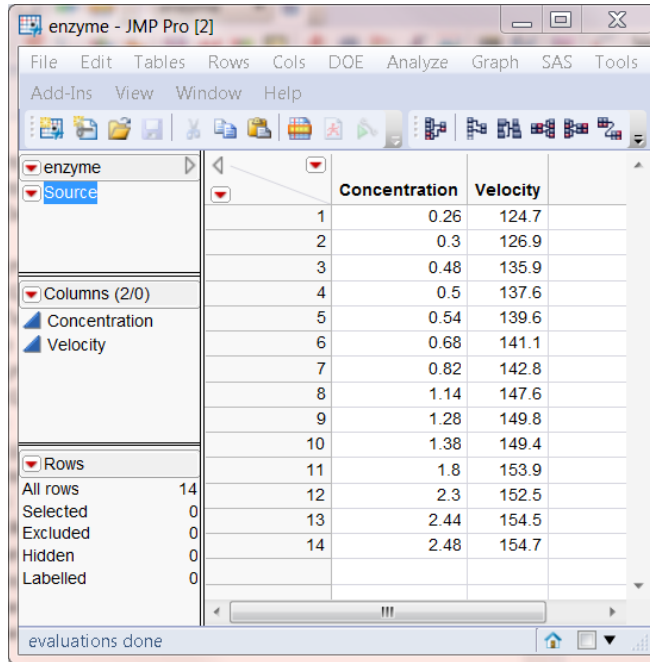
Figure 8: Enzyme Data

To begin, select analyze → Modeling → Nonlinear. Assign Velocity as the Y variable and Concentration as the X variable. After selecting OK, examine the plot of the data shown in Figure 9.
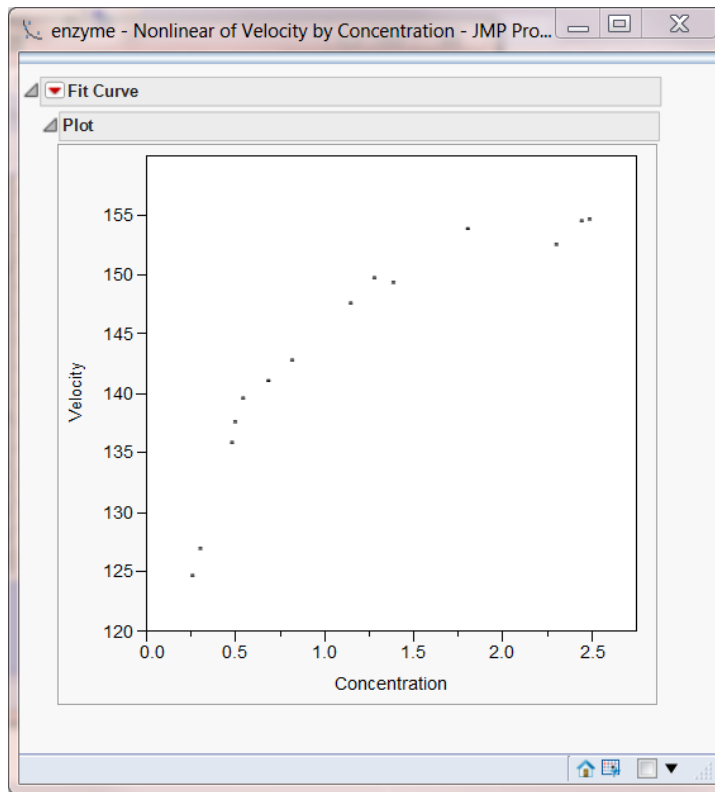

Figure 9: Plot of enzyme data

At first glance, one might consider fitting a quadratic model to this data. While a quadratic model is not nonlinear, such polynomial modes are included in the Fit Curve personality for ease of comparison. Fit a quadratic model by selecting Polynomials → Fit Quadratic from the red triangle menu.
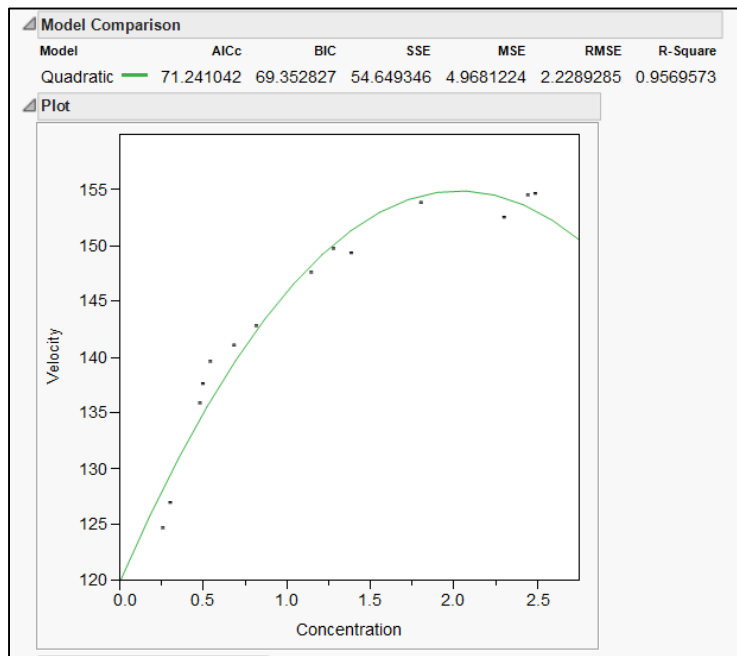

Figure 10: Quadratic fit of enzyme data

The R-Square statistic is quite high. However, a scientist might be concerned about the predicted velocity beginning to decline at a certain point in time. Assuming that this decline is not possible in this circumstance, a growth model might be more appropriate. Fit both mechanistic growth and Michaelis Menten models to the data by selecting Exponential Growth and Decay → Fit Mechanistic Growth (and Fit Michaelis Menten) from the red triangle menu. Examine the results.
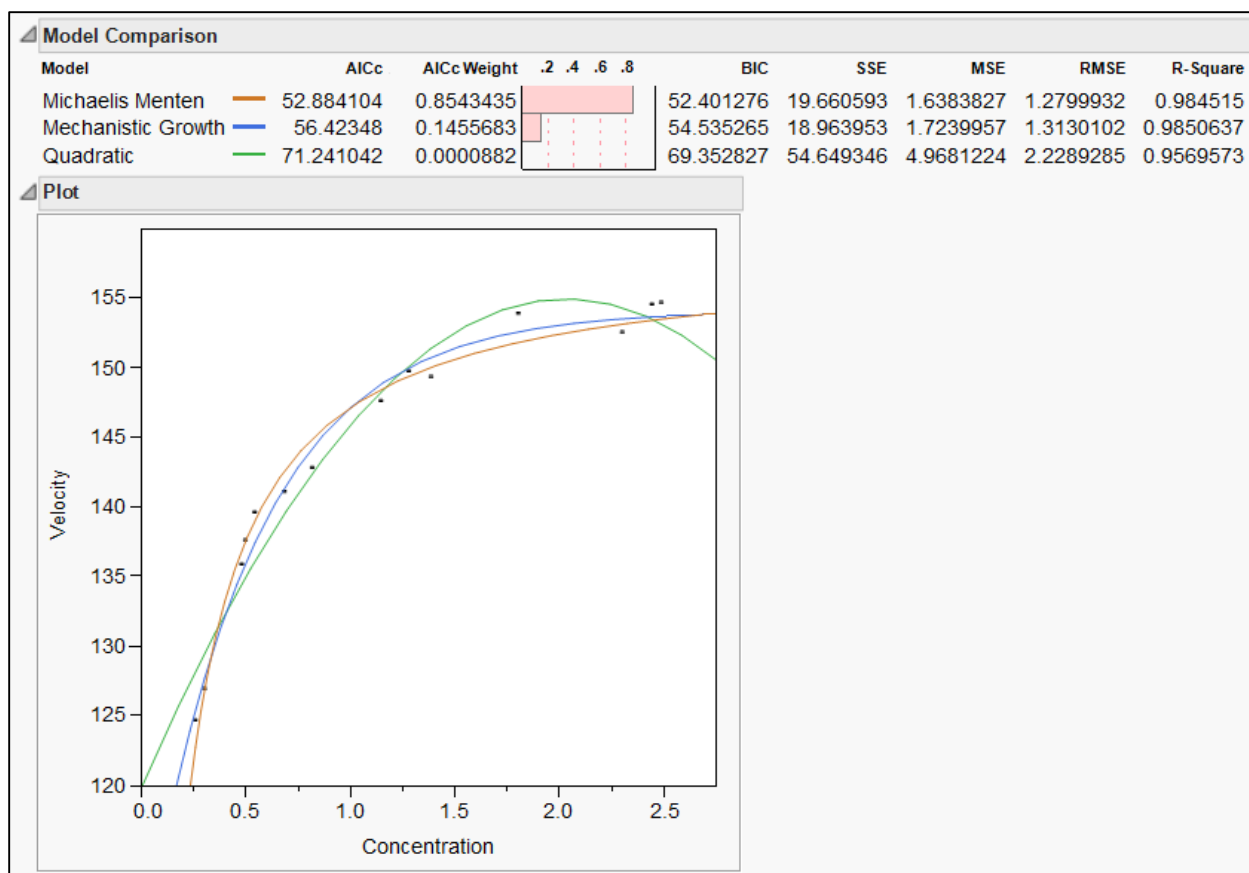
Model Comparison

Model		AICc	AICc Weight	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R-Square
Michaelis Menten	—	52.884104	0.8543435		52.401276	19.660593	1.6383827	1.2799932	0.984515
Mechanistic Growth	—	56.42348	0.1455683		54.535265	18.963953	1.7239957	1.3130102	0.9850637
Quadratic	—	71.241042	0.0000882		69.352827	54.649346	4.9681224	2.2289285	0.9569573

Figure 11: Model Comparisons

The plot now has all three models superimposed. Using either the AICc or BIC statistics the Michaelis Menten model would be considered to be the best model. The AICc weight can also be used for comparison purposes as a measure of how much better one model is compared to the others. The AICc weight is defined as:

$$AICcWeight = \{exp[-0.5(AICc - AICcBest)]\}/AICcSum$$

Where AICcBest is the smallest AICc among the fitted models, AICc is the AICc statistic for the current model, and AICcSum is the sum of the AICc's for all of the fitted models. The AICc weight is the probability that the model is the "right" model given that one of the fitted models is in fact the right model. In this example, there is an 85% chance that the Michaelis Menten model is correct under the assumption that one of the three models fit is the right model.

The Model Library

Although the Fit Curve personality includes a collection of popular models, it is always possible that a user will want to fit a model not available within Fit Curve. The Model Library provides an alternative way to fit a number of commonly used models. As stated earlier, this library provides graphs for many of the models included as well as the capability to use the model graph with the data to set starting values for the parameters. The model library is used to create a model formula column for use with the platform.

The Model Library: Calcium Example

An experiment was conducted by Howard Grimes, Idaho State University, to determine the amount of radioactive calcium in cells. In particular the relationship between the amount of time the cells were in a

suspension and the amount of radioactive calcium in the cells was of interest. The researchers had reason to believe the response would follow a Weibull growth model [6]. One form of the Weibull growth model is included in the model library. To view the model library, select Analyze → Modeling → Nonlinear, then click on the Model Library button. A list of the available models is displayed; when a model is selected, the model formula is also displayed.
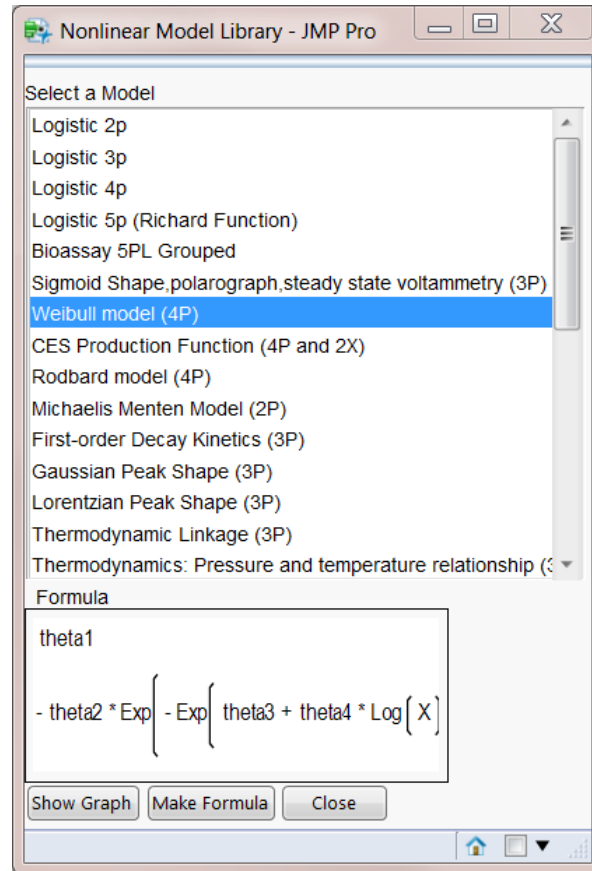

Figure 12: Model Library

To see the general shape of the model, click on the Show Graph button. If you are certain this is the model you want to use without looking at a graph, click on Make Formula. This will prompt you to supply the X and Y variables for this model, in this example time and calcium, respectively. After selecting OK, a graph of the general model form with the points is shown. JMP prompts you to set the starting values for the parameters.

In this particular example, the range used for axes in the graph is much larger than the data. Double-click on each axis to open the axis specification window and change the maximum value and increment for the axis as appropriate. You can then change the parameter estimates by using the slider bar. In this example, previous experiments provide potential starting values: theta1 and theta2 equal to 5, theta3 equal to -2, and theta4 equal to 1. Use the slider or type in these values in the graph window as shown in Figure 13.
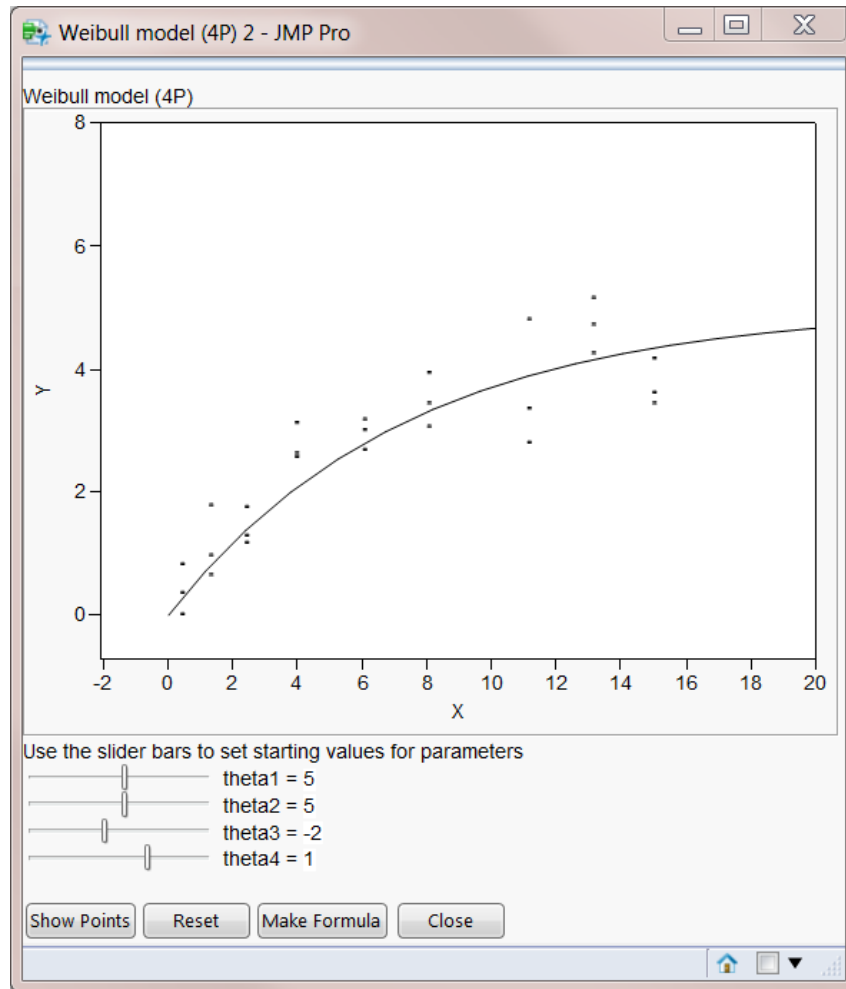
Figure 13: Weibull model

To create a column in the data table with the formula, click Make Formula, then close both the Weibull model window and the Model Library window. Examine the data table to see the column that has been added. You can also view the formula for the column to see both the formula and the starting values for the parameters.
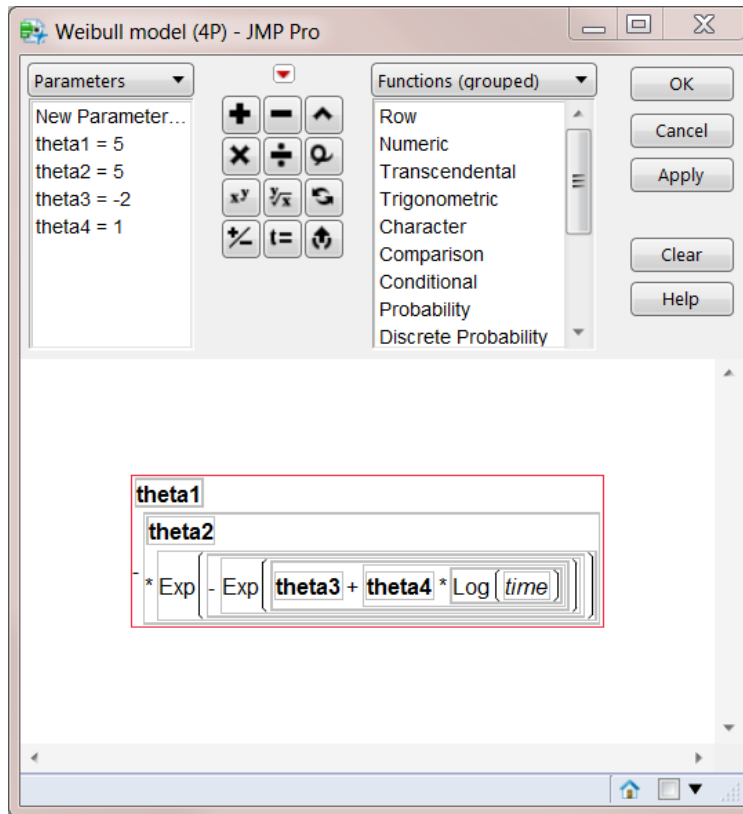
Figure 14: Weibull model formula and parameter estimates

Once the model formula column is created, return to the Nonlinear dialog box and enter the response variable (calcium) and the newly created predictor formula, Weibull model (4P). Select OK. At this point you have an opportunity to view a plot of the data with the model superimposed and make adjustments to settings as desired. Among the settings that can be changed are the maximum number of iterations, the convergence criterion, and the starting values for the parameters. Click Go to start the iterative process of fitting the model.

Examine the results paying careful attention to the convergence message toward the top of the control panel. If the model has not converged, you may choose to select Go again to do additional iterations. Among the numerous options that are available once the model is fit are:

- Save Estimates which saves the current parameter values to the parameters in the formula column.
- Confidence Limits which computes confidence intervals around the parameter estimates.
- Remember Solution which creates a Remembered Models report. Results of remembered models can be compared.

Save Estimates and Confidence Limits are both available as buttons in the results window. Remember Solution is available as an option in the red triangle menu.

Model Formula for Custom Models

A final option for the nonlinear platform available for use when the desired model is not included in the Fit Curve personality or the Model Library is to create a model formula in a data table column. If the model is one that you expect to use often, you might consider adding it to the model library. Information on doing so is contained in the JMP Modeling and Multivariate Methods book in the chapter Performing Nonlinear Regression [7].

12

Custom Model: Calcium Example

Continuing with the calcium experiment, suppose the researcher wants to use a different form of the Weibull model. Specifically, he is interested in the following form:

$$Y = \beta_1 + (\beta_2 - \beta_1)e^{-(\beta_3 X)^{\beta_4}} + \varepsilon$$

This form of the model might be particularly useful as the parameters have a useful interpretation:

β_1 the upper horizontal asymptote

β_2 the y-intercep

β_3 the rate parameter

β_4 the shape parameter. If the shape parameter is greater than one the curve is sigmoidal.

To enter this as a column formula in JMP, the first step is to define the parameters and their starting values. To begin, create a new column in the data table and add a column formula. In the formula editor, change the list choice on the left side from Table Columns to Parameters.
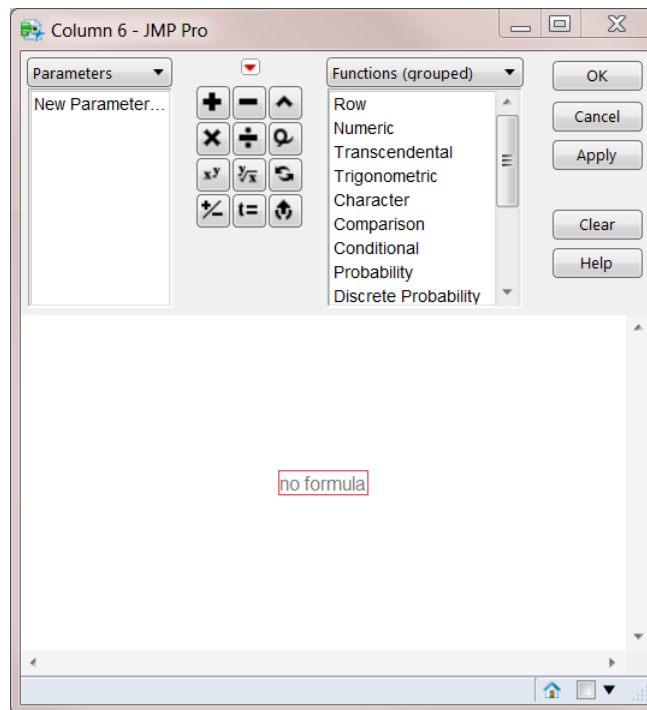
Figure 15: Formula Editor for Parameter Definition

Click on New Parameter. In the dialog box that opens, provide the name of the parameter and the starting value as shown in Figure 16.
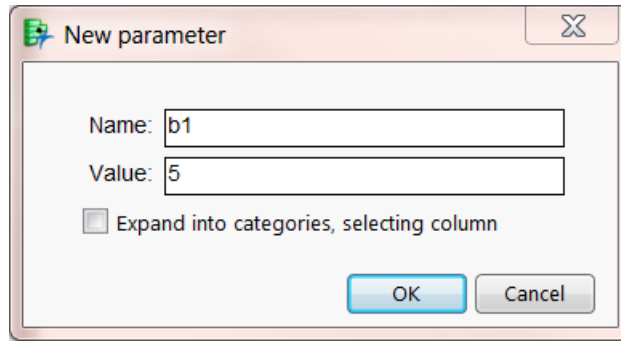
Figure 16: New parameter entry

Repeat the process to enter each parameter in the model along with its starting value.
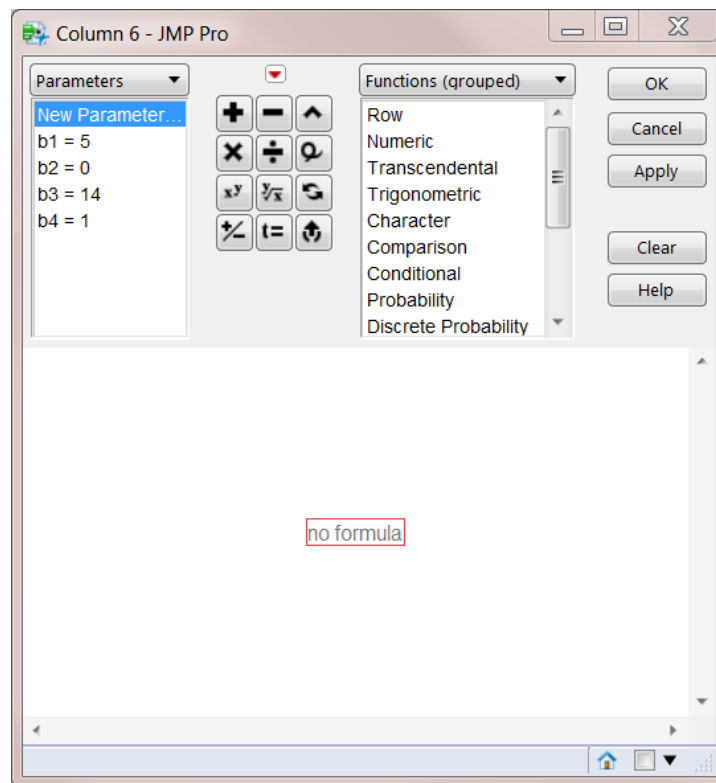

Figure 17: Model parameters

Once the parameters have been defined, the model formula can be entered using the formula editor. You will likely need to toggle between Parameters and Table Columns in order to complete the formula. Note that the parameters will appear as bold entries in the formula.
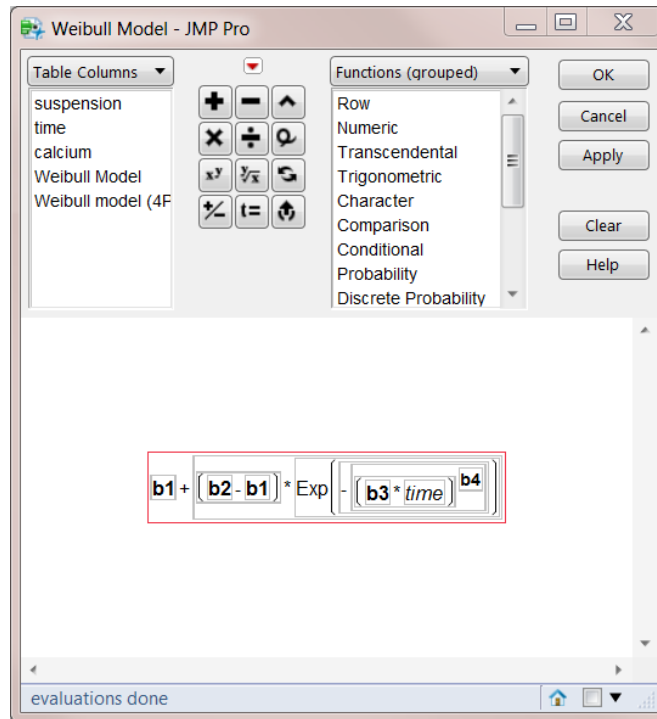
Figure 18: Completed Model Formula

Once the model formula is in the data table with the appropriate parameter starting values, proceed to the nonlinear platform, entering the Y variable and the model formula column created above as the X, Predictor Formula. After the first set of iterations, this model should have converged. Click on the Confidence Limits button to see the limits around the parameter estimates.
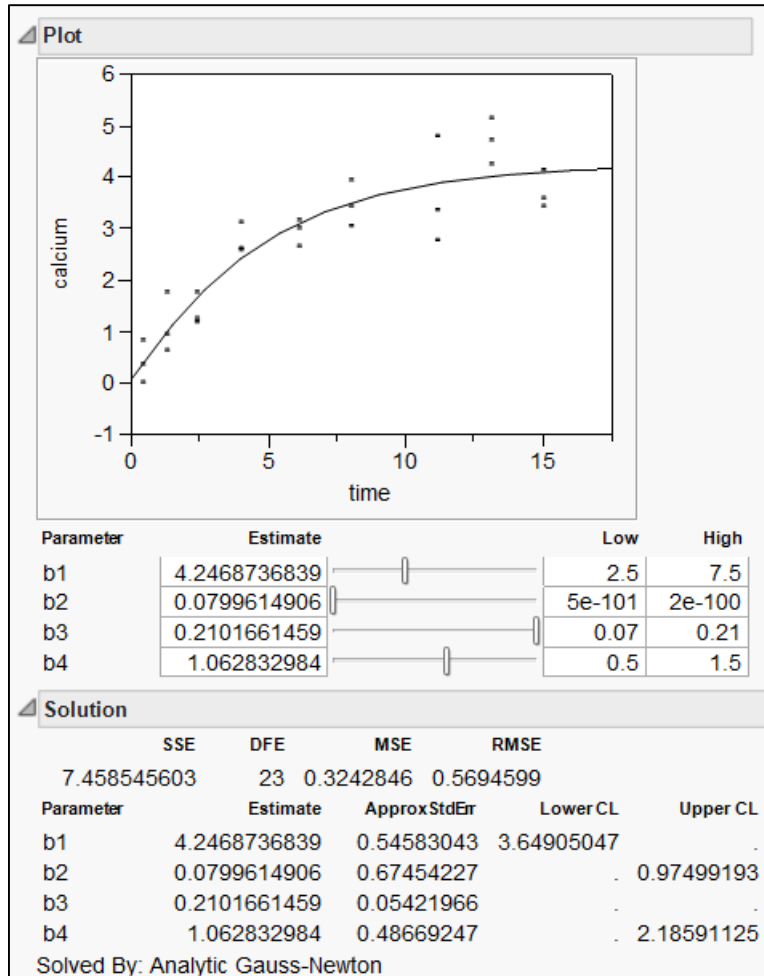
Figure 19: Nonlinear model results

In this case, several of the confidence limits are reported as missing values. The intervals are profile-likelihood confidence intervals, and it is not uncommon that the iterative technique used to obtain these limits fails to successfully converge on limits for the parameter estimates.

In examining these results, the researcher notes that the estimate for parameter b4 is very close to one. Given that b4 represents an exponent in the model, if it is equal to one it has no effect on the computation and does not need to be included in the model. He would like to explore the effect on the model if he fixes b4 at one and reruns the model. He would then like to compare the original model to the one with b4 fixed.

First, remember the current solution by clicking on the red triangle and selecting Remember Solution. Name this model Original. This adds a new section to the results window with fit statistics and parameter estimates for the remembered model. Scroll to the control panel in the results window. Change the value of b4 to one and click in the square to lock that value. Rerun the model by clicking Go.
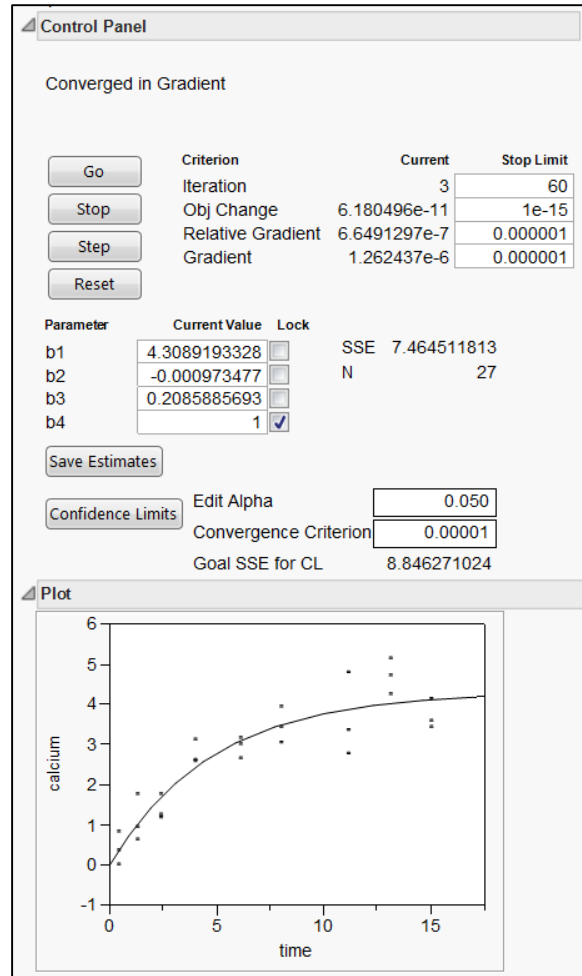
16

Figure 20: Nonlinear results after fixing b4=1

Once again, convergence has been achieved. Remember this solution and then examine the Remembered Models panel. Summary of fit information, the parameter estimates for the new model, and a comparison between the models has been made. In this case, there is insufficient evidence to conclude that the Error Sums of Squares (SSE) of the models are different. This could lead to a conclusion that the simpler model is a better choice in this case.

Once again examining the parameter estimates for the model, the parameter estimate for b2 is very close to zero. The researcher decided to lock b2 at zero and determine the effect on the model. Changing the current value of b2 to zero, placing a lock on the value, rerunning the model, and remembering the solution results in the table shown in Figure 21.

Figure 21: Remembered Models

Again, there is no statistically significant difference in the SSE's of the new model to the previous models. The researcher might choose the simplest model, a two parameter Weibull.

Custom Model: Plant Height Example

As a final example of the power and flexibility of the nonlinear platform, it can be used to fit a segmented model. A group of scientists are considering the effect of plant growth regulators on tulip growth. The goal was to determine an effective method of controlling plant height to facilitate shipment because of the limited amount of chemical available to treat a large quantity of bulbs [8]. Looking at a portion of the data for this study, the researchers want to fit a model to predict the height of the plants based on the milligrams of chemical used per pot. The researchers assume a linear relationship between the two variables. However, there is a point beyond which additional chemical will not further decrease the final height of the plants. Part of the experiment was to determine the value of the point at which the height levels out. Therefore, the model they are interested in fitting is:

$$\text{if } x \leq x_0, \text{ then } y = a + bx$$

$$\text{if } x > x_0, \text{ then } y = c$$

The parameters to be estimated are a, b, and c. In order to solve this type of segmented equation that involves determining both of the lines and the location of the join, x_0, the location of the join needs to be stated in terms of the parameters. As it is assumed that the two lines meet at the join, it follows that:

$$a + bx_0 = c$$

Solving for x_0:

$$x_0 = \frac{c - a}{b}$$

The equations to be used in fitting the model then can be expressed:

$$\text{if } x > \frac{c-a}{b}, \text{ then } y = c$$

$$\text{else } y = a + bx$$

These parameters and the equation can be entered into the formula editor as shown in Figure 22.
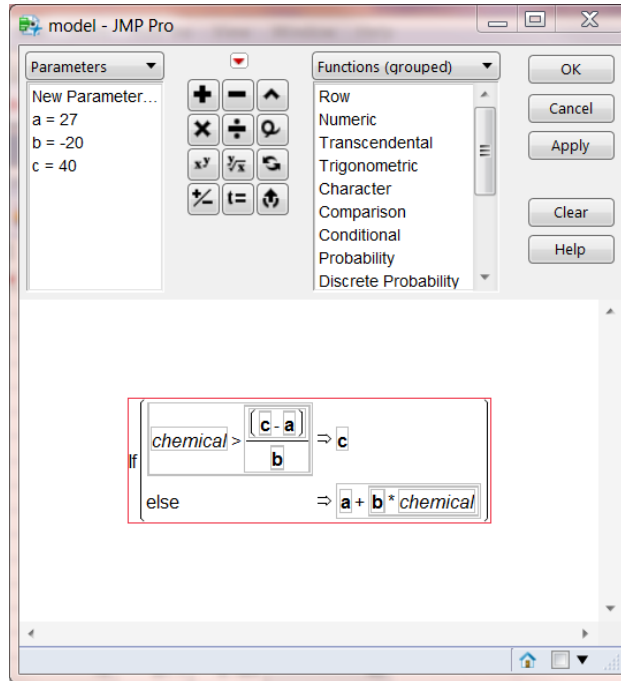
18

Figure 22: Formula for segmented model

In this case, the researchers determined the starting values of the parameters based on previous studies. This model can then be fit using the nonlinear platform placing the formula column in the X, Predictor Formula role and the height in the Y, Response role. Figure 23 shows the results of running this model.
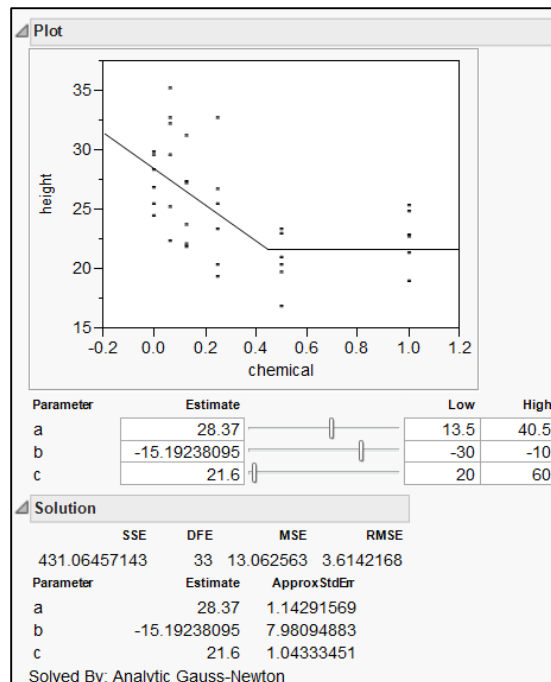

Figure 23: Plant growth model

Conclusion

In JMP, there are three different ways to approach fitting a nonlinear model: Fit Curve, Model Library, and column formula. The classic nonlinear approach, using a column formula, is more powerful and flexible, while the Fit Curve personality is easier to use. For those new to fitting nonlinear models, using the Fit Curve personality may be a good way to ease into nonlinear modeling. Once the user is more comfortable with specifying models, then the full power of the classic approach in JMP can be explored to expand the types of models available.

[1] SAS Institute Inc. 2013. SAS/STAT® User's Guide. Cary, NC: SAS Institute Inc.

[2] Rawlings, John O., Sastry G. Pantula, and David A. Dickey. *Applied Regression Analysis: A Research Tool.* New York: Springer-Verlag, 1998. Print.

[3] "Theophylline FAQ", *National Jewish Health*, May 2009. Web. 9 Aug. 2013. <http://www.nationaljewish.org/healthinfo/medications/lung-diseases/longterm/theophylline-faqs/>.

[4] Nelson, P. R., Wludyka, P. S., and Copeland, K. A. F. (2005) *Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions.* SIAM Press, Philadelphia, PA.

[5] SAS Institute Inc. 2013. SAS/STAT® User's Guide. Cary, NC: SAS Institute Inc.

[6] Data used with the permission of the researcher, Howard Grimes, Idaho State University.

[7] JMP® 10 Modeling and Multivariate Methods. Cary, NC: SAS Institute Inc.

[8] Krug, Brian A. et al. "Comparison of Flurprimidol to Ancymidol, Paclobutrazol, and Uniconazole for Tulip Height Control". HortTechnology April – June 2005: 370-373. Print. Data is used with permission.