The data set includes the following pieces of information:

• Column 1: Observed dependent variable

• Column 2: Fitted Residuals

• Column 3: Fitted Values

• The remainder of the columns: Predictors (not all predictor variables used in the model included in the model are included in the included data set).

1. Load and summarize the included data.

2. Determine a reasonable measure of model quality, and use the included data to compute and report this statistic.

3. Do you consider it a good model? Why or why not?

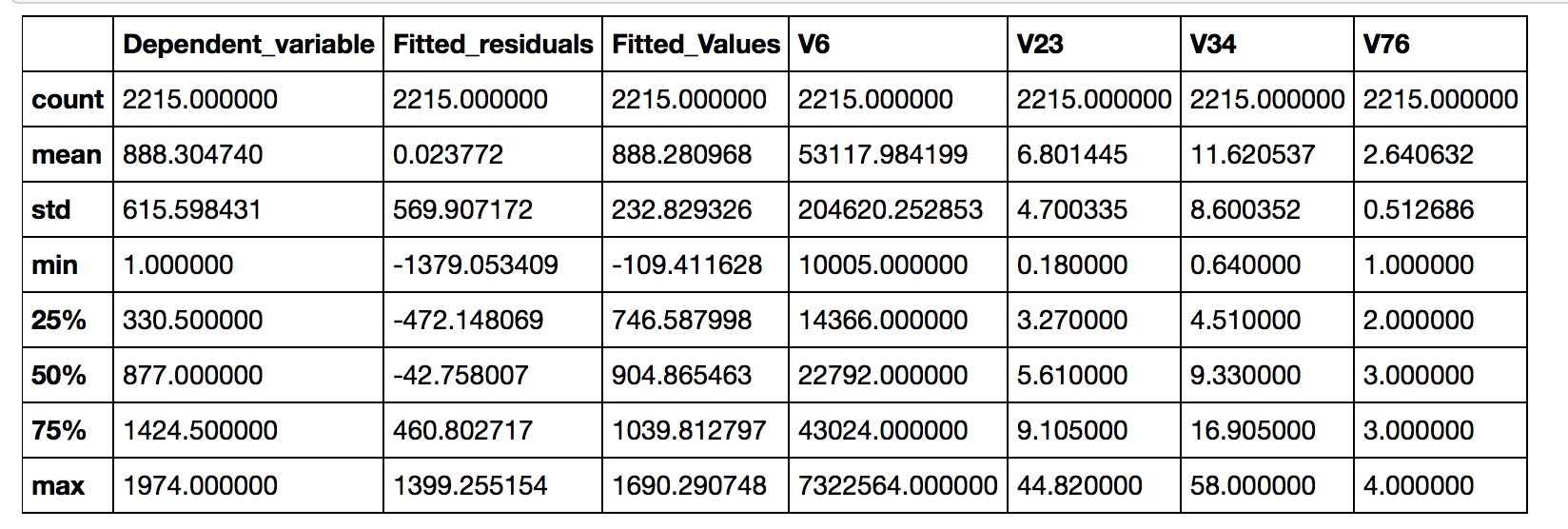
4. Examine the columns of predictor variables. Note any predictors that should be transformed, dropped or interacted to improve the model.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1. Load & summarize data

Data was given in xlsx format, first converted into cvs format for further processing.

Summary (Statistics)



1. Model

Linear Regression using Ordinary Least Square Method

OLS is one of the most common methods for estimating the unknown parameters in a linear regression method. It minimizes the difference between the observed response and the response predicted by the linear approximation of the data.

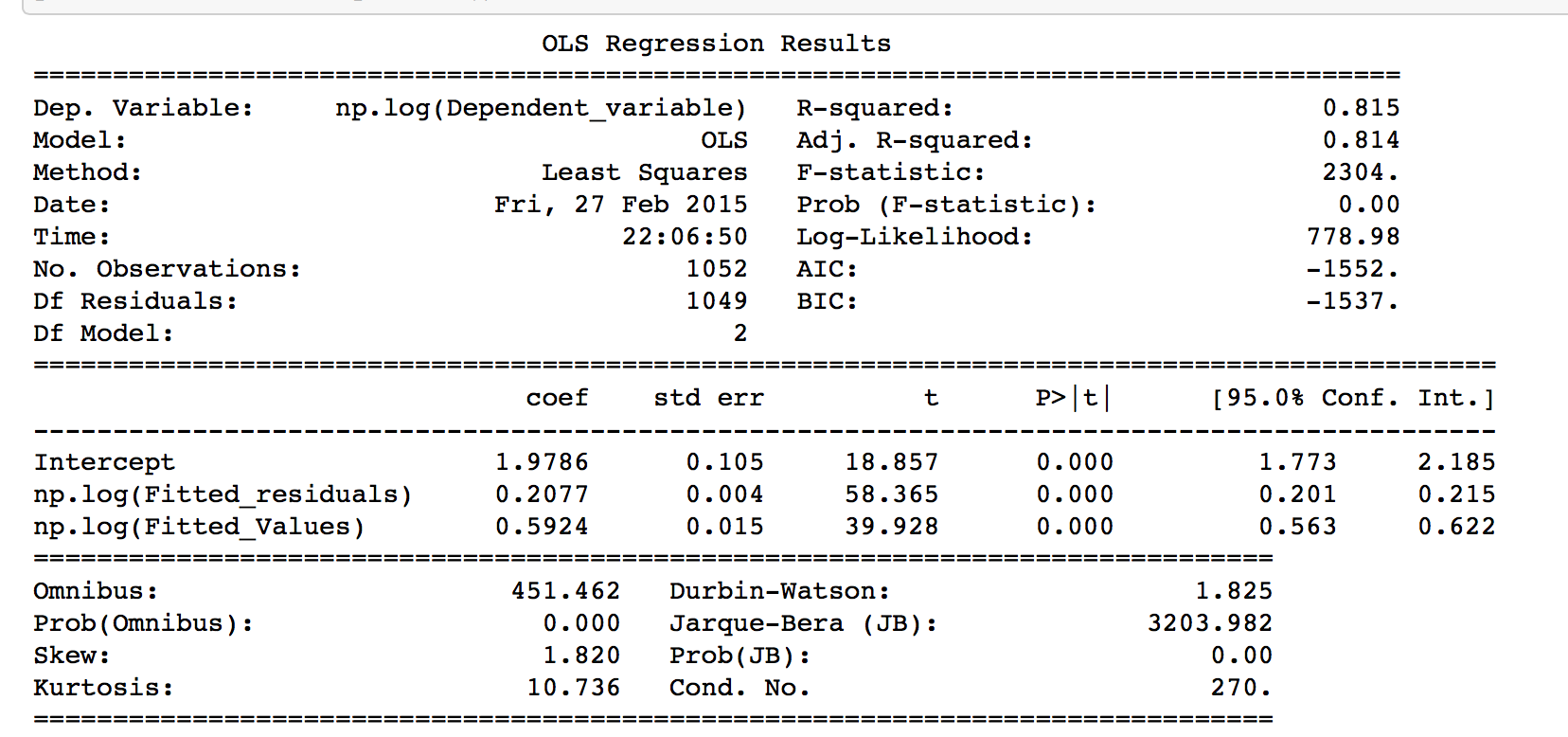
3. OLS Method

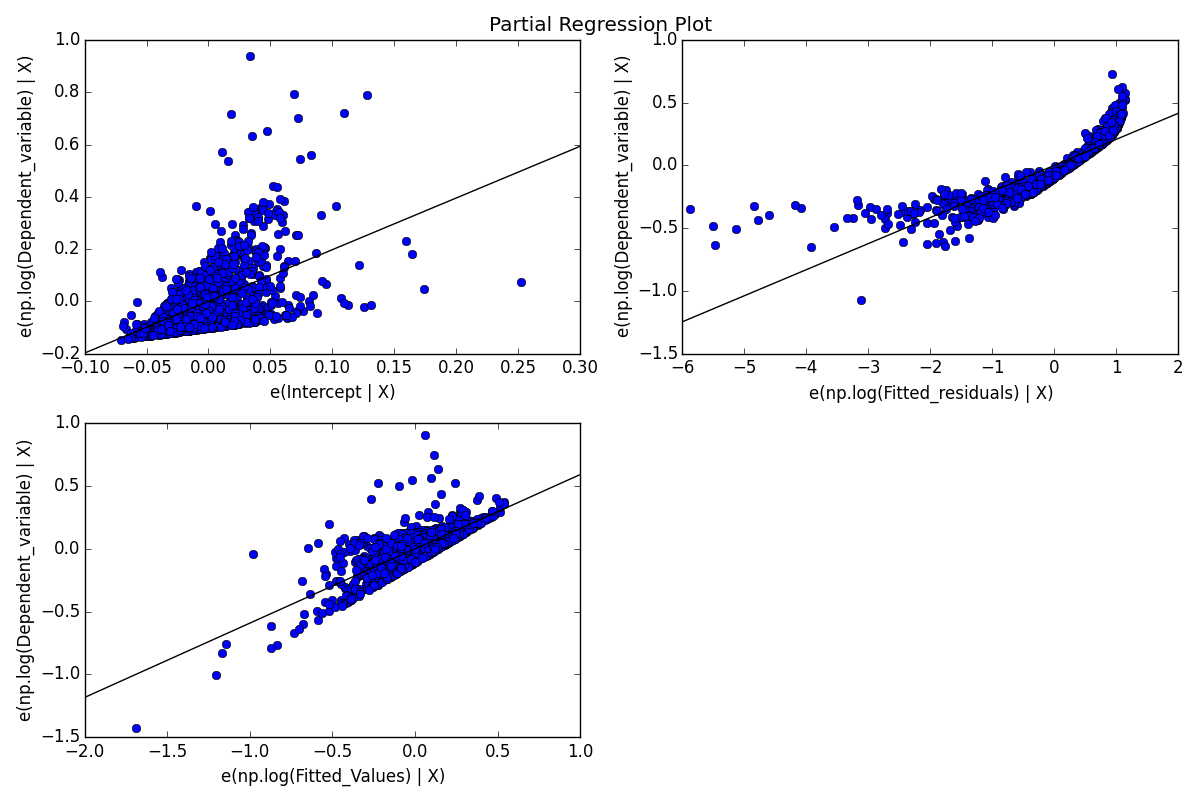
Since OLS minimizes what is known as “noise”, or the deviation of the actual data from our model, which is the straight line through the data points. It can be used for multiple variables and determine the line of best fit for a model.

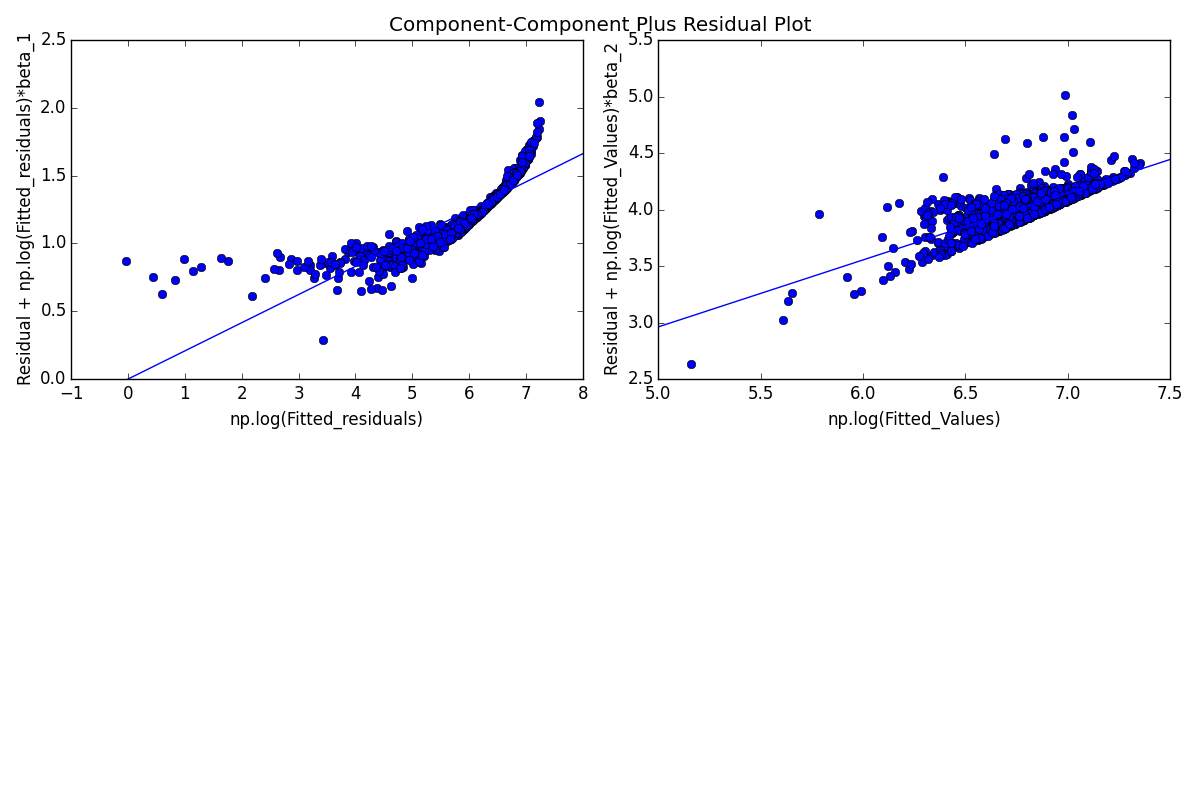
However, some of the values of the residuals are negative in sign while others are positive. If we sum the residuals, positive values will cancel out negative values so the sum will not accurately reflect the total amount of error. To get rid of this problem we square the residuals before we add them together.

It’s a good method to establish a baseline. The estimator is unbiased and has minimum variances.

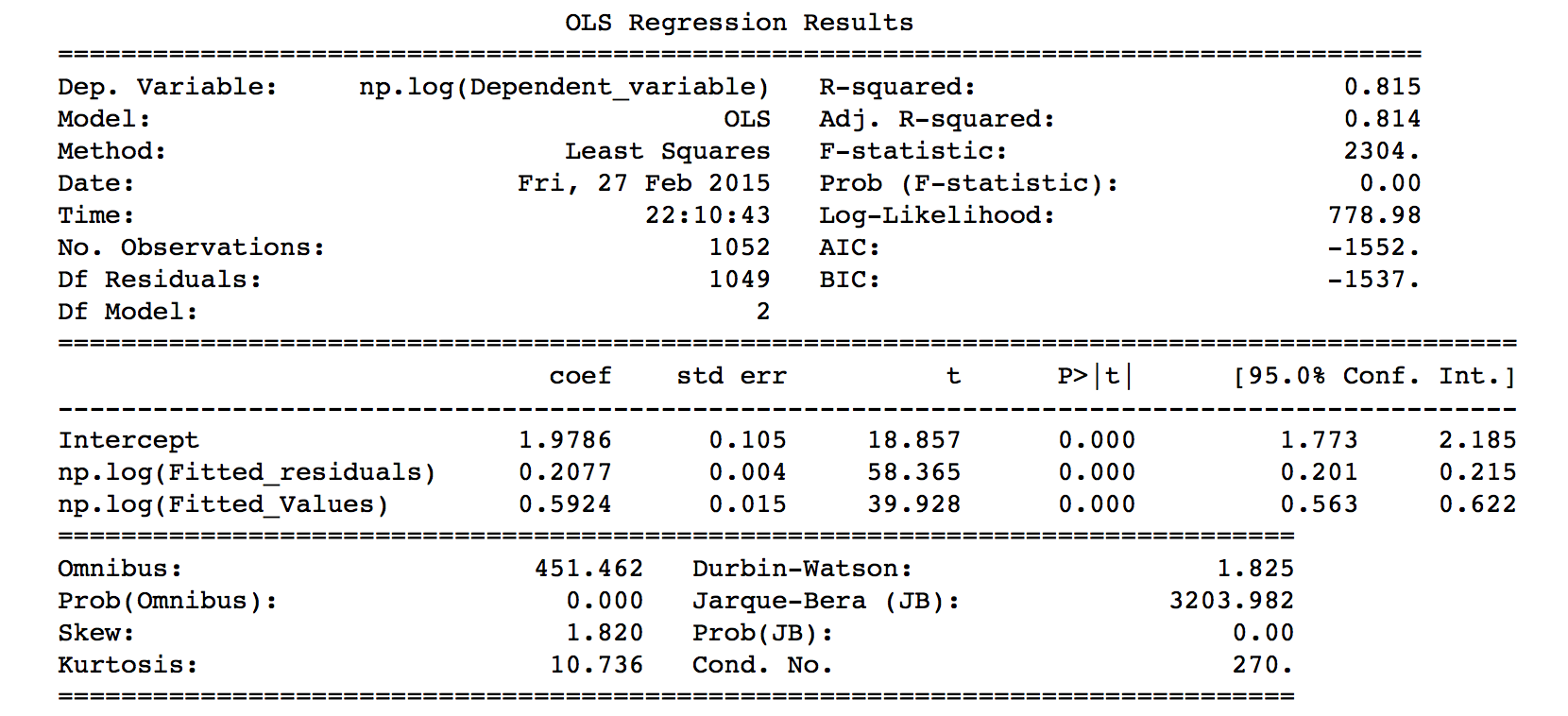
Results : considering 'Dependent\_variable ~ Fitted\_residuals + Fitted\_Values'

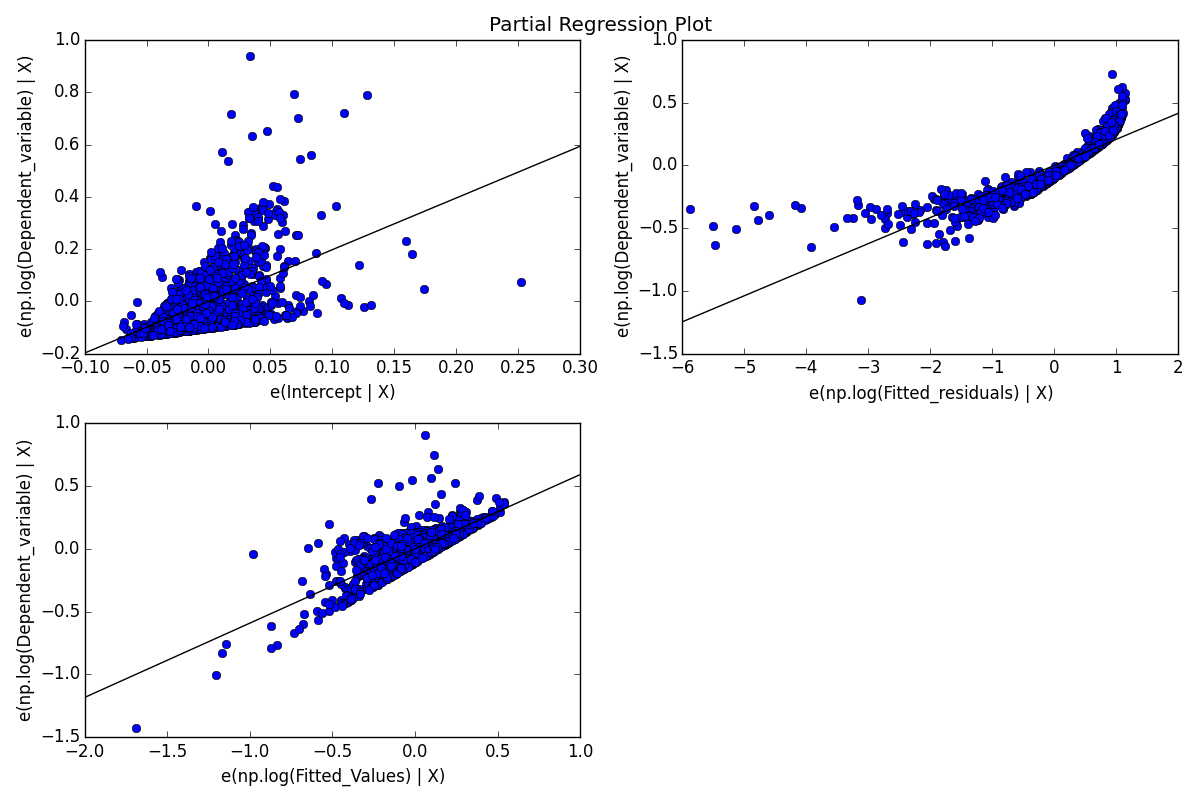


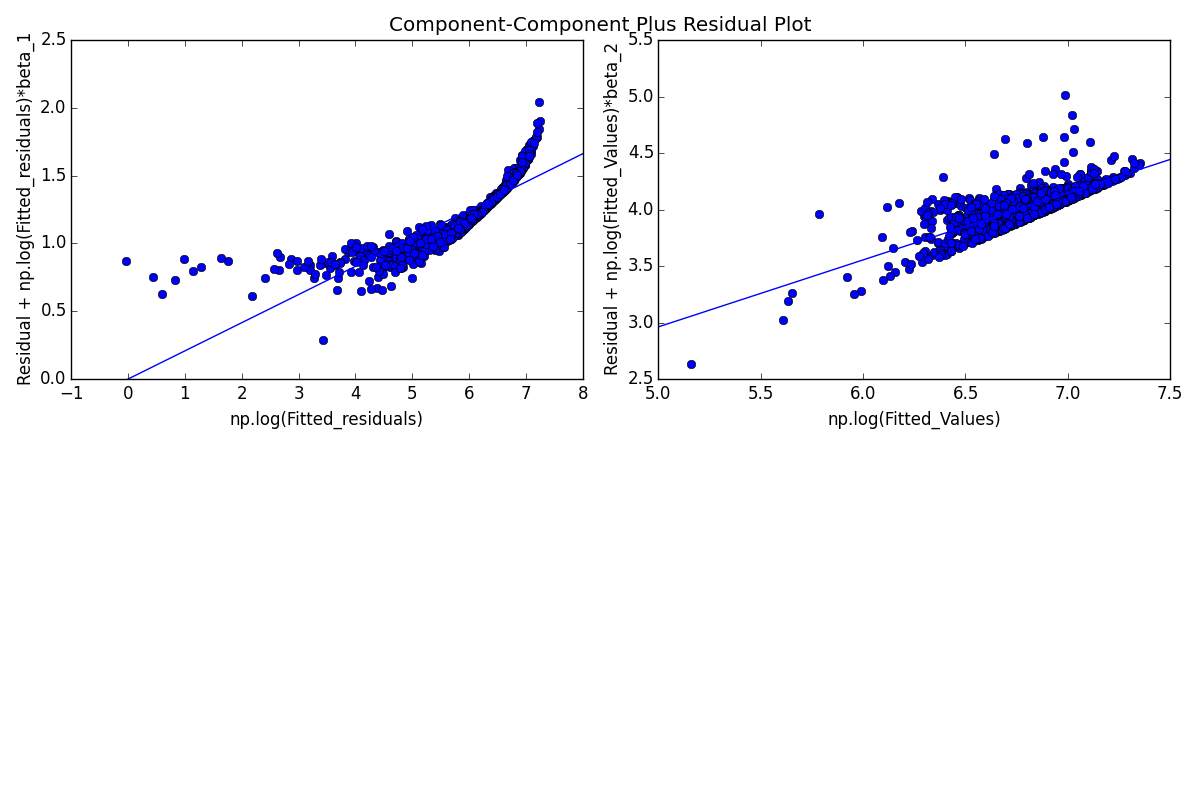




1. Including the predictor column variables







Using the subset data approach doesn’t list any change in output.

b) using standalone values has a significant change on intercept values

