Subject: Re: Regression fit and random noise
Posted by Phillip Bitzer on Thu, 28 Mar 2013 23:18:50 GMT
View Forum Message <> Reply to Message

The (linear) correlation coefficient, r, is a measure how well the independent/dependent variables are correlated. For perfectly correlated data, r = 1, and the data plots as a straight line with positive slope. Perfectly anit-correlated data has a r=-1, and the data plots as a straight line with negative slope. Uncorrelated data has r=0; in this case, the best fit line has a slope of zero (imagine data points that are scattered with no perceptible trend). (You're dealing with the multiple correlation coefficient, but the concept is similar. There's a nice discussion in Bevington, among other places. BTW, the multiple correlation coefficient can be shown to be a linear combination of the linear correlation coefficients for each variable x_i. Further, the linear correlation coefficient can be used to assess the usefulness of a predictor in the model.)

In your case, setting noise ratio = 0 should provide the same value as if no noise was present because no (artificial) noise is present! As you increase noise_ratio, you're essentially "destroying" the correlation, in a manner of speaking. I bet if you crank up noise_ratio far enough you can get essentially uncorrelated data.

Be careful when you speak of a "good fit" - there ways to qualify what is a good fit (for example, using the chi squared value to test the null hypothesis). Depending on the SNR, the model will still be a "good fit" to the (noisy) data.

Ultimately, the answer to your question lies in the underlying statistics - there isn't (shouldn't be?) anything wonky going on in IDL.

Hope this helps!