

My stat teacher said I was supposed to plot the data

Posted on Aug 16, Posted by [Tom Reilly](#) Category [Forecasting](#)

Your Stat teacher told you a lot of things. Mostly wrong. The way Intro stat classes go is that they start with the wrong things and slowly as you evolve up the level to a PHD they finally start telling you how to do it right. It starts off with decomposing a series into Seasonality, Trend, Levels. They dabble in this and that (ie exponential smoothing, trend, trend squared, logs). Then they break out the surprise a few years into the Stat degree process and tell you that the errors need to be N.I.I.D. Hold on. So, everything you taught me about exponential smoothing and Holt-Winters violated the gaussian assumptions? How about your current software? Does it verify that the model or does it just fit a model from a list? I want my money back. I will refer you the [Meat Loaf](#) song to pick your spirits up here.

Every good statistician will tell you that you should plot your data. That might work fine when you have a couple of series, but no so when you have thousands. It might not work so well even when you have just a few. The reality is that it would take a very very strong analyst to tease out a model that detects the usual from the unusual or "signal and noise" that exists in data.

The process of identifying model that works can take you down many many paths often ending in a dead end. So, the process is iterative and long. Statisticians can spend a lot of time to do this and still end up with a half-successful model/forecast. There might be a level shift in the data due to legislation, competition, etc that may exist in the data that you might not even realize or two? Where does this level shift exist?? How to find it? In a word: An algorithm that iterates. An algorithm that tries a bifurcates the data to identify these level shifts. Or two. Or three.

We open up text books (yes even newly published ones) that disappoint, websites, posts on the blogs/discussion groups and see very simple approaches being brought to try and solve very nuanced data problems. I spoke someone at a Conference who had been out of the forecasting world for a bit and came back and said that she felt that things hadn't become more simple and not for the better. The 80/20 rule doesn't apply here. You can do better than an "B". You can get an "A" on your report card with a little more effort.

We see software that even rigs the game so that the model/forecast seems better as it says it fit the last withheld observations well. What happens when there is a level shift or an outlier in that withheld period? Well then you have a model that predicts outliers well.

See more about how little lies your teacher told you like taking LOGS and tricks get you into trouble [here](#)

Tags: Tagged in: [time series box-jenkins acf pacf level shift plot](#)