

IBM just launched SPSS Modeler 18. Let's look underneath the hood, but why?

Posted on Mar 20, Posted by [Tom Reilly](#) Category [Forecasting](#)

IBM released version SPSS Modeler 18 recently and with it a [30 day trial version](#) .

We tested it and have more questions than answers. We would be glad to hear any opinions(as always) differing or adding to ours.

There are 2 sets of time series examples included with the 30 day trial.

We went through the first 5 "broadband" examples that come with the trial that are set to run by default. The 5 examples have no variability and would be categorized as "easy" to model and forecast with no visible outliers. This makes us wonder why there is no challenging data to stress the system here?

For series 4 and 5 both are find to have seasonality. The online tutorial section called "Examining the data" talks about how Modeler can find the best seasonal models or nonseasonal models. They then tell you that it will run faster if you know there is no seasonality. I think this is just trying to avoid bad answers and under the guise of it being "faster". You shouldn't need to [prescreen](#) your data. The tool should be able to identify seasonality or if there is none to be found. The ACF/PACF statistics helps algorithms(and people) to help identify seasonality. On the flipside, a user may think there is no seasonality in there data when there actually is so let's take the humans out of the equation.

The broadband example has the raw data and we will use that as we can benchmark it. If we pretend that the system is a black box and just focused on the forecast, most would visually say that it looks ok, but what happens if we dig deeper and consider the model that was built? Using simple and easy data avoids the difficult process of admitting you might not able complicated data.

The default is to forecast out 3 periods. Why? With 60 months of data, why not forecast out at least one cycle(12)? The default is NOT to search and adjust for outliers. Why? They certainly have many varieties of offerings with respect to outliers, but makes me wonder if they don't like the results? If you enable outliers only "additive" and "level shift" are used unless you go ahead a click to enable "innovational", "transient", "seasonal additive", "local trends", and "additive patch". Why are these not part of the typical outlier scheme?

When you execute there is no audit trail of how the model go to its result. Why?

You have the option to click on a button to report "residuals"(they call them noise residuals), but they won't generate in the output table for the broadband example. We like to take the residuals from other tools and run them in autobox. If a mean model is found then the signal has been extracted from the noise, but if Autobox finds a pattern then the model was insufficient...given Autobox is correct. :)

There is no ability to report out the original ACF/PACF being reported. This skips the first step for any statistician to see and follow why SPSS would select a seasonal model for example 4 and 5. Why?

There are no summary statistics showing mean or even number of observations. Most statistical tools provide these so that you can be sure the tool is in fact taking in all of the data correctly.

SPSS logs all 5 time series. You can see here how we don't like the [kneejerk movement](#) to use logs.

We don't understand why differencing isn't being used by SPSS here. Let's focus on Market 5. Here is a graph and forecast from Autobox

