Handle with Care - R Packages - Are they business ready?

Posted on Oct 09, Posted by Tom Reilly Category Forecasting

Does free software do what you thought it would?

For ANOVA and t-tests, the R packages are just fine as this type of statistics is pretty basic with not a lot of moving parts, but when you are trying to model time series data or to do pattern recognition with that data is much much more difficult.

Is there "version" control with the R software packages? Yes, there seems to be. Errors are documented and tracked in the Change log section. Take a good close look at this log and the number of changes and the changes made.

Statistical forecasting software has been found to have different forecasts for the identical model and different estimates of the parameters. Bruce McCullough from Drexel University has spent a large part of his statistical career publishing journal articles that debunk forecasting software and their errors. Bruce first railed on Excel's inability to be trusted as a reliable source for statistics. Others have taken up that same cause with Google's Spreadsheets.

A paper by Yalta showed problems with ARIMA modeling benchmarks in some packages and showing Autobox, GRETL, RATS, X-12-ARIMA to be correctly estimating models. The references at the bottom of that paper show the main papers in this area of research, if you are interested. Many of them McCullough's.

At a recent meeting with customers and prospects, the topic of whether R packages could be used for Corporate analysis came up. We can tell you that it is being used for Corporate and

personal and Dissertations and on and on. We shared our experience with someone testing out the auto.arima forecasting package which had a model which was flawed. Models are debatable for sure, but having "buggy" software is not just acceptable for a business or even research environment as bad forecasting has bad consequences. We would like to help you in your evaluation of your models. One way for you to do this is to take the residuals from your forecasting software model and enter it into a trial version of AUTOBOX (http://www.autobox.com/cms/index.php/30day). If AUTOBOX finds a model, other than a constant, then you can conclude that your model missed a piece of information and that you chose wrong with your current tool. Sometimes software is worth what you pay for it.

Most software has releases every year or every other year. Extensive testing is performed on benchmarks to identify errors and prove that the software is stable. Most software uses Regression testing to identify and correct for issues. If a version gets created every other month or week can you trust it to run your Enterprise or even Dissertation?

Tags: Tagged in: cran ets auto.arima gretl rats