

Information Overload: The Risk Documentation Challenge

By Tony Hughes, February 8, 2019

<https://www.garp.org/#!/risk-intelligence/technology/all/a1Z1W000004PbidUAC>

For banks, one of the more onerous aspects of regulatory supervision is documentation. Regulators demand detailed descriptions of the models being used and banks have responded by the bucket-load.

To cite one example, my team reviewed some PPNR models for a CCAR bank a couple of years back and found that each model was accompanied by a 200-page tome. To put this into context, this was about four pages of text for every data point that was available to the researchers to construct the models.

Why do bank analysts persist in producing doorstop-sized model documents? Put simply, they do so because generating an extra 100 pages is cheaper for banks than being officially reprimanded by a regulator. A poorly-documented model can carry a disproportionately large reputation risk, since it gives the impression that the bank has something egregious on its books that it desperately needs to hide.

Supervisors, meanwhile, encourage verbosity by equating documentation length with transparency. It takes a lot of skill to read a concise description of a model and conclude that the output is sound.

Burgeoning documentation can have very real consequences for model quality. One way to fill those additional pages is to run every conceivable test that may be applied to a given model. In such cases, managers are taking a risk that some of the test results will actually be believed, possibly forcing the abandonment of a sound, fit-for-purpose specification.

A perfect model should fail roughly 5% of the diagnostic tests applied to it; if a model passes 100 tests with no failures, you should either suspect fraud or some outrageous form of data strip-mining.

There are many situations in medicine where screening tests are not advised, and similar rules should pervade the world of risk modeling. The specific rules should depend upon the purposes to which the model will be put.

If I am building a stress testing model, for example, I primarily care about medium-term forecasting properties and how well the model responds to different economic scenarios. I also care about the

stability of the model – whether the data are likely to continue to behave in the manner they have exhibited in recent years or decades.

To these ends, many tests are irrelevant. You can draw a thick line through tests of residual normality and most forms of heteroscedasticity. You can substantially discount serial correlation, unless it is so extreme that it threatens the stability of the forecasts (suggesting the presence of other, deeper problems with the model). If you are concerned about inference in the presence of any of these issues, high-quality robust statistics are usually available to aid your quest.

The stuff that matters relates to forecast accuracy – mean-squared error against a challenger – and issues of parameter or model specification stability. This latter category includes various forms of stationarity/unit root testing, though the researcher must pay very careful attention to the size and power properties of the tests being used.

A doctor would never prescribe a test with no power, and neither should you. If the test is little better than a coin flip, it should not be believed and does not deserve two pages of a bank's model documentation.

A special mention should be made of multicollinearity. A few years ago, a colleague and I did some [research](#) that showed that stress test models perform better if this “problem” is actively ignored.

Put simply, such models often suffer from too little multicollinearity. This result is not particularly surprising: the standard textbooks will tell you that correlated regressors are basically irrelevant when it comes to forecast performance. Our exploration went further than the traditional literature, concluding that concern for such phenomena actually hampers the performance of the model when it is specifically designed for stress testing.

Despite this, multicollinearity appears to be the bane of model validators, prompting acres of unnecessary (or even harmful) text in model documents.

Parting Thoughts

So, what should be writ?

If the entire modeling team succumbs to food poisoning after the holiday party, obviously the replacement team should be able to pick up the strands. This means that the data dictionary should be well maintained, and any code used to generate the forecasts should be archived with a carefully constructed knowledge base that describes the process used to run the models.

In terms of justifying the modeling choices made, describing the theory and running necessary diagnostic tests, a bare-bones academic style of writing should be adopted. If a particular test is mission critical, it should be included, even if the sample is small.

For a PPNR model, built with fewer than four score and seven data points, a four-to-eight-page paper should suffice; any longer and you are probably placing too much weight on flaky statistics whose signals should not be trusted. For more material loan-level PD models, where many thousands of observations are usually available, 15 to 20 well-crafted pages would probably be ample.

With more data, you have more degrees of documentation freedom. Tests that use more data are likely to have a modicum of power, allowing them to be implemented and reported and then trusted by the reader of the document.

There are two ways to obfuscate – say nothing or divulge everything. The next time you see a three-variable regression backed by 200 pages of text, you should start to wonder where the bodies are buried.

Tony Hughes is a managing director of economic research and credit analytics at Moody's Analytics. His work over the past 15 years has spanned the world of financial risk modeling, from corporate and retail exposures to deposits and revenues. He has also engaged in forecasting of asset prices and general macroeconomic analysis.