

## The Limits of AI in Banking

**This article was originally published on GARP's Risk Intelligence website on Friday, September 21, 2018**

---

*By Tony Hughes*

Let's say I recently ventured to a Himalayan mountaintop. There I met a strange hermit who bestowed on me a precious gift: the specification of the true model for predicting credit card default.

There were two catches: (1) she said nothing about how the model would perform in a recession; and (2) she emphasized that my gift would be lost if I divulged its secrets to any living soul. When asked to clarify, she said that I could never share the model's secrets with bank examiners and model validation teams.

After the encounter, I got home, made some forecasts, waited, and found that the model was as promised. While I couldn't specifically predict defaults, my predictions would, in the long run, be far more accurate than those produced by other analysts.

Given these parameters, would a bank be interested in my modeling services? Even though I'd have no meaningful documentation, I think most banks — after an extensive proof of concept — would be willing to give me a chance.

This situation, which I concede may be apocryphal, is not a million miles removed from the promise of machine learning and artificial intelligence. These new modeling tools can often be shown to perform better than traditional methods. The downside is that the rationale behind decisions made by an algorithm tends to be indecipherable to the average bank executive or regulator, and even the analyst who actually designed the algorithms in the first place.

## **Gaining Credibility**

In the past, “black box” solutions have generally been viewed with suspicion in the banking industry. The point of credit models was to enable bank executives to gain greater insight and understanding of the nature and performance of their portfolios, helping them make better decisions about what risks were worth taking.

Now, it seems, the “machine learning” moniker is gaining credibility and, as long as the new model works demonstrably better than the status quo, it'll have a good shot at being implemented at a bank regardless of its complexity.

In cases where the success or failure of a model-based solution can be precisely determined, such as assessing the credit-worthiness of loan applicants, it's hard to imagine people-based decision systems surviving for much longer.

For example, I remember working with a risk analyst in the late 2000s who was asked to assess the performance of a mortgage application review team. Applications with middling credit scores were parceled off to this team for review, after which prospective clients were either accepted or rejected. What the bank discovered was that using just the model-based score, with a strict cut-off, led to higher revenue and lower credit losses. Even if the team worked for free, in other words, it still would have been preferable for the staff involved to be redeployed or laid off.

## **Stress Testing Complexities**

In cases where the best model is harder to determine — more strategic problems like stress testing and capital planning spring immediately to mind — the benefits of black box techniques will likely be elusive. Stress test models are, after all, notoriously difficult to invalidate.

The point of the stress test model is to predict behavior during a banking system crisis, but we have had only one such event during the past 20 years. We can show that our model exhibits rising losses in a 2008 back-test, but no one can say whether the Great Recession was a “typical” stress event in any meaningful sense.

If we could measure stress test model performance precisely, and if an artificially intelligent model was found to perform substantially better than the alternative, there would be no valid argument against its

application. An examiner may bemoan his lack of understanding of the technique, but he would be unable to deny its superior performance.

In the real world, given the difficulty in truly validating stress test models, we must instead content ourselves with the fig leaf of model interpretability.

### **Uncertain Future**

Of course, if artificially intelligent systems were smart enough, separate stress testing models would be unnecessary. If the machine had seen it all before (recession, famine, revolution and war) and developed the smarts to both gauge the likelihood and calculate the capital needs of the bank in the unlikely event, the executive's only role would be to feed the dog tasked with preventing human intervention in the prediction process.

However, even if we were armed with data on every banking interaction in human history, we would still be millennia removed from this point. Such events are just too strange and infrequent for us to allow data-based methodologies to be central in crisis prediction and mitigation.

In a world where AI models completely took over, the other key role for executives would be deciding when and how to check the machines, and, if needed, switch them off. This could happen — if external conditions change or if customers or competitors adapt to exploit weaknesses in the artificially intelligent system.

I suspect this will be the next big challenge for researchers. We know that new techniques predict accurately under current conditions. But can we now demonstrate the circumstances in which the AI model will one day break?

*Tony Hughes is the managing director of economic research 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.*