Before #BigData, Let's Confront #BadData

Gathering complete and accurate data is critically important and painfully challenging. Even so, don't trust the data completely. (Suggested sub-title / deck)

Two critical lessons of recent years – all years, really – are that financial models are inherently fallible and that all organizations must "scrub" their internal data for completeness and accuracy. Beyond models and internal data, there lurks a larger, darker hazard. Information and data that swirl around the globe and enter your firm's models, databases or employees' brains have widely varying levels of accuracy and credibility. Just as astute risk professionals interpret model results with skepticism and caution, so too must we refrain from trusting without reservation the external data and information we use.

Don't trust models, don't trust data

Many experts propound the worthy and valid warning that financial models are, at best, merely approximations of reality. The common phrase "model risk" captures this sentiment. In an earlier *Quant Perspectives* column (*How to Build Disastrous Financial Models*), we advised caution in creating and employing models and developed guiding principles. "Good faith" in the crafting and running of models is the most essential principle.

As with models, the generation and interpretation of data require "good faith." Data and information of external sources that enter your firm have unknown context and quality. Do not trust these external inputs without review and validation. The suppliers and vendors to automakers and large retailers must undergo the careful scrutiny of their large customers. By the same token, financial institutions should perform deliberate diligence on their data suppliers.

Data doesn't lie?!

Physicists have an old joke that goes like this: "Nobody believes a theory except the person proposing the theory. Everybody believes an

Joe Pimbley

experiment except the person performing the experiment." Like all good jokes, it's funny because it holds so much truth. Healthy (or even pathological) skepticism is the hallmark of good science. Always assume new ideas and theories and models and experimental results are wrong. Let the proponent provide persuasive evidence and argument to convince you otherwise. Do not simply believe what you're told or shown.

Science has worked out its first line of validation. For a theory, the theorist must *make a prediction* by proposing experiments and measurements that have never been performed in which the results would show clearly that the new theory is superior to existing theories. If there's no prediction or no successful real-world confirmation of the prediction, then the theory is nothing more than one person's musing.

Scientific protocol demands the validation of experiments also. Instead of describing this review, let me go back to the joke and explain that the experimentalist knows best how uncertain her techniques can be, how difficult it is to tune and calibrate equipment and apparatus and how challenging it is to control all the extraneous factors that might distort and contaminate the environment of her measurements.

There's the old saying that "data doesn't lie." But "data does lie" in the sense that it can be wrong or invalid or misleading in the relevant context. Purveyors of data may have intentional or unintentional bias or simply be reckless in their methods. Examples of recent years include LIBOR settings, bids and offers of HFT ("high-frequency trading") algos and mortgage loan data. Deliberately false data underpins all Ponzi schemes.

Good measurements get corrupted

Success will even poison otherwise "good" measurements. Consider FICO scores for U.S. consumers. This numerical assessment of borrower creditworthiness is important, relevant and arguably accurate in past time periods. Having a high FICO score helps consumers gain access to lending in the form of credit cards and auto and mortgage loans. Perceiving this connection, financial advisors suggest that consumers take deliberate actions (such as opening unneeded loan accounts) to improve their FICO scores that

Joe Pimbley

have little bearing on their actual ability to repay loans. As a consequence of the visible importance of FICO, the scores themselves lose value as creditworthiness indicators. Financial underwriting methods that employ FICO scores should therefore adjust and adapt.

Another example is the quarterly GDP ("gross domestic product") measurement of all economies. GDP and other economic data become "political footballs" and "weapons of misinformation" for a government that wishes to convince citizens and foreign investors of its competence (in other words, essentially all governments). Too many financial analysts believe stated numbers, such as GDP, or information in the form of pronouncements, such as "all large banks are solvent," without considering the motive to mislead of the purveyors of the data and information.

What is the responsible role for data?

Gathering, validating, managing and communicating data are critical to financial decision-making and risk management. But such data proficiency is just one element. If there are proponents of "Big Data," "artificial intelligence," or "machine learning" who claim that data algorithms alone govern the financial world or any other complex human system, we disagree.

As we argued in *How to Build Disastrous Financial Models*, the two best uses of models are (i) the insights one gains from creating the models and (ii) the models' quality-control exercise of testing the data. We must build and interpret models in "good faith" with the mission of finding the *right* answer rather than an answer that simply achieves a short-term goal.

The best role for data is to serve as input for judgment- and concept-based executive decisions. When the data is "too good to be true" or otherwise in conflict with valid concepts of the business and unbiased human judgment, we must channel the experimentalist and ponder every reason the data itself might be wrong.

Joe Pimbley, FRM is a financial consultant in his role as Principal of *Maxwell Consulting, LLC*. * His expertise includes enterprise risk*

Joe Pimbley

management, structured products, derivatives, investment underwriting, training, and quantitative modeling. Joe is co-author of "[Banking on Failure – Fixing the Fiasco of Junk Banks, Government Bailouts, and Fiat Money](#)."

Joe Pimbley