Journal of Risk Finance, Vol. 8, No. 4, 2007.

EDITORIAL

Thoughts on the Scientific Method:

Part 2 – Frequentist Fecklessness

Michael R. Powers*

This editorial is the second of a two-part critique of the SM *as it is commonly practiced*. In the first part,[1] I addressed the problem of inconsistencies in the way the SM is applied, and argued that these inconsistencies may lead researchers to ignore important phenomena, especially when the phenomena are of a subtle and/or transient nature. In the present installment, I consider problems of inefficiency and inertia caused by the SM's collectivist, frequentist orientation, and show how these problems may be avoided by a more individualist, Bayesian approach.

We begin by recalling from Part 1 the following conventional outline of the "scientific method":

(1) The development of a null hypothesis (H_0) summarizing the relationship to be tested.

(2) The identification of an observable phenomenon that is predicted by the underlying hypothesis, and not otherwise explicable.

(3) The design of an experiment comprising a reasonably large number of individual observations, each of which reflects the presence/absence of the identified phenomenon.

(4) The replication of the experiment by several independent researchers.

* Editor, *Journal of Risk Finance*; Professor and Co-Director, Advanta Center for Financial Institutions Research, The Fox School, Temple University; e-mail: michael.powers@temple.edu.

Circular Reasoning?

Turning to the SM's philosophical underpinnings, one quickly encounters what appears to be a problem of circular logic. This is because the SM, as a process of human understanding, is entirely observational in nature, and at the same time, its intellectual justification seems to arise exclusively from its observed successes. In other words, to place credence in the SM, one must already accept the SM as a means of generating belief.

This apparent shortcoming is a fundamental problem of *British empiricism*, the philosophy that the SM formalizes. Developed by the late renaissance writers John Locke, George Berkeley, and David Hume, empiricism argues that human knowledge derives from the observation of objects in the external world through the body's senses (induction), and that internal reflection is useful primarily in identifying logical relationships among these observables. In his famous treatise, *An Enquiry Concerning Human Understanding* (1748), Hume acknowledges the issue of circularity as follows:

> We have said that all arguments concerning existence are founded on the relation of cause and effect; that our knowledge of that relation is derived entirely from experience; and that all our experimental conclusions proceed upon the supposition that the future will be conformable to the past. To endeavor, therefore, the proof of this last supposition by probable arguments, or arguments regarding existence, must be evidently going in a circle, and taking that for granted, which is the very point in question.[2]

Hume ultimately sidesteps this problem by arguing that the intellectual justification for relying on observations is simply human "custom" or "habit". "All inferences from experience, therefore, are effects of custom, not of reasoning," he writes; and subsequently: "All belief of

matter of fact or real existence is derived merely from some object, present to the memory or senses, and a customary conjunction between that and some other object."[3]

I mention the issue of circularity not because I believe it to be a true shortcoming of the SM, but rather because it shows how easily many of today's science devotees will accept the SM based upon an uncritical enthusiasm akin to religious zeal. Surely, one who bases a belief in the SM solely upon its successful track record is no more "scientific" than one who bases a belief in creationism solely upon the Book of *Genesis*. Tautology is tautology.

So how does one escape the SM's circularity problem? Chronologically, humanity had to wait only slightly more than thirty years from the publication of Hume's cited work to a clever resolution provided by the German philosopher Immanuel Kant. Famously stating that he was roused from his "dogmatic slumber" by reading Hume's work, Kant spent ten years of his life trying to reconcile the arguments of empiricism with those of *rationalism*, a philosophy propounded by writers such as René Descartes, Benedict de Spinoza, and Gottfried Wilhelm von Leibniz. In contrast to empiricism, rationalism argues that human knowledge can be derived from internal reflection (deduction), and that the experience of the external world is useful primarily for providing practical examples of internal ideas.

The principal fruit of Kant's ten-year labor was *The Critique of Pure Reason* (1781), in which he proposed the philosophy of *transcendental idealism*. This philosophy argues that certain types of knowledge, such as mathematical reasoning and the understanding of time and space, are built into the human mind. Subject to these *a priori* conditions, human beings learn about the world through observations as they are processed by, and conform to, the *a priori* conditions. Using this type of approach, it is a simple matter to remove the circularity of the SM by arguing that acceptance of the SM is just part of the mind's *a priori* construction – for

example, one could argue that we innately believe that a principle that holds true at one particular point in time and space also should hold true in neighboring points (i.e., during later repetitions of the same experiment). We then are free to employ the SM to our hearts' content.

Feckless Frequentism vs. Bespoke Bayesianism

My primary concern with the SM's philosophical framework is its conventional formulation in terms of frequentist hypothesis testing. The decision to select the frequentist approach over its Bayesian counterpart is generally made implicitly, and for the unstated reason that, since the SM is supposed to represent a collective decision-making process for all of humanity, it would be impermissible for different individuals to employ their own individual prior distributions over the unknown parameters. This idea was expressed clearly by Ronald A. Fisher in his book *Statistical Methods and Scientific Inference* (1956):

> As workers in Science we aim, in fact, at methods of inference which shall be equally convincing to all freely reasoning minds, entirely independently of any intentions that might be furthered by utilizing the knowledge inferred.[4]

While this approach seems quite reasonable in certain situations – for example, the measurement of particle masses in physics or melting points in chemistry – there are other contexts – such as financial planning and decision making – in which such a process is both slow and inefficient.

Consider, for example, the two hurricane-prediction models discussed in a recent editorial (see Powers, 2006), and suppose (rather fantastically, of course) that the simple "technical" model proposed by the author represents the current state of human knowledge, whereas the more sophisticated "fundamental" analysis of William Gray represents a brand new and purportedly superior alternative. Then, under the SM, the null hypothesis would be given by

$H_0 \equiv$ "Dr. Powers' model is best for predicting the number of Atlantic Basin hurricanes in a given year," and the alternative hypothesis would be $H_1 \equiv$ "Dr. Gray's model is best for predicting the number of Atlantic Basin hurricanes in a given year."

Putting ourselves in the shoes of an independent academic meteorologist who has just embarked upon the study of hurricane frequency, it is fairly clear how we should proceed with our work. Having just learned of Dr. Gray's model, and how it was shown statistically to be a better historical predictor of the numbers of Atlantic Basin hurricanes using a given collection of historical data, we would look for ways to replicate Dr. Gray's analysis using different sets of data. Perhaps we would seek alternative measurements of the same input values used by Dr. Gray for the Atlantic Basin, or perhaps we would compare the two models in a more limited geographical context (e.g., among hurricanes making landfall in the United States). In either case, after conducting our study we would publish our results, thereby providing more or less support to Dr. Gray's approach, and helping to move forward the cause of good science.

Now, however, suppose that we are not an academic researcher, but rather the director of a risk-management and emergency-planning agency for a state or municipality along the United States' Gulf Coast. Suppose also that we have been in this job for a number of years, and have witnessed first-hand the relative inaccuracy of Dr. Powers' forecasts as they were published year after year. Given the serious responsibilities of our position, we might find ourselves torn between wanting to embrace wholeheartedly Dr. Gray's model (because of its apparently greater accuracy) and sticking with Dr. Powers' model (because it represents the *status quo*, and as a public official, we cannot be too capricious in our decision making). As a consequence, we may decide to continue using Dr. Powers' forecasts in our planning, but to incorporate some information from Dr. Gray's as well; for example, we may use a weighted average of the two

forecasts, or use Dr. Gray's forecast only if it provides a more conservative (i.e., larger) predicted frequency.

In a third scenario, suppose that we are a hedge-fund manager interested in purchasing financial securities as part of an overall investment strategy, and that we are considering buying weather derivatives traded by the Chicago Mercantile Exchange for Atlanta, Georgia. To understand these instruments better, we do a little research into weather prediction, and soon come across both Dr. Powers' and Dr. Gray's hurricane forecasts. Because of our relative lack of expertise in this area, we contact a number of reputable meteorologists, and strangely, keep hearing things like: "Dr. Powers is just some insurance guy who doesn't know anything about weather systems," and "Frankly, I don't understand how Dr. Powers' model came to be so well established." As a result of this additional information, and given that our motive is solely to do the very best we can on behalf of our fund's investors, we decide to ignore Dr. Powers' model completely (i.e., to reject H_0 in favor of H_1).

From the above scenarios, one can see that R. A. Fisher's ideal of an SM that is "equally convincing to all freely reasoning minds, entirely independently of any intentions" is largely an impossible dream. While it may work reasonably well for academic researchers who are able to proceed at a leisurely pace and who are uncomfortable relying on private (and therefore subjective) information, it becomes more problematic for those in need of prompt answers to address important time-sensitive issues, as well as for those with private information who do not see any need to be constrained by published results.

The question arises, therefore, whether or not it is reasonable to embrace, explicitly, an alternative SM – let us call it a "personalized scientific method" (PSM) – in which one is free to: (1) advance the "current state of knowledge" at one's own pace (i.e., to select levels of

significance with which one is most comfortable, possibly higher than those used by most academic researchers); and (2) incorporate private information through the mechanism of Bayesian analysis.

Clearly, such a program of acquiring knowledge would generate a tremendous amount of chaos if it were to replace the SM. With every researcher "on a different page" from his or her colleagues, there would be no general standard for distinguishing facts from illusion and truth from wishful thinking. Recalling the inconsistency problems with the SM listed in the previous editorial, we can see that both the "apples to oranges" and "moving targets" problems would be greatly aggravated by a PSM, as researchers abandoned any pretense of consensus in the selection of experiments to conduct and allowed their private beliefs to be unduly influenced by "fishing" for statistical significance.

On the other hand, there would be some benefits of a PSM. For one thing, it likely would counteract the inconsistencies of censorship and "double standards" mentioned previously, as researchers would be free to incorporate "negative" results into their private beliefs and to offset bias in the selection of alphas with their own discretion in that regard. And perhaps the best argument for a PSM already has been hinted at: Since large numbers of researchers – like our public official and hedge-fund manager – already use one, why not simply acknowledge it?

Explicitly embracing a PSM does not have to mean abandoning the conventional SM. Researchers could continue to follow the SM as before, but simply inject more opinion, speculation, and subjective judgment into their analyses and conclusions. In fact, this is exactly what professional actuaries do in the field of insurance, where limited data often necessitate intellectual compromise by the averaging of forecasts and estimates from different sources.[5] Interestingly, insurance actuaries were some of the earliest researchers to employ Bayesian

analyses in their work, and it is undoubtedly this Bayesian influence that has allowed them to recognize formally the usefulness of private (subjective) information. In a world with an accepted PSM, researchers who found the new approach offensive naturally would be free to express their opinions, and to organize scholarly journals that adhere to the strictest standards of the SM (which, as we know, is no mean feat).

One of the supreme ironies of modern intellectual life is the great chasm that exists between rigorous science, on the one hand, and established religion, on the other. In today's world, the two opposing camps have reached an uneasy truce under which they officially profess to respect each other's domains, and even allow safe passage for some of their respective adherents to travel back and forth across the bridge that separates them. However, anyone who tarries on the bridge too long – or, God (Darwin?) forbid! – jumps off, is immediately branded a "quack" or "crackpot". A PSM would permit people to explore this middle ground without fear of ridicule.

So what are we waiting for?

References

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*.

Hume, D. (1748), *An Enquiry Concerning Human Understanding*.

Kant, I. (1781), *The Critique of Pure Reason*.

Powers, M. R. (2006), "Catastrophe forecasting: seeing 'Gray' among the 'black boxes'," *Journal of Risk Finance*, Vol. 7, No. 5, pp. 458-462.

Powers, M. R. (2007), "Thoughts on the 'scientific method': part 1 – ignorance through inconsistency," *Journal of Risk Finance*, Vol. 8, No. 3, pp. 209-213.

---

[1] See Powers (2007).
[2] See Hume (1748), Section IV, Part II.
[3] See Hume (1748), Section V, Part I.
[4] See Fisher (1956), Chapter IV.
[5] These techniques are often called "credibility methods," because weights are assigned to the different sources of information based upon their relative "credibilities", or intrinsic levels of believability to the decision maker. This terminology reflects a subjective Bayesian orientation.