Triumph of the Empiricists: *The Birth of Financial Data Science*

Joseph Simonian and Frank J. Fabozzi

Joseph Simonian
is the director of quantitative research at Natixis Investment Managers in Boston, MA.
joseph.simonian@natixis.com

Frank J. Fabozzi
is a professor of finance at EDHEC Business School in Nice, France, and the editor of *The Journal of Portfolio Management*.
frank.fabozzi@edhec.edu

The methodological foundations of contemporary econometrics were laid in the aftermath of a debate that was epitomized by Tjalling Koopmans' (1947) critical review of Arthur Burns and Wesley Mitchell's (1946) *Measuring Business Cycles*. In the review, "Measurement Without Theory," Koopmans, who was a member of the theory-focused Cowles Commission, argued that economic data cannot be properly interpreted without the benefit of well-hewn economic assumptions. The target of the review was not only Burns and Mitchell's book, but also the empiricist econometric methodology employed by the National Bureau of Economic Research, which Koopmans felt was overly preoccupied with devising techniques for measuring economic data at the expense of the development of the theory necessary to draw robust economic conclusions. In the review of Burns and Mitchell's book, Koopmans defines empiricism as a scientific methodology in which decisions about "what economic phenomena to observe, and what measures to define and compute, are made with a minimum of assistance from theoretical conceptions or hypotheses regarding the nature of the economic processes…" The motivating belief that drives Koopmans' argument is a committed philosophical realism regarding economic phenomena. Just as natural science assumes that physical and biological phenomena are regulated by natural laws, Koopmans assumes that economic phenomena are governed by their own set of immutable laws. If this is the case, then the job of the economist is to discover truths about economic reality in the same way that a physicist discovers (or is often assumed to discover) truths about physical reality.[1]

In the years following the theory versus measurement debate, as economics' theoretical footing was being reified, the field was also being increasingly formalized—to the extent that, by the early 1980s, the philosopher of science Alexander Rosenberg could confidently state that economic theory is most appropriately viewed not as a science, but as a branch of mathematics (Rosenberg 1983). In Rosenberg's characterization, economics abstracts away from actual human interaction and posits a set of basic assumptions from which it derives a formally impressive yet empirically empty set of conclusions. He ultimately argued that economics should be treated as "somewhere on the intersection between pure and applied axiomatic systems," whose findings may not correspond to any facts in the world but that are

[1] Koopmans forcefully argued his case and gave the impression that it is impossible to justify an empirically robust theory at all, given that you seemingly need to have a well-grounded theory to comprehend empirical evidence in the first place.

nevertheless interesting from an intellectual standpoint. Although Rosenberg's account of economics may be viewed as somewhat extreme and not reflective of how most economists view themselves and their profession, it nevertheless brings to the fore the extent to which economics is viewed by many as a largely theoretical endeavor. That being said, we do not need to subscribe in whole to Rosenberg's argument to recognize that a broadly theoretical approach has become dominant in economics—econometrics in particular—so much so that today we may succinctly summarize the primary beliefs that drive contemporary econometric practice as follows:

1. The goal of econometrics is to discover well-defined economic processes, mechanisms, and structures.[2]
2. Modern probability theory and statistical inference are indispensable tools in the definition and discovery of economic phenomena.[3]
3. An econometric methodology founded on points 1 and 2 can produce reliable economic forecasts, which can fruitfully be applied in business and policymaking.

Although econometrics is anchored toward the ideology of philosophical realism and strict adherence to the tenets of probability theory, as the quotation from Koopmans indicates, at any given time, the degree to which scientific methodologies are theory-laden may vary. Moreover, scientific frameworks in practice generally differ not by their choice of *either* a purely empiricist or realist methodology, but by the degree to which a given methodological program is guided by empirical considerations. Where economics has erred, in our opinion, is in allowing the pendulum to swing too far in favor of theory. In the physical sciences, although the tension between more theoretically and more empirically inclined methodologies exists, experiments are nevertheless considered indispensable tools for validating or invalidating theories. Experimental tools in the physical sciences are of course better developed than in economics, for a variety of reasons.[4] With the advent of data science, however, we believe that economics now possesses a tool with which economic theories can be tested in a more robust manner, using new and richer datasets. Accordingly, financial data science is well positioned to reorient financial econometrics toward a more empirical stance, a methodological position that was in fact advocated in an argument almost as old as Koopmans'.

THEORY, SHMEORY: AN INSTRUMENTALIST VIEW OF ECONOMICS

At around the same time that Koopmans was arguing in defense of economic realism and the importance of theory, another well-known economist, Milton Friedman (1953), presented an argument in favor of an empirical approach to economics. The strain of empiricism Friedman defended is usually labeled *instrumentalism* (although Friedman never mentioned the term) and emphasizes the predictive role of science, downplaying science's role as an unassailable arbiter of "reality."[5]

In the 20th century, different forms of instrumentalism were championed by a wide variety of thinkers, from Pierre Duhem (1914) to John Dewey (1916, 1938). Today, instrumentalism is an influential methodology in the physical sciences (Torretti 1999). In contrast to Koopmans, Friedman viewed assumptions as tools to be employed in the production of reliable forecasts. As such, in Friedman's instrumentalism, a theory need only be sufficiently coherent if it leads to successful predictions. This view of theory thus eschews, or at least radically downplays, its explanatory role and instead relegates it to a device to frame and guide the process of prediction, or as Friedman put it, "to serve as a filing system for organizing empirical material." Indeed, under an instrumentalist view of economics, the truth or falsity of the axioms and postulates of an economic theory is less relevant than the degree to which a theory facilitates successful prediction.

[2] This belief is perhaps best exemplified by econometrics' preoccupation with causality, a highly complex, not to mention metaphysical, concept that has been a major focus of philosophical analysis for centuries. For a sample of some of the extensive literature on causality in economics, see Haavelmo (1943), Simon (1953), Granger (1969), Hicks (1979), and Hoover (2001).

[3] For a classic statement and argument of this view, see Haavelmo (1944).

[4] To name just two, the ability to conduct closed experiments and to study subject matter that behaves more or less mechanistically gives the physical sciences the ability to confidently draw conclusions from experiments in a way that has hitherto been impossible in economics.

[5] For a review of Friedman's instrumentalism and its critics, see Boland (2016).

Contemporary econometrics' emphasis on theory versus prediction has been detrimental to the ability of the field to produce models with reliable forecasting ability, an outcome that explains its relative lack of influence on economics as a whole versus more "pragmatic" empirical work that generally proceeds with a broad theoretical stance but without a theoretical "straightjacket" (Summers 1991). This paucity of influence on the field as a whole is true even for some of the most popular econometric models, such as the Dynamic Stochastic General Equilibrium (DSGE) class of models, which have shown themselves to be unexceptional forecasting tools both in absolute terms and in relation to much simpler frameworks (Edge and Gurkaynak 2010; Edge, Kiley, and Laforte 2010). Why is this the case? It is surely not due to lack of sophistication on the part of the builders of these elaborate models. To the contrary, it may be due to their "square-in-the-circle" attempts to build predictive models within the strict confines of often complex econometric theories, rather than conforming theories to empirical findings. This approach to model building is, in addition to being less useful from a practical standpoint, also the antithesis of scientific practice; natural scientists, in general, evaluate and refine theories through empirical observation, not the other way around.[6]

FINANCIAL ECONOMETRICS FOR THE 21ST CENTURY

We believe that financial data science represents an advancement over the traditional econometrics toolbox. As a scientific endeavor, data science combines statistics and computing in an effort to uncover patterns in information that can then be used to assist decision-making. Although data science employs statistical concepts, its methodological approach is decidedly instrumentalist and is open to using any type of quantitative method, heuristic, or technique in so far as it is useful in producing accurate predictions and informed decisions, regardless of strict adherence to the tenets of any theory. The instrumentalist orientation of data science is precisely what makes it so useful for applications to investment research, a pursuit that is valuable only if it leads to practical results, namely the improvement of individuals' and institutions' financial well-being.

That said, we believe that financial data science is a discipline in its own right, and not merely the application of data science methods to finance. We hold this view for at least three reasons. First, finance brings with it a unique set of problems and puzzles that distinguish it from standard applications of data science, especially those in the natural sciences. The challenges that practitioners face in devising trading strategies, asset allocation, and financial risk management, for example, all require specific solutions. Second, financial time series possess unique characteristics that reflect their origins in human action and intentionality. The defining properties of financial time series such as volatility clustering, momentum, and mean reversion are prime examples of this. Third, modeling agents, especially the collective agents that constitute "the market," is an extremely challenging problem that demands specialized techniques. For these three reasons, we believe it would be a mistake to think that financial data science is merely one area of applied data science.[7]

Just as we believe it is a mistake to consider financial data science as simply a subset of data science, we likewise believe that it is a mistake to consider financial data science as a branch of financial econometrics. Rather, it would be more accurate to describe financial data science as encompassing traditional financial econometrics and expanding it with new techniques and a new orientation. Although financial data science brings its own set of formal tools to the analysis of time-series, cross-sectional, and panel data, it also brings with it a mathematical arsenal capable of dealing with disparate types of data—both structured data, which is the terrain of traditional econometrics, and unstructured data, such as textual and visual information. Moreover, financial data science has a distinctly applied and hence empirical orientation, dispensing with unnecessary theoretical machinery and abstraction in favor of methods designed to adequately frame and solve real-life problems. Its methodological orientation thus places it, and by extension finance as a whole, closer to engineering than to pure science.

From a historical standpoint, the emergence of financial data science represents both a resurgence of

[6] A classic example of this process is given by the set of experiments designed to verify the theory of special relativity (see, e.g., Robertson 1949).

[7] For a similar argument, see López de Prado and Israel (forthcoming).

instrumentalism as a scientific methodology in financial econometrics and, because of the introduction of a multitude of new analytical techniques, an enhancement of the pragmatic empiricism mentioned earlier. By prioritizing successful prediction and usable results, financial data science promises to bring financial econometrics more in line with mainstream scientific practice and, in doing so, takes up the mantle in defending it and economics as a whole against critics who charge that economics is not a "real science." That financial data science is being increasingly recognized as an indispensable part of investment research is a testament to its practical value and a triumph of the empiricism on which it is founded.

ACKNOWLEDGMENTS

The authors would like to thank Marcos López de Prado for his valuable comments on an earlier draft of this article.

REFERENCES

Boland, L. A. "Reading and Misreading Friedman's 1953 Methodology Essay." In *Milton Friedman: Contributions to Economics and Public Policy*, edited by R. Cord and D. Hammond, pp. 541–560. Oxford: Oxford University Press, 2016.

Burns, A., and W. Mitchell. *Measuring Business Cycles*. New York: National Bureau of Economic Research, 1946.

Dewey, J. *Essays in Experimental Logic*. Chicago: The University of Chicago Press, 1916.

——. *Logic: The Theory of Inquiry*. Oxford: Holt, 1938.

Duhem, P. *The Aim and Structure of Physical Theory*. Translated from the second edition by P. P. Wiener. Princeton: Princeton University Press, 1914 (1954).

Edge, R. M., and R. S. Gurkaynak. 2010. "How Useful Are Estimated DSGE Model Forecasts for Central Bankers?" *Brookings Papers on Economic Activity* 41 (2): 209–259.

Edge, R. M., M. T. Kiley, and J. Laforte. 2010. "A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced Form Models, and a DSGE Model." *Journal of Applied Econometrics* 25 (4): 720–754.

Friedman, M. "The Methodology of Positive Economics." In *Essays in Positive Economics*, pp. 3–43. Chicago: University of Chicago Press, 1953.

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424–438.

Haavelmo, T. 1943. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica* 11 (1): 1–12.

——. 1944. "The Probability Approach in Econometrics." *Econometrica* 12 (supplement).

Hicks, J. *Causality in Economics*. Oxford: Basil Blackwell, 1979.

Hoover, K. D. *Causality in Macroeconomics*. Cambridge: Cambridge University Press, 2001.

Koopmans, T. C. 1947. "Measurement Without Theory." *The Review of Economics and Statistics* 29 (3): 161–172.

López de Prado, M., and R. Israel. 2019. "Beyond Econometrics: A Roadmap Towards Financial Machine Learning." *The Journal of Financial Data Science* 1 (1): 99–110.

Robertson, H. P. 1949. "Postulate versus Observation in the Special Theory of Relativity." *Reviews of Modern Physics* 21 (3): 378–382.

Rosenberg, A. 1983. "If Economics Isn't Science, What Is It?" *Philosophical Forum* 14 (3/4): 296–314.

Simon, H. "Causal Ordering and Identifiability." In *Studies in Econometric Method*, edited by W. C. Hood and T. C. Koopmans. New Haven, CT: Yale University Press, 1953.

Summers, L. H. 1991. "The Scientific Illusion in Empirical Macroeconomics." *The Scandinavian Journal of Economics* 93 (2): 129–148.

Torretti, R. *The Philosophy of Physics*. Cambridge: Cambridge University Press, 1999.

Disclaimer

This editorial reflects the current opinions of the authors, which are not those of Natixis Investment Managers.