Journal of Economic Perspectives—Volume 10, Number 1—Winter 1996—Pages 105–120

Macroeconomics and Methodology

Christopher A. Sims

This essay begins with a sketch of some ways I find it useful to think about science and its uses. Following that, the essay applies the framework it has sketched to discussion of several aspects of the recent history of macroeconomics. It considers skeptically the effort by some economists in the real business cycle school to define a quantitative methodology that stands in opposition to, or at least ignores, econometrics "in the modern (narrow) sense of the term." It connects this effort to the concurrent tendency across much of social science for scholars to question the value of statistical rigor and increasingly see their disciplines as searches for persuasive arguments rather than as searches for objective truth. The essay points to lines of substantive progress in macroeconomics that apparently flout the methodological prescriptions of the real business cycle school purists, yet are producing advances in understanding at least as important as what purist research has in fact achieved.

Science as Data Reduction

Advances in the natural sciences are discoveries of ways to compress data concerning the natural world—both data that already exists and potential data—with minimal loss of information. For example, Tycho Brahe accumulated large amounts of reliable data on the movements of the planets. Kepler observed that they are all on elliptical orbits with the sun at a focus, thereby accomplishing a sharp data

■ *Christopher A. Sims is Professor of Economics, Yale University, New Haven, Connecticut.*

compression.[1] Newton found the inverse-square law, allowing still further compression[2] and also allowing the same formula to organize existing data and predict new experimental or practical data in areas remote from the study of planetary motion.

Economics aims to accomplish the same sort of thing in relation to data on the economy, but is less successful. Whatever theory economists use to characterize data, the actual data always contain substantial variation that is not captured in the theory. The quality of the theory's characterization of the data tends to deteriorate as we extend it to data remote in time, location or circumstances from the data from which the theory was initially developed.

This view, treating science as data reduction, may sound oversimplified, but it is in fact a flexible metaphor that should not be controversial. The contentious issues should concern what "data" are to be characterized and what constitutes a "compression."

It was once common for economists to think of the scientific enterprise as formulating testable hypotheses and confronting them with data. True hypotheses would survive the tests, while false ones would be eliminated. The science-as-data-compression view lets us see the limits of this hypothesis-testing view. The latter is dependent on the idea that there are true and false theories, when in fact the degree to which theories succeed in reducing data can be a continuum. The theory that planetary orbits are ellipses is only approximate if measurements are made carefully enough. It does not seem helpful to say that therefore it is false and should be rejected. Furthermore, "theories" can be so complex that they do not actually allow important data reduction, even though a naive hypothesis-testing approach might accept them as "true." More commonly, theories can differ less in whether they pass tests of match with the data than in the degree to which the theories are themselves simple. Planetary motions could be predicted quite accurately before Kepler; Kepler nonetheless had a better theory.

A good theory must not only display order in the data (which is the same thing as compressing it), it must do so in a way that is convincing and understandable to the target audience for the theory. But this does not mean that a successful scientific theory is understandable by many people. In fact, the most successful scientific theories are fully understood by very few people. They are successful because of institutions and conventions that support the recognition of specialized expertise and its perpetuation by rigorous training.

So though an effective theory must be persuasive, its persuasiveness cannot be determined entirely by examining the theory itself. One has to look also at who the accepted experts are, what kinds of arguments they are trained to understand and

[1] Kepler's theory allowed data on the position of a planet that before required four coordinates (three spatial, one temporal) for each of N observed points to be reproduced with very high accuracy (from an economist's point of view, at least) with two coordinates (arc-length along the ellipse, time) for each of the N data points, plus the five numbers necessary to characterize the ellipse in 3-space.

[2] Newton's theory allowed nearly the same accuracy with a single coordinate, time, together with the location and velocity vectors of the planet relative to the sun at some base time.

approve. And it is part of the continuing task of the discipline to assess what arguments its members *ought* to be trained to understand and approve.

Priesthoods and guilds—organizations of people with acknowledged expertise, training programs and hierarchical structure—are the imperfect social mechanisms by which bodies of knowledge are perpetuated. Modern science and economics are special cases.[3] In understanding methodological disputes, it helps to bear in mind that the discussion is part of the workings of such an institution.

Limits of the Analogy Between Economics and Physical Sciences

Most natural sciences give a much less important role to probability-based formal inference than does economics. Since economics seems closer to natural sciences than are the other social sciences, in that economics makes more use of mathematically sophisticated theory and has more abundant data, why should it not also be less in need of statistical methodology? Examining the differences among sciences in a little more detail, we can see that probability-based inference is unavoidable in economics and that in this economics resembles related sciences, whether social or natural.

Economists can do very little experimentation to produce crucial data. This is particularly true of macroeconomics. Important policy questions demand opinions from economic experts from month to month, regardless of whether professional consensus has emerged on the questions. As a result, economists normally find themselves considering many theories and models with legitimate claims to matching the data and predicting the effects of policy. We have to deliver recommendations or accurate description of the nature of the uncertainty about the consequences of alternative policies, despite the lack of a single accepted theory. Because noneconomists often favor one policy or another based on their own interests, or prefer economic advice that pretends to certainty, there is an incentive for economists to become contending advocates of theories, rather than cool assessors of the state of knowledge.

There are natural sciences that share some of these characteristics. Astronomers can't do experiments, but they have more data than we do. Cosmology is short of relevant data and has contending theories, but is not pressed into service on policy decisions. Epidemiology is policy relevant and has limits on experimentation, but some kinds of experimentation are open to it—particularly use of animal models. Atmospheric science has limited experimental capacity, but in weather forecasting has more data than we do and less demand to predict the effects of policy. In modeling the effects of pollution and global warming, though, atmospheric science begins to be close to economics, with competing models that give different

[3] This is of course an oversimplification, for rhetorical effect. For a more nuanced discussion of how modern science relates to and emerged from priesthoods and guilds, see Ben-David (1971).

policy-relevant answers. But in this area atmospheric science does not have methodological lessons to teach us; I would say if anything the reverse is true.

Axiomatic arguments can produce the conclusion that anyone making decisions under uncertainty must act as if that agent has a probability distribution over the uncertainty, updating the probability distribution by Bayes' rule as new evidence accumulates.[4] People making decisions whose results depend on which of a set of scientific theories is correct should therefore be interested in probabilistic characterizations of the state of evidence. Yet in most physical sciences, such probabilistic characterizations of evidence are rare. Scientists understand the concept of standard error, but it seldom plays a central role in their discussion of results. In experimental sciences, this is due to the possibility of constructing an experiment in such a way, or continuing it to such a length, that standard errors of measurement are negligible. When this is possible, it certainly makes sense to do it.[5]

In nonexperimental sciences with a great deal of data, like some branches of astronomy or atmospheric science, data may be plentiful but not suited to resolve some important outstanding theoretical issue. An interesting example is the narrative in Lindzen (1990, sec. 9.1) of the development of the theory of atmospheric tides—diurnal variations of barometric pressure. For a long time in this field, theory and data collection leapfrogged each other, with theory postulating mechanisms on which little data were available, the data becoming available and contradicting the theory and new theory then emerging. Because the amount of data was large and it was error ridden, something like what economists call reduced-form modeling went on continually in order to extract patterns from the noisy data. Even at the time Lindzen wrote, the best theory could not account for important features of the data. The gaps were well documented, and Lindzen's narrative closes with suggestions for how they might be accounted for. There is no formal statistical comparison of models in the narrative, but also no account of any use of the models in decision making. If they had to be used to extrapolate the effects of interventions (pollution regulations, say) on atmospheric tides, and if the consequences were important, there would be no way to avoid making assumptions on, or even explicitly modeling, the variation the theories could *not* account for: it would have to be treated as random error.

In clinical medicine and epidemiology, statistical assessment of evidence is as pervasive as it is in economics. A treatment for a disease is a kind of theory, and when one is compared to another in a clinical trial, the comparison is nearly always statistical. If clinical trials were cheap, and if there were not ethical problems, they could be run at such a scale that, as in experimental science, the uncertainty in the results would become negligible. In fact, though, trials are expensive and patients cannot be given apparently worse treatments once the better therapy has acquired

[4] See, for example, the first two chapters of Ferguson (1967) or chapters 2 and 6 of Robert (1994).

[5] For an extended analysis of why some natural sciences in practice consider only "objective" probabilities and make little formal use of probabilistic inference despite the validity of the axiomatic foundations of Bayesian decision theory, see Burks (1977).

a high probability, even though near certainty is not yet available. Epidemiology therefore often must work with nonexperimental data that produce difficulties in causal interpretation much like those facing economics. The debate over the evidence linking smoking and cancer has strong parallels with debates over macroeconomic policy issues, and it was inevitably statistical. Biological experiments not involving human subjects were possible in that case, though, and for macroeconomic policy questions, there is seldom anything comparable.

In other social sciences, there has recently been a reaction against formal statistical methodology. Many sociologists, for example, argue that insistence on quantitative evidence and formal statistical inference forces field research into a rigid pattern. Close observation and narrative description, like what has been common in anthropology, is advocated instead (for example, Bryman, 1988). A few economists also take this point of view. Bewley (1994) has undertaken work on wage and employment adjustment, using interviews with individual firms, that is close to the spirit of the new style in sociology.[6]

The coincident timing of attacks on statistical methods across disparate social sciences is probably not an accident. But the common element in these attacks is not a unified alternative approach—those advocating anthropological-style field research are criticizing statistical method from an almost precisely opposite point of view to that of purist real business cycle theorists. Instead, the popularity of the critiques probably arises from the excesses of enthusiasts of statistical methods. Pioneering statistical studies can be followed by mechanical imitations. Important formal inference techniques can be elaborated beyond what is useful for, or even at the expense of, their actual application. Indeed, insistence on elaborate statistical method can stifle the emergence of new ideas. Hence a turning away from statistical method can in some contexts play a constructive role. Anthropological method in field research in economics seems promising at a stage (as in the theory of price and wage rigidity in economics) where there are few theories, or only abstract and unconvincing theories, available and informal exploration in search of new patterns and generalizations is important. A focus on solving and calibrating models, rather than carefully fitting them to data, is reasonable at a stage where solving the models is by itself a major research task. When plausible theories have been advanced, though, and when decisions depend on evaluating them, more systematic collection and comparison of evidence cannot be avoided.

The pattern of variation across disciplines in the role of formal statistical inference reflects two principles. First, formal statistical inference is not important

[6] Another source of skepticism about econometrics may be the side effects of the Lucas critique of econometric policy evaluation. Nothing in the explicit logic of that critique suggests that probabilistic inference is in itself invalid or problematic. It criticizes a particular way of modeling macroeconomic policy interventions in a particular class of models. But the crude summary of its message—"econometric models are useless for policy evaluation"—no doubt contributed to the broader tendency of economists to question econometric method. In Sims (1987), I argue that the original formulation of the Lucas critique is itself logically flawed.

when the data are so abundant that they allow the available theories to be clearly ranked. This is typical of experimental natural sciences. Second, formal statistical inference is not necessary when there is no need to choose among competing theories among which the data do not distinguish decisively. But if the data do not make the choice of theory obvious, and if decisions depend on the choice, experts can report and discuss their conclusions reasonably only using notions of probability.

All the argument of this section is Bayesian—that is, it treats uncertainty across theories as no different conceptually from stochastic elements of the theories themselves. It is only from this perspective that the claim that decision making under uncertainty must be probabilistic can be supported. It is also only from this perspective that the typical inference problem in macroeconomics—where a single set of historically given time series must be used to sort out which of a variety of theoretical interpretations are likely—makes sense (Sims, 1982).[7]

The Rhetoric of Economics

Any economist who uses "rhetoric" in an article these days usually is reflecting at least implicitly the influence of McCloskey's (1983) antimethodological methodological essay and subsequent related writing. This work in part reflected, in part instigated, an impatience with demands for technical rigor that emerged not only in the attitudes of the real business cycle school purists, but also in some macroeconomists of quite disparate substantive views. McCloskey wanted economists to recognize that in their professional writing, even at its most academic or scientific, they were engaged in persuasion. The essay identified and analyzed some of the rhetorical tools specific to economic argument, as well as the way economists use more universal tools. My own viewpoint as laid out above is consistent with McCloskey's in a number of respects. Both recognize that theories are not "true" or "false" and are not "tested" in single decisive confrontations with data. Both recognize that one can legitimately prefer one theory to another even when both fit the data to the same degree. Both reflect a suspicion of orthodoxy, hierarchy and methodological prescriptions as potential tools of priestly resistance to change.

But McCloskey's enthusiasm for identifying rhetorical devices in economic argument and encouraging rhetorical skill among economists risks making us soft on quackery. For example, a simple theory is preferable to a complicated one if both accord equally well with the data, making the simpler one a more thorough data compression. Thus I agree with McCloskey that a naive hypothesis-testing model of how theories are evaluated is a mistake. But a simple theory may gain adherents for

[7] It should be noted that this point of view implies a critical stance toward some recent developments in econometric theory, particularly the literature on hypothesis testing in the presence of possible nonstationarity and cointegration, and is in this respect aligned with real business cycle purists (Sims, 1988; Sims and Uhlig, 1991).

other reasons—it may appeal to people with less training, who want to believe that a theory accessible to them is correct, or the evidence of its poorer fit may not be understandable without requiring rare technical skills; or the simple theory may fit the political or aesthetic tastes of many people. Convincing people that a simple theory is better than a more complicated one by appeal to something like these latter sources of support can be rhetorically effective in that it persuades people, and it may be done with admirable skill. But it is bad economics. Indeed, while I agree with McCloskey that recognizing rhetorical devices in economic discourse and analyzing their effectiveness is worthwhile, my first reaction on recognizing a persuasive type of argument is not enthusiasm, but wariness.

Economics is not physics. Science in general does not consist of formulating theories, testing them against data and accepting or rejecting them. But we can recognize these points without losing sight of the qualitative difference between modern science and classical or medieval natural philosophy: modern science has successfully created agreement that in scientific discourse certain types of apparently persuasive arguments are not legitimate. The only kind of argument that modern science treats as legitimate concerns the match of theory to data generated by experiment and observation. This means that sometimes badly written, difficult papers presenting theories that are aesthetically, politically or religiously displeasing are more persuasive to scientists than clearly written, easily understood papers that present theories that many people find inherently attractive. The fact that economics is not physics does not mean that we should not aim to apply the same fundamental standards for what constitutes legitimate argument; we can insist that the ultimate criterion for judging economic ideas is the degree to which they help us order and summarize data, that it is not legitimate to try to protect attractive theories from the data.

We *can* insist on such standards, but it is not at all inevitable that we will do so. Because economics, like other social sciences, does not achieve the clean successes and consensuses of the natural sciences, there can be disagreement not only about which theories are best, but about which modes of argument are legitimate. The standard that theories need to confront data, not be protected from it, is itself in constant need of defense, and its implications are regularly in dispute among economists who believe themselves committed to it. While McCloskey's argument and analysis can in part be seen as a useful part of this process of defending scientific standards in economics, some of the influence of McCloskey's writing has been malign.

Though it does not show up often explicitly in written literature, I encounter with increasing frequency in one-on-one professional argument an attitude I think of as rhetorical cynicism. A bit of it appears in the original McCloskey essay. For example, McCloskey (p. 489) cites with admiration the Friedman and Schwartz *Monetary History* in a paragraph that ends with, "what was telling in the debate was the sheer bulk of the book—the richness and intelligence of its arguments, however irrelevant most of the arguments were to the main point." It was perhaps not as clear in 1983 when McCloskey wrote as it is now that monetarism was on its way to

the same macroeconomic limbo as Keynesianism, but even so, did McCloskey mean to suggest that it is good that economists were persuaded by irrelevant arguments? Or that we should admire rhetorical techniques that succeed in persuading the profession even if we ourselves can recognize that they should not persuade? I think many economists now see themselves as experts in persuasion as much as experts in substantive knowledge. They are willing to use arguments they know are flawed without explaining the flaws or to cite evidence they know could be shown to be misleading, for the sake of rhetorical effectiveness. There have always been economists who became, sincerely or cynically, uncritical apologists for particular viewpoints. The recent phenomenon is somewhat different. Economists seem to be telling themselves a story like this: a bold article, free of technical detail beyond a respectable minimum, is more likely to be cited frequently than a more cautious one that carefully defines the limits of its results. Leaving it to someone else to write carping technical critiques generates more citations and is not irresponsible if one thinks of one's role as like that of a lawyer in adversarial court proceedings.

As is probably apparent, my own opinion is that whatever the value of viewing economics as rhetoric, that view of economics should remain secondary, with the view of economics as science, in the sense that it is an enterprise that holds theory accountable to data, remaining primary. It then follows that if economists are to communicate about the central questions of the discipline, they will need the language of statistical inference.

The Real Business Cycle School

One way to characterize macroeconomics is as that branch of economics that makes drastic simplifications for the sake of studying phenomena—determination of the price level, the business cycle, economic growth—that inherently require analysis of general equilibrium. It is therefore natural and promising that macroeconomists, as computational power expands, are exploring methods for using previously intractable dynamic, stochastic, general equilibrium (DSGE) models. This phase of research, in which people examine which kinds of models are manageable and interesting and sharpen methods of numerical analysis, shares some characteristics with "normal science" as Kuhn describes it: textbooks are written (Sargent's (1987) *Dynamic Macroeconomic Thoery*, Stokey, Lucas and Prescott's (1989) *Recursive Methods in Economic Dynamics*), researchers pose and solve puzzles, and there is a general sense of powerful methods being extended to cover new areas of application.

This activity has critics in the profession. The models are still too stylized and too remote from fitting the data to provide reliable guides to policy. Since considerable intellectual energy is going in to exploring them nonetheless, economists with strong interests in current policy, or high rates of time discount, or a reluctance to invest in learning the newly fashionable analytic methods, are ready to argue

that research of this kind should not be supported or taken seriously. Not surprisingly, the people who go ahead into this area of work despite the arguments against it develop some intellectual armor against such attacks. Much of this armor is visible more in informal interactions than in published writing. It is therefore valuable to have the armor displayed, even if in a form more rigid than most economists working in the area would probably be comfortable in, in the Kydland and Prescott essay in this issue.

The argument seems to be that what dynamic, stochastic, general equilibrium modelers in economics are doing not only resembles Kuhn's normal science, it is normal science. Macroeconomists are said to have available a "well-tested," or "standard" theory. They do (computational) "experiments." These experiments usually result in "established theory becoming stronger," but occasionally discover an extension of the existing theory that is useful, and thereby "established theory" is "improved."

But these analogies with established physical sciences are strained. The neoclassical stochastic growth model that Kydland and Prescott put forth as the foundation of dynamic, stochastic, general equilibrium modeling is legitimately labeled accepted theory in one limited sense. There is an interacting group of researchers working out the implications of models built on this base; within this group the theory is accepted as a working hypothesis. But even within this group there is no illusion that the theory is noncontroversial in the profession at large. Most in the group would not even assert confidently that it is clear that theory of this type will deliver on its promise, any more than did Keynesian simultaneous equations models or natural rate rational expectations models.

What Kydland and Prescott call computational experiments are computations, not experiments. In economics, unlike experimental sciences, we cannot create observations designed to resolve our uncertainties about theories; no amount of computation can change that.

Dynamic, stochastic, general equilibrium modeling has delivered little empirical payoff so far. Macroeconomists have developed a variety of approaches to compressing the time series data using only informal theoretical ideas. The business cycle stage charts of the early NBER business cycle analysts were among the first of these reduced form modeling approaches, but multivariate spectral analysis, distributed lag regression, turning point timing analysis, cross-correlation functions, distributed lag regression, principle component analysis, vector autoregression impulse response analysis and dynamic factor analysis have all seen use. Certain patterns turned up by these analyses have the status of stylized facts: for example, Okun's law (the tendency of employment to lag output), the strong predictive value of interest rate innovations for output and prices, the smoothness of aggregate price and wage movements, the tendency of productivity to fluctuate procyclically, and the strong correlation of monetary aggregates with nominal income and their (Granger) causal priority to it. I think it is fair to say that most real business cycle research has ignored most of the

known facts about the business cycle in assessing the match between DSGE models and the facts. Kydland and Prescott rightly point out that all theories (at least in macroeconomics) are false and that therefore it does not make sense to discard a theory if it fails to fit perfectly. But if a theory fits much worse than alternative theories, that is a strike against it. We may still be interested in a poorly fitting theory if the theory offers an especially dramatic data compression (that is, is very simple relative to the data it confronts) or if it is a type of theory that promises to fit better with further work. But there can be no argument for deliberately shying away from documenting the ways in which the theory does and does not match the data.

This issue is distinct from the question of whether we should employ formal methods of statistical inference. Here the issue is only whether the data is going to be confronted as it exists, in all its density. When Mark Watson (1993), eschewing "formal statistical inference," used an extension of standard tools for examining fit of a time series model by frequency in a Fourier analysis, he allowed us to see that the neoclassical stochastic growth model at the core of the real business cycle approach is very far from accounting well even for what we think of as business cycle variation in output itself. Watson's analysis does not imply that the real business cycle approach should be abandoned. It does suggest that frequency domain analysis or other standard methods of orthogonal decomposition of macroeconomic time series data (like vector autoregression impulse responses) ought to be a standard part of evaluating real business cycle models, with the aim of getting better fit than what Watson found.

Kydland and Prescott appear to argue strongly against using econometric tools in the modern sense of the term. In part this stems from their caricature of formal statistical inference as "statistical hypothesis testing" that will certainly reject any (necessarily false) theory when given enough data. Bayesian critiques of classical hypothesis testing have long made the same kind of point, without rejecting formal statistical inference. Kydland and Prescott also claim, "Searching within some parametric class of economies for the one that best fits a set of aggregate time series makes little sense, because it isn't likely to answer an interesting question." Yet they put forth as an interesting type of question, "How much the U.S. postwar economy would have fluctuated if technology shocks had been the only source of fluctuations?" Surely one approach to such a question would be to construct a parametric class of dynamic, stochastic, general-equilibrium models in which the parameter indexed the contribution of technology shocks to fluctuations and to examine the behavior of model fit as a function of this parameter. Of course it might turn out that model fit was insensitive to the parameter—the model was weakly identified in this dimension—but it might be instead that some sources of impulse response could be ruled out as unlikely, because they implied a poor fit. Showing this would not amount to simply finding the fit-maximizing parameter value, of course, but instead to characterizing the shape of the likelihood. If Kydland and Prescott are objecting only to the idea of ending inference with the picking of the parameter that gives the best fit, they are taking the position of Bayesian or likelihood

principle-based inference.[8] It seems likely, though, that they intend a broader objection to probabilistic inference, in which case they seem to contradict some of their own position on what are interesting research questions.

Kydland and Prescott do approve some probability-based inference. They argue that it is reasonable to look at the theoretical probability distribution that is implied by a model for a set of statistics and to compare this to the corresponding statistics computed from the actual data. But a stochastic model produces a distribution, not a statistic. How are we meant to "compare" a distribution to a statistic? What conclusions might we draw from such a comparison? In this paper there is little guidance as to how we should make or interpret such a comparison.[9] It is perhaps not surprising in the light of Kydland and Prescott's inference aversion that there is little guidance, because these questions are the root out of which all of statistical inference grows.

We can guess what use of such a comparison Kydland and Prescott intend by looking at their discussion of what constitutes "well-tested theory." They consider the neoclassical growth framework a well-tested theory. They say that it gives us confidence in the theory that it implies that when a model economy is subjected to realistic shocks, "it should display business cycle fluctuations of a quantitative nature similar to those actually observed." Another way of putting this, apparently, is that stochastic models based on the neoclassical growth framework produce "normal-looking" fluctuations. For some purposes this kind of language may suffice, but when we need to consider which of two or more models or theories with different policy implications is more reliable, it does not take us very far to be told that we should be more confident in the one whose simulated data is more "normal-looking" or is of a "quantitative nature" more "similar" to the actual data. My view, for example, is that Watson (1993) showed us that the stochastic neoclassical growth model as usually applied in the real business cycle literature produces simulated data that is drastically dissimilar to the actual data. If Kydland, Prescott and I are to have a reasoned discussion about this point, we will have to start talking about what we mean by "similar" and about what alternative models or theories are the standard against which the match of this one to the data is to be judged. Then we will be engaged in statistical inference.

If hyperbolic claims for the research accomplishments of dynamic, stochastic, general equilibrium models and for immunity of such models to criticism for naive econometric method were necessary to sustain the enthusiasm of the participants, Kydland and Prescott might be justified in a bit of forensic exaggeration. But the field is interesting enough to attract attention from researchers without this, and the air of dogma and ideological rigidity these claims attach to the field are an unnecessary burden on it.

[8] The likelihood principle is an implication of a Bayesian approach to inference, but can be argued for on other grounds as well. See Berger and Wolpert (1988).

[9] Actually, in an earlier version there was a little more guidance, and it sounded a lot like simple statistical inference.

Progress in Quantitative Macroeconomics

There is work that is based on dynamic, stochastic, general equilibrium models and that makes serious use of formal methods of probability-based inference. McGrattan, Rogerson and Wright (1993), for example, estimate a standard real business cycle model using maximum likelihood, which could be a first step toward comparing its fit to other types of DSGE models or naive reduced form models. Leeper and Sims (1994) fit a DSGE model that adds to what Kydland and Prescott lay out as the standard neoclassical model a fluctuating relative price of consumption and capital goods and an articulated monetary and fiscal policy sector. The Leeper-Sims model comes close to the fit of a first-order reduced form vector autoregression model on a set of three variables. This is apparently much better than the fit of the versions of the neoclassical model that appear throughout the real business cycle literature, though because the fit of those models is seldom examined in a careful, standardized way, it is difficult to be certain of this. In any case, it is clear that it is becoming quite feasible to produce likelihood surfaces and one-step-ahead prediction residuals for DSGE models, thus providing a basis for comparison of alternative models meant to explain the same data series. There is considerable interest in making such comparisons, and it is essential before the models can become the basis for quantitative policy analysis. It will be happening even if Kydland and Prescott cannot be persuaded to assist in the process.

Other streams of research in macroeconomics are making as much progress, from an empirical point of view, as real business cycle modeling has yet achieved. One is modern policy modeling in the tradition of simultaneous equations modeling. These models have had limited attention from academic macroeconomists because of the concentration of research interest on building equilibrium models. But dynamic, stochastic, general equilibrium models have not been produced at a scale, level of detail and fit that allows them to be used in the actual process of monetary and fiscal policy formation. The result is that the field of policy modeling has been left to economists close to the policy process outside of academia, together with a few academic economists with strong policy interests. The models have developed to include rational expectations and expanded to include international linkages. This has come at a cost, however. Because implementing rational expectations at the scale of these models has been seen as computationally difficult, the stochastic structure of the models has become even more stylized and unbelievable than when I wrote about the then-existing models in Sims (1980b).

A good example of work on this line is Taylor (1993). The larger of the two models in that book uses consumption and investment functions loosely based on dynamic optimizing theories and incorporating expectations of future income explicitly. It has wage adjustments that are sluggish and depend on expectations of the future. It has international linkages that use expectational interest rate parity conditions. In these respects it is an advance beyond the state of the art in 1980. But because the optimizing theory is used only informally, the model contains some structural anomalies. Further, equations of the model are estimated one at a time,

in some cases using statistical methods that are incompatible with the model's stochastic specification. The work is nonetheless promising, because it maintains the simultaneous equation modeling tradition of dense contact with the data and because it seems on the verge of substantial further progress. Recent developments in computing power and standardization of solution methods for linear rational expectations models make it seem likely that Taylor could apply to the full-scale model in the latter part of his book the more internally consistent statistical methodology he uses on a smaller example model in the first chapter. For the same reason, he should be able to connect his model more completely to a theory of dynamic optimizing behavior. It would thus more closely approach the real business cycle school's level of internal consistency and interpretability, reattain or improve upon the standards of statistical fit of the original simultaneous equation modeling tradition and still preserve the scale and detail needed for application to policy analysis.

Another distinct stream of research, which may look more significant to me than it should because of my own involvement with it, uses weakly identified time series models to isolate the effects of monetary policy. This style of work, in contrast with standard simultaneous equation modeling, begins with careful multivariate time series modeling of the data, developing evidence on prominent regularities. Restrictions based on substantive economic reasoning are imposed only as necessary to interpret the data and always with an eye to avoiding distortion of the model's fit.

Work in this style began with Friedman and Schwartz, who challenged conventional Keynesian thinking by displaying evidence of the strong correlation between monetary aggregates and income and of timing relationships, both in time series data and in particular historical episodes, that suggested causality running from money to income. I showed that beyond correlation and timing, there was a one-way predictive relationship between money and income (Sims, 1972). This implied that money was causally prior to nominal income according to the same definition of a causal ordering that underlies putting the "causally prior" variable on the right-hand side of a regression.

From this point onward, most of this literature focused on using reduced-form models that summarized the data by showing how all variables in the system are predicted to change following a surprise change in any one variable. This in effect breaks the variation in the data into mutually uncorrelated pieces, helping to isolate the major regularities. The surprise changes are called "innovations," and the predicted patterns of change are called "impulse responses." I showed in Sims (1980a), following work by Y. P. Mehra (1978), that short-term interest rate innovations absorbed most of the predictive power of money innovations for output when the interest rates were added to a multivariate system. This undermined the monetarist claim that exogenous disturbances to the money stock were generated by policy and a dominant source of business cycle fluctuations, but left open the question of how the interest rate innovations should be interpreted. Starting in the mid-'80s with work by Bernanke (1986), Blanchard and Watson (1986) and Sims (1986), the

informal identification arguments that had been used in this literature were supplemented with formal ones, adaptations to this weakly identified context of the methods used in the simultaneous equations literature. Bernanke and Blinder (1992) argued for identifying federal funds rate innovations as policy shocks, supporting their time series analysis with institutional detail. I showed (1992) that the stylized facts that this literature was codifying and interpreting were stable across a range of economies outside the United States, but that in certain respects these facts did not accord well with the interpretation of interest rate innovations as policy disturbances. I had already showed that the main anomaly—the "price puzzle" in which inflation rises after an apparent monetary contraction—would disappear under a particular set of identifying assumptions for U.S. data (Sims, 1986). Other researchers joined in trying to delineate the range of identifying assumptions consistent with the data (Christiano, Eichenbaum and Evans, 1994; Gordon and Leeper, 1995; Sims and Zha, 1995). Other recent work has extended the models to open economy environments with interesting results (Eichenbaum and Evans, 1993; Soyoung Kim, 1994; Soyoung Kim and Roubini, 1995; Cushman and Zha, 1994).

This literature has advanced knowledge in several ways. It has established firmly that most of the observed variation in monetary policy instruments—interest rates and monetary aggregates—cannot be treated as exogenously generated by random shifts in policy. (Incredibly, real business cycle school attempts to introduce monetary variables into dynamic, stochastic, general equilibrium models still often contradict this by now elementary business cycle fact.) It has given us a clearer quantitative picture of the size and dynamics of the effects of monetary policy. It has shown us that our knowledge about the size of these effects is still uncertain and that, because monetary contraction is rarely a spontaneous policy decision, the apparently eloquent fact that monetary contractions are followed by recession is hard to interpret.

The work suffers from some rhetorical handicaps. It cannot be understood without some familiarity with time series and simultaneous equations modeling ideas. Though the interpretations that it puts forward are influenced by rational expectations arguments, they find no need to make formal use of them. The work has proceeded incrementally, with no single paper or idea having dramatically changed people's thinking. The conclusions the work leads to tend more strongly to undermine naively confident interpretations of the data than to provide technical support for any simple policy position.

Some macroeconomists seem to have the impression that because this literature has not used dynamic optimization theory or rational expectations explicitly, and because it has found a version of simultaneous equations modeling essential, it is part of a tradition that we now know to be obsolete. In fact, this literature aims at minimizing reliance on ad hoc modeling conventions of both the traditional simultaneous equations style and the new dynamic, stochastic, general equilibrium style, in order to focus cleanly on the central issue of identifying the distinction between variation generated by deliberate policy action and variation generated by

disturbances outside of the policy process.[10] The literature is probably now reaching a level of maturity at which it will pay to open up connections to DSGE models (by examining a range of more tightly restricted identifications) and to large-scale policy models (by considering larger and internationally linked versions of the models). Soyoung Kim (1994) is a step in the latter direction, and Sims (1989), Jinill Kim (1995) and Leeper and Sims (1994) are steps in the former direction.

Conclusion

Empirical macroeconomists are engaged in several promising lines of work. They are also engaged in making strained analogies between their work and the natural sciences and in classifying work in styles other than their own as outdated or mistaken based on its methods, not its substance. Since there is also a tendency in the profession to turn away from all technically demanding forms of theorizing and data analysis, it does not make sense for those of us who persist in such theorizing and data analysis to focus a lot of negative energy on each other. All the lines of work described in the previous section, including real business cycle modeling, are potentially useful, and the lines of work show some tendency to converge. We would be better off if we spent more time in reading each others' work and less in thinking up grand excuses for ignoring it.

References

Ben-David, Joseph, *The Scientist's Role in Society.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Berger, James O., and Robert L. Wolpert, *The Likelihood Principle.* 2nd ed., Hayward, Calif.: Institute of Mathematical Statistics, 1988.

Bernanke, Ben, "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy,* Autumn 1986, *25,* 49–99.

Bernanke, Ben, and Alan Blinder, "The Federal Funds Rate and the Channels of Monetary Transmission," *American Economic Review,* September 1992, *82,* 901–21.

Bewley, Truman, "A Field Study on Downward Wage Rigidity," processed, Yale University, 1994.

Blanchard, Olivier, and Mark Watson, "Are All Business Cycles Alike?" In Gordon, R. J., ed., *The American Business Cycle.* Chicago: University of Chicago Press, 1986, pp. 123–56.

Bryman, Alan, *Quantity and Quality in Social Research.* London and Boston: Unwin Hyman, 1988.

Burks, Arthur W., *Chance, Cause, Reason.* Chicago and London: University of Chicago Press, 1977.

Christiano, Lawrence, Martin Eichenbaum, and C. Evans, "The Effects of Monetary Policy

[10] More detailed discussion of why the "Lucas critique" does not cast doubt on this work appears in Sims (1982, 1987). The discussion takes us back to consideration of the meaning of "structure" in economic modeling, as in Hurwicz (1962).

Shocks: Evidence from the Flow of Funds," *American Economic Review,* 1994, *82*:2, 346–53.

Cushman, David O., and Tao Zha, "Identifying Monetary Policy in a Small Open Economy Under Flexible Exchange Rates: Evidence from Canada," processed, University of Saskatchewan, 1994.

Eichenbaum, Martin, and C. Evans, "Some Empirical Evidence on the Effects of Monetary Policy Shocks on Exchange Rates." NBER Working Paper No. 4271, 1993.

Ferguson, Thomas S., *Mathematical Statistics: A Decision Theoretic Approach.* New York and London: Academic Press, 1967.

Friedman, Milton, and Anna Schwartz, "Money and Business Cycles," *Review of Economics and Statistics,* February 1963, *45*:1, Part 2, 32–64.

Gordon, David, and Eric Leeper, "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification," *Journal of Political Economy,* 1995, *102*:6, 1228–47.

Hurwicz, Leonid, "On the Structural Form of Interdependent Systems." In Nagel, Ernest, et al., eds., *Logic, Methodology and Philosophy of Science.* Stanford: Stanford University Press, 1962, pp. 232–39.

Kim, Jinill, "Effects of Monetary Policy in a Stochastic Equilibrium Model with Real and Nominal Rigidities," processed, Yale University, 1995.

Kim, Soyoung, "Do Monetary Policy Shocks Matter in the G6 Countries?," processed, Yale University, 1994, 81–117.

Kim, Soyoung, and Nouriel Roubini, "Liquidity and Exchange Rates: Evidence from Identified VARs," processed, Yale University, 1995.

Leeper, Eric, and Christopher Sims, "Toward a Modern Macroeconomic Model Usable for Policy Analysis," *NBER Macroeconomics Annual,* 1994, 81–117.

Lindzen, Richard S., *Dynamics in Atmospheric Physics.* Cambridge: Cambridge University Press, 1990.

McCloskey, Donald, "The Rhetoric of Economics," *Journal of Economic Literature,* June 1983, *21,* 481–517.

McGratten, Ellen R., Richard Rogerson, and Randall Wright, "Household Production and Taxation in the Stochastic Growth Model." Federal Reserve Bank of Minneapolis Working Paper No. 521, October 1993.

Mehra, Yash Pal, "Is Money Exogenous in Money-Demand Equations," *Journal of Political Economy,* 1978, *86*:2, Part 1, 211–28.

Robert, Christian, *The Bayesian Choice.* New York: Springer-Verlag, 1994.

Sargent, Thomas J., "Autoregressions, Expectations, and Advice," *American Economic Review,* May 1984, *74,* 408–15.

Sargent, Thomas J., *Dynamic Macroeconomic Theory.* Cambridge: Harvard University Press, 1987.

Sims, Christopher A., "Money, Income and Causality," *American Economic Review,* September 1972, *62*:4, 540–52.

Sims, Christopher A., "Comparing Interwar and Postwar Business Cycles: Monetarism Reconsidered," *American Economic Review,* 1980a, *70*:2, 250–57.

Sims, Christopher A., "Macroeconomics and Reality," *Econometrica,* January 1980b, *48*:1, 1–48.

Sims, Christopher A., "Scientific Standards in Econometric Modeling." In Hazewinkel, M., and A. H. G. Rinnooy Kan, eds., *Current Developments in the Interface: Economics, Econometrics, Mathematics.* Dordrecht, Boston, and London: D. Reidel, 1982, pp. 317–37.

Sims, Christopher A., "Are Forecasting Models Usable for Policy Analysis?," *Quarterly Review, Federal Reserve Bank of Minneapolis,* Winter 1986, *10*:1, 2–16.

Sims, Christopher A., "A Rational Expectations Framework for Short-Run Policy Analysis." In Barnett, W. A., and K. Singleton, eds., *New Approaches to Monetary Economics.* Cambridge: Cambridge University Press, 1987, pp. 293–308.

Sims, Christopher A., "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control,* June/September 1988, *12,* 463–74.

Sims, Christopher A., "Models and Their Uses," *American Journal of Agricultural Economics,* May 1989, *71,* 489–94.

Sims, Christopher A., "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review,* June 1992, *36,* 975–1000.

Sims, Christopher A., and Harald Uhlig, "Understanding Unit Rooters: A Helicopter Tour," *Econometrica,* November 1991, *59,* 1591–99.

Sims, Christopher A., and Tao Zha, "Does Monetary Policy Generate Recessions?," processed, Yale University, 1995.

Stokey, N., and Robert E. Lucas, with Edward Prescott, *Recursive Methods in Economic Dynamics.* Cambridge: Harvard University Press, 1989.

Taylor, John, *Macroeconomic Policy in a World Economy.* New York and London: W.W. Norton, 1993.

Watson, Mark, "Measures of Fit for Calibrated Models," *Journal of Political Economy,* December 1993, *101,* 1011–41.