



## The Scientific Illusion in Empirical Macroeconomics

Lawrence H. Summers

*The Scandinavian Journal of Economics*, Vol. 93, No. 2, Proceedings of a Conference on New Approaches to Empirical Macroeconomics. (Jun., 1991), pp. 129-148.

Stable URL:

<http://links.jstor.org/sici?sici=0347-0520%28199106%2993%3A2%3C129%3ATSIIEM%3E2.0.CO%3B2-K>

*The Scandinavian Journal of Economics* is currently published by The Scandinavian Journal of Economics.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/sje.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# The Scientific Illusion in Empirical Macroeconomics

*Lawrence H. Summers\**

Harvard University and NBER, Cambridge MA, USA

## Abstract

It is argued that formal econometric work, where elaborate technique is used to apply theory to data or isolate the direction of causal relationships when they are not obvious *a priori*, virtually always fails. The only empirical research that has contributed to thinking about substantive issues and the development of economics is pragmatic empirical work, based on methodological principles directly opposed to those that have become fashionable in recent years.

## I. Introduction

Many macroeconomists and most econometricians believe and teach their students that (i) empirical work in macroeconomics should concentrate on identifying "deep structural parameters" characterizing preferences and technology; (ii) the best empirical work in macroeconomics formally tests substantive hypotheses rigorously derived from economic theory; (iii) sophisticated statistical technique can play an important role in sorting out causation in systems with many interdependent variables. These beliefs constitute the core of what I regard as the scientific illusion in empirical macroeconomics.

This paper argues that formal empirical work which, to use Sargent's (1987, p. 7) phrase, tries to "take models seriously econometrically" has had almost no influence on serious thinking about substantive as opposed to methodological questions. Instead, the only empirical research that has influenced thinking about substantive questions has been based on

---

\*I have been helped in writing this paper by comments received from Olivier Blanchard, Stanley Fischer, Richard Freeman, Mervyn King, Robert Lucas, Jeff Miron and Ken Rogoff. Many of my teachers and colleagues have influenced my attitudes on the matters treated here, but I alone bear responsibility for the views expressed. This is a revised version of a manuscript originally presented at the NBER Macroeconomics Annual Meeting in 1987.

methodological principles directly opposed to those that have become fashionable in recent years. Successful empirical research has been characterized by attempts to gauge the strength of associations rather than to estimate structural parameters, verbal characterizations of how causal relations might operate rather than explicit mathematical models, and the skillful use of carefully chosen natural experiments rather than sophisticated statistical technique to achieve identification.

These views may seem extreme. But I invite the reader to try and identify a single instance in which a "deep structural parameter" has been estimated in a way that has affected the profession's beliefs about the nature of preferences or production technologies or to identify a meaningful hypothesis about economic behavior that has fallen into disrepute because of a formal statistical test.

Now consider two issues where today's macroeconomics textbooks present a radically different picture than did the macroeconomics textbooks of the 1960s — the long-run neutrality of inflation and the relative importance of monetary and fiscal policies in affecting economic behavior. Changes in opinion about inflation neutrality resulted from theoretical arguments about the implausibility of money illusion that became compelling when inflation increased and unemployment did not decline during the 1970s. Formal statistical tests contributed almost nothing. Surely, *A Monetary History of the United States* (1963) had a greater impact in highlighting the role of money than any particular econometric study or combination of studies. It was not based on a formal model, no structural parameters were estimated, and no sophisticated statistical techniques were employed. Instead, data were presented in a straightforward way, to buttress verbal theoretical arguments, and emphasis was placed on natural experiments in assessing directions of causality.

The argument of this paper is in very much the same spirit as the attack on modernism in McCloskey's (1985) treatment of *The Rhetoric of Economics*. Like McCloskey, I emphasize persuasiveness as a crucial criterion in evaluating empirical work. My argument here is also similar to critiques of conventional econometric approaches put forward by Leamer (1978). While Leamer and others, notably David Hendry, have been at pains to develop methods for ensuring that econometric results are robust, my argument is somewhat more nihilistic in emphasizing qualitative rather than quantitative conclusions from econometric work.

In order to avoid misunderstandings, it will be helpful at this point to delimit this paper's argument in a number of respects. First, my quarrel is not with the goal of estimating deep parameters and seriously evaluating precisely formulated hypotheses but with its feasibility. Attempts to make empirical work take on too many of the trappings of science render it uninformative. Second, the issues treated in this paper are largely

independent of the substantive issues separating New Classical and Keynesian macroeconomists. Research in the New Classical tradition like that of Mehra and Prescott (1985) where formal statistical tests are eschewed, falls within the pragmatic approach advocated here, whereas Keynesian research like that of Rotemberg (1983) or Blanchard (1986) takes the formal statistical viewpoint that this paper criticizes.

Third, diversification is good in research strategy as in most other things. Undoubtedly the line I have drawn between the research strategies I praise and condemn is too bright. I am only insisting that the approach to methodology normally offered to students in graduate school is seriously deficient in slighting approaches to empirical work that have been enormously productive.

The remainder of the paper is organized as follows. Section II offers some general observations on the very different roles played by formal econometric work in economics and experimental and observational work in the natural sciences. Section III uses recent research directed at estimating deep parameters and at identifying causal relationships between macroeconomic time series as examples in discussing systematic reasons why so much formal econometric work in economics has so little impact. Section IV describes the pragmatic approach to empirical research in economics and discusses some of its many successes. Section V offers some tentative observations on the relationship between theory and empirical work in economics. I argue that the failure of empirical work to serve up stylized facts in a usable form is an important reason for the sterility of so much economic theory. Section VI concludes.

## **II. Empirical Work in Economics and the Natural Sciences**

It is interesting to contrast the relationship between theoretical and formal econometric work in economics with the relationship between theory and experiment that prevails in other sciences. While reading lay expositions of major breakthroughs in the physical and biological sciences, one is immediately struck by the crucial role of empirical work in stimulating theory in these sciences.<sup>1</sup> For example, Weinberg (1977) describes the crucial role that Hubble's observation, whereby stars appeared to recede from the earth at a rate proportional to their distance from it, played in the development of the Big Bang Theory. And he describes how theories had to be overhauled to accommodate empirical observations on the uniformity of background electromagnetic radiation in the universe and on the chemical

---

<sup>1</sup>Two brief and accessible books whose evidence supports the discussion in the text are Weinberg (1977) about the beginning of the universe and Smith (1986) about modern developments in biology.

composition of stars. Smith (1986) describes the crucial role of fossil studies, fruit fly experiments, and molecular studies of the DNA of different creatures in shaping the current understanding of the process of evolution.

Theoretical particle physicists wait anxiously to see whether experimentalists will be able to identify the particles their theories postulate.<sup>2</sup> The image of an economic theorist nervously awaiting the result of a decisive econometric test does not ring true. The negligible impact of formal econometric work on the development of economic science is manifest in a number of ways.

First, the major writings of leading economic theorists, those oriented to both micro and to macroeconomic issues, contain almost no reference to econometric studies. A perusal of several volumes of *The Journal of Economic Theory* reveals no references to econometric studies. Kreps's (1990) hugely successful textbook *A Course in Microeconomic Theory* defines the purpose of economic theory as gaining "a better understanding of economic activity and outcomes". Yet of more than nearly 200 studies that it references, only two contain any econometric work, and neither of those studies seeks to estimate a structural parameter. A random sampling experiment suggests that the 200 articles and books Kreps references taken together contain less than a dozen references to econometric articles.

Things are only slightly different in macroeconomics. Lucas's (1987) book surveying business cycle theory includes few references to the results of econometric estimation. The empirical studies that he does reference are simulation exercises such as the work of Kydland and Prescott (1982) and relatively informal discussions of the properties of data such as Romer (1986). Sargent's (1987) treatise on macroeconomic theory does refer to a number of econometric studies in a way that makes clear how minimal is their contribution to the development of theory. A typical reference notes without amplification regarding the results that "X has examined econometric implications of this equation". Solow's (1970) masterful exposition of growth theory does not contain a single reference to an econometric study, though great emphasis is placed on stylized facts revealed by straightforward data analysis. Tobin's (1980) forceful defense of Keynesian economics makes no reference to econometric tests that support his conclusions.

Second, the informal "rules of the game" governing econometric research in economics bear out the suspicion that most of it has little influence. Recent audits of econometric research that uncovered pervasive

---

<sup>2</sup> Indeed, the current (as I am writing) issue of *The Economist* describes the theoretical physicists' response to the recordings made by neutrino detectors in the wake of the recent supernova.

problems have attracted a good deal of attention; see e.g. Dewald *et al.* (1986). Perhaps more disturbing than the problem of replication is the fact that, except for the purpose of audit, none of the work examined was considered worth replicating. In the natural sciences, investigators rush to check out the validity of claims made by rival laboratories and then to build on them. Such efforts are very rare in economics. I find it implausible that this is a consequence of the accuracy and robustness of the conclusions of econometric studies. Instead, the absence of replication attempts for most empirical work are a consequence of the fact that the results are rarely an important input to theory creation or the evolution of professional opinion more generally.

Those few econometric studies that have generated numerous attempts at replication and the evaluation of robustness, have usually had as their goal demonstrating a qualitative proposition rather than estimating a structural parameter or formally testing a hypothesis. An excellent example is Feldstein's (1974) study of the impact of Social Security on private saving, and the subsequent literature it spawned.

Finally, take a longer perspective on the development of economic science. Think about the papers from the 1950s or 1960s that are remembered today. Many are purely theoretical and are remembered primarily because they enriched the language of economic argument. Samuelson's (1958) introduction of the overlapping generations model is an example. Many others are remembered for their empirical generalizations — perhaps informal theories would be a better term. *A Monetary History of The United States* (1963), Modigliani and Brumberg's (1955) and Friedman's (1957) work on the consumption function are examples. Solow's (1957) low estimate of the contribution of capital formation to economic growth is another one, although his contribution was methodological as well as substantive. It is difficult to think today of many empirical studies from more than a decade ago whose bottom line was a parameter estimate or the acceptance or rejection of a hypothesis.

Given the tremendous professional investment in econometric work, it is natural to ask why it has so little impact in either the short or long run. I take up this question in the next section. Then in Section IV, I try to assess what does change economists' beliefs.

### **III. Why Does Formal Econometric Work Prove so Unpersuasive?**

The two dominant approaches to formal econometric work in macroeconomics are the "deep parameter" approach of Sargent and his followers and the "VAR" approach of Sims and his followers. Both seek to be scientifically rigorous, and both have involved significant methodological innovation. Researchers in both these traditions seem to me to suffer from

the scientific illusion, and to confuse methodological with substantive advance. Such an argument is difficult to make in the abstract.<sup>3</sup> I therefore proceed by considering in some detail prominent papers in each of these genres.

### *Seeking Deep Parameters*

Perhaps the most influential recent work on deep parameter estimation has been Hansen and Singleton's (1982, 1983) treatment of the relationship between consumption and asset pricing. This work embodies the virtues most prized by those committed to making economics more "scientific". Hansen and Singleton attack an important problem area — the relationship between consumption behavior and asset pricing. They seek to estimate truly structural parameters. Their work is methodologically innovative. They rely on new estimation techniques to build a bridge between a fully articulated stochastic theory and data. Given the dominant professional view of what constitutes high quality empirical work, it is easy to understand why their paper was awarded the Frisch medal as the outstanding paper published in *Econometrica* over a several year period. Precisely because it is so outstanding an example of formal econometric work, it is instructive to consider its deficiencies. For if, as I shall argue, it is deeply flawed, it is likely that the entire genre it represents is as well.

One interpretation of Hansen and Singleton's research is that the primary goal is to test a particular hypothesis regarding the pricing of assets. Here they do provide an answer to a question — is a model of a representative consumer with an additively separable utility function and constant relative risk aversion and...rejected statistically by data for the sample period 1959–81? But it is hard to see that this is a very interesting question. The "model" is rejected by the observations that the consumption of different consumers is not perfectly correlated and that a large fraction of the population holds zero wealth, or by the observation that money is held despite being a dominated asset. Hansen and Singleton's model, like any theory, is literally false.

The interesting question is how accurate it is as an approximation to reality for the purpose of making different types of predictions or understanding different types of behavior. Their  $J$  statistic sheds no light at all on this question. It is obvious that with enough data, the  $J$  statistic will attain any given level. Observing its level tells the reader as much about the amount of relevant variation there was in the data Hansen and Singleton studied as it does about the validity of the hypothesis under consideration.

---

<sup>3</sup>The argument here is largely an application of general considerations, developed in Leamer's (1978) brilliant book, to specific targets.

The point here is a much more general one. Without some idea of the power of statistical tests against interesting alternative hypotheses and/or some metric for evaluating the extent to which the data are inconsistent with a maintained hypothesis formal statistical tests are uninformative.

Science proceeds by falsifying theories and constructing better ones. Falsifications of hypotheses based on overidentifying restrictions of the kind provided by Hansen and Singleton are unenlightening in two senses. First, they provide little insight into whether the reason for the theory's failure is central to its logical structure or is instead a consequence of auxiliary assumptions made in testing it. For example, should asset returns be evaluated on a pre or post-tax basis? Are the consumption needs of children equivalent to those of adults? Is clothing a durable good? Any test of the representative consumer model involves a test of whatever assumption is made about these issues and a dozen similar ones. While in principle it would be possible to explore a range of possible assumptions, this type of "data mining" is usually condemned by those who favor formal approaches to empirical work.

Second, suppose that the theory is rejected for reasons that do involve the details of empirical implementation. The fact of rejection gives little insight into the direction in which the theory should be modified. Consider this question. What empirical or theoretical approach to asset pricing has been stimulated by the specifics of Hansen and Singleton's results? Following their work, others have sought to estimate models more general than theirs, but the direction of generalization has not to my knowledge been influenced at all by the specifics of their results. Nor have their rejections provided the impetus to new theoretical developments regarding asset pricing. As I discuss below, the informal approach taken by Mehra and Prescott (1985) has had a much greater influence on the development of theories of asset pricing.

An alternative interpretation of Hansen and Singleton's work is that they were seeking to estimate "deep" parameters describing the representative consumer's behavior. As things turned out, the fact that their data rejected their model was taken to mean that they had not in fact found deep parameters. But let us imagine that the data had failed to reject their model. Alternatively imagine that in the future some model with more general preferences than those postulated by Hansen and Singleton passes tests of overidentifying restrictions. Would anyone take seriously the deep parameter estimates that resulted? I doubt it. Revealed behavior can enlighten us about the extent to which Hansen and Singleton's results are taken seriously in two ways.

First, taking seriously deep parameter estimates describing tastes and technology would presumably involve regarding them as being in some sense more stable than the "mongrel" parameters contained in traditional



Keynesian models. Lucas (1976) highlighted the fact that Keynesian models rarely were estimated using all the available data as the theory of economic policy on which they were based would suggest. He took this as a kind of “revealed behavior” argument that the parameters being estimated were not truly structural. While many investigators in a more pragmatic empirical tradition, Robert Gordon most prominently, have sought to examine the robustness of conclusions using prewar as well as postwar data, and using data from other countries, neither Hansen and Singleton nor their followers have yet ventured off the Citibank data tape.

A second way of evaluating the extent to which Hansen and Singleton’s results are taken seriously is to ask whether anyone would use them in making predictions about the effects of policy interventions. Lucas’s (1976) famous critique of econometric policy evaluation treated the effects of temporary tax cuts on consumption. His point was that the permanent income hypothesis implied that the marginal propensity to consume out of a tax cut depended on its likely duration. On Lucas’s view as I understand it, empirical research concerned with predicting the effects of tax cuts should concentrate on estimating the parameters of the utility function of the representative consumer. Once it is known, the marginal effect of any change in policy can be computed.<sup>4</sup> Hansen and Singleton have estimated the utility function parameters the importance of which Lucas stressed. Would anyone genuinely interested in predicting the effects of a tax cut make use of their estimates of representative consumers’ utility functions?

To my knowledge, no one has used these estimates for this or for any other purpose. It is not difficult to understand why. The uncertainties about the discount rate of the representative consumer upon which Hansen and Singleton’s empirical work can potentially shed light are dwarfed by uncertainties about other nontaste variables.<sup>5</sup> For example, many consumers may be unable to borrow money at any interest rate because of information and enforcement problems associated with debt contracts. These consumers will spend all of the proceeds of a tax cut. Others may lack the information processing skills necessary to distinguish the reasons for changes in their take-home pay and so will not be influenced by the temporary character of the tax cut. Still others may revise their expectations about future income and so alter their spending on the basis of their own view, or one acquired from the newspaper about the tax cut’s likely impact on the economy.

---

<sup>4</sup> This statement is not quite right because of general equilibrium considerations. In order to actually calculate the effects of a fiscal policy change on consumption for example, it would also be necessary to know its impact on interest rates which presumably depends on technology parameters.

<sup>5</sup> From here on, I am assuming that the tax cut is matched by spending cuts so I do not have to worry about any expected future tax changes.

None of these uncertainties can be addressed by estimating the utility function of the representative consumer. Nor, I suspect, will anyone modify their view about the empirical importance of these considerations when and if someone gets a representative consumer model to fail to reject its overidentifying restrictions. I find it hard to escape the conclusion that Hansen and Singleton's work creates an art form for others to admire and emulate but provides us with little new knowledge.<sup>6</sup> *Mutatis mutandis*, the same is likely to be true of any work which seeks to test a highly restricted and surely incorrect structure using elaborate methods which do not shed light on the cause of any deviations of data from theory.

### *Isolating Economic Structure*

While efforts to link specific stochastic theories to data represent an important component of formal econometric work, there exists a very different econometric tradition as well. This tradition seeks to use statistical technique to pin down the direction of causal relationships in situations where a great deal of simultaneity is present by using techniques which place little or no structure on the data. Vector autoregressions represent the most prominent example of recent work in this tradition. Sims (1980) provides the most powerful argument in favor of work in this tradition. His argument has both a destructive and a constructive component. The destructive component stresses the heroism of the assumptions that are made in many efforts to impose structure on data and questions the validity of any results that emerge. I find it convincing. I find the constructive part of the argument much less convincing in holding out the hope that robust conclusions can emerge from work that makes few identifying assumptions.

Again, the point is probably best argued by example. I consider the work of Bernanke (1986) on the relationship between money and output because I agree with Robert King's observation that it is an "admirable piece of normal science" even if I do not share his view that the paper is "a tribute to macroeconomics". Bernanke's thoughtfulness, care and attention to detail is apparent. His scientific objectivity in considering his own theory and that of others is exemplary. As with Hansen and Singleton's work, I single Bernanke's work out because it is an example of excellent research within the tradition it represents.

Bernanke's paper does not really reach a substantive conclusion that could possibly change the views of any serious observer of the economy.

---

<sup>6</sup>I regard the fact that Hansen and Singleton's paper was awarded the Frisch medal even after it was recognized that it contained major data errors which affected the substantive conclusions as corroborating this view.

His conclusions state that he has provided evidence against the hypotheses that credit is irrelevant to cyclical fluctuations and that real shocks explain all cyclical fluctuations. Even these judgments are deemed to be "tentative". The only firm conclusion reached is that structural interpretations of VARs are very sensitive to the model one assumes and that future research using VARs should take this into account. It is hard to see how such findings can stimulate new theoretical developments or bring about improvements in our ability to predict, control or explain economic events.

Of course Bernanke could not have known before undertaking his research that the results would be so inconclusive so the last paragraph is not entirely fair. It is instructive therefore to consider what possible results Bernanke could have obtained that might have led to clear conclusions. Since Bernanke is entirely persuasive in challenging standard VAR interpretation procedures, I will focus on his preferred procedure of imposing just enough structural restrictions to provide exact identification. Evidence on the role of credit comes in effect from an analysis of the contemporaneous partial correlation between unforecastable movements in credit and output after holding constant unforecastable military spending, real money balances, and a postulated structural disturbance. The evidence that credit matters is that this partial autocorrelation is significantly positive.

Apart from a myriad of statistical questions that could be raised, obvious questions of economic logic come up. Wouldn't one expect banks to expand their lending activities at times when asset values are up because future business prospects appear bright? If real money properly measured determines real GNP, but there are measurement errors in the real money data as is likely given the proliferation of money substitutes, wouldn't one expect any other variable that moved with GNP to enter an equation for it in a significant way? These considerations suggest that a finding that credit matters hardly proves that it has an active role in causing cyclical fluctuations.

Nor would a finding that credit did not enter Bernanke's equations be evidence that an exogenously caused credit disturbance would have no effect on the economy. Is it not reasonable to expect that if the economy appeared to be overheated, the Federal Reserve would pursue policies to contract credit in order to reduce subsequent values of GNP? Might not credit tend to go down as output went up because firms had more internally generated funds even if *ceteris paribus* credit promoted output growth?

Remarkably, Bernanke never considers interest rates and credit variables in the same set of equations. The only empirical work he reports that considers interest rates as part of a treatment of the monetary

mechanism ignores the role of credit, even though he concludes that credit is important! No explanation is given. I suspect that demands of Bernanke's approach to identification force this rather serious compromise.

The evidence Bernanke regards as suggestive regarding real business cycle models comes from analyses of whether innovations in quarterly measures of base and inside money estimated net of their contemporaneous responses to price and output and postulated to be orthogonal to nominal interest rate disturbances contribute to the forecast variance of real output.

Stepping back from the details of his model, it is hard to see what a correlation like this can prove. If money mattered a great deal and the Fed used it to partially stabilize future movements in output forecastable on the basis of variables like the leading indicators that are not included in the VAR, a negative relationship between money and future output would show up. If money was exogenous, a positive relationship would appear. A zero relationship could easily appear if these two considerations offset each other, even if money had potent effects on economic activity. Alternatively, if money did not matter at all but was allowed to passively accommodate forecastable future GNP movements, an investigator using Bernanke's methods might find it to be statistically very important.

These examples suggest that investigators like Bernanke, who believe that they can learn something new about causal relationships without introducing information beyond that contained in time series whose properties have already been studied thousands of times, are shadow boxing with reality. As the foregoing observations suggest, their identifying assumptions are obviously implausible once stated in English. Formalism and the attendant matrix algebra serves primarily to obscure the futility of the exercise in which they are engaged.

How then can one hope to learn something about economic structure and directions of causation? I argue below that skillful exploitation of natural experiments that provide identifying variation in important variables represents the best hope for increasing our empirical understanding of macroeconomic fluctuations. While lacking the scientific pretension of an explicit probability model, careful historical discussions of events surrounding particular monetary changes, such as those provided by Friedman and Schwartz (1963) persuade precisely because they succeed in identifying relevant natural experiments, and describing their consequences. In the next section I discuss the role of natural experiments in successful empirical economics in more detail as I consider how pragmatic, informal empirical work has greatly advanced economic science.

#### IV. Why Does Pragmatic Empirical Work Have So Great an Impact?

While formal econometric work has had little impact on the growth of economic knowledge, I believe that any history of macroeconomics would have to give great weight to informal pragmatic empirical approaches to economic problems. I think of Friedman's (1957) and Modigliani and Brumberg's (1955) treatments of the consumption function as obvious examples. Other examples might include Friedman and Schwartz (1963), Solow (1957) and Denison (1967); Phillips (1958) whose empirical finding has had profound if largely unanticipated effects; and the large body of work summarized in Fama (1970) on the stochastic properties of speculative prices.

These successful pieces of pragmatic empirical work have three elements in common that distinguish them from most empirical work that strives to be scientific.<sup>7</sup> First and foremost, in each case, the bottom line was a stylized fact or collection of stylized facts characterizing an aspect of how the world worked rather than parameter estimates or formal tests of a point hypothesis. In Friedman's and Modigliani's case, the point involved the importance of savings as a device for smoothing consumption. In Fama's case it was the impossibility of making money by trading stocks on the basis of publicly available information. The conclusion could prove to be persuasive or unpersuasive, but the reader was not in doubt that there was one.

Second, pragmatic pieces of empirical work produce regularities of a kind that theory can seek to explain. Modigliani's finding that wealth entered the consumption function in an important way, as well as corroborating his theory, was suggestive for theories regarding the operation of fiscal and monetary policies. Friedman and Schwartz's work on money's effects continues to be important in spurring theoretical developments. Phillips' finding has gone through a number of interpretations but it has stood as a reality that any theory of the business cycle must confront.

Third, successful pieces of pragmatic empirical work have no scientific pretense. They start from a theoretical viewpoint not a straitjacket. No single test is held out as decisive. Many different types of data are examined. Mayer (1972) counts 16 different types of evidence adduced by Friedman in support of the permanent income hypothesis. No single episode in *A Monetary History* was held out as decisive. No single test reported in Fama's survey proved or disproved anything but a persuasive pattern emerged from the totality. Solow did not treat his estimate as a

---

<sup>7</sup> I find it interesting that the much less extensive empirical studies reported in Lucas (1973) and Lucas (1980a) have most of the characteristics noted below and few of the trappings of rigorous science.

decisive measurement of capital's contribution to economic growth but only as suggestive of its role.

Consider the example of early empirical work on the efficient markets hypothesis suggesting the martingale properties of stock prices. An investigator saddled with the goal of creating a general equilibrium asset pricing theory would have found it much harder to apprehend this empirical observation. In the same way, he would have been unlikely to construct a theory that had as a natural test a comparison of the performance of professional money managers and market averages. Yet such tests, which bore out the efficient markets hypothesis to a very great extent, surely stand as one of the most important empirical conclusions of financial economics.

Perhaps the best way to begin an argument, stressing the effectiveness of pragmatic empirical work, is to consider the problems addressed by the formal work deemed unsatisfactory in the previous section. One such problem was predicting the response of consumption to a change in its determinants such as a temporary tax cut. Here informal work can and has contributed a good deal. First, it can learn the lessons of history. Surely the most relevant knowledge for predicting the effects of a temporary tax cut today is knowledge of the effects of previous temporary tax cuts. Similarly, history can be informative about the effects of announcements regarding future tax policies.<sup>8</sup>

Second, it can examine other relevant natural experiments. For example, there is a sizeable literature surveyed in Mayer (1972) regarding the effects of windfalls on consumption. While there is controversy regarding the issue of whether or not the marginal propensity to consume out of windfalls was lower than that out of more permanent changes in income, there is a consensus that the marginal propensity to consume on nondurables out of windfalls was well in excess of 15 per cent.

Third, pragmatic empirical work can report relevant information and apply common sense. For example, it is surely relevant to an evaluation of the importance of liquidity constraints to know what fraction of a given tax cut will be given to persons who hold no financial assets. Since such persons might be unable to borrow, it would be informative to know also how much of a spike there was at zero in the distribution of net worth.<sup>9</sup> I would also find survey results regarding the fraction of the population that was aware that a tax reform bill passed recently to be informative.

---

<sup>8</sup> See Poterba and Summers (1986) for a first look at this question.

<sup>9</sup> The available data indicate that a large fraction of the population, (greater than half) have essentially no liquid assets and that there is a substantial spike very close to zero. I am not aware of any calculation of the fraction of the population weighted by income or by tax payments with no liquid assets.

Perhaps an even better example of how much more enlightening informal empirical work is than formal work is the area of asset pricing. It is instructive to contrast Hansen and Singleton's test of the consumption capital asset pricing model with the much more informative results of Mehra and Prescott's informal calibration exercise. Hansen and Singleton's bottom line is that their model and the associated stands they took were rejected by the data. It does not point in any particular direction. Mehra and Prescott's work on the other hand highlights the ways in which the data are inconsistent with theory — the spread between the return on stocks and bonds is greater than is consistent with reasonable assessments of the extent of risk aversion. By highlighting a fact, it points towards the development of alternative theories.

At the same time, the nature of the argument makes clear the robustness of the conclusion. Reading Mehra and Prescott's paper, unlike Hansen and Singleton's, it is clear that making different assumptions about separability and time aggregation, or allowing a role for consumer durables is not going to be helpful in salvaging the model under consideration.

An issue closely related to the ones treated by Hansen and Singleton and Mehra and Prescott is the volatility of asset prices. A large literature starting from Shiller's (1981) seminal paper has debated whether stock prices are more volatile than can be rationalized on the basis of fluctuations in fundamental values. It is hard to identify people whose mind has been durably changed by this research. In contrast, two studies, French and Roll (1986) and Shleifer (1985), neither of which are pretentiously scientific in the modern way or make use of elaborate techniques seem to me to force a major change in standard approaches to the theory of asset pricing. Each points to a striking, robust empirical conclusion that theory ought to be able to account for.

French and Roll show that the market is more volatile during trading hours than during nontrading periods, even after controlling for the amount of news received during both periods. That is, over periods when the same amount of public news (statistically) comes out, the market moves by an almost proportionately greater amount when it has been open a larger fraction of the time. This belies the hypothesis that market movements entirely reflect rational responses to news about fundamentals.

Shleifer's result is that when an exogenous event with no information content (joining the S&P 500) leads some investors to increase their demand for a stock, its price moves significantly. That is, the demand curve for securities is not horizontal. This also suggests that price movements are caused by something other than rational responses to publicly available information.

It is too early to know how theory will accommodate these findings. Perhaps they will lead to new models of privately informed traders. My

own suspicion is that the result will be theories of how asset prices are determined when some traders are not rational in the conventional economics usage of the term. The crucial point is that these findings are very likely to have an effect on the theories economists use to understand asset pricing. The reason these studies have been and will be influential is that they have brought new information to bear on important issues. It is new information not new technique that leads to new insights in empirical economics.

Consider as a final example the problem of the role of money in cyclical fluctuations. Reading Friedman and Schwartz's treatment of the 1937 downturn is surely more convincing as a demonstration that financial disturbances have important effects on real economic activity than any Granger causality test. Histories of postwar business cycles like those provided by Eckstein and Sinai (1986) and Wojnilower (1980) that highlight the credit crunches precipitated by Federal Reserve actions that preceded every economic downturn surely is more suggestive for theory than a comparison of the predictive power of a "money" and "credit" variable in a multivariate equation. Evidence like Mussa's (1986) demonstration that across dozens of countries and a number of decades, floating exchange rates dramatically increase real exchange rate variability are much more convincing in suggesting that purely nominal changes can have real effects than are the cross correlations of any particular set of blips no matter how carefully selected.

In most of the examples considered in this section researchers presented an empirical regularity that was sufficiently clear cut that formal techniques were not necessary to perceive it. It is difficult to believe that any of the research described in this section would have been more convincing or correct if the author had begun by laying out some sort of explicit probability model describing how each of the variables to be studied should evolve within a specific pseudo-world. Conversely, it is easy to see how a researcher who insisted on fully articulating a stochastic pseudo-world before meeting up with data would be unable to do most of the work described in this section.

## **V. Tentative Reflections on the Role of Theory**

I have devoted most of this paper to issues of empirical research strategy rather than to the role of theory. This reflects a judgment about my comparative advantage, the existence of Friedman's (1946) persuasive attack on purely formal theory, and a recognition of the fact that macroeconomics is an empirical science. This section concentrates on aspects of the relationship between theory and empirical work. I have already argued



that a great deal of empirical work is too closely tied to highly specific theories. Here I argue that much of macroeconomic theory is excessively divorced from empirical observation and that as a consequence it is taken more seriously than it deserves to be. In large part this is the result of the failure of empirical work to deliver facts in a form where they can be apprehended by theory.<sup>10</sup>

Modern scientific macroeconomics sees a (the?) crucial role of theory as the development of pseudo world's or in Lucas's (1980b) phrase the "provision of fully articulated, artificial economic systems that can serve as laboratories in which policies that would be prohibitively expensive to experiment with in actual economies can be tested out at much lower cost" and explicitly rejects the view that "theory is a collection of assertions about the actual economy". These attitudes lead to the sort of purely formal theorizing which Friedman saw as futile, and increasingly to egregious failures of economic prediction.

While Lucas's definition does not require it, a great deal of the theoretical macroeconomics done by those professing to strive for rigor and generality, neither starts from empirical observation nor concludes with empirically verifiable prediction. A good example is provided by the large (now slightly dated) theoretical literature on monetary growth and the effects of inflation on capital intensity. While early contributions could be seen as extended arguments in the language of mathematics for the empirical prediction that inflation would increase capital intensity, it is difficult to see the empirical content of many subsequent studies. The typical approach is to write down a set of assumptions that seem in some sense reasonable, but are not subject to empirical test (e.g. money enters the utility function) and then derive their implications and report them as a conclusion. Since it is usually admitted that many considerations are omitted, the conclusion is rarely treated as a prediction.

These exercises create potential analogue systems. They examine their properties. Frequently conclusions are reached about what policies would be desirable in an economy like that modeled. However, an infinity of models can be created to justify any particular set of empirical predictions. And I suspect that there is a meta-theorem that any policy recommendation can be derived from some model of optimizing behavior. What then do these exercises teach us about the world? True, each fully worked out system defines laws of motion for its endogenous variables and so can in principle be "taken seriously econometrically" and confronted with data. But, as I have already argued, this tends not be a fruitful or memorable

---

<sup>10</sup> In this section I deal only with work that purports to be in at least an interim sense a final product. Presumably the work directed at developing techniques of analysis should be judged by the results of applying the techniques.

exercise. In fact only a small fraction of theoretical work is ever applied empirically in this way. If empirical testing is ruled out, and persuasion is not attempted, in the end I am not sure these theoretical exercises teach us anything at all about the world we live in.

Given my skepticism about decisive formal econometric tests of hypotheses, it should be clear that in urging that theory should generalize facts and make predictions I do not refer only or even primarily to evidence provided by econometric analyses. Indeed, the empirical facts of which we are most confident and which provide the most secure basis for theory are those that require the least sophisticated statistical analysis to perceive. A good example of theory drawn from experience and logic is Barro and Gordon's (1983) theory of inflation as arising out of considerations of dynamic consistency. This theory explains how policymakers come to pursue policies leading to higher long run average rates of inflation than they or the public would prefer given their benefits. This empirical regularity cannot be demonstrated econometrically but that does not detract from the theory's power. The theory also makes predictions — for instance that societies will seek out those who are more inflation averse than the representative citizen to determine monetary policy — which are amenable to empirical analysis if not formal statistical testing.

My argument so far has been that theoretical research divorced from the problems of empirical generalization and prediction is unlikely to be fruitful. There is however a still greater danger however in research directed at achieving internal consistency starting from first principles without explicit regard for empirical observation. It is all too easy to confuse what is tractable with what is right. There is a tendency to reason that since the world must be consistent, and since all known full-blown models derived from optimizing behavior share a common prediction, that prediction must have some validity. This form of illogic is a modern development. The factor price equalization theorem has rarely been treated as evidence that measurements showing wages to be different across countries were flawed. Yet, Prescott (1986) has recently responded to discrepancies between the predictions of his theory and the data by asserting that "theory is now ahead of measurement".

Reliance on deductive reasoning rather than theory based on empirical evidence is particularly pernicious when economists insist that the only meaningful questions are the ones their most recent models are designed to address. Serious economists who respond to questions about how today's policies will affect tomorrow's economy by taking refuge in technobabble about how the question is meaningless in a dynamic games context abdicate the field to those who are less timid. No small part of our current economic difficulties can be traced to ignorant zealots who gained influence by providing answers to questions that others labeled as

meaningless or difficult. Sound theory based on evidence is surely our best protection against such quackery.

## VI. Conclusions

This paper has made the case that pragmatic empirical work has contributed a great deal to the development of economics just as experimental and observational work have played a key role in the natural sciences. I have argued that formal econometric work where elaborate technique is used to either apply theory to data or to isolate the direction of causal relationships where they are not obvious *a priori* virtually always fails. If this argument is accepted it suggests that in evaluating empirical work we should begin by asking different questions than the ones usually posed. Instead of considering the methods employed, we should ask whether the fact reported is an interesting one that affects our view of how the economy operates. Does it affect our belief about a substantive question?<sup>11</sup>

All too often researchers, referees and editors fail to ask these scientific questions. Instead, they ask the same questions that jugglers' audiences ask — Have virtuosity and skill been demonstrated? Was something difficult done? Often these questions can be answered favorably even where no substantive contribution is being made. It is much easier to demonstrate technical virtuosity than to make a contribution to knowledge. Unfortunately it is also much less useful.

Just as not all demonstrations of virtuosity contribute to knowledge, most empirical work that actually contributes to knowledge does not display the author's capacity for statistical pyrotechnics. Good empirical evidence tells its story regardless of the precise way in which it is analyzed. In large part, it is its simplicity that makes it persuasive. Physicists do not compete to find more and more elaborate ways to observe falling apples. Instead they have made so much progress because theory has sought inspiration from a wide range of empirical phenomena.

Macroeconomics could progress in the same way. But progress is unlikely as long as macroeconomists require the armor of a stochastic pseudo-world before doing battle with evidence from the real one.

---

<sup>11</sup> Consider an example of the application of this test. A large literature has attempted to evaluate the importance of liquidity constraints using Euler equation techniques. There have been few indications that the results have been imprecise relative to researchers' expectations. Yet I do not think the professional consensus regarding the short run marginal propensity to consume is very different today than it was 10 years ago. Nor do I believe that the degree of uncertainty attached to it has declined very much. I find it difficult to believe that those undertaking studies in this tradition today really believe that their individual study will change this situation.

## References

- Barro, Robert J. & Gordon, David B.: A positive theory of monetary policy in a natural rate model. *Journal of Political Economy* 91 (4), 589–610, Aug. 1983.
- Bernanke, Ben: Alternative explanations of the money-income correlation. *Carnegie Rochester Conference Series on Public Policy* 25, 49–101, Autumn 1986.
- Blanchard, Olivier: Wages, prices and output: A structural investigation. Mimeo, MIT, 1986.
- Denison, Edward F.: *Why Growth Rates Differ*. Brookings Institution, 1967.
- Dewald, William G., Thursby, Jerry G. & Anderson, Richard: Replication in empirical economics: The *Journal of Money, Credit and Banking* Project. *American Economic Review* 10 (4), 587–603, Sept. 1986.
- Eckstein, Otto & Sinai, Allen: The mechanisms of the business cycle in the post-war era. In Robert J. Gordon (ed.), *The American Business Cycle: Continuity and Change*, Studies in Business Cycles, Vol. 25, University of Chicago Press for the National Bureau of Economic Research, Chicago, 39–122, 1986.
- Fama, Eugene F.: Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, 383–417, May 1970.
- Feldstein, Martin: Social security, induced retirement and aggregate capital accumulation. *Journal of Political Economy*, 905–926, Sept./Oct. 1974.
- French, Kenneth & Roll, Richard: Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, 5–26, 1986.
- Friedman, Milton: Lange on price flexibility and employment — A methodological criticism. *American Economic Review*, 613–31, Sept. 1946.
- Friedman, Milton: *A Theory of the Consumption Function*. Princeton University Press, 1957.
- Friedman, Milton & Schwartz, Anna: *A Monetary History of the United States 1867–1960*. Princeton University Press for the National Bureau of Economic Research, 1963.
- Hansen, Lars Peter & Singleton, Kenneth J.: Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50 (5), 1269–86, Sept. 1982.
- Hansen, Lars Peter & Singleton, Kenneth J.: Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy* 91 (2), 249–65, March 1983.
- Kreps, David M.: *A Course in Microeconomic Theory*. Harvester Wheatsheaf, New York, 1990.
- Kydland, Finn & Prescott, Edward: Time to build and aggregate fluctuations. *Econometrica* 50 (6), 1345–71, 1982.
- Leamer, Edward: *Specification Searches*. John Wiley & Sons, 1978.
- Lucas, Robert E.: Some international evidence on output-inflation tradeoffs. *American Economic Review*, 326–34, June 1973.
- Lucas, Robert E.: Econometric policy evaluation: A critique. In Karl Brunner & Alan Meltzer (eds.), *Carnegie Rochester Conference Series on Public Policy* 1, 19–46, 1976.
- Lucas, Robert E.: Methods and problems in business cycle theory. *Journal of Money, Credit and Banking* 12 (4), Part 2, 696–715, Nov. 1980a.
- Lucas, Robert E.: Two illustrations of the quantity theory of money. *American Economic Review*, 1980b.
- Lucas, Robert E.: *Models of Business Cycles*, Basil Blackwell, Oxford, 1987.
- Mayer, Thomas: *Permanent Income, Wealth and Consumption*. University of California Press, 1972.
- McClosky, Donald: *The Rhetoric of Economics*. University of Wisconsin Press, Madison, 1985.
- Mehra, Rajnish & Prescott, Edward: The equity premium: A puzzle. *Journal of Monetary Economics* 15 (2), 145–62, 1985.

- Modigliani, Franco & Brumberg, Richard: Utility analysis and the consumption function: An interpretation of cross section data. In Kenneth Kurihara (ed.), *Post Keynesian Economics*, George Allen and Unwin, 1955.
- Mussa, Michael: Nominal exchange rate regimes and the behavior of real exchange rates: Evidence and implications. *Carnegie Rochester Conference Series on Public Policy* 25, 117–215, Autumn 1986.
- Phillips, A. W.: The relation between unemployment and the rate of change of money wages in the United Kingdom, 1861–1957. *Economica*, Nov. 1958.
- Poterba, James & Summers, Lawrence H.: Finite lifetimes and the effects of budget deficits on national savings; 1986; forthcoming in *Journal of Monetary Economics*.
- Prescott, Edward C.: Theory ahead of business-cycle measurement. *Federal Reserve Bank of Minneapolis Review* 10 (4), 9–22, Fall 1986.
- Romer, Paul: Crazy explanations for the productivity slowdown. In Stanley Fischer (ed.), *NBER Macroeconomics Annual*, 1986.
- Rotemberg, Julio: Aggregate consequences of fixed costs of changing prices. *American Economic Review* 73, 433–6, 1983.
- Samuelson, Paul A.: An exact consumption loan model of interest with or without the social contrivance of money. *Journal of Political Economy*, Dec. 1958.
- Sargent, Thomas: Rational expectations, econometric exogeneity and consumption. *Journal of Political Economy* 86 (4), Aug. 1978b.
- Sargent, Thomas: *Dynamic Macroeconomic Theory*. Harvard University Press, 1987.
- Shiller, Robert: Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 421–36, June 1981.
- Shleifer, Andrei: Do demand curves for stock slope downwards? *Journal of Finance*, July 1985.
- Sims, Christopher: Macroeconomics and reality. *Econometrica* 48 (1), 1–48, Jan. 1980.
- Solow, Robert: Technical change and the aggregate production function. *Review of Economic Statistics*, 312–20, 1957.
- Solow, Robert: *Growth Theory: An Exposition*. Oxford University Press, 1970.
- Smith, John Maynard: *Problems of Biology*. Oxford University Press, 1986.
- Tobin, James: *Asset Accumulation and Economic Activity*, Basil Blackwell, Oxford, 1980.
- Weinberg, Steven: *The First Three Minutes*. Basic Books, 1977.
- Wojnilower, Alfred: The central role of credit crunches in recent financial history. *Brookings Papers on Economics Activity* 2, 277–326, 1980.