Is Applied Monetary Policy Analysis Hard?

Jon Faust

(202) 452-2328
faustj@frb.gov
http://e105.org/faustj

Federal Reserve Board
Washington, DC 20551

March 2005
(revised March 28, 2005)

Abstract

In this paper, I argue that applied monetary policy analysis is hard. In particular, all of our models are grossly deficient relative to the ideal, and this cannot be corrected in the medium term. This view has important implications for answering the Simsian question of whether any given change in policy analysis methods is progress or regress. As an application of these ideas, I assess the potential role of dynamic stochastic general equilibrium (DSGE) models to policy analysis. If DSGE models are *modified* so they can play all the functions as old-style models, progress or regress may be unclear. If these models are used in the roles for which they are best suited, they can make a vital positive contribution.

Keywords: DSGE, identification, forecasting
JEL classification:

I would like to thank David Bowman, Dale Henderson, and Eric Leeper for valuable insights. I thank Chris Erceg, Luca Guerrieri, and Chris Gust for the same reason but also for invaluable help with the SIGMA simulations.

What sort of modeling should support day-to-day monetary policy choices?

The economics profession continues to have a great deal to say on this topic, and central bankers spend considerable resources contributing to and remaining abreast of this research. There is surely no other area of economic policymaking where professional economists are so heavily represented on the staffs and policymaking bodies. For example, the Federal Reserve System alone employs over 300 Ph.D. economists. Economists with distinguished academic careers are heavily represented on the FOMC. Central banks around the world sponsor myriad conferences bringing together academic and central bank economists.

In a simplistic view, the agents at central banks would optimally incorporate all relevant information. If so, answering the question of how policy analysis should best be conducted would require little more than summarizing current practice.

Sims (2002) applied this approach, interviewing economists at several central banks about the conduct of policy analysis. His conclusion about the progress in the 25 years since "Macroeconomics and Reality" was negative:

> Econometrics and macroeconomics were active research areas during the 70s, 80s, and 90s, and one might therefore have hoped that there would be clear progress as we moved from the early simultaneous equations models, to MPS and the RDXs, thence to the current QPM and FRBUS model. But if there is progress, it certainly isn't clear, and my own view is that the changes in these models over time have by and large been more regress than progress. (p.23)

This outcome might seem surprising: from the standpoint of the 1970s critiques it might have seemed like there was no where to go but up. As King put it,

> Taken together with the prior inherent difficulties with macroeconometric models, these two events [stagflation and publication of Lucas's criticism] meant that interest in large-scale macroeconometric models essentially evaporated. (King, 1995, p.72)

Of course, King was talking about academic interest; as Sims documents, at least at the Fed and some other central banks, interest continued more-or-less unabated.

It would be easy to explain this state of affairs by positing irrationality on the part of central bankers or academics. In my experience, both groups find some appeal in this explanation, but they tend to disagree regarding which is the nuttier group. An optimizing account might be preferred.

This paper takes up what I will call the Simsian question that we should ask of any marginal change in short-run policy analysis: is it progress or regress? The first part of the paper presents a semi-formal case regarding the nature of progress on policy analysis, and gives rise to some constructive suggestions on answering the Simsian question. The second part applies this perspective to the recent interest at central banks around the world in bringing dynamic stochastic general equilibrium (DSGE) models into the policy process.

A sketch of the argument is as follows. Monetary policy analysis is hard; progress is likely to be slow; and there may be no consensus regarding whether any given step represents progress at all. These observations may strike some as mundane and others as too overwhelming to be of use. This paper argues for a middle ground. The first part develops the foundation for a particularly thoroughgoing version of the claim that monetary policy analysis is hard. In stark form, all our policy analysis models are grossly deficient relative to the ideal and will be so for the indefinite future.[1] I use the term *grossly deficient* in part for dramatic effect, but in part to emphasize my meaning that models are not currently in a small neighborhood of the ideal.

From a practical standpoint, the claim that a model is grossly deficient is not a criticism—it is the starting point of any serious discussion. Further, comparison to the ideal, while important, may not shed much light on the magnitude or even the direction of a marginal change. While I try to make these arguments in a stark and semi-formal way, they are not new; I suspect they were something close to the norm at one time and are clearly represented in contributions of, for example, Friedman and Solow. While this perspective seems to be far less prominent now, it is a strong thread running through Sims's work.

[1] By *ideal* here I mean roughly that the characterization of the business cycle and policy problem embedded in the models is sufficiently good that little judgment is required in getting from the model to a an arguably optimal outcome in practice.

A concrete implication of this view is that model-based research that is put forward as having relatively direct practical implications for policy should, as a matter of course, contain two elements: i) an itemization of the gross deficiencies of the model, ii) some reasoning as to how those deficiencies bear on the practical merits of the advice offered. I will give some illustrations of this approach.

It is a particularly apt time to review standards for progress in practical policy analysis. Policy analysis at central banks may be undergoing the most fundamental change observed since the advent of the first large-scale econometric models. Many central banks around the world and the IMF have constructed or are in the process of constructing dynamic stochastic general equilibrium (DSGE) models for use in the policy process.[2]

In the second part of the paper, I apply the perspective provided in the first by asking the Simsian question of DSGE models. I conclude that these models can make an important marginal contribution to certain dimensions of policy analysis, but cannot yet seriously compete with other forms of analysis in other important dimensions. The largest risk we face in applying advances in DSGE modelling to practical policy is that key benefits may be sacrified in attempting to force the new models to meet all the roles played by the old models. If so the DSGE models' contribution to progress may at best be inscrutable

1 A Sketch of the Problem with 1970s Policy Analysis

The critiques of 1970s policy analysis models contain some elements of broad consensus and others of disagreement. There seems to be general agreement that the models were not grossly deficient as rough characterizations of the reduced form relations in the data. Sims (1980) and Lucas (1981) emphasized this at the time and the view has persisted (e.g., King, 1995). I suspect that Sims and Lucas gave

[2] For example, Bank of Canada: Murchison et al. (2004), Ambler, et al. (2004); Bank of England: Harrison, et al. (2005), ECB: Smets and Wouters (2003), Christiano, et al. (2004), Coenen and Straub (2004); Fed (Erceg, et al. 2005); IMF (Bayoumi, 2004); Riksbank: Aldofson (2005a,b). This list is still under construction.

the reduced form fit of these models a bit of a free pass, however, in order to get on to deeper or more intriguing critiques. Hendry (1980, 1985) argues that many problems could be traced to mundane specification problems detectable before the major breakdown using tools of the time. Pagan (2002) suggests that this view may have been important in explaining how central banks responded to the breakdown of their models.

Lucas argued that the cavalier treatment of expectations was at the root of the problem, and a rational expectations model might solve the problem. The Lucas Critique more narrowly construed argued that the decision rules of forward-looking agents depend on parameters of the policy process; analyses of alternative policies that ignore this fact are useless. Lucas further argued that a mistaken belief in a long-run exploitable tradeoff between inflation and unemployment was one example of the critique in action, and that this belief played an important role in the policy mistakes of the 1970s. It is universally agreed that the work of Friedman, Lucas, and Phelps made a vital contribution in cementing the notion of long-run neutrality in the consciousness of the profession.

For this paper, the implications of the Lucas critique for short-run policy analysis more generally are crucial. Lucas seemed to argue that traditional models were fatally flawed:

> More particularly, I shall argue that the features which lead to success in short-term forecasting are unrelated to quantitative policy evaluation, that the major econometric models are (well) designed to perform the former task only, and that simulations using these models can, in principle, provide *no* useful information as to the actual consequences of alternative economic policies. (emphasis in orig.; 1981, p.105)

Sims (e.g., 1980, 1987) argued that the Lucas critique was not a particularly important impediment to useful short-run policy analysis. Rather, in "Macroeconomics and Reality," Sims argued that the problem with conventional policy analysis models was a familiar one, even at that time: failure to apply the lessons of the Cowles Commission regarding identification.

Finally, more recently, Orphanides (2004) has emphasized that mismeasurement of the natural rate of unemployment or output gap was a major contributor to the policy mistakes at the time. The Fed overestimated the degree of slack in the economy and provided too much stimulus.

The profession has not yet reached consensus on the relative contributions of these various problems to the mistakes of the 1970s. All of the culprits, however, fit under the broad umbrella of failing to recognize and make proper allowance for the limits of our understanding. While one clear lesson is that we needed better understanding, Sims's review of modeling a quarter century later questions whether progress has been realized. In the next section, I begin laying out a perspective on how this could happen and how we might do better.

2 Analysis of Hard Problems: The Abstract Case

The view I am advocating is a traditional one. Formalizing a bit may be helpful, and fortunately we can borrow the tools from theoretical computer science and from a few applications in economics. Rust (1997) provides a brilliant introduction to, and perspective on, these issues in economics. Rust's main emphasis is on prospects for advances in basic research, however. My purpose in bringing in this literature is to draw out some practical lessons for those who want to contribute to the making of monetary policy, say, next week or next year.

Begin with two relatively straightforward claims: First, some decision problems are currently too hard to *solve formally*. Second, *formal* methods outside the class of *formal solutions* sometimes prove useful in such cases.

I hope that these claims are uncontroversial in the minimal sense that we can all envision nontrivial interpretations of them that are correct. The decision problem of a doctor diagnosing and prescribing has not been formally solved— medical decisions are not backed by an optimality proof. Despite this fact, medical science has evolved techniques that I believe could reasonably be called formal; these techniques have arguably advanced medical practice.[3]

Slightly greater precision may be useful; thus, define,[4]

> A decision problem is *formally solved* if an optimal decision rule is derived based on a description of the environment and criterion functions of the agents.

That is, a formal solution comes from a model with microfoundations. We can partition the study of an unsolved problem into two types: A) the formal solution of simplified versions of the problems, and B) all others—approaches, whether formal or not, that do not involve literally formally solving problems.

2.1 Hard problems and progressive approaches

The progressive research strategy under the type-A analysis involves solving a sequence of ever more realistic problems until the problem is for practical purposes solved. This approach has great appeal, but as any well-trained economist knows, whenever we specify such a sequence, it is important to examine the convergence properties.

Without analysis of the convergence properties of the type-A agenda, there is no guarantee that the type-A approach progresses in any simple or direct sense.[5] In some fields, such as theoretical computer science, it is common to evaluate the likely convergence properties of the type-A sequence. Certain high dimensional problems are conjectured to be intractable, and in generating practical policy advice on these problem type-B approaches are also pursued.

As a familiar example, consider the decision problem of a chess player. Formally describing the decision-problem in chess is relatively straightforward. Simplified versions of the game (e.g. endgame problems) are regularly solved. However, because

[3] To put a sharper point on things, up until recently, medicine had a good understanding of the microfoundations of very few diseases. Further, up until recently and, perhaps beyond, only a small portion of all medical judgment was backed by carefully controlled studies. Sims (1996) makes a similar point.

[4] While I give one notion of formal solution, in a more formal context many more could be proposed (e.g., Schaeffer and Lake, 1996).

[5] The words *simple or direct* are important here. Type-A work may be vital, say, in informing the judgments that may be involved in type-B work.

the number of states of the full game (on the order of 10^44) is large and no reliable methods for radically collapsing the analysis is known, progressively solving more complete versions has no prospect of progressing to fully the optimal solution.

The computer science literature provides an elaborate classification of the complexity and tractability of formally solving certain problems. For a good presentation of these issues, see Garey and Johnson (1983). The concepts fit our discussion of convergent sequences quite nicely, as they are based on how the number of computational steps in solving the problem grows with an index of the size (in our case, the realism) of the problem. A standard approach to showing *de facto* intractability is to demonstrate that a problem is computationally equivalent to those in some class widely conjectured to be intractable. The most familiar set conjectured to be intractable is "NP-hard" problems.[6]

Formal definition of NP-hard is not crucial here, but an example may clarify. A familiar NP-hard problem is the traveling salesman problem: find the shortest route among N points. The number of possible routes grows rapidly with N and no efficient way of finding and verifying the full optimum is known. The proof that this problem is NP-hard leads to the conjecture that no efficient method for collapsing the analysis of these routes exists; high-dimensional versions of the problem are taken to be intractable.

The point here is simple: if the goal is to generate advice on real world decision problems, it is sensible to first ask if the problem is *hard*, where for the purpose of discussion in this paper, I define *hard* as follows:

> A decision problem is *hard* if formally solving a sequence of progressively more realistic problems—the type-A approach—is unlikely to yield a formal solution to the practical problem, at least on any relevant horizon.

For *hard* problems, the type-A approach may not be of much direct use on any short horizon.

[6] Loosely, these are problems for which it is conjectured that the solution complexity cannot be bounded by a (deterministic) polynomial.

2.2 Successful examples of Type-B analysis

The mere fact that a decision problem is *hard* does not imply that there are better routes to improving practical decisionmaking than the type-A approach. In practice, however, there are many successful examples of formal type-B modeling. The example of medicine was discussed above, but some examples closer to monetary policy may be of interest. Type-B work on practical scheduling problems, which are closely related to the travelling salesman problem, has arguably been useful.[7] Many of these approaches might be described as approximate formal solution by numerical optimization.

Examples with an even looser connection to formal solution may be more provacative. The decision problems in checkers, chess, and air combat are each viewed as unamenable to type-A progress.[8] A common, and arguably successful, approach to modeling these problems is the use of a method resembling dynamic programming and resembling the practical strategy Svensson (2004) describes for monetary policy. In particular, at each point when a decision is to be made, one forecasts the state of the problem a few steps into the future under various decisions. One assigns a value to each of the forecasted states; the decision with the highest value is chosen. This is type-B analysis because the value function is entirely *ad hoc*. For example, the possible states of the system may be partitioned in some rather arbitrary way and unrelated value functions used depending on which partition the current state lies in.[9]

Using the language of monetary policy analysis, we may say that policy simulation models based on these type-B principles are regularly used in improving the judgments of actual policymakers. For chess and checkers, policy rules based on these methods has arguably attained a status on par with the best subjective policy

[7] See, e.g., the special issue of *Operations Research* dedicated to "stochastic and dynamic models in transportation," Jan./Feb. 1993.

[8] For interesting discussion in these areas, on chess, see, e.g., Marsland and Schaeffer, 1990; on checkers, Schaeffer and Lake, 1996 and Schaeffer, et al. 1992; and on air combat, see, e.g. Hämäläinen 2002. What I present here is obviously a gross simplification.

[9] Much of the art in this area is efficiently evaluating many branches of the problem tree. For an interesting discussion of these issues, see Schaeffer, 1989

rules of experts.[10]

2.3 Some Implications

Generic weaknesses of type-A and type-B strategies are obvious. Formal solutions of simpler problems will give technically correct insights, but the value of these insights for the practical problem may be difficult to discern. Moving outside the class of such formal solutions may allow one to consider a more realistic description of the problem, but the lack of analytical foundations implies that the value of the insights obtained may be murky at best.

For clarity, however, let me emphasize some type-B gross deficiencies. It should be no more than a graduate-school exercise to show that type-B approaches such as those mentioned above may be subject to the Lucas critique, may not be internally consistent, and may lead to dynamic instabilities (of a dramatic variety in the case of air combat) or other bad properties that would not be present in a fully optimal solution. Thus, it may be trivial to show that type-B analysis is grossly deficient *relative to the ideal*.

The main point of this discussion, though, is that for *hard* problems, neither approach has an obvious advantage. Empirically, it seems clear that both are highly productive. For hard problems, any practical insights from type-A work must involve an element of type-B work—the analysis required to bridge the gap between the type-A characterization and the practical problem.

Gross deficiencies may be easier to identify than resolve. In a world of finite development resources, such problems should go on the list with other known gross deficiencies of the analysis. The optimizing type-B analyst should attempt to apply resources efficiently to the resolving or at least making allowances for these deficiencies.

[10] It is commonly believed that checkers has been solved. The senses in which this is true or false is at the heart of this discussion. Checkers has not been solved from a type-A perspective. However, decision rules based on type-B reasoning have been refined to the point that they arguably dominate judgmental play. In practical fields, this is one definition of *solved*. For others, see Schaeffer and Lake, 1996.

3 Is Monetary Policy Analysis Hard? Should Economists Give Up?

Translated to these terms, Lucas correctly argued that the monetary policy analysis in the 1960s and 1970s was type-B analysis, that the logical foundations of that analysis were a mess, and that important policy mistakes had followed from the deficiencies. This led to a dramatic shift in the profession toward type-A analysis as represented in DSGE modeling. King (1995) lays out the progressive approach of the DSGE agenda in more-or-less the same terms I used above. The approach started solving simple models and attempting to match a few features of the data. The models progressively became more complex and capable of fitting more features.

Is this approach progressive in any simple sense? Lucas, in his own words, "went out on a limb" (1981, fn. 18, p. 237) and conjectured that it is:

> I think it is fairly clear that there is nothing in the behavior of observed economic time series which precludes ordering them in equilibrium terms, and enough theoretical examples exist to lend confidence to the hope that this can be done in an explicit and rigorous way. To date, however, no equilibrium model has been developed which meets these standards and which, at the same time, could pass the test posed by the Adelmans (1959). My own guess would be that success in this sense is five, but not twenty-five years off. (1981, p. 234)

As I have framed it, a broad interpretation of this conjecture entails the conjecture that the business cycle modeling problem is not *hard*. While the conjecture may have been reasonable at the time, there clearly is no consensus in the profession that the ensuing decades have seen the predicted success.

With the benefit of hindsight, I argue that monetary policy analysis and the business cycle modeling that would support it may in fact be *hard*. This section gives a bare-bones argument that this is a theoretically coherent position, that it is a formalization of a traditional perspective, and that policymakers such as Alan Greenspan arguably take this perspective.

3.1 Monetary policy analysis may be *hard*

As Rust (1997) notes, there are a few applications of complexity theory to problems resembling macroeconomics. For example, Spear (1989) and Board (1994) show senses in which reaching rational expectations in equilibrium in a simple decentralized economy is a *hard* problem.[11] If agents cannot reach rational expectations equilibrium, it seems unlikely that central bank economists will be able to solve the optimal policy problem.

Although these papers use tools that may be unfamiliar, the source of the result is painfully familiar. In pursuing formal solutions of macro models, we do not stray far from the cases of representative agents and firms, common information, no learning, and perfect foresight. As anyone working in this area knows, substantially loosening any of these assumptions can make the problem intractable using current methods (e.g., Sims, 1998). In my view, it is at best an article of faith to suppose that monetary policy analysis is anything but *hard*.[12]

While I have made this argument with reference to formal results regarding complexity, the view was probably conventional wisdom at one time. Hayek (e.g., 1989) makes the general case. Solow argued the following in the inaugural Hicks Lecture:

> But suppose economics is not a complete science . . . , and maybe even has very little prospect of becoming one. Suppose all it can do is help us to organize our necessarily incomplete perceptions about the economy, to see connections the untutored eye would miss, to tell plausible stories with the help of a few central principles. Suppose, in other words, that economics is a 'discipline, not a science. Those are Sir John's words. . . . In that case what we want a piece of economic theory to do is precisely

[11] Spear shows that incomplete information may render the problem intractable even without the polynomial bounding of computing; Board shows that, even with complete information, *probably approximately correct learning* is *hard* for arguably relevant cases. That is, even getting a close approximation to the true economy with high probability is computationally intractable.

[12] Note that I am not arguing that these problems are unsolvable, only that the solution may not come any time soon. Applicants of intractability results in all fields emphasize that the results are for the most part conjectures, depend on the formalization of the problem, and further humans have solved many problems conjectured to be intractable by one of these standards. Rust (1997) disucsses these points in the economic context and Fraenkel (2002) similarly discusses them regarding the analysis of games.

> to train our intuition, to give us a handle on the facts in the inelegant American phrase. (1984,p.15)

As evidenced below, Friedman arguably based his case for the k-percent rule on similar beliefs.

Statements of policymakers such as the Feds Chairman Greenspan also are consistent with them believing that the monetary policy problem is *hard*

> Despite the extensive effort to capture and quantify these key macroeconomic relationships, our knowledge about many of these important linkages is far from complete and in all likelihood will always remain so. (Greenspan, 2003, pp.1–2)

> In implementing a risk-management approach to policy, we must confront the fact that only a limited number of risks can be quantified with any confidence. And even these risks are generally quantifiable only if we accept the assumption that the future will replicate the past. Other risks are essentially unquantifiable— representing Knightian uncertainty, if you will. . . .(p.5)

3.2 Is type-B analysis of macroeconomic policy meaningless?

While type-B analysis has arguably been of some value in some other contexts, it may be that it would not be valuable for monetary policy analysis. As noted in Section 1, Lucas argued that traditional approaches can provide *no* useful information. He further argued,

> In situations of [Knightian] risk, the hypothesis of rational behavior on the part of agents will have valuable content, so that behavior may be explainable in terms of economic theory. In such situations, expectations are rational in Muths sense. In cases of [Knightian] uncertainty, economic reasoning will be of no value. (1981, p.224)

These quotes come from the height of the rational expectations revolution, and my point is not to critique the rhetoric of revolution. Given a quarter century of hindsight, however, I am arguing for a reappraisal of the view that the type-A approach is the exclusive route to progress. In this spirit, I note that these statements are not supported by formal analysis—no axioms were specified under which economic reasoning was shown to be of no value. Sims was one of the earliest advocates of the view that the Lucas Critique was not a fatal impediment to short-run policy analysis, and he has elaborated the view in a number of papers (1980,1987,2002).[13]

3.3 Progressive Type-B Analysis of Monetary Policy

Friedman's argument for the k-percent money growth rule is a beautiful example of type-B analysis of monetary policy. Because the optimality properties of such a rule have been much studied, one might forget that Friedmans justification was based not on optimality, but on ignorance. For example, while he conjectured that this rule would perform "surprisingly well," Friedman noted,

> It is not perhaps a proposal that one would consider at all optimum if our knowledge of the fundamental causes of cyclical fluctuations were considerably greater than I, for one, think it to be; it is a proposal that involves a minimum reliance on uncertain and untested knowledge. (1948, p.263)

He continued with a fairly thorough discussion of the main dangers in the proposal, including, "The proposal may not succeed in reducing cyclical fluctuations to tolerable proportions.... I do not see how it is possible to know now whether this is the case." (p.264) In his own foray into type-B analysis, Lucas endorsed Friedman's program, concluding,

> The program would (I think on this there is no serious professional disagreement) *fully* protect the economy against sustained inflation. It would *fully* insure against the kind of monetary collapse which was so important a factor in the early stages of the Great Depression... (1981, p. 257)

In retrospect, very few economists today would advocate the k-percent rule. Arguably the greatest advance in monetary policymaking since the 1970s, however, is the emergence of a strong consensus that monetary policy should first and foremost

[13] Leeper and Zha (2003) give a thorough practical perspective on this point. Others have argued against the empirical importance of the Lucas Critique in various contexts from a time series perspective (e.g., Ericsson and Irons, 1995).

protect the economy against sustained inflation and monetary collapse. This view is clearly reflected in the inflation targeting framework. Indeed, one could argue that the inflation targeting framework maintains Friedmans limited ambitions, but represents a further concession to our ignorance beyond that in his proposal. The k-percent rule is rigid and its presumed good properties rest, essentially, on velocity being relatively stable. In contrast to the rigidity of the k-percent rule, the inflation targeting framework is a “snug fitting garment” (Bernanke, et al., 1999) that allows the central bank to bend as needed in, perhaps, unforeseen ways, to attain the goal of nominal stability. In a notable piece of type-B reasoning, Svensson (e.g., 2004) attempts to put some structure on this policy approach.

Inflation targeting advocates would, I think, be reticent to claim that it fully protects the economy from inflation and deflation; rather, they would probably argue that it may perform surprisingly well in this regard. Of course, the final chapter on inflation targeting is not yet written, but in my view, this chain of developments is a clear triumph of progressive work in the type-B tradition.

A second and related area of type-B progress is the growing consensus that central banks should be as transparent as is reasonably possible. Like the cases for the k-percent rule and inflation targeting, the case for transparency does not rest on an optimality proof. Indeed, the skeptic would argue that the main lesson from the transparency literature is that, in a distorted economy, enhancing transparency does not unambiguously enhance welfare.[14] The most persuasive economic argument for transparency, in my view, comes instead from a type-B perspective:[15] in a complicated economy, agents are likely to find opaque policy confusing and central banks, in turn, will find analysis of policy effects confusing. Being clear has the prospect of reducing unintended consequences stemming from avoidable confusion.

Of course, we can illustrate something akin to this point in a model with micro-foundations; transparency can increase the Kalman gain on certain optimal inference

[14] This argument is reviewed more completely in Henderson and Faust, 2004.

[15] As many have argued, a strong argument can also be made for transparency on democratic grounds. Another persuasive economic argument is that transparency may in some cases reduce the time consistency problem. For one discussion of these, see Faust and Henderson, 2004.

problems of the private sector as in, e.g., Faust and Svensson (2001). This sort of work invites one to consider the optimal level of transparency— a level that often turns out to be less than full transparency.

A stronger, but less well-founded, case for transparency comes from the Friedman perspective above. In the face of our limited knowledge, optimizing transparency is beyond our ability; minimizing confusion seems like a good rule of thumb. The evolution of transparency at the Fed has arguably been driven, in significant part, by this perspective.[16]

The introduction of robust control to macroeconomics is also an important contribution in type-B analysis, as it involves stepping away from exact optimization. It is beyond the scope of this paper to review either the more abstract or applied contributions in this regard;[17] whether or not formal robust control is in some sense *the answer* remains questionable (e.g., Sims, 2001b).

3.4 Barbarians, Gates, Camels, and Tents

Thus far, I have argued the modest claims that our models may be grossly deficient in important ways, that they may remain so for the indefinite future, and that, in light of this, work outside the type-A tradition is important. Put differently, Friedman (1948) may not have exhausted the valuable analysis that could flow, not from solving models, but from recognizing their limits. Modest as this claim may be, it might (and probably should) provoke fears that the barbarians are at the gates, or at least that the camels nose is under the edge of the tent. Solow recognized this problem, noting,

> I hope that no one will fall into the error of thinking that this low-key view of the nature of economics is a license for loose thinking. Logical rigour is exactly as important in this scheme of things as it is in the more self-consciously scientific one. (Solow, 1984, p.15)

[16] See the discussion in the February, 1994 FOMC transcript. For example, President Broaddus stated, "But there are risks of not doing this [announcement]. If there were any confusion tomorrow going into the weekend or this thing gets played out in the *New York Times* on Saturday and Sunday or on CNN, I think we would have a real mess." (Federal Reserve Board, 1994).

[17] See, e.g., Hansen and Sargent (2003) the special issue of *Macroeconomic Dynamics*, vol. 6, 2002, and e.g., Levin, et al. 1999, for a nice application.

To amplify this view, let me emphasize my own view that the profession has made much practical progress in our ability to reason about dynamic macroeconomic questions over the past 20 years. While there have also been many advances in empirically documenting dynamic features, our improved ability to coherently discuss these is due in large part to the dogged pursuit of formalism in the Type-A tradition.

Solow's main point in the Hicks lecture was that the IS/LM model represented the trained intuition of the macroeconomic profession. He argued that, when confronted with a macro problem, the typical economist's thought processes were channeled through some IS/LM-like framework. As Sims (2000) argues, reasoning about dynamic macro questions with the IS/LM model is not easy. Arguably, the trained intuition of the profession is now, or will soon be, closer to some DSGE model. This is a significant advance. Overall, type-A work is vital, and I am not even arguing that excessive resources have been directed to that effort.[18]

3.5 Implications

The key claims of the first part of the paper are these. Our policy analysis models are grossly deficient relative to the ideal, and are likely to remain so. In this context, a large element of type-B analysis is needed—at least to bridge the gap between model and practical policy. It sells the profession short to say we have nothing to say, or even nothing formal to say, on this topic.

In my view, progress would be facilitated if any model-based practical policy advice were evaluated in light of an itemization of the gross deficiencies and informed analysis of how the deficiencies might or might not temper the appeal of the advice. By *practical policy advice*, I mean advice being put forward as relatively directly applicable in practice.[19]

[18] To venture a small distance out on a limb, I do suspect that greater resources could productively be applied to type-B analysis. I am less clear on the least productive current activity that might be reduced to pay for this.

[19] Thus, I am not suggesting that papers constituting basic research meet this standard. Policy implications reported in such work come with an explicit or implicit introduction to the effect, "If we were to take the model literally, which I am not advising, here are the policy implications."

Friedman and Sims provide us outstanding applications of these principles. Friedman's advocacy of the k-percent rule came with a thorough discussion of the potentially fatal problems and gaps in the analysis. Sims's advocacy of VAR work in "Macroeconomics and Reality" came with a listing of issues to be resolved and the sober and prophetic warning that,

> A long road remains, however, between what has been displayed here and models in this [VAR] style that compete seriously with existing models on their home ground—forecasting and policy projection. (1980, p.33)

It is important to note that evaluating progress in a world of multiple gross deficiencies is subtle. Lessons from the welfare theory of the second best may be of value in this endeavor. In the face of multiple distortions, we know that rectifying any single distortion need not increase, and may decrease, welfare. Judging modeling advances in a world of multiple deficiencies is similarly complex. Clear thinking on this issue, however, can be facilitated clear discussion of the complete set of deficiencies. In the second part of the paper, I apply this perspective to the DSGE models that are coming into use at central banks.

4 DSGE models for policy analysis: a type-B perspective

I discuss two deficiencies of current DSGE models: inscrutable identification and counterfactual risk premia. Then I attempt to apply a broad type-B view to answer the Simsian question of whether these models represent progress or regress.

4.1 Inscrutable identification

We now have business cycle models with microfoundations that show many characteristics we believe to be important in reality. For now, let us concede that, as far as fitting the reduced form, these models have attained a status on a par with the models of the 1970s. Of course, this was the starting point of the critique in "Macroeconomics in Reality." To translate this critique to the DSGE context, it is not sufficient that we derive a model with *arbitrary* microfoundations that *happens* to match the reduced form. To be clearly better than old-style models in terms Sims's classic critique, we need models with microfoundations that are arguably correct. This point has not been lost on Sims,[20] of course, who argues,

> But we need to remain aware that there are many potential ways to generate price stickiness and non-neutrality. Similar qualitative aggregate observations may be accounted for by mechanisms with contradictory implications for welfare evaluation of monetary policy. (2001a,p.5)

Sims further notes (1998, p.11), "Exact, formal, complete identification based on mistaken assumptions can lead to error fully as bad as inexact identification." On this same point, Wallace (2001) argues that getting the *right* frictions in the model (in his case, to generate the demand for money) may lead to fundamentally different answers to causal questions than come from the use of more convenient specifications. While Wallace's topic of choice is money, I could extend this section by attempting to document the claim that there is no sector of current DSGE models for which a consensus of sector experts supports the microfoundations.[21] In this section, I try to analyze how DSGE modellers avoid the problem with incredible identification that has plagued macroeconometricians, leading to a constructive suggestion for moving the discussion forward.

The econometric literature has struggled mightily to convincingly identify the effects of various kinds of shocks in macro.[22] What is the DSGE solution? DSGE models are often strongly overidentified—even in their linearized form. For example, there may be sufficiently few shocks in the model so that the data are stochastically singular. As King (1995) notes, the study of models that would be rejected outright by standard tests has greatly complicated—and at times led to a breakdown of—discussion of model adequacy between econometricians and "quantitative modelers." Special tools like those of Watson (1993) evolved to facilitate this discussion.

[20] Leeper, 2005 makes a similar point.

[21] For example, on consumption, see Carroll (2001), on sticky prices, see Bils and Klenow (2004), and so on.

[22] For my take on the myriad problems, see Faust, 1998

Perhaps we need an analogous way to discuss the reliability of the identification of DSGE models, which in traditional terms may be grossly overidentified. The essence of the identification problem as laid out by the Cowles Commission is this: two structures with different causal interpretations of the data can be observationally equivalent—that is, imply the same reduced form. Before putting one structure forward as true, the others need to be ruled out. A convenient feature of the analysis is that the family of observationally equivalent structures we need to rule out is trivial to characterize. In an N-equation linear system, this family is generated by pre-multiplying the system by any full rank matrix.[23] Thus,

$$AX_t = BX_{t-1} + \varepsilon_t$$

is observationally equivalent to $A^*X_t = B^*X_{t-1} + \varepsilon_t^*$, where $A^* = DA$, $B^* = DB$ and $\varepsilon^* = D\varepsilon$, for any full rank D.[24]

As DSGE modelers encountered persistence in the data that was hard to match, the exploration was naturally extended to allow various ways of making things "sticky." These include various forms of adjustment costs on the magnitude and/or rate of change of adjustment, non-time-separable preferences of various sorts, brute constraints on decisions such as price setting, and limits on knowledge and learning.

This modeling strategy looks might be described as searching a set of restrictions until one finds a subset the data find amenable. This bears a close family resemblance to the econometric approach that the Cowles Commission and Sims warned us against.[25]

The proper standard for credible identification in the quantitative macro setting is probably something like the following. Specify a broad family of modifications to

[23] Obviously, as I have stated it here, some of these equivalent structures only involve a change in units and not in causal structure; N of the restrictions may be viewed as picking the units.

[24] This neat and convenient feature may be more apparent than real, as recent discussions of weak identification make clear. In finite samples, we should be interested in ruling out all the alternative causal structures that, loosely speaking, fit nearly as well as the proposed structure. Characterizing this set can be more difficult.

[25] That is, searching for a set of zero restrictions on lags that are both not rejected and, if imposed, would meet the rank condition for identification.

the basic model that one is willing to entertain.[26] A particular causal story inferred from the data should be viewed as credibly-identified only if there is no model in the acceptable family that that fits the data similarly well by the chosen criterion yet yields a different causal story. This standard modifies the familiar Cowles Commission standard in two ways. First, in this context, the natural family of alternatives may not be as obvious as in the linear model case. Second, the criterion of "fit" is vague enough to allow either econometric or quantitative modeling criteria.

Work that is in this spirit exists.[27] The important fact from the type-B perspective is that these studies usually inspect a very small family of alternatives. The VAR literature has taught us the perils of this approach. For example, nearly every identified VAR paper contains the obligatory paragraph or footnote that says, essentially, "We tried a few other identifications, and causal inferences were pretty much the same." When thorough search of the arguably relevant space of alternative causal structures is pursued, however, many of these robustness claims are overturned (e.g., Faust, 1998; Faust and Rogers, 2003).

It is natural that the DSGE literature has proceeded by investigating a limited number of ways to generate persistence: the literature is not on the wrong path. Until we have attained some identification standard like the one described above, however, we should view the status of our causal claims from these models to be roughly on a par with the models Sims so appropriately criticized in the past. Thus, we might argue that DSGE models do not show clear progress on identification and, indeed, may show regress.

An slight amendment is in order, however. Sims described the identification of policy models as "incredible;" I prefer the term "inscrutable." One cannot readily discern the reasoned basis for preferring one causal structure over others. Having observed the evolution of these models for many years, I believe that the design phase of the models is not considered complete until the model gets the *right answer* to key

[26] Loosely speaking, specify an acceptable set of frictions.

[27] For example, Guerrieri (2005) examines the performance of various contracting assumptions in accounting for inflation persistence; Milani (2005) examines how learning versus other mechanisms perform in generating such persistence.

questions. Thus, the actual identifying assumptions are the set of beliefs that the model must match. While the foundations of these beliefs may be murky, if beliefs they be, forcing the policy models to match them is surely a good idea. (Note that this is a purely type-B conjecture and would be difficult to prove.)

The switch from "incredible" to "inscrutable" has an important practical implication: it is easier to improve on "inscrutable." The *right answers* that identify the model could be codified and imposed in some way on the model.[28] The work of Christiano, Eichenbaum, and Evans (2005) illustrates one way this can be done; the impulse response to a monetary policy shock is specified and taken as given; the DSGE parameterization is chosen to give the right answer.[29] Clarifying the identification in this way could make more productive the efforts to constructively critique the models by both central bank staff and outside observers. Absent such clarification, outside assessment may be limited mainly to speculation about whether the identification is incredible or inscrutable.

4.2 Gross deficiencies in fit

In the previous subsection, we took as given that current DSGE models characterize the reduced form acceptably well. Of course, very serious questions remain about the ability to capture many important features of the data.[30] The particular problem I will use as an illustration is currency risk premia; I could make the same point with risk premia reflected in the term structure of interest rates.

One reason I take risk premia as my example fit problems is that they raise

[28] Note that this explicitly rejects that notion that at this point in time we can parameterize macro models, choose the parameters based, say, on micro studies, and then learn the causal effects we want in our policy models. We simply are not there yet. As we see the need to add ever more frictions to generate the observed patterns in the data, we seem to have gone beyond the point where the required parameters can be obviously connected with some real analog. Of course, the calibration approach may still be a vital part of more basic research; but it will not currently generate adequate policy simulation models.

[29] The particular impulse response in that work is taken from the VAR literature and treated as fixed. These details are not crucial to the main point in the text.

[30] As with critiques regarding the appropriate frictions, everyone may have their favorite deficiency in fit. For a review of some of these and some implications for policy, see Bowman and Doyle, 2002 or Leeper, 2005.

an issue that is the polar opposite to those raised above. With identification, the problem is that we may have multiple ways to provide microfoundations that imply similar fit to the data. In contrast, we seem to have no widely accept microfoundations for matching the behavior of certain risk premia. These premia may, however, be at the core of monetary policy analysis questions.

We typically study linearized versions of DSGE models in which there are no risk premia. Thus, uncovered interest rate parity (UIP) holds.[31] For later reference, it is useful to define the UIP relation as,

$$E_t s_{t+1} - s_t = i_t - i_t^*$$

In an obvious notation, the one-period expected change in the (log) exchange rate is equal to the home minus foreign one-period interest rate differential, and s is stated as the price of foreign currency. We will also talk about a UIP deviation defined as

$$\xi_t = i_t - i_t^* - (E_t s_{t+1} - s_t)$$

UIP fails badly empirically, especially in U.S. data. UIP implies that forward rates are equivalent to expected future short rates, but time series examination of dollar exchange rates, and many others, generally finds a negative correlation between expected changes implied by forward premia and subsequent changes in exchange rates (at horizons such as a few months). This is known as the forward premium bias in the exchange rate literature.[32] More generally, in what I view as a classic of type-B analysis, Meese and Rogoff (1983) showed that all the professions and central banks fundamentals-based models performed worse than a random walk model in predicting exchange rate variation.

Before picking on DSGE models, it is important to note that UIP failure represents a significant irritant in all forms of practical policy analysis. At the Fed, any

[31] Solving the models to a higher order of approximation leads to essentially the same result for relevant parameterizations, as risk premia turn out to be quite small.

[32] For a review, see Engel, 1996. For recent estimates regarding the term structure and UIP, see Chernenko et al. 2004. UIP seems to hold approximately at very short horizons, Chaboud and Wright (2003), and perhaps at very long horizons, Chinn and Meredith (2001).

fundamentals-based analysis of the exchange rate (which will be roughly consistent with UIP) tends to be tempered with a strong dose of the Meese-Rogoff medicine. Of course, policymakers may differ on whether to go with the data or with the theory. For example, at the Bank of England (1999),

> Some [Monetary Policy] Committee Members were inclined towards UIP, in part because they could see no compelling reasons relating to risk to assume that the exchange rate will depreciate at a slower rate than implied by interest differentials. (Bank of England, 1999)

It is a classic type-B question to resolve which of these is a more sensible approach.[33]

As a brief aside, I note that this same issue applies to the term structure of interest rates. That is, while the expectations theory fails badly in the data it holds in the DSGE models and analysts regularly reason based on it. As Blinder observed,[34]

> Yet everyone—and here I mean analysts, market participants and central bankers alike—continues [despite the evidence] to "read" the market's expectations of future short rates from the yeild curve, as if doing so made sense. I find it hard to explain why everyone is doing what everyone knows to be wrong.... Perhaps the reason is that no one has offered a convincing alternative interpretation of the term structure. (1997, p.16)

Current DSGE models embed this same problem.

The fact that the data present features we have no good way to account for raises special problems for DSGE-based policy analysis. I will illustrat these issues using the Fed's SIGMA model. This model has been developed in the International Finance Division, primarily by Erceg, Guerrieri, and Gust (e.g., 2004). The model is, I believe, an outstanding example of the new class of DSGE models emerging Central Banks around the world.

I hope that a brief discussion of SIGMA will suffice for my purposes. The version of the model that produced the results below has two countries, with the home country calibrated to be about one-quarter of the world. The openness to trade of the

[33] Pagan (2003) further discusses the UIP problem as it manifests itself at the Bank of England.
[34] For recent evidence on the predictive power of the term structure, see Chernenko, et al.,2004.

home country roughly approximates that of the U.S. As in Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2003), the model has several features to add stickiness: habit persistence in consumption, costs of adjusting investment, and wage and price setting governed by Calvo contracts. Additionally, there are costs of adjusting imports and some *Keynesian,* or rule-of-thumb consumers. The model is calibrated conventionally and has various good properties, including producing a reasonable impulse response to a monetary policy shock.

Figure 1 shows the reaction of key variables to what I will call a baseline monetary policy shock. The nominal one-quarter interest rate (left panel) rises by a percentage point; through the endogenous persistence generated by the Taylor rule for policy (which contains the lagged rate) and the model economy, the interest rate gradually decays back to baseline. Because feedbacks to the other economy are not strong, the foreign interest rate is not much affected, so the interest differential mirrors the home rate. The economy responds roughly in line with conventional views (right panel). Both output and inflation fall, the maximum effect is delayed a few quarters, and the variables return to baseline.

These dynamics obey UIP. One might argue for ignoring UIP deviations based on the fact that *conditional UIP* could hold. That is, whatever the source of the unconditional failure of UIP, monetary policy shocks might not contribute to the variance of ξ, the deviation from UIP. As Eichenbaum and Evans (1995) demonstrated, this assertion can be tested in the VAR setting. Faust and Rogers (2003) extend this idea by searching to see if there exists *any* reasonable way to identify a policy shock in conventional VARs that is roughly consistent with conditional UIP. [35] They find that there is not: there seem to be substantial, persistent deviations from UIP in response to any shock that looks roughly like a policy shock.[36]

To shed some light on the importance of UIP deviations, Erceg-Guerrieri-Gust implement an *ad hoc* UIP shock in SIGMA. In particular, they *artificially* impose

[35] See the paper for a detailed discussion of the meaning of "any reasonable way." The minimalist criteria mainly involved sign restrictions on impact effects of the shock.

[36] Note that there is no claim here that this fact creates a puzzle from an efficient markets perspective. We are only claiming that conventional DSGE models do not produce such a result.

that ξ is a persistent AR(1) process. One can interpret the approach as positing the existence of a proportional transactions cost between the sending and receipt of cross-border interest payments. The revenues are rebated in a lump-sum fashion.

Figure 2 compares the effects of our baseline policy shock with an alternative composite shock constructed as a linear combination of the baseline shock and a small, but persistent, UIP deviation (or transactions cost) shock. The magnitude of the UIP deviation (ξ) can be seen in the lower left panel of fig. 2: it is very persistent at about 16 basis points. This size UIP shock is roughly consistent with the Faust-Rogers findings. The resulting path of the home interest rate is not much different from before, but the path of the interest differential is now a bit different from the path of the home rate (top left panel).

Because the home economy is calibrated to the U.S., a relatively closed economy, the path of output under our composite shock is not much different from the baseline. There is independent interest at the Fed (especially at present) about the effect of shocks on the trade balance (top right) and value of the home currency (bottom left), call it the dollar. These variables behave quite differently under the two shocks. In the baseline the trade balance initially deteriorates; after 6 quarters, the effect is largely gone. In the composite shock, there is an immediate improvement in the trade balance and the effect grows for many quarters. The nominal exchange rate appreciates initially in the baseline shock and then depreciates toward a new long-run, slightly appreciated, value. In the composite shock, the exchange rate instead depreciates initially. It then further depreciates to a new depreciated long-run value.

The internal consistency of these results can be described starting with the behavior of relative inflation (lower right; note the longer time horizon). The paths are similar in the two cases, with inflation at home initially falling in relative terms. The curve is shifted up slightly for the composite shock. By purchasing power parity, the change in the long-run nominal value of the dollar is essentially given by the (negative of the) sum of relative inflation over the course of the shock. For both policy tightening shocks, home inflation is initially pushed down relative to foreign, but this situation is eventually reversed. The reversal is small for the baseline shock, and the total effect on net inflation is negative, implying long-run appreciation. By shifting the relative inflation curve up a bit, the tail comes to dominate, the net effect becomes positive, and the long-run effect on the dollar is depreciation. Having nailed down the long-run effect, the time-path of the nominal exchange rate is given by the interest differential (net of any risk premium). Thus, under each shock the exchange rate must depreciate along the path as home rates rise relative to foreign.[37] This gets us to the initial effects on the nominal exchange rate. Because the exchange rate initially depreciates in the composite shock, the trade balance immediately improves. The rise in exports places some upward pressure on home inflation. This reconciles the upward shift in the net inflation curve, which is where we started.

One idea illustrated here is obvious: the data show UIP deviations; the core model does not; this may matter. In SIGMA one can impose a UIP deviation, but the cleanest way to interpret the source of the deviations is as a transactions cost; this is not put forward as a realistic account. Our results might be different if the UIP deviations arose in a more plausible way. Further, the composite shock still obeys the expectations theory of the term structure. Clearly, the failure of UIP and the expectations theory could be related in ways that matter for the questions of interest.

The second point is a bit more subtle. It has long been noted that in simultaneous equations modeling, single-equation or other limited-information methods might be preferred to fully simultaneous methods in presence of misspecification.[38] Full systems methods may tend to spread the effects of the specification error throughout the system in unwelcome ways. One might prefer to confine the effects.

With the old-style models, users became expert at imposing various type-B beliefs (that is, judgments) about how to resolve puzzles. Further, the potential chan-

[37] The differential net of premium ultimately changes sign but is quite small in both shocks so that the exchange rate reverses a bit of the depreciation.

[38] I am unaware (but would like to know) if some individual deserves credit for this bit of wisdom. The basic undergraduate and graduate texts I learned from both make the point (Maddala, 1977; Amemiya, 1985).

nels for misspecification in one area were limited and could be managed.

The glory of the DSGE models is their internal consistency and general equilibrium nature. Especially in the face of deficiencies for which we simply know no good fix, however, this rigid internal consistency can be a mixed blessing. This is especially so if one wants to perform practical policy analysis.

I should emphasize that there are offsetting benefits of the DSGE structure. For example, it is relatively straightforward to explore the effects of alternative assumptions in these models. In the big old-style models it is essentially impossible to pose and address questions like, "What if price stickiness arises in a different way?"

4.3 The role of DSGE more generally

While I have provided some discussion of two deficiencies of current DSGE models for policy analysis, this in no way resolves the Simsian question regarding the marginal contribution these models might make to the policy process. As noted above, we are in a second-best world, and evaluation of marginal progress is a subtle question that requires consideration of the full set of problems and constraints. In this case, I think the discussion requires clarity on the roles models play in our world of incomplete understanding.

From the perspective of the ideal, we probably should perform all of our analysis (forecasting, policy analysis, etc.) using a single, internally consistent and economically coherent model of the world. Central bankers regularly lament that this is not currently possible (e.g., Stockton 2002) and appeal to a suite of models. Pagan (2002, 2003) discusses many trade-offs faced by model designers giving rise to a production possibilities frontier. He questions whether the suite of models we actually observe is mainly driven by sensibly choosing points on the frontier, however. I present a slightly different perspective on the roles models play motivated in large part by the thoughtful discussions of Pagan and Stockton.

1. Refining Economic Intuition. Refining our intuition is an essential, perhaps the essential, role of fully articulated theory models. Policymakers are likely to find this role most essential in the face of nonstandard shocks (where intuitions are not well refined), especially shocks that have effects depending fundamentally on expectational effects and general equilibrium effects. It is no surprise that the development of the Fed's SIGMA model got a large boost and its earliest applications in examining the effects of productivity shocks of uncertain persistence.

To be most effective in this regard, the model needs to be large enough—have enough shocks and margins that shocks might affect—to examine nonstandard topics that arise. On the other hand, as King (1995) emphasized, if the model is to be of help in refining our intuitions, it also must be small enough that one can understand it. DSGE models are, in my view, ideally suited for playing this role and are already doing so.

2. Reduced form forecasting. Forecasting is central to policymaking. Various gross deficiencies of DSGE models in fitting the data may make them less than ideal for this role. One solution to this problem is to add shocks—perhaps with their own dynamics—and frictions until the fit is acceptable. Clearly, this will compromise the effectiveness of the model in function (1).

As Sims (1980, 2002) and Bernanke (2005) emphasize, forecasting models arguably should incorporate information from a very large number of variables. Much work (e.g., Stock and Watson, 2002, 2003a) lends empirical credence to this view. Additionally, Sims (2002) suggests that the good properties of the Fed's Greenbook forecast are due to the staff's ability to incorporate a broad range of information about the state of the economy at the time of the forecast. Expanding the DSGE models to incorporate a very large set of variables and to have the hooks for injecting nonstandard information may further compromise the DSGE model's ability to meet function (1).

Finally, a large body of empirical work directly or indirectly casts doubt on the view that the best forecast will, in practice, come from a coherent model. For example, the large forecast-averaging literature suggests that the average of a large number of diverse forecasts often dominates using any single forecast.[39] Of course, the average of the forecast from a diverse group of models need not itself be representable as the output of a coherent model.

It is important to note that the empirical regularity in the forecast averaging literature is more bizarre than it sounds initially. The result is that simple averaging outperforms many more sophisticated methods of "optimal" combination. In a number of contexts, thinking harder does not, empirically, seem to help.[40]

For all these reasons, we have little reason to suppose that any single model, and especially any easily comprehensible DSGE model, will provide the basis of the best reduced-form forecast. The first two roles of models are obviously in conflict. Sims (1998) provides an excellent discussion of the practical conflict here and insightful (type-B) conjectures about how to make the best of things. Before turning to balancing these issue, I propose a third role that is arguably the most important role played by large models at the Fed.

3. Repository of a baseline view of the world. Forecasting and policymaking, more generally, at the Federal Reserve is heavily judgment based.[41] As the FOMC works largely by consensus, the policy adopted is intended to reflect the collective judgment of the 19 members of the FOMC.[42] There are well-known problems of human judgment generally and of the judgment of groups and of experts, in particular. It is beyond the scope of this paper to fully review the literature on these topics. To give a flavor, however, I mention three issues that may be important: i) how an issue is framed may play an important role in the decision that is reached (Tversky and Kahneman, 2000), more specifically, ii) when presented a brute fact (say, a five percent chance of recession), judgment is affected by the story that comes with it

[39] See e.g., Granger (1989) and Granger and Newbold (1974). For an application near central banking, see Stock and Watson, 2003b. More recently, Bayesian model averaging in economics shows promise, e.g., Wright, (2003a,b).

[40] Of course, this result cannot, in principle, be true generally. Hendry and Clements (2002) discuss, e.g., the problems with averaging encompassed models.

[41] This is arguably true of other central banks as well, but I will confine my comments here to the case with which I am most familiar.

[42] This is despite the fact that only a subset are officially voting at any point in time.

(and consumer confidence is down) (Tversky and Kahneman, 2002), iii) experts and others may not reliably assess their past successes and failures of past judgments in assessing the reliability of current judgments (Griffen and Tversky, 2002; Dawes et al. 2002).

It is arguably the case that a *repository model* could be especially useful in light of these observations.[43] The large model, can help frame discussion similarly at successive policy meetings. It can be the repository of a standard, slowly evolving set of stories underlying the forecast. It might also help in calibrating views of past success and failure.

I suspect that this role as a repository of a baseline view may be one of the most important played by the Feds large models. Two comments are particularly relevant to the topic of this paper. First, I see little reason to suppose that the model best for fulfilling this role will also be best in fulfilling the other two roles. Indeed, this role may sit between the other two. For framing discussion and providing reasons, it is useful that the repository model have more structure than, say, a reduced form forecast based on averaging forecasts from mutually inconsistent models. On the other hand, policymakers may have important (type-B) judgments about some issues that remain puzzles from the standpoint of current DSGE models. If these are central to the issues at hand, then such DSGE models may not be an adequate repository of the standard view.

Second, it may be important that the repository model codifies, in some crude way, how economists think about the economy (at least a relevant subset of them). As noted above, in this view, old-style models are clearly becoming less useful; a model incorporating important aspects of DSGE-style models may, and perhaps should, come to dominate.

[43] Put another way, Blinder (1997, p.5) argues that policymaking at the Fed is "far too situational." Perhaps the large models are used as a partial offset to this problem.

4.4 The Simsian Question: A tentative bottom line and constructive suggestion

Models play at least three roles at central banks. Given the current deficiencies in our knowledge, the best model from the perspective of any one role taken in isolation is unlikely to be the best from the narrow perspective of the others. Indeed, optimizing the models for these separate roles independently would almost surely give rise to models that are, at least in some respects, mutually inconsistent.

In this context, what sort of marginal contribution can DSGE models make? My own suspicion is that it would do more harm than good to toy with the models in an effort to make them fulfill all three roles. This would likely lead to many compromises and perhaps create an incentive simply to paper over certain gross deficiencies. Only if central banks take this path are we likely to reach a future point at which it appears there has been regress, or even a point at which progress has become inscrutable. If central banks use DSGE models to contribute along the lines of their greatest strengths—loosely speaking, helping us understand how conventional economic reasoning plays out in a given context—the contribution will be very positive.

There is a more general point here. As noted above, central bankers regularly lament the fact that they cannot perform their analysis using a single, internally consistent framework. DSGE modeling may on some long horizon be the source of such a framework. In the short-run, internal inconsistency is often easier to identify and criticize than to *beneficially* resolve. To paraphrase Emerson, a foolish consistency is the hobgoblin of little models.

This argument leads me to one main suggestion. For short-term advances in practical policy analysis, effort spent clarifying the roles of models and optimizing models in those roles may be at least as important as efforts spent developing the grand unified model.

In my experience, the role that gets the least direct attention is simply characterizing the reduced form. Critics often pass over the reduced-form fit on their way to more enticing subjects. The shift to DSGE models with their fit problems could move us further in this direction. The tendency of central banks to publish only a *conditional forecast* based on a counter-factual path for policy further complicates the study of forecast properties. More careful analysis of the strengths and weaknesses of an unconditional forecast could lead to more rapid advances in our understanding of the reduced form.

5 Conclusions

Initial readers of this paper astutely pointed out the lamentably few concrete implications—despite my best attempts to paper over this gross deficiency. My best excuse in this regard is that in the main, I wanted to put forward and illustrate a perspective that I hope could lead to a more rapid and constructive evolution of policy analysis at central banks.

Short-run monetary policy analysis is hard; all of our current models are grossly deficient relative to the ideal. In this view, those of us involved in practical monetary policymaking would be well-advised to maintain a list of the outstanding gross deficiencies. In this endeavor, Sims has, for many years, been the central banker's best friend.

The biggest correctable mistake in the 1970s was failure to critically evaluate and account for gaps in our knowledge. That our models are grossly deficient may not be a correctable problem, but our ability to account for and guard against failures stemming from those deficiencies may be improved.

Two simple steps will arguably help. First, we should welcome expansion of the list of gross deficiencies. Second, the deficiencies should be addressed in an efficient manner in light of the full set of constraints, including those imposed by the other deficiencies. In light of multiple deficiencies, as in the theory of the second best, it may be hard to judge whether a marginal change is an improvement. In making this judgment simple comparisons with the ideal may not be very helpful. Ultimately, these judgments must be based on what I have called type-B analysis: they cannot be purely model based.

From the perspective of helping central banks improve short-run policy analysis, the greatest correctable problem with academic research is failure to thoroughly pursue the two steps just mentioned. First, there seems to be a tendency to avoid a frank discussion of gross deficiencies. Such discussion, I believe, is viewed as *negative*, where the italics indicate a deep, but ill-defined, sense of distaste. Perhaps the divisive nature of macroeconomics over the last 25 years has also reduced the tendency for any group to frankly air its dirty laundry. Second, I think there has been a reduced tendency, relative to the past, to move outside the model and engage economic judgment to assess practical value of the work in light of the deficiencies. Looking back, Friedman's (1948) proposal for monetary and fiscal stability sets a remarkably high standard in this regard.

Absent greater academic focus on these issues, central bankers—and a few friends like Sims—are left to provide these bits of analysis on their own. Speaking for myself as a central banker, we could use some help.

References

Adolfson, M., S. Laséen, J. Lindé and M. Villani, Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through, Sveriges Riksbank Working Paper no 179. March 2005a.

Adolfson, M., S. Laséen, J. Lindé and M. Villani , The Role of Sticky Prices in an Open Economy DSGE Model: A Bayesian Investigation, *Journal of the European Economic Association*, Papers and Proceedings, 2005b, forthcoming.

Ambler, Steve, Ali Dib, and Nooman Rebei, Optimal Taylor Rules in an Estimated Model of a Small Open Economy, Bank of Canada Working Paper, 2004-36, 2004.

Amemiya, Takeshi, *Advanced Econometrics*, London: Basil Blackwell, 1985.

Bank of England, *Monetary Policy Report*, Nov. 1999, Bank of England: London.

Bayoumi, Tamin, GEM A New International Macroeconomic Model, Occasional Paper 239, International Monetary Fund, 2004.

Bernanke, Ben, Monetary Policy Modeling: Where Are We and Where Should We Be Going? in *Models and Monetary Policy: Research in the Tradition of Dale Henderson, Richard Porter, and Peter Tinsley,* Jon Faust, Athanasios Orphanides and David Reifschneider, eds., Federal Reserve Board: Washington, 2005, forthcoming.

Bernanke, Ben, Laubach, Thomas, Mishkin, Frederic, and Posen, Adam, *Inflation Targeting,* 1999, Princeton: Princeton University Press.

Bils, Mark and Pete Klenow, Some Evidence on the Importance of Sticky Prices, *Journal of Political Economy,* Oct. 2004, vol. 112, no. 5, pp. 947–985.

Blinder, Alan, Distinguished Lecture on Economics in Government: What Central Bankers could learn from Academics—and Vice Versa, *Journal of Economic Perspectives,* vol. 11, no. 2, Spring 1997.

Board, Raymond, Polynomially Bounded Rationality *Journal of Economic Theory,* vol. 63, iss. 2, 1994, pp.246–70

Bowman, David, and Brian Doyle, New Keynesian, Open-Economy Models and Their Implications for Monetary Policy, in *Price Adjustment and Monetary Policy,* Canada: Bank of Canada, 2002

Carroll, Christopher, A Theory of the Consumption Function, With and Without Liquidity Constraints, *Journal of Economic Perspectives,* vol. 15, no. 3, Summer 2001.

Chaboud, Alain and Jonathan Wright, Uncovered Interest Rate Parity: It Works, but Not For Long, *Journal of International Economics*, forthcoming, 2003.

Chernenko, V. Sergey, Krista B. Schwarz, and Jonathan H. Wright, The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk, IFDP no. 2004-08, Federal Reserve Board, 2004.

Chinn, Menzie and Guy Meredith, Testing Uncovered Interest Rate Parity at Short and Long Horizons, manuscript, University of Wisconsin, 2001.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy, *Journal of Political Economy*, vol. 113, no. 1, Feb. 2005, pp. 1–45.

Christiano, Lawrence J., Robert Motto, Massimo Rostagno, The Great Depression and the Friedman-Schwartz Hypothesis, NBER Working Paper 10255, Jan. 2004.

Coenen, Günter, and Roland Straub, Non-Ricardian Households and Fiscal Policy in an Estimated DSGE Model of the Euro Area, manuscript, ECB, 2004.

Dawes, Robyn, David Faust, and Paul Meehl, Clinical versus Actuarial Judgment, in *Heuristics and Biases, the Psycology of Intuitive Judgment,* Thomas Gilovich, Dale Griffen, and Daniel Kahneman eds., Cambridge University Press: Cambridge, 2002.

Engel, C. The forward discount anomaly and the risk premium: a survey of recent evidence, *Journal of Empirical Finance* 3, 123–191.

Erceg, Christopher, Luca Guerrieri, Christopher Gust, SIGMA: A New Open Economy Model for Policy Analysis, manuscript, Federal Reserve Board, 2003.

Ericsson, Neil R and Irons, John S., The Lucas Critique in Practice: Theory without Measurement, *Macroeconometrics: Developments, tensions, and prospects,* 1995, pp.263–312, Recent Economic Thought Series. Boston, Dordrecht and London: Kluwer Academic,

Faust, Jon and Henderson, Dale, Is Inflation Targeting Best Practice Monetary Policy? *Federal Reserve Bank of St. Louis Review,* vol. 86, No. 4. 2004

Faust, Jon and Rogers, John H., Monetary Policy's Role in Exchange Rate Behavior *Journal of Monetary Economics,* vol. 50, iss. 7, 2003, pp.1403–24.

Faust, Jon and Svensson, Lars E. O., Transparency and Credibility: Monetary Policy with Unobservable Goals, *International Economic Review,* vol. 42, iss. 2, 2001, pp.369–97.

Faust, Jon, The Robustness of Identified VAR Conclusions About Money, *Carnegie-Rochester Conference Series on Public Policy,* vol. 49, 1998, pp.207–244.

Federal Reserve Board, Transcript FOMC Meeting, Feb 3-4, 1994.

Fraenkel, Aviezri, Complexity, Appeal, and Challenges of Combinatorial Game Theory, *Theoretical Computer Science,* vol. 313, no. 3, pp. 417–425.

Friedman, Milton, A Monetary and Fiscal Framework for Economic Stability, *American Economic Review,* vol. 38, no. 3, 1948, pp.245–264.

Garey, M. R. and Johnson, D. S., *Computers and Intractability: A Guide to the Theory of NP-Completeness,* 1983, New York: W. H. Freeman.

Granger, C., Combining forecasts — twenty years later, *Journal of Forecasting*, vol. 8, 1989, pp. 167-174.

Granger, C., and P. Newbold, Experience with statistical forecasting and with combining forecasts, *Journal of the Royal Statistical Society* 1974.

Greenspan, Alan, Opening Remarks, Monetary Policy and Uncertainty: Adapting to a Changing Economy, Federal Reserve Bank of Kansas City: Kansas City, 2003.

Griffen, Dale and Amos Tversky, The Weighing of Evidence and the Determinants of Confidence, in *Heuristics and Biases, the Psycology of Intuitive Judgment*, Thomas Gilovich, Dale Griffen, and Daniel Kahneman eds., Cambridge University Press: Cambridge, 2002.

Guerrieri, Luca, The Inflation Persistence of Staggered Contracts, *Journal of Money, Credit, and Banking,* 2005 (forthcoming).

Hämäläinen, Vesa-Pekka, *A Decision Analytic Simulation Approach to a One-on One Air Combat Game*, Mat-2.108, Helsinki University of Technology, 2002.

Hansen, Lars, and Thomas Sargent, *Robust Control and Economic Model Uncertainty*, manuscript, 2003.

Harrison, Richard, Kalin Nikolov, Meghan Quinn, Gareth Ramsay, Alasdair Scott, and Ryland Thomas, *The Bank of England Quarterly Model*, Bank of England, 2005.

Hayek, F.A., The Pretense of Knowledge (reprint of 1974 Nobel Prize Memorial Lecture), *American Economic Review* vol. 79 no. 6, 1989, pp. 3–7.

Hendry, D.F., Monetary Economic Myth and Econometric Reality, *Oxford Review of Economic Policy*, vol. 1 no. 1, Spring 1985, pp. 72–84.

Hendry, D.F., Econometrics—Alchemy or Science? *Economica*, vol. 47, No. 188, Nov. 1980, pp. 387–406.

Hendry, D.F. and Clements, M.P., Pooling of Forcasts, *Econometrics Journal*, vol.5, 2002, pp. 1–26.

King, Robert, Quantitative Theory and Econometrics, *Federal Reserve Bank of Richmond Economic Quaterly*, Summer 1995, pp.53–105.

Leeper, Eric M., Discussion of 'Price and Wage Inflation Targeting: Variations on a Theme by Erceg, Henderson and Levin' by Matthew B. Canzoneri, Robert E. Cumby and Behzad T. Diba, in *Models and Monetary Policy: Research in the Tradition of Dale Henderson, Richard Porter, and Peter Tinsley,* Jon Faust, Athanasios Orphanides and David Reifschneider, eds., Federal Reserve Board: Washington, 2005, forthcoming.

Leeper, Eric M. and Zha, Tao, Modest Policy Interventions, *Journal of Monetary Economics,* vol. 50, iss. 8, 2003, pp.1673–1700.

Levin, Andrew, Volker Wieland, and John C. Williams, The Robustness of Simple Monetary Policy Rules under Model Uncertainty, in John B. Taylor, ed., *Monetary Policy Rules*, Chicago: University of Chicago Press, 1999.

Lucas, Robert, *Studies in Business-Cycle Theory,* 1981 MIT Press: Cambridge.

Maddala, G.S., *Econometrics*, McGraw Hell: New York, 1977.

Meese, Richard, and Ken Rogoff, Empirical EXchange Rate Models of the 1970s: do they fit out of sample? *Journal of International Economics*, vol. 14, pp. 3–24.

Milani, Fabio, Expectations, Learning, and Macroeconomic Persistence, manuscript, Princeton, 2005.

Murchison, Stephen, Andrew Rennison, and Zhenhua Zhu, A Structural Small Open-Economy Model for Canada, Bank of Canada Working Paper, 2004-4, 2004.

Orphanides, Athanasios, Monetary Policy Rules, Macroeconomic Stability, and Inflation: A View from the Trenches, *Journal of Money, Credit, and Banking,* vol. 36, iss. 2, 2004, pp.151–75.

Pagan, Adrian, *Report on modeling and forecasting at the Bank of England*, London: Bank of England, 2003.

Pagan, Adrian, What is a good macroeconomic model for a central bank to use? manscript, Austrailian National University, 2002.

Rust, John, Dealing with the Complexity of Economic Calculations, manuscript, University of Maryland, 1997.

Schaeffer, Jonathan, Joseph Culberson, Norman Treloar, Brent Knight, Paul Lu, and Duane Szafron, A World Championship Caliber Checkers Program, *Artificial Intelligence*, v. 53 no. 2-3, 1992, pp. 273–290.

Schaeffer, Jonathan, and Robert Lake, Solving the Game of Checkers, *Games of No Chance*, v. 29, 1996, pp. 119-133.

Sims, Christopher, The Role of Models and Probabilities in the Monetary Policy Process, *Brookings Papers on Economic Activity, 2002,* iss. 2, 2002, pp.1–40.

Sims, Christopher, Comments on Papers by Jordi Galí and by Stefania Albanesi, V.V. Chari, and Lawrence J. Christiano, manuscript, Princeton, 2001a.

Sims, Christopher, Pitfalls of a Minimax Approach to Model Uncertainty, manuscript, Princeton, 2001b.

Sims, Christopher, Whither ISLM, manuscript, Princeton, 2000.

Sims, Christopher, Macroeconomics and Methodology, *Journal of Economic Perspectives,* 10, Winter 1996, pp. 105–120.

Sims, Christopher, Projecting Policy Effects with Statistical Models, manuscript, Princeton, 1988.

Sims, Christopher, A Rational Expectations Framework for Short Run Policy Analysis, in *New Approaches to Monetary Economics,* W. Barnett and K. Singleton eds, Cambridge University Press: Cambridge, 1987, pp.293–310.

Sims, Christopher, Macroeconomics and Reality *Econometrica,* vol. 48, iss. 1, 1980, pp.1–48.

Smets, Frank and Wouters, Raf, An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area, *Journal of the European Economic Association,* vol.1 no.5, 2003, pp. 1123–1175.

Spear, S.E., Learning rational expectations under computability constraints, *Econometrica*, vol. 57, 1989, pp.889–910.

Solow, R. M., Solow, Robert, Mr. Hicks and the Classics, *Oxford Economic Papers,* vol. 36, iss. 0, 1984, pp.13–25.

Stock, James and Mark Watson, Forecasting Output and Inflation: The Role of Asset Prices,*Journal of Economic Literature*, 41, 2003a, pp. 788-829

Stock, James and Mark Watson, Combination Forecasts of Output Growth in a Seven-Country Data Set, *Journal of Forecasting*, forthcoming, 2003b.

Stock, James and Mark Watson, Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association,* vol. 97, December 2002, pp. 1167–1179

Stockton, David, What Makes a Good Model for the Central Bank to Use?, manuscript, Federal Reserve Board, 2002,

Svensson, Lars E.O., Monetary Policy with Judgment: Forecast Targeting, manuscript, princeton, 2004.

Tversky, Amos, and Daniel Kahneman, Rational Choice and the Framing of Decisions, in *Choices, Values, and Frames*, Kahneman and Tversky, eds. Cambridge University Press: Cambridge, 2002.

Tversky, Amos, and Daniel Kahneman, Extensional versus Intuitive Reasoning, in Heuristics and Biases, the Psycology of Intuitive Judgment, Thomas Gilovich, Dale Griffen, and Daniel Kahneman eds., Cambridge University Press: Cambridge, 2002.

Wallace, Neil, Whither Monetary Economics?, *International Economic Review*, vol. 42, no. 4, pp. 847–869, Nov. 2001.

Watson, Mark W., Measures of Fit for Calibrated Models, *Journal of Political Economy,* vol. 101, iss. 6, 1993, pp.1011–41.

Wright, Jonathan H.,Forecasting U.S. Inflation by Bayesian Model Averaging, Federal Reserve Board, IFDP 2003-780, September 2003a.

Wright, Jonathan H., Bayesian Model Averaging and Exchange Rate Forecasts Federal Reserve Board, IFDP 2003-779, September 2003b.

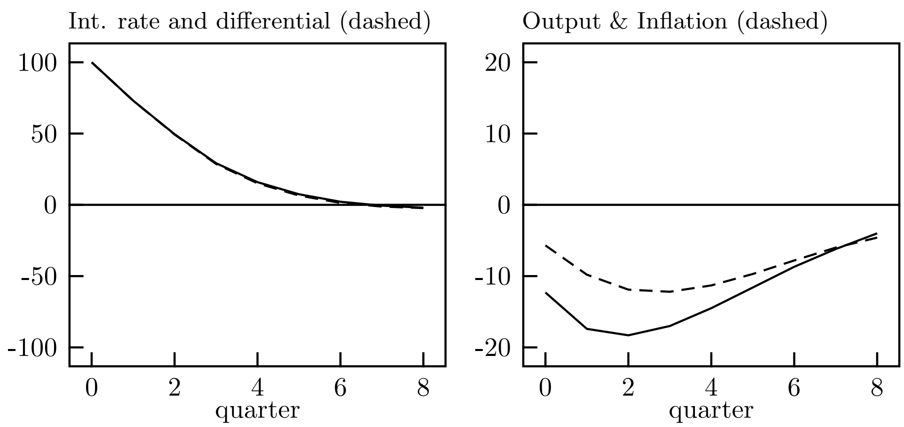

FIGURE 1: Impulse responses to baseline monetary policy shock. Magnitudes are in basis points.

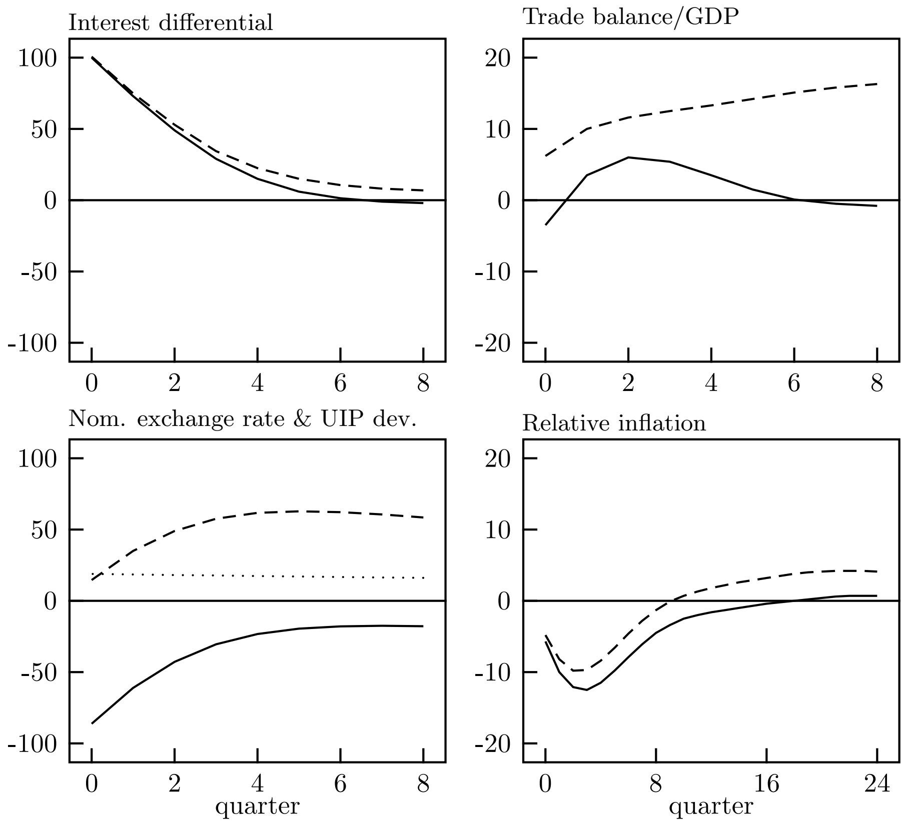

Figure 2: Impulse responses to baseline monetary policy shock (solid) and composite monetary policy shock (dashed). Magnitudes are in basis points. The composite shock has UIP deviation, ξ, shown in dots on lower left.