DSGE models in a second-best world of policy analysis

Jon Faust

(410) 510-7614
faustj@frb.gov
http://e105.org/faustj

Johns Hopkins University
Baltimore, MD 21218

March 2005
(revised August 27, 2008)

Abstract

Large-scale macroeconometric models have for 40 years been workhorses of policy analysis at central banks—this despite the bumpy patch that accompanied their adoption. DSGE models are now replacing the old macroeconometric models. This paper is about avoiding the bumpy patch this time. The paper argues that the new models are on the whole no better, and are in some dimensions worse, than the old models by the standards of the critiques of the 1970s models. The paper argues for a more pragmatic standard and illustrates some tools for enhancing the practical value of the models.

Keywords: DSGE, identification, forecasting
JEL classification:

I would like to thank Abhishek Gupta, Luca Guerrieri, Dale Henderson, Eric Leeper, and Lars Svensson for valuable insights and Marco DeNegro and Jesper Linde for the same and for going above and beyond the call of duty in helping me analyze DSGE models they helped develop.

The 1960s were an exciting time. Most notably, an impressive new kind of macroeconometric model was entering central banking, and cutting-edge central banks were beginning to analyze policy as an problem of optimal control. The Dec. 1965 edition of *Time*, a popular U.S. news magazine, has Keynes on the cover, and is almost giddy in tone over the successes of countercyclical policy. Indeed, one gets the impression that the future of the business cycle might be rather dull: '[U.S. businessmen] have begun to take for granted that the Government will intervene to head off recession or choke off inflation.'

By the revealed preference of central bankers, the new econometric models were a lasting success. The original models and their direct descendents remained workhorses of policy analysis at central banks for the next forty years or so. Were it not for the role the models played in the tragic economic events of the 1970s, this would be a very happy tale of science translating into advances in practical policymaking.

We are again in exciting times, with a new breed of policy analysis model entering central banking from academics. Cutting-edge central banks are beginning to analyze monetary policy as an optimal control problem. For the first time since the mistakes of the 1970s, *science* is gaining the upper hand in some discussions of the art and science of monetary policymaking (e.g., Mishkin, 2007). At a recent central banking conference, I heard a senior central banker lament that the modern strategy of model-based flexible inflation targeting might render central banking rather dull.

As was true of the large-scale macroeconometric models, the new dynamic stochastic general equilibrium (DSGE) models coming into use at central banks represent a major advance over what came before. There are plentiful opportunities for practical policy to benefit from the recent scientific advances in modelling.

This paper is, however, about how to avoid the bumpy patch that occurred soon after adoption of last generation of models. I argue that the process of moving major scientific advances into policy is prone to such mistakes. More specifically, I argue that the mistakes flow, in part, from failure to carefully appraise what has actually been achieved in the science, failure to carefully identify the source

of practical policymaking benefits, and failure to pay attention to the limits of the new science. The unsurprising implication is that we should carefully evaluate the practical merits of scientific advances.

Unfortunately, the contentious nature of macroeconomics, in general, and policy modeling, in particular, make it especially challenging for the policymaking world to reliably assess the practical merits of advances. Before critiquing the models, I attempt to untangle some of the confusing threads debates over policy modeling, leading to two important perspectives from which to evaluate the new models.

The bottom line from these critiques is that, from the perspective of the prominent critiques 1970s modelling, the new models are no better than existing models. The new models are clearly worse in some important dimensions. In concrete terms, optimal policy analysis in the new models rests on foundations that, on balance, are no more solid than analysis using the previous generation of models.

I argue that the critiques of the 1970s set an unrealistic standard, however. From what I argue is a more practical perspective the models represent an vital advance. In the final section of the paper I provide some suggestions about how to maximize the benefits from the new science.

1 Preliminaries: science, policy, and academic critique

1.1 A medical tragedy

Fleming's 1928 discovery of the antibiotic properties of penicillin revolutionized the science of infectious disease. The expanding array of antibiotics over the following decades led to amazing decreases in mortality and morbidity from these diseases [e.g., Lewis, 1995].

Indeed, by the 1970s, many authorities were declaring the problem solved, or nearly so. William Stewart (2008), the U.S. surgeon general, is quoted (Upshur, 2008) as saying that we had wiped out bacterial infection in the U.S. Nobel Prize winner Macfarlane Burnett with David White (1972, p. 263) speculated that, 'the

future of infectious disease... will be very dull.'

Of course, these predictions have been radically wrong and many infectious diseases are making a major comeback [e.g., Lewis, 1995; Upshur, 2008]. The emergence of multi-drug resistant bacteria is a major problem in hospitals and elsewhere. Many failed to take note of the adaptability of bacteria—a sort of bacterial Lucas critique—and a slowed pace of discovery of new antibiotics.

The fact that the experts made a bad prediction does not make this a tragedy; two additional factors do. Cautious observers, well aware of the potential problems from the start, proposed policies to avoid these problems.

In his Nobel lecture Fleming (1945, p. 93) noted that it 'is not difficult to make microbes resistant to penicillin in the laboratory by exposing them to concentrations not sufficient to kill them....' In the concluding passages of his lecture he detailed the dangers that have since transpired. In practice, actual medical policy looks more like an optimal strategy described by Fleming for *building* nasty bugs than avoiding them.

The second tragic factor is the revolution that did not take place. Around 1850, Ignaz Semmelweis demonstrated the best defense against bacterial transmission in hospitals: hand washing. While this finding was largely undisputed and the underpinnings became more solid over the next 150 years, the handwashing lesson went largely ignored. An editorial by William Jarvis in *the Lancet* (1994, p.1312) entitled 'Handwashing—The Semmelweis lesson forgotten?' accompanied the summarized yet another recent study on the subject: '[Health care workers] in intensive care units and in outpatient clinics, seldom wash their hands before patient contacts.'

Of course, the two tragic factors interact. The misuse of antibiotics along with the failure to wash hands in hospitals probably played a significant role in making hospitals the incubators of nasty bugs (e.g., Jarvis, 1994; Stone, 2000).

If you have visited a hospital lately, you know that the handwashing revolution is now finally underway–at least 50 years, if not 150 years, too late. Why? Among the most important barriers stated in studies is that doctors, in particular, are so

busy bringing the patients the benefits of modern science that they simply forget the mundane step of hand washing.

These medical policy mistakes share some features with the mistakes of macroeconomic policy in the 1970s. In both cases, I think the broadest lesson is that it is perfectly possible for policy to dissipate the benefits of scientific breakthroughs through excess optimism over what has actually been achieved and insufficient attention to mundane limitations.

1.2 Critiques of macro policy modelling in macro

The remainder of the paper takes it as given that careful attention to the details of what has been achieved in science is important if we are to maximize the practical policy benefits and minimize the risks from implementing scientific advances. This claim should not, I think, be too controversial; and we might simply move to a critique of the new DSGE models. The history of critique of policy models in macro has some recurring and very confusing themes, however, and it is important that we sort these out before continuing.

Policy modeling as it is practiced today arguably had its roots in Hick's creation of the IS/LM model. Later, Hicks severely critiqued the IS/LM model, arguing that it was wholly inadequate for the task at hand. Of course, the model remained a pedagogical workhorse, and arguably remained at the core of large-scale macroeconometric models. This outcome led some to study 'the strange persistence of the IS/LM Model.'[1]

In the 1960s, large scale macroeconometric models were constructed. Following the policy debacles of the 1970s, Lucas completely rejected the entire class of models, arguing,

> 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

[1] This is the title of a special issue of the History of Political Economy, 2005.

> 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)

Lucas advocated creation of a new class of rational expectations, equilibrium models that would be more suitable. Until an acceptable model in this class was realized, economic reasoning was of no value whatsoever:

> In situations of 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 Muth's sense. In cases of uncertainty, economic reasoning will be of no value. (1981, p.224)

These views had a strong impact in academics. As King, notes,

> 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. (1995, p.72)

Of course, when King argues that interest in large-scale macroeconometric models evaporated, he is referring to interest on the academic side. Despite the crushing critiques, these models had, from an academic perspective, a strange persistence, remaining a workhorse of policy analysis at least until recently.

Sims had severely critiqued the original models in his seminal work 'Macroeconomcs and reality,' arguing that the identifying assumptions were simply 'incredible.' He took up the 1990s updates:

> [O]ne might therefore have hoped that there would be clear progress as we moved from the early simultaneous equations models, to MPS and the RDX's, 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. (2002, p.23)

The Bank of England recently introduced its new BEQM model (Harrison, et al., 2005), a hybrid of DSGE elements and econometrics. Sims's (2006) verdict: worse

still.[2]

How can it be that models that are of no value whatsoever and can generate little or no academic interest, remained prominent in the policy process?. How can it be that Central Banks have presided over several decades of monotonic decline in model quality?

Of course, it is possible that either Lucas and Sims or the central bankers were simply misguided. My preferred explanation, though, is that the two sides are evaluating the models along different dimensions. If we are to carefully assess the merits of the new DSGE models, it is important that we understand both sets of criteria. The criteria underlying the Sims and Lucas critiques are well known; the alternative criteria are worth developing further.

1.3 A traditional view of macro modelling

As noted above, the history of crushing, but ineffectual, critique of macro policy modeling may have begun with Hicks's rejection of the IS/LM model. Solow provided a possible explanation in defending younger Hicks against older Hicks in the inaugural Hicks lecture in Oxford (1984):

> 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.... In that case what we want a piece of economic theory to do is precisely to train our intuition, to give us a handle on the facts in the inelegant American phrase. (1984,p.15)

This perspective starts with the presumption that our best characterization of the relevant issues is far from complete and that our models and policy ambitions should be tailored to this fact. Hayek (1989) makes this argument in general terms in his Nobel lecture, and Milton Friedman's case for the k-percent money growth rule was clearly based in this perspective.

[2] Sims argues that while models like FRB/US can at least be interpreted as consistent with a coherent probability model, BEQM cannot.

Because the optimality properties of the k-percent rule have been much studied, one might forget that Friedman's justification was based not on optimality, but on the fact that we could not possibly derive a rule that is optimal in any meaningful sense. Friedman argued,

> 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... (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)

Friedman recognized that, in the face of radically incomplete understanding, our policy recommendations do not rest on the kind of foundations we would prefer to have—we are doomed to be mistake prone. Of course, most macroeconomists now believe that the k-percent rule would not be tolerable.

The Friedman/Hayek/Solow-style perspective that macro policy analysis must be premised on the presumption of substantively incomplete understanding flourishes in central banks. For example, the Fed's Chairman Greenspan argued that,

> 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. (2003, pp.1–2)

I think this perspective is precisely what is required to understand the strange persistence of IS/LM and large-scale econometric models. In short, the models served the modest purpose of helping organize incomplete perceptions, tell plausible stories, and provide a handle on the facts.

Of course, models playing this role are quite far from the modeling ideal that remains the legitimate goal of research. Lucas is probably right that models suitable for the more modest role are of little or no value for the task of performing demonstrably optimal policy analysis. While the models evolved to better meet limited

ambitions, Sims is arguably correct that this evolution took the models farther in some relevant metric from ones that would meet the ideal.

If we are to understand the proper role of the new DSGE models in policymaking, I believe we must critique them from both perspectives. Before conducting my version of these critiques, it is worth exploring the incomplete-model, limited-ambitions perspective more fully.

1.4 Toward a theory of the second best in policy analysis

Lucas asserted that economic reasoning would be of 'no value' in policy analysis outside the context of a rational expectations-style equilibrium. The view that any other form of modelling is *ad hoc* and thereby worthless, or nearly so, remains, in my view, an undercurrent in the macro profession.[3]

Let us concede that the sort of microfounded, optimizing model advocated by Lucas is the ideal. But suppose we also posit a world in which the policymaker has the sort of incomplete understanding of the economy described above. For this case, I think we need a theory of the second best in policy analysis that is akin to that used in welfare analysis. A few lessons that would be part of that theory can be seen in a much simpler context.

Two nations have decided to settle their economic differences with a game of chess; the policymaker's problem is to choose how her nation will play this game. One proposal is to have the world chess champion—a local citizen—play for the country; the alternative is to play an optimal model-based strategy.

The chess champion describes his approach: I consider the board and various aspects about my opponent and come up with that seems likely to be a good move. The optimizer argues that this is *ad hoc*. How do you impose consistency? By what criteria can we judge whether any given move is in fact optimal, or more nearly so, than another?

The optimizer argues that he can prove that the champion's strategy is sub-

[3] Sims 2006 perceives a persistence of this view.

optimal and offers an optimal strategy as an alternative. The policymaker asks, 'You can prove the champion's chess play is suboptimal?' Well, not precisely, he responds, for chess is too complicated. Through incredible advances in algorithms and computation, we have, however, solved an approximation to chess called checkers (note: checkers was formally solved in 2007, see Schaeffer, et al, 2007). We have proved checkers is a tie, and, hence, proved that the champion, who sometimes loses, plays suboptimally. 'Does this proof provide a constructive strategy for play?' the policymaker asks. Well, the computational algorithm behind the proof is not practical for use in real time, so we have developed a feasible, and approximately optimal, strategy for a simplified version of checkers.

The chess champion argues that we have little reason to suppose that an approximately optimal strategy for simplified checkers might not be even close to optimal for the real-world problem. The optimizer concedes that the model has some limitations, but argues that at the very minimum, the world champion should consider the 'optimal policy play' as a baseline in contemplating moves.

Chess has not been formally solved and will not be solved soon: we have no satisfactory rational expectations model of optimal play in chess. Despite this fact, two facts are indisputable: Some individuals are very good at playing chess, and *ad hoc* models of chess now give policy advice that compares favorably with the best human play. To be clear, these *chess policy models* are outside the class of optimization-based, microfounded, rational expectations models.[4] Thus, even in areas such as playing chess for reward, in which rational expectations modelling is clearly applicable and presents no conceptual challenges, such modelling may be intractable.

Perhaps the key empirical lesson suggested by this example is that we have little reason to suppose that, in the face of intractable problems, the best practical model will be some microfounded model simplified to the point of tractability. It is manifestly false that *ad hoc* models outside the microfounded class are of no practical

[4] Loosely speaking, the programs are based on a set of *ad hoc* scoring algorithms that have been tuned for empirical success in particular contexts. See, e.g., Marsland and Schaeffer, 1990

value. For the economic problem of winning chess games, the best policy advice from *ad hoc* models currently performs exceedingly well and may soon dominate all alternatives. Such models are, however, of no direct value in the legitimate enterprise of deriving formally optimal strategies.

There are endlessly many reasons why making monetary policy is not directly analogous to a chess play. Perhaps the most telling is that the chess champion can perfect the skill through playing and studying thousands of games. While we do not fully understand this learning/optimizing process, one can imagine that through frequent repetition, humans approximate optimality better than any but the most recent of algorithms.

The next example is a bit closer to monetary policy in some ways. It is a targeting problem, the conditions are complex, and repetition, experimentation, and trial-and-error learning is limited.

The policymaker is about to finance a voyage of discovery sometime in the 15^{th} century. She is choosing between navigation by a method that has formally been proven optimal and captain who uses an *ad hoc* alternative. The captain describes his approach as a 'look at everything strategy:' the captain looks at weather, sea conditions, sea depth, what sea birds have been observed recently, and forms his best estimate of location. The captain notes that this approach has been used with some success in the past—he has succeeded in one long voyage. A few others have as well; but many practitioners have, admittedly, found themselves, to borrow Lucas's phrase, in way over their heads.

The optimizer proposes optimal targeting of the destination—proven to be optimal under the assumptions of constant ocean current and wind speed. Upon the objection that these assumptions are quite special, the optimizer notes that the assumptions are actually quite general: they may be viewed as a first-order approximation to arbitrary smooth problems, and a graduate student is generalizing the analysis to navigation in spaces of arbitrary dimension. The sea captain continues to object that the optimal approach is too simplistic. The obvious response: 'it

takes a model to beat a model.'

Like many bits of aphoristic wisdom, this statement is, in my view, either tautologous or wrong. It is true that in the important search for the best microfoundations-based model, it takes one to beat one. It is plainly false, however, that in the search for the best guide for practical policy, a microfoundations-based model can only be beaten by another. Indeed, in the chess example just given, the best *ad hoc* human play and the best *ad hoc* models would quite literally beat all microfoundations-based models.

1.5 Implications

From these extended preliminaries, I see a few tentative lessons.

First, as new science is brought into the policy process, we should be particularly careful to examine the practical merits of the advances. This is not to say that we should simply be skeptical of claims coming from science (although this might be a good idea). Rather, we should be very on what the science has established and particularly clear about which elements of the advances are of practical relevance.

Second, in the area of policy modelling, critiques from two perspectives may be of interest. From the first-best perspective, we can ask whether the advances bring us a complete formal model that produces optimal policy calculations of direct value in policy—or, at least are serious competitors to the *ad hoc* approaches that have been constituted best practice to date.

From the perspective of the second best, we ask different questions. In this perspective, our knowledge is incomplete and we do not yet have a model meeting the first-best standards. We ask of a model mainly that it be useful in structuring our thinking.

These two perspectives need not conflict, but as the history of policy critiques shows, what is arguably progress from one perspective may be regress from the other. We should pay particular attention to which sort of advances are embedded in the new models.

2 Critique of DSGE models from two perspectives

The DSGE literature has brought myriad advances in basic economics, techniques of dynamic optimization, computations, and Bayesian methods. Perhaps with a bit of hyperbole, let us accept that these models provide the magnitude of advance that antibiotics brought to medicine. To help avoid the mistakes of medicine, in this section, we critique DSGE models as currently implemented from both the first-best and second-best perspectives.

Because much of the progress in the DSGE literature appears focused on achieving the first-best ideal of a microfounded model, one might suppose that the new models have great practical merits in this regard. This is a fundamental misreading of what has been achieved to date.

2.1 A brief history of DSGE models

Following the failures of the 1970s, Lucas laid out a roadmap for a new class of models with microfoundations that would be less prone to such failure. In particular, the models would begin with explicit statement of objectives and the information sets for all agents and of the constraints they face. Then equilibrium behavior is derived as the result of explicit constrained optimization problems.

In 1981, Lucas laid out the roadmap,

> 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)

The modeling efforts began with Kydland and Prescott's (1982) Nobel Prize winning work; notable contributions include (Chrisitiano, et al., 2001,2005; Erceg, Henderson, Levin, 2000; Greenwood, Hercowitz, Huffman, 1988) It did not take

long, however, to recognize that the task would take considerably longer than five years. A number of new technical tools were needed, but the main roadblock was that it proved difficult to specify explicit individual decision problems in such a way that the aggregate dynamics matched the kind of persistent co-movement that we associate with the business cycle. In short, producer and consumer behavior tended to adjust too quickly to new information in the early models.

Modelers began to look for the sorts of constraints that would generate persistent dynamics. For obvious reasons, the general class of constraints that would do the trick are known as 'frictions,' and to a large extent, the development of DSGE models became a broad-ranging search to discover a set of frictions that, when layered onto the conventional core model, might pass Adelman-type tests of reproducing realistic dynamics. As of the turn of the century, we were arguably beginning to produce realistic dynamics.

In what was a major set of advances, Smets and Wouters (2003, 2007), building most specifically on work of Christiano, Eichenbaum and Evans, added a larger set of persistent exogenous shocks to the core model than had previously been typical, employed a large set of promising frictions,[5] specified a diffuse prior over the parameters, and then applied a Bayesian estimation scheme. The resulting posterior met various criteria of fit to 7 macro variables—criteria that had previously been impossible to attain. In particular, forecasts using the DSGE model compared favorably to certain well-respected benchmarks.

DSGE models that follow approximately this recipe are being formulated and coming into use at central banks around the world. Notably, a version of the Smets-Wouters model is used at the ECB, and a model that is similar in form, called Ramses (e.g, Adolfson, et al. 2006, 2007), is now used by the Swedish Riksbank.

Once one has a formal model, it is natural to perform optimal policy computations. This project was initiated in the 1970s, but largely died with the 1970s problems. The new DSGE models have a much more sophisticated treatment of

[5] Sticky wages and prices, sticky adjustment of capacity utilization, investment adjustment cost; habit formation in consumption.

expectations and other features, which make optimal policy computations more complicated analytically. There have been many important advances in the study of optimal monetary policy in DSGE models (e.g. Woodford, 1999, 2000, 2001, 2003). Until recently, there has been little work on the way optimal policy calculations might be used in day-to-day policymaking. Recently, Adolfson, et al. (2006) has filled this void, showing how to produce optimal policy projections that are the natural analog of the *ad hoc* model projections commonly used in policy discussions at central banks.

2.2 Brief description of the Ramses model

In the applications below, we use the Ramses model. To fix ideas, we describe key elements of the model here. This is a natural choice for a several reasons. It was developed at the Riksbank for policy analysis, it is well-documented in publicly available papers (e.g., (e.g, Adolfson, et al. 2006), and Adolfson et al. (2007) have recently shown how to use the model for practical optimal policy calculations. Such calculations may become part of the policy process at the Riksbank.

The model fits in the general framework described above: a core model with a large number of frictions and exogenous shocks; exogenously specified dynamic structure for the shocks. Ramses is an open economy model based on the standard small open economy assumption that the foreign sector is exogenous. There are 15 observable variables used in the estimation of the model—a one-period interest rate and a real exchange rate; 5 output quantity variables: output, consumption, investment, exports, imports; a real exchange rate, hours worked, real wages; three nominal price measures: two price inflation rates and an investment deflator; and three world variables: output, inflation, and an interest rate. .

There are more exogenous shocks than variables. These include a monetary policy shock, 5 shocks affecting either technology or the substitutability between types of investment goods, two markup shocks (exporters and domestic), a risk premium shock, and two household preference shocks. As for frictions, the model includes

sticky prices and sticky wages (with indexation), part of the wage bill of firms must be financed in advance, there is external habit formation in consumption, and investment adjustment costs. Monetary policy is given by a Taylor-type rule that includes inflation, the output gap, the real exchange rate, and the lagged interest rate.

The model is estimated using Bayesian techniques. The prior for each of about 50 parameters is independent of the others, centered on reasonable and fairly diffuse.

The critique that follows presumes a model with these braod features: multiple frictions and shocks with exogenously specified dynamic form estimated under a diffuse prior.

2.3 Critique from the first-best perspective

Let us take it as given that if policy analysis is to conform to the first-best ideal, the models must overcome the major critiques of the 1970s models: we need a model with microfoundations; a complete probability model for the phenomena at hand that reproduces business cycle features of the data and is based on credible identifying assumptions.[6]

Let us also concede that the models have at minimum attained good enough fit to the data, forecasting properties, and policy implications to warrant serious consideration for use in the policy process. Of course, the models of the 1970s also had good fit and forecasting properties and reasonable policy implications. Merely attaining this standard is no response to the devastating critiques from the first-best perspective. Nor is it sufficient to avoid the mistakes of the 1970s. This section critically evaluates how the new models stand up to three major critiques of the 1970s models.

[6] There are, of course, continuing debates over the importance of some of these critiques. On the Lucas critique, see Ericsson and Irons, 1995 and Sims, e.g., 2006.

2.3.1 The Lucas Critique

One set of advances embedded in these models is found in the family of issues surrounding microfoundations, which are essential if we are to avoid the Lucas critique. Up to now, we have followed the literature in using the term microfoundations rather loosely. At this point, it is important to distinguish two senses of microfoundations. A model has what I will call *weak form* microfoundations if decisions by agents are governed by explicit dynamic optimization problems: the modeler states the constraints, information sets, and objectives explicitly and derives optimal behavior.

A model has strong form microfoundations if, in addition, the formulation of the problem faced by agents is consistent with relevant microeconomic evidence on the nature of those problems, and fixed aspects of the constraints on behavior (parameters, etc.) are specified in terms of features that are reasonably viewed as immutable, or at least not subject to choice by the optimizers.

Whereas the research agenda began as a search for strong form microfoundations, the reliance on well-founded micro and arguably 'deep' parameters gave way, to some degree, to a search to discover what sort of *ad hoc* frictions might work. The publication of the work of Smets and Wouters (2003) may be a reasonable point to mark the end of the search for a model with weak form microfoundations.

From the standpoint of responding to the Lucas critique, however, attaining a model with weak form microfoundation is mainly a promising starting point.

Consider the microfoundations of sticky prices and wages. Of course, sticky prices and wages have always been at the center of the Keynesian story of business cycles. Providing a solid rationale for the stickiness is an important subject for Keynesians. The current DSGE models have no such rationale: they generally exogenously impose that firms can only change prices at certain exogenously chosen points in time—those points may be stochastic (Calvo pricing) or deterministic (Taylor contracts). From the standpoint of a fundamental rationale, this is, at best, a modest advance over Hicks's IS/LM model: instead of fixed prices, a firm's price is fixed until some exogenous process unrelated to economic fundamentals allows

the firm to change them.

Setting aside the heavy-handed form of the pricing friction, one might ask whether at least the parameter determining the exogenous frequency of price adjustment might reasonably be viewed as 'deep.' I have seen no serious argument, however, for the view that the frequency of price adjustment should be seen as a deep parameter'in the economy.[7] A quick check of recent events in Zimbabwe, though, reminds us that price setters are perfectly capable of changing the frequency with which they adjust prices. The microeconomic evidence from more modest inflations is, at best, mixed (e.g., Nakamura and Steinsson, forthcoming).

At least, one might argue, the exogenous average frequency of price adjustment is chosen to be consistent with the microeconomic evidence summarized, e.g., by Bils and Klenow (2004) and Nakamura and Steinsson (forthcoming). Even this is true in only a peculiar and limited sense. I think the best reading of the micro evidence calls into question the validity of assuming a single average rate of price adjustment. The microeconomic evidence overwhelmingly supports the view that different sorts of goods have different average frequencies of price adjustment.

As elsewhere in macro, it is only under very restrictive assumptions—amounting to a form of linearity—that the behavior of an aggregate of heterogeneous agents is best described by the behavior of an individual with average parameters.[8] At this point, we have a good start on the exploration of this heterogeneity (e.g., Carvahlo (2006) and Nakamura and Steinsson (2008)). At minimum the results support the view that the heterogeneity may matter for optimal policy.

We have barely begun to explore several elements of heterogeneity. For example, the microdata show dramatic secular and time-varying rates of relative inflation across broad categories of goods. In the U.S. CPI data, durable goods inflation has been negative for the past 20 years, while inflation on, say, medical care and textbooks has fluctuated closer to 10 percent. Heterogeneous inflation rates could

[7] Leeper (2005) also makes this argument.

[8] Aggregation is generally not even closed: the aggregate need not behave like an individual for any parameter value.

also have important implications for optimal policy.[9]

My goal is not to give an exhaustive analysis of this topic, only to emphasize that the assumption that firms' prices are exogenously fixed for extended periods does not constitute a microeconomic rationale for price stickiness. It is not specified in terms of a plausibly 'deep' parameter, and serious consideration of the micro evidence provides ample reason to question whether this assumption captures relevant features of the data. I will take up related arguments about the assumption of habit formation in consumption below, and we could perform a similar analysis of other frictions.

I am not arguing that the DSGE literature has gone astray. In the search for a model with strong-form microfoundations, achieving a plausible DSGE model with weak-form microfoundations is a major achievment, setting the stage for assault on the larger goal. From a practical policy perspective, however, the current implementation of these models is far from meeting the standard laid out by Lucas.

2.3.2 The Sims critique

Probably the second most prominent critique of 1970s models is Sims's argument that the causal structure of the models rested on 'incredible' identifying restrictions. Sims began his critique, as Lucas did, by conceding reasonable fit and forecasting properties of the models. Echoing the Cowles commission, Sims argued that, from among all the different structural models that would imply similarly good fit and forecasts, one must pick the one with the correct causal structure if one hopes to perform reliable policy analysis. Because these alternative structures fit the data similarly well, the choice must be made on *a priori* grounds. Sims argued that, in the large macroeconometric models, this choice was based on criteria that were simply 'incredible.'

[9] For example, in simple sticky price models, the central bank can eliminate the sticky price distortion by targeting sticky prices. In a world of optimal choice of frequency of adjustment and heterogeneous inflation rates, the stickiest price, all else equal, would simply be that on the good with inflation rate closest to zero. This contains no information about how to minimize the distortion on goods with nonzero inflation rates.

As for the new models, Sims 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)

This view is worth explicating further. Let us focus on the transmission mechanism of monetary policy, a key feature to get right in any policy analysis model. Identifying the dynamics of how changes in the policy interest rate spread to the economy, causing changes in aggregate income and inflation, is a long-standing and unsettled problem in macro—the profession has struggled mightily with this issue for decades. The new DSGE models add another voice to the chorus, resolving the issue by imposing zero mass on certain frictions and arbitrarily varying (but diffuse) amount of mass on others.[10]

It would be very difficult, perhaps incredible, to argue that the identifying restrictions implicit in this approach resolve the long-standing controversy. Sims (2008) makes a similar case.

2.3.3 The Hendry critique

In discussing the Lucas and Sims critiques, we have stipulated that the fit of the DSGE models is adequate. Sims and Lucas made the analogous concession in critiquing the macroeconometric models of the 1970s.

Hendry (e.g., 1980, 1985) argues that the 1970s models simply did not fit. The models showed a glaring inability to account for arguably important features of their estimation samples. Pagan (2002) argues that this helps rationalize why the central banks refined the models rather than abandoning them after the breakdowns in the 1970s.

How do the DSGE models fare under this critique? There are many aspects of fit of DSGE models we could critique. I will focus on one. The models are fit to a

[10] I say *arbitrarily varying* to emphasize that the diffuse prior is not in any meaningful sense flat; the prior discriminates among different formulations, but in a way that is unmotivated.

very small set of variables (say, less than 20); the older macroeconometric models were fit to a much larger set of data. The DSGE models have implications for a much larger set of variables than are used in estimation, however. Perhaps most prominently, the models have implications for the entire term structure of interest rates, yet only short-term interest rates are used in the empirical analysis.

The expectations theory of the term structure holds (or almost holds) in the models and this theory is known to be grossly inconsistent with the data—especially the U.S. data. To put it most contentiously, we have discovered one way to 'fit' the dynamics of the quantity of investment (which is included in the analysis): move the long-term interest rate in arbitrary, counterfactual ways. We may echo Hendry in stating that these models show a glaring inability to account for arguably important aspects of the data. In this regard, the models are, almost by design, unambiguously worse than the large macroeconometric models of the past.

2.3.4 Bottom line from a first-best perspective

The first-best modeling ideal is clear and stringent. We now have in place a set of tools and a tractable set of models that might allow us to attain the goal.

It would be foolish, however, to conclude that, in demonstrating the generic viability of DSGE models, the science has produced a model that actually has the desired properties. What has been provided is a model with largely *ad hoc* microfoundations, unstudied and arbitrary identifying assumptions. The models have grossly counterfactual implications for key variables that have been left out of the analysis, such as long-term interest rates. From the standpoint of actually overcoming the 1970s critiques, it is difficult to say whether the current models represent progress or regress. What the models clearly represent, in my view, is a jump to a new and very promising starting point from which to assault the goal.

My main conclusion will be that the practical merits of DSGE models resides mainly in the promise of further development. Before turning to that development, it is important to emphasize that even the existing DSGE models have an important

role to play in the policy process.

2.4 Existing DSGE models from the perspective of the second best

As outlined by Solow, in the face of incomplete understanding and intractable problems, models can play a role in organizing our thinking, placing structure on our interpretation of the data, and training our professional intuition. I see at least three advantages of the current DSGE models from this perspective.

First, even weak-form microfoundations have advantages. In a model with microfoundations, weak or strong, economic reasons for any outcome can be traced to root fundamental causes. Thus, whether or not the microfoundations are *right*, the models allow us to sharpen economic intuitions about how basic economic mechanisms operate. Over the nearly 20 years I spent at the Fed, I observed a considerable increase in the sharpness with which dynamic economics was discussed, and I think this would have been hard to attain had many participants in the process not sharpened their skills using DSGE models.

Second, for some questions, existing DSGE models may be the best models we have. Policymakers are sometimes presented with questions that so thoroughly involve expectational and general equilibrium effects that traditional models and ways of thinking are of little use. I believe the first major presentation of DSGE results to the Board of the Fed came in analyzing the differential properties of expected and unexpected, asymmetric productivity shocks in the open economy.[11] Of course, the Fed was concerned that this was a relevant case in the late 1990s, but sorting out the complicated mix of effects on consumption, labor, output, external borrowing, and the exchange rate would have been very difficult without a DSGE model. Despite the arguments from the first-best perspective given above, the models provided a very useful role in the second-best goal of 'getting a handle on the facts.'

Eggertson and Woodford's (2004) analysis of the role of expectations in optimal policy behavior when nominal interest rates are near zero provides second example

[11] This work was performed by Erceg, Guerrieri, and Gust using the Fed's SIGMA model and is documented in Erceg, et al. 2003.

of a problem that would be difficult to study without the internally consistent treatment of expectations in DSGE models. For similar reasons, DSGE models make it possible to get a basic handle on the practical merits of discretionary policy versus, say, commitment under the timeless perspective (e.g., Adolfson, et al. 2006). This issue is beginning to be discussed in public speeches by monetary policy makers (Bergo, 2007).

Third, the initial generation of economists trained in the DSGE tradition are now at least at mid-career. Essentially all the new economists entering central banking have been trained in this tradition. Whether for good or for ill (I think it is for good), DSGE macro is rapidly becoming the way macroeconomists structure their thinking. For many questions, even the current generation of DSGE models may facilitate a more productive policy discussion than is possible using one of the traditional big macro models, which, in my experience, simply perplex many of those trained after, say, 1985 or 1990.

These advantages of the current implementations of DSGE models are admittedly modest from a first-best perspective. They are all in the family of training our intuition, illuminating broad principles, and providing a language for discussion. I think these benefits are substantial from the second-best perspective. There are, in my view, much bigger gains attainable through further development of the models.

3 Exploiting the policymaking potential of DSGE models

The broad, largely pedagogical role just described does not include taking very literally the model's optimal policy implications. Indeed, it does not include any of the roles in day-to-day assessment of the economy and policy that have been served by the old macroeconometric models. In this section, we explore the degree to which the new models should supplant the old in these roles.

In the second-best perspective, we would like models to help structure our cur-

rent analysis, but if they are to do so effectively, it is important that we understand the degree to which the models broadly reflect our understanding of the economy—limited as they may be. In this regard, I think it is useful to contrast the traditional model formulation with that described for DSGE models. Unfortunately, the traditional development approach is *ad hoc*, opaque, and difficult to characterize. I observed, but did not participate directly in, the development of the Fed's new models (FRB/US, FRB/Global) introduced in 1995. The process involved heavy involvement and of economists and policymakers at every level level of the organization. I think it is accurate to say that the model development phase did not stop until the relevant group of decisionmakers agreed that the model met the second-best-style goals sufficiently well.[12]

In contrast, the formulation and estimation of the current generation of DSGE models looks more like an attempt to purge, or at least to minimize, the effects of prior judgment. The specification of the model involves a great many largely *ad hoc* decisions: what margins to add shocks to? What decisions to put frictions on, what form should these take? The estimation is then based on a diffuse prior over the parameters of a large, imperfectly understood model with a large, and weakly justified, set of frictions and driven by a large, and weakly justified set exogenous shocks.

Before the new DSGE models should supplant more traditional models, policymakers should probably consider questions like the following: Is the model broadly consistent with your view of the business cycle? Of the transmission mechanism of monetary policy? If the model produces a result that conflicts with your intuitions, would you be more likely to question the model or your intuition? These are all essentially ways of asking whether the formal posterior computed for the model is a reasonable characterization of the poliymakers' actual (partial and imperfect) posterior at the end of the analysis.

I am not criticizing the standard DSGE modelling approach as a contribution to

[12] This is consistent with the descriptions of these issues in, Reifschneider et al. 2005, and Stockton 2002.

basic research. Indeed, in this role the approach may be preferred. By purging the analysis of any one set of expert beliefs, the research can demonstrate the generic viability of the class of models. However, imposition of beliefs that might have been seen as cheating or bias in establishing the scientific result, is precisely the expert judgement we want to bring to bear in policymaking.

Of course, one glory of the Bayesian approach is that it allows for a coherent and systematic melding of expert judgement and data. Bringing formal Bayesian tools to bear in incorporating judgement presents a great opportunity to put these models on a much stronger footing than that of predecessors.

In this section, I argue for and illustrate a particular blend of model checking. The formal Bayesian tools, for the most part, are standard. For example, John Geweke (2005) presents a wide array of general tools for evaluating model adequacy. The particular emphasis is based on Geweke's (2007) recent suggestions about inference in incomplete models. The results are worked out more fully in Gupta (2008) and Faust and Gupta (2008).

At the outset, I emphasize that the builders of the DSGE models have critiqued these models themselves (e.g,, DSSW, 2007; Del Negro and Schorfheide, 2007,2008; Schorfheide, Sill, and Kryshki, 2008), stating boths strengths and weaknesses and developing many useful diagnostic tools that are complementary to what I propose.

3.1 Inference about features of interest

In the second best perspective, we have certain beliefs and we would like a model to help structure our thinking, helping us make connections we might not otherwise have made, etc. If the model surprises us, we might alter our understanding, but we might decide that the surprise is an artifact of some unsavory feature of the model that we had not noticed or had not yet found a way to fix.

In my view, as a new model is brought into the policy process, it would be best to have a solid sense of which aspects of the model are to be taken relatively seriously and which are the unsavory bits. A key step, in this view, is to follow the Adelmans

in listing some beliefs explicitly.

For example, many of us (in my experience) have fairly strong priors about some basic business cycle properties of data. Since Granger's (1966) classic work identifying the 'typical spectral shape' of a macro variable, it has been standard to view the business cycle in frequency-domain terms. One common approach is to partition the spectrum into low frequency, business cycle frequency, and high frequency variation. In any model, we can compute the share of the variance of each variable attributable to fluctuations in each category.

In this way, we can evaluate whether the model distributes variance across the spectrum in a way that roughly matches the data. Of course, this has been a major problem for DSGE models. As noted above, there is nothing technically special about at frequency domain measures: the point here is to build a list of important features for the model to meet.

Consider a more structural example. Historically, central bankers and academics have been concerned about the long, and potentially variable, lags in the response of the economy to monetary policy shocks. In practical discussions, one regularly hears statements from central bankers that policy does not have it main effects for up to a year. Of course, a linearized model will not produce variable lags (except as sampling fluctuation), but we can assess whether the lags are long. For example, we might compare the impact effect of a policy shock and the magnitude of the shock at other horizons relative to the impact effect.

Finally, we may have core beliefs about the deficiencies of current DSGE models. For example, a key problem in DSGE models has been that agents in the model seem to be too willing to substitute at the margins. This is what motivates habit formation, adjustment costs, and persistent shocks to marginal conditions. Thus, one might want to focus on, say, the correlation between, say, interest rates and consumption or investment.

3.2 Measures

Formally, all of the population features I will discuss will be a function of the DSGE model parameter. That is, if θ is the vector of DSGE model parameters, then the features can be written as nonstochastic functions of θ, say $\gamma = f(\theta)$.

There are several natural items to investigate regarding these γs. Since the prior reported in standard work is diffuse and largely arbitrary, it may be useful to investigate the implied marginal prior for the γs. This could reveal whether the formal prior as specified conforms at all to our actual priors about the business cycle. Further, we might also discover that the formal prior, which is diffuse in a particular sense, might be dogmatic about particular features of interest.

Of course, neither θ nor γ is directly observable, and it can be important to anchor the analysis in a discussion of observables. There will generally be one or more natural sample analogs of the features. Let me call any such measure $hat\gamma(Y_T)$ when the available sample is Y which includes T observations. These functions of the data correspond to Box's (1980) checking functions. The prior over θ, in conjunction with the model, will also imply a density, called a prior predictive density, for $\hat{\gamma}$. This density reveals range of results for $\hat{\gamma}$ we should expect to see in any given sample, given the prior and model.

Obviously, a primary factor affecting the prior predictive density is sample size. In small samples that is not very informative about γ, we expect the prior predictive density for $\hat{\gamma}_T$ to be quite diffuse even if the prior for γ is not.

Finally, we might also consider the posterior predictive density for sample features, $\hat{\gamma}$. The posterior from one exercise is, of course, the prior for the next, so the computation and interpretation of the posterior predictive density mirrors that of the prior predictive.

Computation of all these measures is fairly trivial (see, e.g., Geweke, 2005) when the model has been estimated using a Monte Carlo-based method. All the inputs for the calculations are natural outputs of the estimation process. For details on the specific measures in this paper and a sketch of the computational issues, see the

Appendix.

3.3 Example

First, consider whether the variance of output growth and inflation in the model is distributed across, low, business cycle, and high frequency cycles in a way that is consistent with the data. Figure 1 shows that the model is a great success in this regard for both variables in all three frequency ranges. For inflation, the prior, posterior, and data all correspond: most variance in inflation in the sample is at low frequency and the model can accommodate that. For output growth, the data lead to a substantial shift from the prior to the posterior. In a full analysis, we would continue this examination with the other variables and considering the coherence of the variables.

For illustration purposes, turn to the second topic: long and variable lags in the transmission mechanism. Figure 2 gives the prior and posterior densities for the impulse response to a policy shock at 2 horizons—on impact and after 4 periods. The shock raises the annualized interest rate 25 basis points on impact. The prior and posterior for the impact effect on output nearly correspond and are centered on a one-for-one effect: 25 basis points on the interest rate *on impact* gets you 25 about 25 basis points on the annualized quarterly growth rate. After a year, the negative effect on the growth rate has grown, about doubling at the mean. For inflation, the impact effect is similar, the effect does not grow as much, however. I suspect that the nearly one-for-one impact effect with only modest expansion over the next year does not capture the conventional wisdom of many policymakers. Some structural VAR work (e.g., Faust, 1998) suggests that small changes in the identified impact effect of policy shocks can be associated with large changes in other aspects of the model.

Finally, turn to the response of the economy at the intertemporal margin (Fig. 3). In the prior, consumption growth is fairly strongly negatively correlated with interest rates. The posterior is shifted toward lower negative correlation, but the

correlation remains much stronger than in the data. Both the prior and posterior for the correlation between consumption and investment are centered on zero: higher interest rates lower consumption, but presumably may raise or lower investment. The data, however, show a strong positive correlation. Some of the dissonance here may be more apparent than real. The prior predictive density for the consumption, investment correlation (lower right) is widely dispersed, indicating that under the prior, the available sample size is sufficiently small that a wide range of sample values are consistent with the model/prior. The consumption-interest rate correlation remains problematic from this perspective.

Once again, the particulars I have chosen to discuss are not important. My intent is simply to illustrate the selection and examination of some broad features of interest.

3.4 More formal uses and incomplete models

The methods discussed so far take no explicit account of the second-best presumption that our best models may be incomplete. Geweke (2007) notes explicit Bayesian inference generally presumes existence of a complete probability model for the phenomena in question and that performing inference under an explicit assumption of incomplete understanding has been a long-standing topic in the Bayesian framework. Geweke proposes a way to bring in the notion of incompleteness that fits nicely with the second-best perspective discussed here.

Our idea of an incomplete model is essentially that analysts may have beliefs about certain macroeconomic phenomena or features, without having a complete coherent model of those phenomena. One natural interpretation of this state of affairs is that the analyst has a prior directly over the joint distribution of certain $\hat{\gamma}$s.[13]

In a standard Bayesian comparison of two complete models, A and B, we might compute the Bayes factor in favor of A: $p(Y_T|A)/p(Y_T|B)$, where the p terms are

[13] As emphasized by Geweke, this prior need not even be coherent in the sense that the prior is consistent with any complete model.

the marginal likelihood of the sample Y_T, given the model.

In the case where model B is incomplete, Geweke argues for considering the Bayes factor, $p(\hat{\gamma}(Y_T)|A)/p(\hat{\gamma}(Y_T)|B)$ that is, the marginal likelihood of the relevant observables. Loosely speaking this is the natural analog of the complete model comparison confined to the behavior of the features of interest. If the Bayes factor in favor of the formal model is favorable, we might conclude the model passes one test of adequacy as a representation of our beliefs about the feature. If not, we might go back to the drawing board.

Overall, merely explicitly specifying features of interest and examining the model from the standpoint of those features, as done in the example, is a step in the direction of brining expert judgement into the process in a systematic way. Going further and attempting explicitly codify the beliefs about the features in an incomplete model offers important opportunities for further gains.

4 Conclusion

For the first time since the early 1970s, the times again are exciting for those of us interested in the art and science of practical policy analysis. The new class of DSGE models and rapidly expanding set of tools provide every prospect that policy analysis can be put on a much sounder basis than ever before.

The advances in DSGE modelling to date, however, have come at the more basic end of the research spectrum. Excitement over these incredibly important advances should not blur our vision over what of practical relevance has actually been achieved. While we have generated a family of models that broadly matches business cycle features, existing implementations of these models are little better, and may be worse, than the models of the 1970s from the standpoint of the major critiques of those older models. We have a promising jumping off point for assaulting those problems, but considerable work remains to be done.

I have argued for a perspective of the second best in policy analysis. In this perspective, the critiques of the 1970s models set too high a standard—that is, a

standard we may not soon meet. In the meantime, policy analysts need models that help structure their thinking and data analysis and train our professional intuition. This perspective has a long tradition in macro, and, in my view, this perspective explains the major role 1970s-style macroeconometric models have continued to serve at central banks for the 25 years after the major mistakes.

DSGE models are rapidly becoming an important tool in this modest role. The biggest opportunity for advance, in my view, is in the area of checking the senses in which these models are and are not consistent with conventional wisdom about business cycles and the role of policy. Greater clarity on these issues will greatly enhance the important roles these models are already serving.

Appendix

This appendix describes the computation of the numbers reported.

For any model parameter, θ, we need to compute $\gamma(\theta)$ and $\hat{\gamma}(Y_T(\theta))$, the latter being the sample statistic in a sample of size T when the data are generated according to θ.

We want the statistics to have well-defined definitions outside the context of the model, so the γs computed are (a numerical approximation to) the population statistics for a pseudo-true VAR approximation to the model with parameter θ. For any θ, we solve for the dynamics of the observables in the model using code provided by the authors of the model. Then we generate one draw of 10,000 observations for all variables, estimate a VAR(4) and then compute and save the γs implied by properties of that VAR. Our sample moments are also computed as the properties of an analogous VAR, but estimated using the actual sample data. Thus, for $\hat{\gamma}(Y_T(\theta))$, we generate the relevant sample and compute the VAR and then the implied γs.

For the impulse responses, there is no sample analog. The impulse responses are those implied directly by the model parameter and computed using code provided

by the original authors.

For the frequency domain statistics, we partition the spectrum at $2\pi/32$ and $2\pi/4$.

For the prior distributions of γs, we take a large number of draws from the prior distribution and compute and save the γs for each. The resulting histograms are what is reported. For the posteriors, since the models were originally estimate using Monte Carlo methods, we can follow the original estimation scheme to get a set of posterior draws for the model parameters. Once again, we compute the γs for each, and report a histogram of the results.

References

Adolfson, M., S. Laséen, J. Lindé and M. Villani, Evaluating an Estimated New Keynesian Small Open Economy Model, Journal of Economic Dynamics and Control, 2007,forthcoming.

Adolfson, M., S. Laséen, J. Lindé and Lars E.O. Svensson, Optimal Monetary Policy in an Operational Medium-Sized DSGE Model, manuscript, Sveriges-Riksbank 2006.

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.

Bergo, Jarle, Interest rate projections in theory and practice, Speech by Deputy Governor Jarle Bergo at the Foreign Exchange Seminar of the Association of Norwegian Economists at Sanderstlen, January 26, 2007. http://www.norges-bank.no/Templates/Article___64910.aspx

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

Box, G.E.P. Sampling and Bayes Inference in Scientific Modelling and Robustness (withdiscussion), Journal of the Royal Statistical Society, Series B, 1980, 143, 383–430.

Burnet M., White D., Natural History of Infectious Disease, 4th edn Cambridge: CUP, 1972, p 263).

Carvalho, Carlo, 2006, Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks, Berekely Journals in Macro, Frontiers, 2:1.

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, Martin Eichenbaum, and Charlie Evans (2001). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. Federal Reserve Bank of Cleveland Working Paper.

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.

Del Negro, Marco, and Frank Schorfheide, 2008, Forming Priors for DSGE Models (And How it Affects the Assessment of Nominal Rigidities),

Del Negro Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters, 2007 On the Fit and Forecasting Performance of New Keynesian Models Journal of Business and Economic Statistics , with discussions and rejoinder, 25(2), 123-162.

Del Negro, Marco, and Frank Schorfheide, 2007, Monetary Policy with Potentially Misspecified Models NBER Working Paper 13099

Eggertsson, Gauti, and Michael Woodford, Optimal Monetary and Fiscal Policy in a Liquidity Trap, NBER International Seminar on Macroeconomics, 2004.

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

Erceg, Chris, Dale Henderson, and Andrew Levin (2000). Optimal Monetary Policy with Journalof MonetaryEconomics, Staggered Wage and Price Contracts. 46,pp.281–313.

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 Abhishek Gupta, 2008, Evaluating DSGE model adequacy: an incomplete models approach, in progress.

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

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.

Geweke, John, Bayesian Model Comparison and Validation, manuscript, University of Iowa, 2007.

Geweke, John, Contemporary Bayesian Econometrics and Statistics, New York: Wiley, 2005

Granger, Clive W.J. (1966), The Typical Spectral Shape of an Economic Variable, Econometrica, Volume 34, Number 1 (January), 150–161.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory Huffman (1988). Investment, Capacity American Economic Review, Utilization and the Real Business Cycle. 78(3), pp. 402–417.

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.

Gupta, Abhishek, A forecasting metric for evaluating DSGE models, in progress.

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.

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

Kydland, Finn and Edward Prescott (1982). Time to Build and Aggregate Fluctuations. Econometrica, November, 50(6), pp. 1345–1370.

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.

Lewis Ricki, 1995, The Rise of Antibiotic resistant infections, FDA Consumer Magazine, Sept. http://www.fda.gov/Fdac/features/795_antibio.html

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

Nakamua, Emi, and Jon Steinsson, Five Facts About Prices: A Reevaluation of Menu Cost Models, Quarterly Journal of Economics, forthcoming.

Nakamua, Emi, and Jon Steinsson, 2008, Monetary Non-Neutrality in a Multi-Sector Menu Cost Model,manuscript.

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.

Reifschneider, D.L., D.J. Stockton and D.W. Wilcox (1997): Econometric Models and the Monetary Policy Process, *Carnegie Rochester Series on Public Policy*, 47, pp.1-37.

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, Neil Burch, Yngvi Bjornsson, Akihiro Kishimoto, Martin Muller, Robert Lake, Paul Lu, Steve Sutphen, 2007, Checkers Is Solved, Science, Sept., 317:5844, 1518– 1522.

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

Schorfheide, Frank, Keith Sill, and Maxym Kryshko, 2008, DSGE Model-Based Forecasting of Non-Modelled Variables

Sim, Christopher, 2006, Improving Policy Models, manuscript, Princeton.

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.

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

Stone, S.P., Hand Hygiene—the case for evidence-based education, Journal of the Royal Society of Medicine, 94, June 2001, 278–281.

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.

Upshur, Ross, 2008, Ethics and Infectious Disease, Bulletin of the World Health Organization (BLT), 86:8, August 2008, 577-656

Woodford, Michael, Optimal Monetary Policy Inertia, August 1999.

Woodford, Michael, Optimal Interest-Rate Smoothing, Review of Economic Studies 70: 861-886 (2003).

Woodford, Michael, Pitfalls of Forward-Looking Monetary Policy, American Economic Review 90(2): 100-104 (2000).

Woodford, Michael, The Taylor Rule and Optimal Monetary Policy, American Economic Review 91(2): 232-237 (2001).

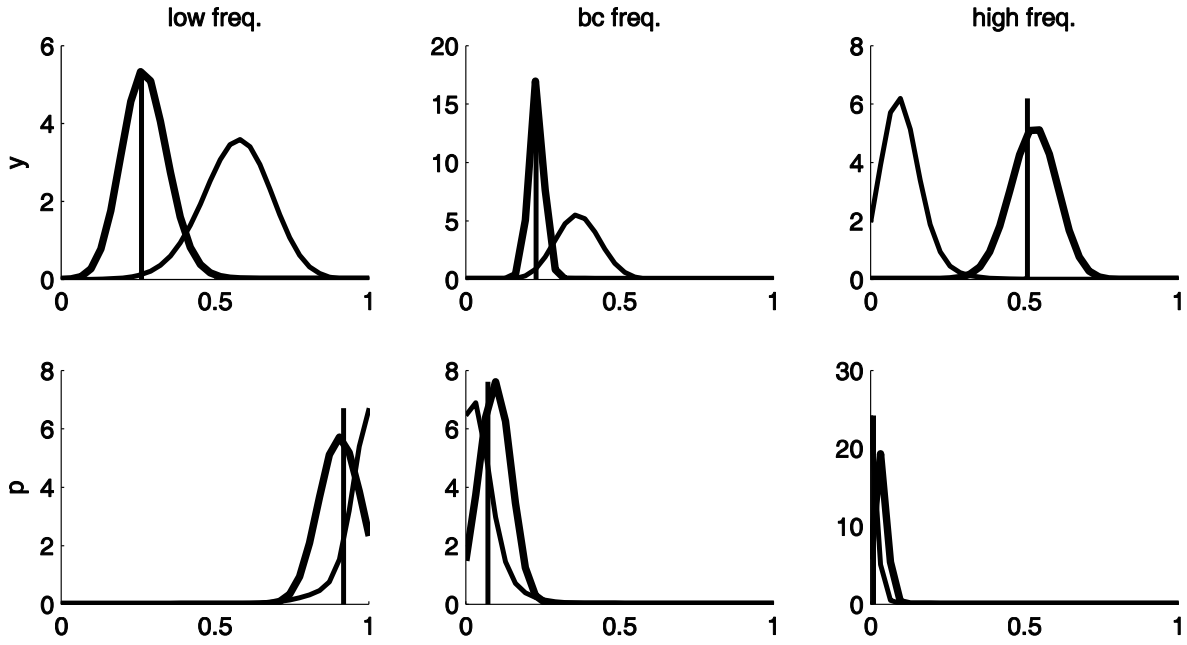

Figure 1: Prior, posterior, and sample values for share of variance of output growth (y) and inflation (p) at low, business cycle and high frequencies. Note: prior is thin; posterior is thick, sample is vertical.

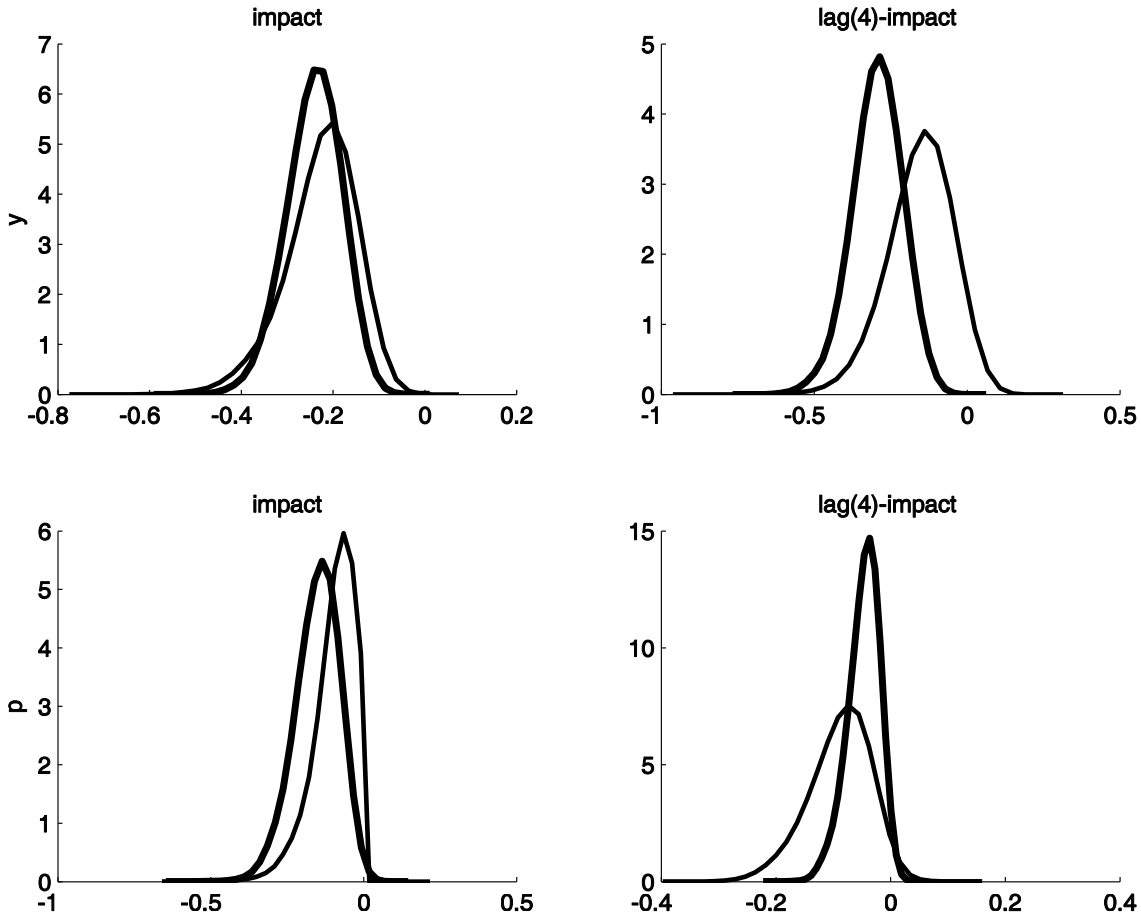

Figure 2: Prior and posterior for the response of output growth (y) and inflation (p) to monetary policy shock on impact and after 4 quarters. The shock raises the interest rate 25 basis points on impact. The responses are in annualized percentage points. The left column gives the impact effect; the right column gives the effect after 4 periods minus the impact effect. Prior is thin; posterior thick.

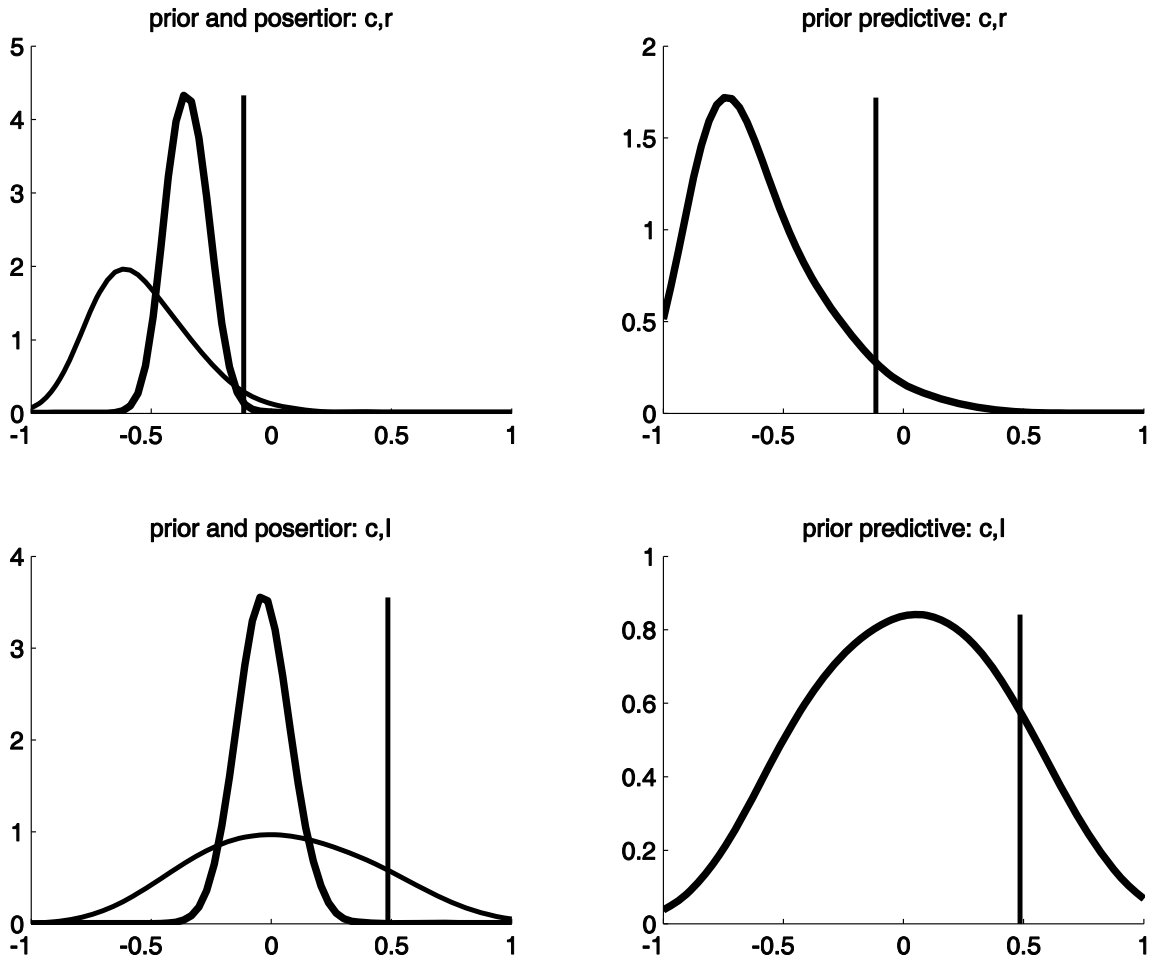

Figure 3: Prior, posterior, and prior predictive densities and sample value for the unconditional correlation between consumption growth and interest rates (c,r), and consumption and investment growth (c,I). On left , prior is thin, posterior thick, sample vertical; on right prior predictive is given, sample vertical.