The new macro models: washing our hands and watching for icebergs

Jon Faust[1, 2]
Professor, Econ. Dept., Johns Hopkins University, Baltimore, USA

The 1960s were an exciting time – at least for macroeconomic modelers. An impressive new kind of macroeconometric model was entering central banking, and cutting-edge central banks were beginning to analyze policy as a problem of optimal control. The December 1965 edition of *Time*, the popular U.S. news magazine, has Keynes on the cover, quotes the experts of the day extensively, and is almost giddy in tone regarding 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 practice of central bankers, the new econometric models of the 1960s were a long-term 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 scientific advance.

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

[1] The views in this article have evolved over many years and have greatly benefited from myriad discussions, especially those with John Geweke, Dale Henderson, Eric Leeper, John Rogers, Chris Sims, Lars Svensson, and Jonathan Wright.

[2] Louis J. Maccini Professor of Economics and director of the Center for Financial Economics, Johns Hopkins University.

I suspect that boredom is not currently the greatest concern of central bankers anywhere. When organizers of the Riksbank Conference on Refining Monetary Policy asked me to write a paper about the proper role of model-based, optimal policy calculations in real-world policymaking, the topic seemed to be at the forefront of technical issues facing the most advanced central banks.[3] This issue has at least momentarily faded in importance – models by their very nature have a limited domain of applicability, and most of us would agree that the current versions of the new macro models are not built to analyze a complete breakdown in credit markets. The role of model-based optimal policy calculations remains an important one, however, and the current turmoil presents a sobering yet informative backdrop against which to discuss the issue.

I am optimistic about the role the new macro models can play in the policy process once the crisis subsides. The point of this paper, however, is to discuss how we can minimize the risk of repeating the startup mistakes that were associated with bringing online the macro models of the 1960s. In this paper, I invite policymakers, central bank staff, and other concerned parties to consider two claims:

1. The adoption of new technologies, models, and ways of thinking is often accompanied by catastrophic and avoidable mistakes.
2. Answering some hard-nosed, common sense questions about the new macro models may help us maximize the benefits and minimize the risk of catastrophe.

To put the point a bit more aggressively: It would be very foolish to forgo the immense benefits that can come from rapid adoption of the new macro models; it would be similarly foolish to ignore the lessons of history regarding catastrophic mistakes that often accompany such advances.

The issues are too large and complex to be fully developed and documented in this brief article; this article is mainly meant to entice the reader to consider these points and to provoke further discussion and study of their merit. I first give some cautionary tales of technical advance, and draw some tentative lessons. I then attempt to clearly describe the sort of hard-nosed questions we should be asking of the new models. While I offer of few of my ideas on the answers to those questions, my answers are not the point. My hope is that concerned parties will ask and then answer questions like these for themselves.

Finally, let me note that this project was initiated on behalf of organizers of a Riksbank conference, despite the fact that that I have been

[3] The conference was held September 5 and 6, 2008.

critical of some aspects of inflation targeting at the Riksbank and elsewhere. The particular macro model I use to illustrate some points below is a version of the Riksbank's Ramses model. I could not have completed the work without an immense amount of help from the developers; these economists – Jesper Linde and others – went out of their way to help me, knowing that the point of my work was to invite policymakers to ask hard questions about the value of the model. This all is testament to the commitment to transparency and open, honest discussion of difficult issues that, in my view, is one unambiguously positive aspect of the inflation targeting framework. The Riksbank, in my experience, is unsurpassed in its commitment to this hallmark of modern central banking.

1. Advance and catastrophe

History suggests that bringing new technologies into expert practice is often accompanied by catastrophic error. Of course, some mistakes might be an inevitable part of applying new ideas. People make more mistakes when they are new to an idea than they do after considerable experience. What I will discuss is a different kind of mistake that is not inevitable. In particular, we often see the following pattern: a new idea is adopted and experiences some initial success; inflated optimism arises among experts regarding what has been achieved; traditional cautions are neglected; catastrophe follows; after a period of recovery, the new idea settles into its more modest but rightful productive place.

I am not new in making these observations. The ancient Greeks wrote of this elegantly under the heading of hubris. Jumping forward a few centuries, Fenn and Raskino (2008) state a 5-phase 'hype cycle' for how society, in general, reacts to new technology: 1. Technology Trigger, 2. Peak of Inflated Expectations, 3. Trough of Disillusionment, 4. Slope of Enlightenment, 5. Plateau of Productivity. While the `hype cycle' is meant to characterize a media-driven societal dynamic, the elements are very close to what I argue regularly accompanies the transfer of scientific advances into practice by expert practitioners. Perhaps the point is that experts are subject to some the same tendencies as other mortals.

The simplest example of the dynamic I am describing is that surrounding the Titanic – unquestioned advances in ship building, inflated optimism about the magnitude of the advance, neglect of traditional cautions, catastrophe, and finally the technological advances settling in as part of a general improvement in ship building. As noted in the introduction, I see this same dynamic in the adoption of new models in the 1960s, but before returning to that case, consider a case from medicine.

1.1 Antibiotics and hand washing

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 striking decreases in mortality and morbidity from these diseases (e.g., Lewis, 1995).

By the 1970s, some authorities were declaring the problem of infection to be solved, or nearly so. William Stewart, the U.S. surgeon general, is quoted (Upshur, 2008) as saying that we would wipe 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. 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.

Two additional factors highlight the ways in which this is a case of a sort of expert hype cycle. First, cautious observers were well aware of potential problems with antibiotics. 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 warned of problems that might come from antibiotic misuse in practice. His hypothetical discussion reads like an astute prediction of the path medicine subsequently took.

The second tragic factor involves a 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 ever more solid over the next 150 years, the hand washing lesson went substantially ignored. An editorial by William Jarvis in *The Lancet* (1994, p.1312) entitled 'Handwashing – The Semmelweis lesson forgotten?' summarized one recent study on the subject: '[Health care workers] in intensive care units and in outpatient clinics, seldom wash their hands before patient contacts.' Why? Studies state that one of the most important barriers is that doctors are so busy bringing patients the benefits of modern science that they simply do not take the mundane step of hand washing.

Of course, the misuse of antibiotics and the failure to wash hands in hospitals interact: the pair may have played a significant role in making hospitals the incubators of nasty bugs (e.g., Jarvis, 1994; Stone, 2000).

As you probably have noticed if you have been in a hospital recently, the hand washing revolution in hospitals is now well underway, arguably, 150 years late.

1.2 MACRO MODELING IN THE 1960S AND RISK MODELING IN THE 1990S

The adoption of new macro models arguably demonstrates a similar dynamic. Unquestioned advances in modeling were associated with modest successes in the 1960s, and were part of excessive optimism on the part of many experts over what had been achieved. The December 1965 edition of *Time* magazine quoted in the introduction provides a clear view of the tenor of certain experts at the time; Lucas (1981) broadly documents and pillories the hubris of the times. In my view, this optimism was accompanied by the abandonment at many central banks of traditional cautions about inflation and debasing the currency. We all know the catastrophe that followed.

It is too early to fully understand the role of modern risk modeling in the current financial crisis, but public information about the collapse of the insurance giant AIG suggests that excessive confidence in risk models for predicting losses on credit default swaps played an important role. Former Chairman Greenspan of the Fed (2008) concludes:

> The whole intellectual edifice, however, collapsed in the summer of last year because the data inputted into the risk management models generally covered only the past two decades, a period of euphoria. Had instead the models been fitted more appropriately…

we might not be in the current mess. The thought that small, unrepresentative samples may lead to unreliable inference is not, to use a Wall Street term, rocket science: this is a major point in any good undergraduate course in applied econometrics. Advising modelers to carefully attend to sample adequacy is the econometric equivalent of advising doctors to regularly wash their hands.

1.3 TENTATIVE CONCLUSIONS

These examples follow a pattern: Excess optimism – Titanic unsinkable, infection defeated, business cycle tamed, swaps will never default – paired with what looks *ex post*, at least, like failure to heed common wisdom--watch out for ice bergs, wash your hands regularly, keep your eye on inflation, check if your sample is representative. Experts may, it seems, be capable of excessive faith in the merits of technological advance – faith that seems to overrule conventional expert wisdom or common sense in the area in which they work.

The tales just given are not proof of anything, of course. They are meant only to motivate taking seriously some modest advice: when experts come bearing a miraculous new technology, ask hard-nosed questions about what has actually been achieved.

2. Macro models, old and new

In the remainder of the paper, I articulate the sort of hard-nosed questions I think we should ask of the new macro models as they enter the policy process. I start with the collapse of the last generation of models.

Robert Lucas won a Nobel prize in part for his critique of the models of the 1960s and 1970s:

> 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)

As noted by King, Lucas's critique, along with the events of the day, had devastating effect:

> 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)

Lucas argued that what was needed was a new kind of model in which macroeconomic behavior was derived as the equilibrium outcome of dynamic optimization by rational agents. Lucas set us on a path to creating what have become known as dynamic stochastic general equilibrium (DSGE) models. When we can model behavior as a rational response to risk, Lucas argued, we are on solid ground; otherwise, *economic reasoning* itself is worthless:

> 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)

Let us concede that the Lucas ideal is indeed the legitimate and ultimate goal of macro modeling.[4] This might lead one to believe that the first hard-nosed question we should ask of the new models is, 'Do the models meet the Lucas ideal?

This is, however, the wrong question, in part, because the obvious answer is 'no'. To see this, we need a brief history of DSGE modeling.

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. Equilibrium behavior is then derived as the result of explicit constrained optimization problems. In 1981, Lucas put it this way:

> 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) [of fitting basic facts of the business cycle]. 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

4 Many would debate this point, especially in the details, but these issues are not essential to the argument here.

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 the Lucas-Adelman-type tests of reproducing realistic dynamics.

By of the turn of the century, we were arguably beginning to produce models with 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 an acceptable model has been formulated, it is natural to perform optimal policy computations. This project was initiated in the 1970s, but largely died when the models were abandoned academically. The new DSGE models have a much more sophisticated treatment of 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. If

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

we are to use the models in this way, it is natural to ask whether we have attained the Lucas ideal.

2.2 DO THE NEW MODELS HAVE SOLID MICROFOUNDATIONS?

The essence of the question about achieving the Lucas ideal is whether we have replaced *ad hoc* behavioral assumptions of the old models with economic behavior that is derived as an equilibrium response of optimizing agents. In the profession, a short-hand for this question is, 'Do we now have solid microfoundations?'

The profession uses the term *microfoundations* fairly informally, but it is important to be clear on this matter. 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.

Note that turning a model with *ad hoc* assumptions about behavior into one with weak-form microfoundations is conceptually trivial: just replace the *ad hoc* assumptions on behavior with *ad hoc* technological constraints. Instead of assuming that agents behave in a certain way, we specify constraints such that the only reasonable optimizing choice is that they behave in the way formerly assumed.

Of course, this cannot represent (much) real progress,[6] and one might suppose that the profession would recognize the limited value of this step. As we shall see, however, current DSGE models in key respects take this approach.

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

Whereas the DSGE research agenda began as a search for strong-form microfoundations, the reliance on well-founded micro and arguably fixed parameters gave way, to a significant degree, to a search to discover what sort of *ad hoc* frictions might make the model fit. In my view, 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.

[6] Even this minimal step may provide a building block for further model development.

What has actually been achieved? I will focus on one aspect of behavior arguably at the core of the models: sticky prices and wages. Of course, sticky prices and wages have always been at the center of the Keynesian story of business cycles. At least since Lucas's arguments it has been clear that providing a solid rationale for the stickiness is an important project for Keynesians.

Whereas old models simply assumed that prices are sticky, the new models allow the firms to optimize in the setting of prices. The firms are, however, subject to the technological constraint that they can only change their price when an oracle tells them they can. Imagine each firm has a beacon in its business office, which generally shows red; it periodically flashes green and at that point the firms can change prices. The beacon turns green at random times unrelated to economic fundamentals.

While this assumption has proven extraordinarily productive in practical modeling terms, it is obvious that it provides no rationale for stickiness. Relative to old models, we have replaced an *ad hoc* assumption about behavior with an *ad hoc* constraint essentially forcing firms to behave as formerly assumed.

Setting aside the heavy-handed form of the assumption, one might ask whether at least the parameter determining the frequency with which the beacon turns green might reasonably be viewed as a fixed and immutable economic fact as required for solid microfoundations. Of course, there is no such argument,[7] and if one wants some contrary evidence, a quick check of recent events in Zimbabwe confirms that firms are perfectly capable of changing the frequency with which they adjust prices. Moreover, are we really confident that, in the current economic crisis, firms will wait for their beacon to blink green before lowering prices?

From the standpoint of the Lucas critique, one might at least hope that the exogenous average frequency of price adjustment in the models 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. The microeconomic evidence overwhelmingly supports the view that different sorts of goods have different average frequencies of price adjustment. While heterogeneity dominates the data, we have barely begun to explore this topic (see e.g., Carvahlo (2006) and Nakamura and Steinsson (2008)). At this point it is clear that there is no strong support for the microfoundations of calibrating the model to a single average frequency of price adjustment.

The assumption that firms' prices are exogenously fixed for extended periods until a beacon blinks does not constitute a microeconomic ration-

---

7 Leeper (2005) also makes this argument.

ale for price stickiness; it is not specified in terms of a plausibly fixed parameter; and serious consideration of existing theory does not resolve how to condense the heterogeneous micro data into a single frequency of price adjustment.

One could continue this analysis with other aspects of the microfoundations (as in Faust 2005, 2008). In this paper, though, my object is mainly to invite the reader to ask in a hard-nosed way whether we have met the Lucas Ideal. In my view, the answer is clear.We have made immense progress in attaining weak form foundations; we are, however, probably closer to the end of the beginning than the beginning of the end in the construction of a model with strong form microfoundations.

2.3 GIVEN OUR STATE OF KNOWLEDGE, THE LUCAS QUESTION IS THE WRONG QUESTION

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 achievement, setting the stage for assault on the larger goal.

From a practical policymaking perspective, however, as we await ultimate success, there are other questions we should be asking. Here we run up against a stubborn view in the profession, which seems to be rooted in Lucas's emphatic argument that models are of no value outside the class of models he was advocating. This view is not only wrong, it has become quite dangerous. It has created a worrisome urge to declare some sort of victory in overcoming the Lucas critique This pressure probably accounts for the tendency in some parts to view the sort of microfoundations just discussed (blinking beacons, etc.) to be solid microfoundations. Declaring false victory – over icebergs, infectious disease, or the Lucas critique – is surely one way we start down the path to catastrophic error.

Thus, I think it is important that we set aside the view that policy modeling is valueless unless we meet the Lucas ideal. While Sims (e.g., 2006) and others have taken up this case before, it seems to have gone largely unappreciated that nothing in Lucas's Nobel prize winning critique contained a proof that the critique rendered economic modeling *valueless*. I suspect that Lucas's absolutist claims were simply hyperbole of the sort that marked all sides during the violent upheaval in the profession that was the rational expectations revolution. As the new DSGE models enter the policy process, though, it is time we re-examine the value of less-than-ideal models.

3. Less-than-ideal DSGE models: a more pragmatic standard

How should we assess the value of models that do not meet the ideal? Lucas's brilliant statement of the ideal does not help us much here, and this subject has not received vigorous debate in the profession in part because of the absolutist view that anything less than the ideal must be worthless. In this section, I attempt to resurrect a more traditional perspective on macro modeling, and to articulate the sort of hard-nosed questions I think we should be asking of less-than-ideal models used in the policy process.

3.1 PRAGMATIC AMBITIONS IN MACRO

In the inaugural Hicks lecture in Oxford (1984), Solow laid out a case for limited modeling ambitions in macro. He did this in the context of defending young Hicks's IS/LM model against older Hicks's outright rejection:

> 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)

Hayek (1989) makes the same argument in general terms in his Nobel lecture, and in 1948, Milton Friedman's case for the k-percent money growth rule was clearly based in this perspective.[8] Because the optimality properties of the k-percent rule have been much studied, one might forget that Friedman's original 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 stated,

> 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\ldots (1948, p.263)

[8] In later writings, Friedman sometimes takes a harder line.

Continuing with a fairly thorough discussion of the main dangers in the proposal, he concluded, '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 this view, we have not attained a model in which the implied optimal policies are ones we can feel confident will, in any meaningful sense, be optimal in practice. We should aspire, then, to design well-behaved policy in light of our conceded inability to design meaningfully optimal policy. How do we appraise models for use in this project?

3.2 DSGE MODELS AND LAB RATS

The question of how best to use an admittedly flawed and incomplete model in policy is a subtle one. While discussion of this topic in macro has been somewhat stunted, one can find some guidance in other fields. One interesting parallel comes from regulatory policymaking regarding human exposure to potentially dangerous chemicals. Monetary policy and toxicological policymaking share an important feature: in neither case is it acceptable to simply run experiments on the actual target population. We do not randomly change monetary policy to learn its effects on people's spending; nor do we randomly expose them to chemicals to find out what makes them ill. Thus, we find ourselves forming policy based on models.

Policymakers in environmental and pharmaceutical toxicology understand that one would ideally make policy based on a model with biological microfoundations matching the human case. But humans are large, complex dynamic, general equilibrium systems; and we currently have no ideal model. Instead, regulators turn to imperfect models in the form of nonhuman mammals: we check how the chemical works in, say, rats as a basis for drawing conclusions about its potential toxicity for humans. Like the DSGE model, rats match a large number of the stylized facts regarding the human system; still, they do not constitute an ideal model of a human.

What is strikingly different from the case in macro, however, is that in toxicology there is a robust discussion of what sort of framework should be used for drawing conclusions based on a less-than-ideal model.[9] A joint working group of the U.S. EPA and Health Canada conducted a detailed study of the human relevance of animal studies of tumor formation. They summarized their proposed framework for policy in the following four steps:

[9] For example, a scholar google search on 'human relevance' and 'animal studies' or 'in vivo studies' turns up hundreds of studies. Examples are Cohen, et al. 2004 and Perel, et al. 2006, and Zbindin, 1991.

1. Is the weight of evidence sufficient to establish the mode of action (MOA) in animals?
2. Are key events in the animal MOA plausible in humans?
3. Taking into account kinetic and dynamic factors, is the animal MOA plausible in humans?
4. Conclusion: Statement of confidence, analysis, and implications. (Cohen, et al., 2004)

In the first step, we get clear about the result in the model. The remaining steps involve asking serious questions about whether the *transmission mechanism* in the model – to borrow a monetary policy term – plausibly operates similarly in the relevant reality. This process is inherently judgment based, and unavoidably subject to error,[10] but an active literature exists deriving and assessing ways to refine this process.

This discussion dovetails nicely with Solow's perspective discussed above. Even in the face of incomplete understanding, models can play an important role in organizing our thinking, placing some structure on our interpretation of the data, and helping us 'get a handle on' the facts. The essential element highlighted by the toxicology case is that, crucially, a key part of this reflection is forming a judgment about which features of the model are plausibly shared by the target of the modeling and which are not.

3.3 PRACTICAL QUESTIONS ABOUT DSGE MODELS

DSGE models are incredibly sophisticated. Still there is a substantial gap between a DSGE model of a dozen or so macro variables and the actual economy. Indeed, this gap strikes me as not so different in magnitude from that between lab rats and humans. In the face of this gap, I am advocating that we follow the toxicologists. To paraphrase the framework above: Are key events in the DSGE mode of action of monetary policy plausible in the actual economy? Taking into account kinetic and dynamic factors, is the DSGE mode of action plausible in reality? In more macroeconomic terms: Is the model broadly consistent with our understanding of real world business cycles? Of the transmission mechanism of monetary policy?

If the model's implications surprise us, we have a choice. It might be that we should alter our understanding; alternatively, we might decide that the surprise is an artifact of some implausible feature of the model

10 A large part of the literature documents the mistakes and steps and missteps the field has taken in response.

that we had not previously noticed or had not yet found a way to fix. The issue then, is how to deal consistently with these problem areas in making policy.

Unfortunately, these questions are not trivial to answer. Moreover, the development path of these models makes the answers more opaque. To harshly condense the model development process described above, the Bayesian estimation is based on a largely unmotivated prior belief over the parameters of a large, imperfectly understood model, which has a large and weakly justified set of frictions and is driven by a large and weakly motivated set of exogenous shocks. It is very difficult to determine from this process in which ways the economic mechanisms in the model will reflect reality and in which ways they will not.[11]

It is true that these models have been shown to match some broad aspects of reality. They fit the handful of data series in the estimation sample well and forecast about as well as standard benchmark models. Of course, the 1960s models fit and forecasted well. Lucas and other critics took their task to be explaining why the models contributed to catastrophe despite these facts.[12] Surely the excellent forecasting of the 1960s models helped bring false confidence to the users, a mistake we should avoid this time.

As we bring the new models into the policy process, I think there is no substitute for careful checking of where the mechanisms in the model reflect the common understanding and wisdom of the policymakers and where they do not. The natural way to proceed is by stating a set of beliefs, perhaps corresponding to common wisdom about the macroeconomy, and then comparing those beliefs with the mechanisms in the model.

3.4 ILLUSTRATION BASED ON THE RAMSES MODEL

Given the Bayesian approach to model estimation used in this area, it is natural to use Bayesian tools to perform this sort of comparison. The formal Bayesian tools I use in the following are standard and described in Geweke (2005). I mainly sketch a small portion of a more complete analy-

[11] In contrast, I observed, though did not participate directly in, the development of the Fed's more traditional models (FRB/US, FRB/Global) introduced in 1995. The development process was *ad hoc*, opaque, and difficult to characterize. It involved heavy involvement of economists and policymakers at every level of the organization. Whatever else one says about this highly problematic process, it had one virtue: the model development phase did not stop until the relevant group of decisionmakers agreed that the model broadly reflected the views of the group on key questions about the business cycle and monetary transmission mechanism.This is consistent with the descriptions of these issues in Reifschneider et al. 2005, and Stockton 2002.

[12] Lucas makes this point explicitly in the quote that begins section 2. Sims's (1980) famous critique likewise is based in the fact that good fit notwithstanding, the economic mechanisms in the model lead to bad policy prescriptions.

sis here. The more complete approach is based on Geweke's (2007) recent suggestions about inference in incomplete models and is worked out more fully in Faust (2008), Gupta (2009) and Faust and Gupta (2009).

The illustrative results presented below are based on a version the Ramses model, a model used in the policy process at the Riksbank. The model fits the general framework described above: a core model with a large number of frictions and exogenous shocks, with exogenously specified dynamic structure for the shocks. The model is well documented elsewhere (e.g., Adolfson, et al. 2006); and Adolfson et al. (2007) have recently shown how to use the model for practical optimal policy calculations. It is important to emphasize that the particular version of the model I am discussing is not identical to the one used in the policy process and that these results should be viewed only as illustrative. The suggested evaluation process begins by stating a few core beliefs.

Consider two.

Consumption growth is insensitive to short-term changes in short-term interest rates. Based on data from many countries and time periods, combined with a certain amount of theory that has been built up to explain these facts, many economists believe that aggregate consumption is not very sensitive to short-run changes in short-term interest rates. Indeed, a key problem in DSGE models has been that agents in the model seem to be too willing to substitute between current and future consumption when given a small incentive to do so. This problem explains why habit formation, adjustment costs, and persistent shocks to marginal conditions have been added to the core model. Based on this belief, we might want to investigate what the model says about the consumption growth-interest rate correlation.

Long and variable lags of monetary policy. 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 its 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 simply consider how much the economy reacts in the very quarter a policy is adopted.

In the Bayesian estimation approach used with these models there are at least two questions of interest when we consider economic features such as the two just discussed. The estimation begins with a statement of prior beliefs about the economy.[13] The prior beliefs might be thought

[13] Where prior is meant to mean before considering the data at hand.

of as the personal *biases* one brings to the analysis: the stronger the prior belief, the less subject the belief will be to alteration based on the data. Ideally, the prior beliefs used in model estimation would reflect the actual beliefs of key participants in the process. In practice, this is difficult to implement, so the prior used in estimation is largely arbitrary. Thus, it becomes interesting to ask how the formal prior compares to ones actual prior beliefs and how much the arbitrary formal prior is affecting the results of the analysis.

For the version of the Ramses model we are examining, the formal prior belief regarding the interest rate-consumption correlation is shown by the roughly bell shaped curve in figure 1, top panel. The horizontal axis gives values of the correlation. The height of the curve reflects the prior plausibility of the corresponding correlation value on the horizontal axis – where the curve is highest, the corresponding correlation is assigned higher prior plausibility.

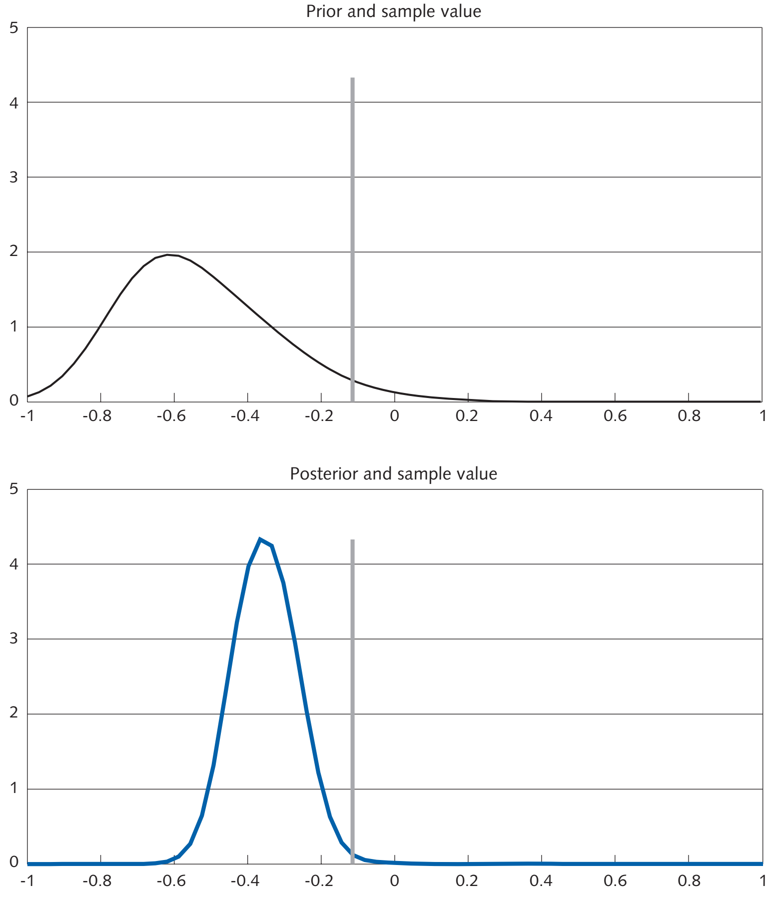
The prior used in estimating Ramses fairly strongly favors a strongly negative contemporaneous correlation (in quarterly data) between the short-term interest rate and consumption growth. Correlation of about –0.6 is most likely in the prior and values near zero are viewed as quite improbable. The correlation value in the estimation sample (vertical line, Fig. 1) reflects the common finding of little systematic relation between these variables.

The Bayesian estimation approach combines the model, the prior belief, and the data to form a new assessment of all aspects of the model, including this correlation. This new assessment, called the posterior belief, is shown in the bottom panel. The posterior still fairly strongly favors a negative correlation with the most likely value around –0.4, and once again values near zero are very implausible.

Thus, this estimation of the Ramses model was based on a strong prior belief that consumption is quite sensitive to interest rates and this prior belief continues to be reflected in the posterior. What should we make of this? This is precisely the challenging question I believe policymakers using this model should confront. Is the low correlation as found in the Swedish data and many advanced economies a fluke? Should we make policy based on the belief in a strong sensitivity of consumption to changes in short-term interest rates? Or should we view correlation as a possibly unfortunate artifact of the model building process – an important difference, as it were, between the laboratory rat and the human?

The analogous examination of long lags in the effects of monetary policy is depicted in Fig. 2, which shows the effect on the growth rate in GDP of a one-quarter percentage point rise in the policy interest rate. To emphasize, the growth effect is in the same quarter as the change in the

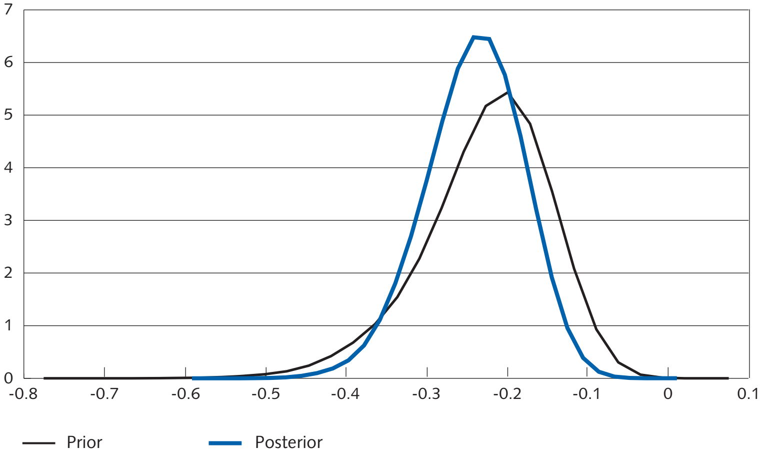
Figure 1. Prior and posterior densities along with the sample value for the contemporaneous correlation between the short-term interest rate and quarterly consumption growth in a version of the Ramses model.
Percent



Source: Author's calculation using computer code provided by Riksbank staff.

policy rate. In this case, the prior and the posterior beliefs for the immediate effect of an unanticipated change in policy roughly correspond. That is, the posterior belief is largely driven by the (largely arbitrary) prior belief. That prior belief puts maximum plausibility on a one-for-one immediate effect of a surprise change in the policy rate. That is, the one-quarter percentage point rate increase immediately gets you a one-quarter point fall in the annualized quarterly growth rate of GDP. This one-for-one immediate effect does not capture the conventional wisdom; it is common to assume that the immediate effect is actually zero. Further, some structural VAR work (e.g., Faust, 1998) suggests that conclusions about the effects of policy may be sensitive to what is assumed about the immediate effects of the policy shock.

Figure 2. Density for the effect of a one-quarter percentage point surprise rise in the policy interest rate on the growth rate in the quarter the cut takes place. A value of –0.25 means that the quarter point rise in the interest rate leads to an immediate quarter point fall in the annualized quarterly growth rate of GDP.
Percent



Source: Author's calculation using computer code provided by Riksbank staff.

What should we make of this result? Should the policy predictions of this model be taken seriously in this dimension, or is this one of the implausible aspects requiring careful translation between the model results and reality? Once again, this is the sort of question that I believe policymakers and other users of these models should be addressing.

The particular 'core beliefs' that I employ as illustrations may not in fact be core beliefs of the reader or of policymakers at central banks, but I hope the point is clear. Nothing guarantees that the economic mechanisms in the model correspond to the ways macroeconomists generally organize their thinking. Hopefully, the two examples given at least suggest that there may be areas of important tension here. Where model and standard thinking conflict, there may be no strong presumption about which should change – on one hand we have myriad unmotivated aspects of the specification of the model and prior beliefs, on the other hand known failures of existing professional wisdom. Before we use these models in the Solow-style mode of helping to organize our thinking and refine our trained intuitions, it seems only sensible that we check first where the models reflect and where they contradict common understanding. This investigation can then provide the basis for building a systematic framework for use in translating between model results and reality.

4. Conclusion

History teaches us that, despite advances in shipbuilding, sea captains should watch out for icebergs and that, despite advances in antibiotics, doctors should wash their hands regularly. To this list, I would add that macro policymakers should explicitly note and make allowance for their less-than-ideal models. As we bring new models into the policy process, we should familiarize ourselves with the most and least plausible parts of the models and then formulate standard ways of accommodating the perceived flaws.

Of course, one way to do this is to simply begin using the models. If the history of modeling has taught us anything, it has taught us that the flaws will become apparent with use. Policymakers and staff can evolve ways to deal with the flaws 'on the fly' as policy is made. This haphazard process, however, is prone to just the sort of policy breakdowns and even catastrophes associated with macro modeling in the 1960s and with risk modeling in financial markets more recently. My argument amounts to little more than advocating a hard-nosed common sense at the outset in bringing these models into the policy process.

Advocates of the new models sometimes react in mild horror to the suggestion that we add a layer of judgment – based in explicit examination of model flaws – to the process of applying the model. The very purpose of the model, in this view, is to remove discretion and ensure consistency and transparency in policymaking. Of course, consistency is important. As the American sage, Ralph Waldo Emerson argued, though, 'a foolish consistency is the hobgoblin of little minds.' The approach I am advocating is intended to help attain a sophisticated consistency: be clear at the outset about model flaws and the ways in which these will systematically be accommodated. The list of flaws will undoubtedly change with use of the model – some problems fixed, new ones discovered – but the framework for use of model results can remain relatively static, consistent, and transparent.

Opponents of the new models sometimes hear in my critique of flawed models a neo-Luddite argument in favor of rejecting the models entirely. In concluding, let me emphasize that, to the contrary, I believe that these models are essential to progress. Over the nearly 20 years I spent at the Fed, I observed a considerable increase in the sharpness with which dynamic economics was discussed – an advance that would have been hard to attain had many participants in the process not sharpened their skills using DSGE models. So long as we incorporate some simple cautions – and wash our hands regularly – I am confident that we are only beginning to obtain the immense policy benefits that can come from further work with these models.

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 Association, Papers and Proceedings, 2005b, forthcoming.

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

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

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

Christiano, L.J., M. Eichenbaum and C.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, L.J., M. Eichenbaum and C.L. Evans (2001), Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. Federal Reserve Bank of Cleveland Working Paper.

Cohen, S.M., J. Klaunig, M.E. Meek, R.N. Hill, T. Pastoor, L. Lehman-McKeeman, J. Bucher, D.G. Longfellow, J. Seed, V. Dellarco, P. Fenner-Crisp, D. Patton, 2004. Evaluating the Human Relevance of Chemically Induced Animal Tumors, Toxicological Sciences, 181–186.

Dawes, R., D. Faust and P. Meehl, Clinical versus Actuarial Judgement, in Heuristics and Biases, the Psycology of Intuitive Judgement, Thomas Gilovich, Dale Griffen and Daniel Kahneman eds., Cambridge University Press: Cambridge, 2002.

Del Negro, M. and F. Schorfheide, 2008, Forming Priors for DSGE Models (And How is Affects the Assessment of Nominal Rigidities).

Del Negro, M. and F. Schorfheide, F. 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, M. and F. Schorfheide, 2007, Monetary Policy with Potentially Misspecified Models NBER Working Paper 13099.

Erceg, C., L. Guerrieri and C. Gust, Sigma: A New Open Economy Model for Policy Analysis, manuscript, Federal Reserve Board, 2003.

Erceg, C., D. Henderson and A.Levin (2000). Optimal Monetary Policy with Journal of Monetary Economics, Staggered Wage and Price Contracts. 46, pp. 281–313.

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

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

Faust, J., 2005. Is applied monetary policy analysis hard? manuscript, John Hopkins University.

Faust, J., 2008. DSGE models in a second-best world of policy analysis? manuscript, John Hopkins University.

Faust, J., and G. Abhishek, 2009. Bayesian Evaluation of Incomplete DSGE models. Manuscript in progress, John Hopkins University.

Fenn, J., and M. Raskino, Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time Harvard Business School Press: Cambridge, 2008.

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

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

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

Greespan, A., Testimony in front of Committee of Government Oversight and Reform, U.S. House of Representatives, October 23, 2008.

Greewood, J., Z. Hercowitz and G. Huffman (1988). Investment, Capacity American Economic Review, Utilization and the Real Business Cycle. 78 (3), pp. 402–417.

Gupta, A., 2009. A forecasting metric for DSGE models. Manuscript, John Hopkins University.

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.

King, R., Quantitative Theory and Econometrics, Federal Reserve Bank of Richmond Economic Quarterly, Summer 1995, pp. 53–103.

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

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

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

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

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

Perel, P., I. Roberts, E. Sena, P. Wheble, C. Briscoe, P. Sandercock, M. Macleod, L.E. Mignini, P. Jayaram and K.S. Khan, 2007. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ, 334, 197.

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.

Schorfheide, F., K. Sill and M. Kryshko, 2008, DSGE Model-Based Forecasting of Non-Modelled Variables.

Sims, C., 2006, Improving Policy Models, manuscript, Princeton.

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

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

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

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

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

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

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

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

Smets, F. and R. Wouters, 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.

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

Stockton, D., 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, L.E.O., Monetary Policy with Judgment: Forecast Targeting, manuscript, Princeton, 2004.

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

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

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

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

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

Zbinden, G., 1991. Predictive value of animal studies in toxicology. Regulatory Toxicology and Pharmacology, 14, 167–177.