

Improving Monetary Policy Models

by

Christopher A. Sims, Princeton University and NBER

CEPS Working Paper No. 128

May 2006

IMPROVING MONETARY POLICY MODELS

CHRISTOPHER A. SIMS

ABSTRACT. If macroeconomic models are to be useful in policy-making, where uncertainty is pervasive, the models must be treated as probability models, whether formally or informally. Use of explicit probability models allows us to learn systematically from past mistakes, to integrate model-based uncertainty with uncertain subjective judgment, and to bind data-based forecasting together with theory-based projection of policy effects. Yet in the last few decades policy models at central banks have steadily shed any claims to being believable probability models of the data to which they are fit. Here we describe the current state of policy modeling, suggest some reasons why we have reached this state, and assess some promising directions for future progress.

I. WHY DO WE NEED PROBABILITY MODELS?

Fifty years ago most economists thought that Tinbergen's original approach to macro-modeling, which consisted of fitting many equations by single-equation OLS

Date: May 26, 2006.

©2006 by Christopher A. Sims. This material may be reproduced for educational and research purposes so long as the copies are not sold, even to recover costs, the document is not altered, and this copyright notice is included in the copies. Research for this paper was supported by NSF grant SES 0350686 and by Princeton's Center for Economic Policy Studies. This paper was presented at a December 2-3, 2005 conference at the Board of Governors of the Federal Reserve and may appear in a conference volume or journal special issue.

and assembling them into a multiple-equation model, had been shown to be internally inconsistent and an inadequate basis for scientific progress in macroeconomics.¹ The basic point, made at length by Haavelmo (1944), is that because in economics our theories do not make exact predictions, they can never be proved inadequate simply by showing that they make prediction errors. In order to allow models to be compared and improved, they must be formulated as probability models. That is, they must characterize the probability distribution of observations, rather than simply make point predictions. For macroeconomic models, this means they must be probability models of the joint behavior of the time series they are meant to explain.

If we intend to use the models in decision-making we have to go beyond Haavelmo. We have to recognize that the models have to be able to address sources of uncertainty that the data are incapable of fully resolving. Large macroeconomic models inevitably contain more free parameters and theoretical ambiguities than the available data can pin down for us. The model is not complete, for decision-making purposes, unless it characterizes these sources of uncertainty. This means that attempts to limit probability statements to areas of uncertainty where the frequentist interpretation of probability is useful cannot be adequate. We need to think of our probability models as characterizing uncertainty from all sources and as capable of integrating uncertain information from sources other than the data — one aspect of what is sometimes called “judgment”.

¹These first few paragraphs cover much the same ground as parts of Sims (2002); ?.

While these points might seem obvious, the recent history of central bank macroeconomic modeling has seemed to ignore them entirely. The models that are in actual use as frameworks for discussion in the policy-making process have abandoned the theoretical framework of the Cowles foundation approach that Haavelmo's ideas set in motion. They have not replaced it with another probability-modeling framework, but rather with a reincarnation of the single-equation fitting approach of Tinbergen. There is no attempt to construct a joint likelihood for the observed time series, and no attempt to assess whether the model's own structure can support the single-equation methods used to estimate it. No model-generated measures of uncertainty play any important role in policy discussions.

At a practical level, we need probability models in order to provide reliable measures of uncertainty about model results and projections, to allow us to compare different models' match to the historical data, and to allow us to learn systematically from our mistakes. This latter point has arisen repeatedly in the experience of central bank modelers. There is a tendency to apply ad hoc repairs or extensions to models, or to switch from one model to another as the focus for policy discussion, in response to perceived mistakes. If we do not have a probability structure in which to assess mistakes and proposed fixes for them, there is a tendency to overcorrect, to continually add degrees of freedom to the model so that overfitting bias grows instead of shrinks as information accumulates.

II. WHY HAVE WE ENDED UP WITH NON-PROBABILITY MODELS?

Here I take up a number of possible reasons for the retreat from probability modeling at central banks.

II.1. The ossification of academic econometrics. The Cowles foundation research was based on frequentist asymptotic distribution theory and Neymann-Pearson statistical testing. It was developed at a time when models with more than 5 equations were difficult to handle computationally. Models at the scale needed for central bank policy analysis have many more variables and free parameters. Frequentist asymptotic theory is an unreliable guide to the uncertainty associated with estimates of models like these. Asymptotically efficient methods, like full information maximum likelihood, have exotic and difficult to compute frequentist small sample properties. Instrumental variables methods, if all apparently eligible instruments are used, often reduce to OLS or something very close to it. The frequentist asymptotic theory had nothing to say about how or why to limit lists of instruments, though there were many ad hoc practical suggestions as to how to do so. The Cowles foundation theory, in short, prescribed methods that were difficult to implement and provided distribution theory that was an unreliable guide to the uncertainty associated with the results.

This situation might have been a challenge to academic researchers, leading to development of new and more appropriate methods. But as the models were reaching their full scale, academic researchers became convinced by a simplified version of the Lucas critique that was taken to imply that econometrics-guided real time policy

analysis was internally contradictory or of trivial importance. Academic research on the econometrics of large scale policy models almost ceased.

Another reason for the lack of attention to the problems of inference in large time series models was the long detour of econometric time series theory into unit root and cointegration theory. The basic insight of this theory, that cointegration is possible and that differencing every nonstationary series leads to loss of information, is important. But from the perspective of likelihood-based inference (as opposed to Neymann-Pearson testing), this is all there is to it. There is no need for preliminary identification of unit root counts and cointegration vectors in order to allow correct inference. Frequentist approaches to inference, however, lead to prescriptions for complicated sequential testing procedures whose small sample properties, in large models, are poorly approximated by the asymptotic distribution theory that nominally justifies them. So not only was there a great deal of intellectual effort expended on theory that is of little relevance to modeling for decision-making, the complexity of this theory, and the impracticality of actually applying it to models of the scale needed for central bank policy, probably inhibited bank research staff from trying to do the multiple-equation econometrics right. Very few were trained in Bayesian inference, which avoids these obstacles.

Of course these things are changing. The number of economists who are familiar with Bayesian reasoning and methods, while still not large, is certainly growing rapidly. Changes in computing technology and associated developments in algorithms and programs have made it practical to cope with probability models

that would have been out of the question even 10 or 15 years ago. And chairman Greenspan's characterization of modern monetary policy as roughly an application of Bayesian decision theory has helped to make all this respectable in central banks.

II.2. Fear of being Lucas-critiqued. There is no reason that incorporating rational expectations need entail retreating from probability-based inference. Indeed one branch of the early rational expectations literature took it as a program for developing improved multi-equation macro models, incorporating new "cross-equation restrictions" that would make estimates sharper. But the branch of the literature that read the Lucas critique as implying econometrics-based policy analysis was trivial or delusional took the steam out of the cross-equations restrictions program. Also, implementing the program required staff with a new kind of technical training, which became available only slowly. Minnesota, where I was in the 70's and 80's, was a center of research and training in rational expectations. Paradoxically, the Blanchard-Kahn approach to solving rational expectations models, which is the foundation of the solution methods for large models, was not a standard part of the curriculum at Minnesota. Students instead became expert at reducing models to one-dimensional equations and at using guess-and-verify methods of solution that are impractical on large systems.

With new Ph.D.'s all expert at criticizing any model in which expectations were not rational, yet increasingly sketchy in their understanding of multiple-equation probability modeling, it is natural that attention turned mainly to generating models

that incorporated rational expectations, even if this forced, because of the limitations on staff research time, retreat from probability-based inference.

Within the last few months I have heard a prominent economist repeat, in a presentation, the canard that economists have nothing useful to say about quarter-to-quarter monetary policy settings, and that in any case these questions are not important. His view was that the only kind of advice we have to give is analysis of policy “rules” — and that inflation-targeting is not a rule. I think this point of view is not as widely accepted as it used to be, but it is still common enough to be a source of worry about the decay of this obstacle.

Though I have explained often before (2002; 1987) the internal inconsistency of this view that the only correct or nontrivial form of policy choice is choice of policy “rule”, and people seem to find the point either obvious or incomprehensible, let me explain it. The extreme view of the fallacy runs like this:

The economy can be thought of as a multivariate stochastic process, every period generating new realizations of the vector of random variables that characterizes the state of the economy. One part of the system that determines the behavior of this stochastic process is the behavior of the central bank. The bank’s behavior can be thought of as described by an equation that maps the current state of the economy into a probability distribution determining monetary policy actions next period. So long as this mapping remains the same, the stochastic process characterizing the behavior of the economy is the same. What

the central bank normally does on a month-to-month or quarter-to-quarter basis is simply to choose particular values for the monetary policy random variable. To be more concrete, if the policy rule is a “Taylor rule” mapping past inflation and past output growth into current Federal Funds Rate values, what the policy authority does is pick values for the residual in the Taylor rule. Doing this does not change the rule itself and thus has no effect on the stochastic process being followed by the economy.

But changing the rule — changing the coefficients in the Taylor rule or replacing the Taylor rule with a policy of setting the rate of growth of reserves, for example — would change the stochastic process followed by the economy. It is only this kind of policy action that has nontrivial effects and only this kind about which economists have anything important to say.

The problem with this view is that it does not recognize that changing the coefficients in the policy rule or adopting a money growth target is a policy action like any other, in the sense that, from the point of view of a rational private agent, it is an uncertain event. Changes in “rule” are then just one more source of random disturbance in the stochastic process the economy follows. Changing the rule does not change the stochastic process itself, it just picks a particular realized value for the “change in rule” random variable.

Of course the Lucas critique did teach us that the effects of policy actions depend on the beliefs of the private sector about the distribution of policy actions. It matters whether the public believes that when inflation starts rising this implies that soon the Federal Funds rate will start rising. It matters whether the public believes that when the Funds rate rises, this signals a likely continuing sequence of further rises, or instead that such rises are temporary. But the same considerations apply to policy rules. Does the public believe that a rule that makes interest react sluggishly to inflation will be maintained if inflation accelerates? Does the public believe that a shift to a stronger and quicker reaction of the Funds rate to inflation will be maintained in a period where energy price rises produce simultaneous inflation and recession?

A conventional approach to policy evaluation recently has been to build a model in which expectations are modeled explicitly. Policy changes are then modeled as changes in the policy sector of the model (usually a single “reaction function”), with private sector expectations treated as based on knowledge of this new policy behavior, at the date it is introduced, no knowledge of it before the date it is introduced, and firm belief that the new behavior will be indefinitely sustained. Such analysis is said to avoid the Lucas critique. But in fact it is precisely a replication of the type of policy analysis that Lucas critiqued. It models a policy action (the change in rule) as if it were non-stochastic from the point of view of the public, ignoring private sector expectations about the action before and after it occurred.

The fact is, we must think of every policy action as stochastic. If this is true, there is no possibility of taking policy actions that change the stochastic process followed

by the economy. But this does not mean that policy actions have no important consequences. Stochastic processes do not have to be stationary. Random disturbances, including policy actions, can change the conditional distribution of future paths of the economy in ways that make permanent, important changes in the levels of variables.

While nihilistic views about quantitative policy advice seem to be dwindling only slowly, inflation targeting tends to go hand in hand with more explicit attention to regularly repeated economic projections and with an expanded role for modeling. The policy cycle in central banks is precisely quarter-to-quarter monetary policy making, and to those engaged in it it appears, correctly, neither easy nor unimportant. This tends to keep things moving in the right direction.

II.3. Preventing misbehavior of long term projections. Data inherently cannot have much to say about a model's dynamics at wavelengths of a quarter the sample size or longer, but projections this far, or nearly this far, into the future are not uncommon in policy scenarios. Models fit by standard invariant methods — in other words, without using prior distributions — will tend to misbehave when asked to make long projections. This tendency becomes worse if system methods have not been used in estimation and if continual ad hoc respecification of individual equations has intensified overfitting. Central bank modelers understandably have looked for ways to get improved long run behavior. Probably beginning with the Bank of Canada's QPM, they have turned toward creating a "core" that has reasonable long run behavior imposed and that is deliberately insulated from influence by the data. My impression

is that the appeal of stabilizing the model's long run behavior, the Bank of Canada's demonstration that the "insulated core" approach worked, and the absence of any established alternative within a framework of frequentist, probability-based, multiple equation modeling, has been a major reason for retreat from such frameworks.

But it is clearly not a good thing in principle to be insulating long run properties of models from the data. Particularly in fully articulated behavioral models, the data will be informative about low frequency properties. The initial promise of the real business cycle framework included the idea that we could integrate the study of growth and business cycle fluctuations in a single model.

Estimated DSGE's still suffer from these difficulties. It is common in estimating these models to preprocess data to remove low frequency variation, either by simply removing means and/or trends, or by differencing, so that the likelihood itself is insulated from low frequency variation. This is done because our usual stable of models does not include any that provide flexible enough low frequency behavior to match observed means and growth rates, while at the same time producing realistic cyclical behavior. For example the actual data show drift in hours per worker, investment as a share of GDP, trade shares in GDP, etc. Our standard setups that deliver homogeneously growing steady states cannot easily match these facts. The paper Edge, Kiley, and Laforge (2005) at this conference is a start at confronting these issues.

Structural VAR's are capable of matching complex low frequency behavior. Their problem is that they are too flexible in this dimension, and that fitting to data doesn't

automatically make a VAR's low frequency behavior emerge as reasonable. The generalized dummy observation methodology discussed below provides a promising approach to addressing this problem.

II.4. The need to decentralize and collaborate. In every central bank, preparing projections in the policy cycle involves groups of people with different types of expertise trading information and iteratively adjusting the projections as they communicate. The equation-by-equation approach to model specification and estimation is convenient for these purposes. Subgroups of staff can be put in charge of particular equations or groups of equations, while projections from the assembled model maintain consistency as the groups make adjustments.

Structural VAR's attempt to make minimal identifying assumptions, and accordingly almost always leave large blocks of equations uninterpreted. Equation-by-equation fiddling with the model is therefore harder to justify. While in some ways this is an advantage, it is certainly a disadvantage in trying to decentralize work on the model..

DSGE's with a rich set of structural shocks and no measurement error do have distinct sectors, but if they are formulated as joint probability models, there is no simple way to estimate the sectors in isolation. Such models should be perfectly usable as a framework for communication and decentralization in the policy cycle. Adjustment of guesses about the likely size of current structural disturbances would be the focus of discussion. They are not amenable to decentralized estimation, but so long as full system estimation is feasible, that is probably a plus.

II.5. The need to incorporate judgment. In models that have unstable long run behavior, judgment is sometimes invoked to set the projections back on track. More commonly, judgment influences projections by bringing in expert knowledge of current information that does not show up in the model's data series. Though the need for levers that allow introduction of judgment is sometimes cited as a reason for using the type of model that is now most prevalent, this kind of judgmental adjustment can be implemented more easily and consistently with models estimated with modern Bayesian methods. The judgmental information is naturally characterized as beliefs, with specified precision, that can be combined with the posterior pdf that emerges from the sample.

III. ACTUAL BIG MODELS IN HEAVY USE

Here I discuss FRBUS, the model in use at the US Federal Reserve Board, and BEQM, the model in use at the Bank of England. They are both subject to many of the same criticisms, but they are different in important ways. Looking at their structures in more detail may help us to understand why models in actual use end up not being probability models.

The BEQM core has been more consistent in applying modern macroeconomic theory than was FRBUS. Some of the criticisms I made earlier (2002) of FRBUS — that it introduces expectations in an ad hoc, equation-by-equation way that has no substantial foundation in theory and that its treatment of government debt and the effects of inflation on it doesn't even properly reflect budget constraints — do not apply to the BEQM's core. In FRBUS, changes in the real value of government debt

induced by inflation or interest rate changes unanticipated at the time of the issuance of the debt do not affect household wealth. These effects are of crucial importance in analyzing the risks of sustained deflation or of sudden changes in exchange rates and interest rates that might arise in a shift of foreign appetite for US debt. In an earlier paper (2001) I pointed out that wealth swings from these sources are substantial. A recent paper by Doepke and Schneider (2004) has shown that potential gains to the US public from a surprise burst of inflation are at historically unprecedented high levels. Central bank models that cannot trace these effects could prove deficient at a time when monetary policy becomes particularly difficult.

BEQM has such wealth effects accounted for properly. On the other hand, it also considers only one-period government debt, so the only route for such effects is via inflation unanticipated one period in advance. At least in US historical experience, wealth effects on holders of longer term debt from unanticipated changes in interest rates are as important.

But BEQM is not just the core. It includes “non-core” equations whose interpretation was still not clear to me after having read through the 200+ page book (Harrison, Nikolov, Quinn, Ramsay, Scott, and Thomas, 2005) describing the model. These non-core equations are in some sense estimated from data. As in FRBUS, the estimation is equation-by-equation, but the core, whose parameters are insulated from the estimation, is used in the estimation process. The relation of core to non-core seems to be different in out-of-sample projections from what it is in the estimation. The paper by Alvarez-Lois, Harrison, Piscitelli, and Scott (2005) for this conference lays out an

internally consistent linearized version of the model that matches my initial intuitive interpretation of it: the core is treated as an unobservable component.

Equation (34) of the conference paper, with some detail suppressed, reads

$$\left(I - \begin{bmatrix} A & 0 & 0 \\ B & C & D \\ 0 & E & F \end{bmatrix} L \right) \begin{bmatrix} y_t^* \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} G & 0 & 0 \\ H & K & M \\ 0 & 0 & N \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \xi_t \\ \nu_t \end{bmatrix} .$$

This is a linear system with two observable vectors, y_t and z_t , and one unobservable, y_t^* . It could be estimated and simulated forward for forecasting by use of the Kalman filter. In the paper on BEQM in this volume, a small model in this format is used for forecasting and some of its parameters are estimated using this approach. Proceeding this way recognizes that we can never know y_t^* exactly, even at dates in the past. The approach implies a measure of fit and a way to compare this model to others in terms of match to the data. If the parameters of the core model, which enter the system matrices nonlinearly, are fixed at calibrated values, the degree of mismatch to the data this imposes could be assessed by estimating, as a standard of comparison for model fit, the model with the elements of the A, B, \dots, N matrices unconstrained.

But there is no indication in the BEQM book that the model is ever dealt with in this internally consistent way. The descriptions in the book of estimation and forecasting procedures contain ambiguities, and appear not to be justifiable by any probability model. The core/non-core (CNC) exercises in the conference paper are with a 4-observable-variable model, and though the paper estimates three models

on similar data series, no comparisons of fit among the models, or to less restrictive reduced form models, are carried out. It is a promising new development that BOE researchers are formulating an internally consistent probability model that they see as embodying their core/non-core approach, and I hope we can look forward to this approach being applied to the full BEQM model and used to evaluate fit.

As with any setup that introduces a gap between observed data and those that enter the behavioral equations of agents, the CNC framework implies serious problems in making policy projections. How does one explain to the policy board a situation where actual GDP or consumption moves in the opposite direction from core values of these variables? How does one decide whether judgmental information from experts about the current value of a variable should apply to y^* or to y ? It appears to me from the BEQM book that there is even a distinction between core and non-core versions of the interest rate that enters the policy reaction function. When the model is used to condition on a particular path for the policy rate, what is set, the core rate or the non-core rate? The usual interpretation, that changes in the policy rate are shocks to the reaction function, can't apply here, as the specification implies that there is no influence of non-core shocks on core variables, while the core model reaction function contains no shocks. It seems likely that projections of policy effects are made by dropping the core reaction function from the system and replacing it with a fixed time path for the rate or with an alternative reaction function. What are policy makers to do with projections in which the core policy rate and the actual rate

follow different paths? This type of question seems to me a serious drawback for such models as vehicles for communication.

FRBUS, though never formally considered by the FRB as a joint probability model for the time series to which it is fit, *could* be so treated. One can imagine estimating FRBUS as a big Bayesian slightly nonlinear autoregressive model. This is true because the model can be thought of as a mapping from time t and earlier data to time t residuals. The methods now in use to fit BEQM do not generate residuals in such a model-consistent way. FRBUS has a little VAR that runs alongside it, providing in-sample forecasts for the historical fit and used for out-of-sample expectations in the most common form of scenario simulation. This VAR's properties provide an implicit check on FRBUS, as well as helping to keep its projections on track in the short run. There is no such companion to BEQM. BEQM has an "exogenous variables model" to generate the predictions of exogenous variables that it uses in both in and out of sample projections, but this is not a VAR, or even a single model.

FRBUS also uses the core/non-core idea, but in FRBUS the core is primarily a set of cointegrating relations among variables that are estimated and imposed as if non-stochastic. Here the justification for insulating the core from data, or at least from data-based measures of uncertainty, is the asymptotic theory that predicts that cointegrating relations can be estimated with $1/T$ rate precision.

IV. BAYESIAN DSGE MODELS

Since the pioneering work of Smets and Wouters (2003) there has been interest at many central banks in producing models that are at the same time believable

probability models of the data and interpretable models of macroeconomic behavior. Because they are probability models, they have the advantages we have discussed above: they can provide usable characterizations of uncertainty about policy projections; they allow us to learn systematically from experience without overfitting; they allow us systematically to compare the fit of different models; and they generate the post-sample probability distributions that are needed for decision-making. Of course all these advantages also characterize Bayesian structural VAR's. The difference is that the new Bayesian DSGE's provide more complete stories about how policy actions produce effects on the economy and about where non-policy disturbances to the economy originate.

While these models are extremely promising, there are several problems that inhibit their rapid adoption by policy-making institutions and create skepticism about them among economists. In some cases the problems are misperceptions. Some economists complain that the models have "too many shocks". Because the data are clearly not singular, any model that fits multiple time series must necessarily have at least as many sources of stochastic disturbance as there are variables being explained. It is possible, as in CNC style models, to separate shocks into economically interpretable ones (like technology shocks) and "measurement error", with the latter treated as not part of the behavioral model. This of course does not reduce the shock count. In my view, shocks labeled measurement error in economics seldom believably have the properties attributed to them — lack of correlation with other variables and irrelevance to prediction of future values of variables in the system.

There is of course often a wide gap between the concepts in our theoretical models and the actual accounting rules that produce our time series data. But these conceptual gaps do not have the usual statistical properties of measurement error. There is no “true” model of the economy in which there is a single representative agent, a single good usable for consumption and investment, etc.. Our observed data are not generated by mistakes in measuring the variables in such a true model. The theoretical models we use are themselves crude approximations to a richer reality. The “measurement error” is for the most part in the models, not the data. If this is the actual situation, proceeding as if in fact the observed data were observations on the variables in the theoretical model, contaminated by random measurement error, can lead to very misleading results.

There are some cases, particularly in monthly data, where estimates of sampling error in macro data based on surveys are available. Often sampling methods generate known dynamics in the sampling error because of sample rotation. In cases where the sampling error is a substantial component of prediction error, we should be modeling it explicitly, and we do too little of this. But such cases are relatively rare. Artificially introducing measurement error as a way to pad out a theoretical model so it is no longer singular is a mistake. The Smets and Wouters approach, in which disturbances are for the most part economically interpretable changes in parameters describing behavior, seems a better starting point. It is not hard to believe that actual economies are subject to disturbances of many types, from many sources, and in fact results from these models bear this out. Business cycles are explained, in

the fitted models, by disturbances from many sources. We do not find the model fits best with two or three dominant sources of disturbance.

So the “too many shocks” criticism of the Smets-Wouters approach does not seem to me justified. A related criticism is the “too many adjustment costs” criticism. To fit the data well, Smets and Wouters needed to incorporate in their model many sources of inertia that other researchers had suggested and investigated. It is not hard to believe that the actual economy has many sources of inertia. The difficulty is rather that there seem to be in principle many choices available as ways to model inertia, that the different approaches may have different policy implications, and that neither the data nor economic common sense gives us much basis for choosing how to model inertia. Inertia from adjustment costs is likely to imply much higher costs of fluctuations than inertia from information-processing delays, for example. Our uncertainty about sources of inertia is a good reason to take the policy implications of such models with a grain of salt, especially when the models are used to generate conclusions about optimal policy based on expected utility of the representative agent or agents.

However, the appropriate response to worries about sensitivity of results to inertia assumptions is not to use models with fewer sources of inertia that (therefore) fit poorly. The appropriate response is to develop an array of models, or one model with an array of options for modeling inertia, that let us explore the extent to which the data can or cannot pin down the nature and source of inertia.

The methods used by Smets and Wouters to estimate their models and assess their fit are relatively new to economics. Their properties when applied to models at the Smets-Wouters scale and higher remain incompletely tested. The Markov Chain Monte Carlo method they use is iterative, and while it is guaranteed to converge (under some regularity conditions), there is no foolproof way to assess its convergence. Applications of these methods in econometrics and finance have in some cases paid too little attention to assessing convergence. These issues are particularly difficult when the methods are used to compare models. Models with any element of simultaneity, which will include most economic DSGE models, are likely to have zeroes at isolated places in the likelihood, and this makes for difficulties in assessing posterior probabilities on models (Sims, Waggoner, and Zha, 2005). The appeal of these methods is that they allow us to estimate models and assess their fit, while no other approach can do so in an internally consistent way. Nonetheless we will require more experience with these methods before they can be a routine part of policy modeling.

DSGE models do not easily match the fit of structural VAR's. The first Smets-Wouters paper showed their model apparently slightly surpassing the fit of Bayesian VAR's, but this was in the context of their having preprocessed the data to remove means and trends, which tilted the results in favor of their DSGE. Efforts to compare fits of DSGE's and structural VAR's that pay careful attention to convergence and do not treat detrended data as if it were unprocessed have often found that the DSGE's

are rather far from the structural VAR's in fit. This should not be a permanent problem. There is no reason in principle that DSGE's with flexible structural dynamics should not eventually fit as well as BVAR's, but for the time being BVAR's have an important role to play as a standard of fit.

Finally, in our enthusiasm for having "story-telling" structural models that actually fit the data, we need to bear in mind that the stories these models tell are false. There is no aggregate capital stock, there is no composite consumption good, not everyone continuously dynamically optimizes, there is no one-dimensional index of technology, and there is no permanent steady-state growth path to which we are certain to converge. These are just a few of the important falsehoods DSGE models are generally based on. This does not mean that story-telling models are not useful; but even when they fit just as well as structural VAR's or similar models that minimize detailed assumptions and thereby forego easy story-telling, we will want to have the less restricted models around as checks on the stories we are telling.

V. CONSTRUCTIVE SUGGESTIONS

Though the remarks above are mostly critical, policy-modeling is likely to be making substantial progress in the near future. The Bayesian Monte Carlo methods whose practicality was demonstrated by Smets and Wouters are likely to be used widely. I hope and expect that the methods will be applied, especially by academic macroeconomists, to models of varying types — with different choices about aggregation, different modeling of sources of inertia, different levels of detail in modeling financial markets, for example. The Smets and Wouters example ought to lead not

to a convergence toward one type of model, but to a proliferation of models that fit well, allowing us to compare theoretical approaches with a new level of discipline from the data.

As I noted above, story-telling models will continue to make detailed assumptions that are known to be only approximate. There will therefore continue to be a role for weakly structured models, like Bayesian structural VAR's (SVAR's). Those models can themselves be made more useful by better incorporating economists' beliefs in the long run stability of the economy, so that projections from SVAR's do not easily produce implausibly wide error bands or explosive long run central forecasts. One recent and widely cited approach to accomplishing this is that of DelNegro and Schorfheide (2004), who use a story-telling type of DSGE model to generate a prior on the coefficients of an SVAR model. It is also possible to take a more straightforward approach, if the aim is simply to induce more stable long run behavior. One can weight the likelihood to penalize unreasonable long run behavior, and this can be given a consistent interpretation as implementing a Bayesian prior. The basic idea is explained in notes available at sims.princeton.edu/yftp/DummyObs. There are technical difficulties in interpreting results obtained with such penalties when models are being compared, but the difficulties are manageable.

Ideally, model comparison and combining of model results can be carried out in a consistent Bayesian framework, using odds ratios on models. This requires, though, that all the models being given serious consideration have been brought to the point

of fitting the data well. As matters stand now, many or most models contain conventional, ad hoc simplifying assumptions to make them manageable. They have not been brought to the point of fitting well, but we have in mind that with a lot of additional work it might be possible to do so. If Bayesian odds ratio methods are applied to a collection of such models, they will tend to put heavy weight on the best-fitting model, even though the differences in fit among the models largely reflect ad hoc simplification that do not seem substantively important. This is not to say that the odds ratios calculations should not be done. Models that fit well are possible, and we will discover how to construct them only by assessing fit, even when the assessment brings discouraging news. There seems to be reason to hope that within a decade or so we will find it normal that competing models all fit well. But in a situation like the present one at most policy institutions, when we are comparing substantively different models whose fit differs for reasons we expect arise from ad hoc simplifications, not the substantive differences, we cannot mechanically apply odds ratios to assess model uncertainty.

Since we are likely to continue for some time, or perhaps indefinitely, to use multiple models, and for the time being they may not fit well, we need to consider how to compare them and to combine or choose from the results they give us. One approach would be to let structural VAR's play the role of "base model" in the framework of Schorfheide (2000). Schorfheide discusses the situation where one would like to use a model or models that are known not to be correct because the base model fits better. His conclusion is that in this case the choice of model and the characterization

of uncertainty about the results should involve evaluating the model in use based on the probability distribution implied by the base model. In practice, a rough way to implement this would be to check that projections from the workhorse model lie within VAR-based error bands. Of course this requires that the type of projection being made is valid within the base model. Evaluating the effects of a range of possible choices for a monetary policy path would then require a structural VAR in which the monetary policy equation or block is correctly identified. The reason for using the workhorse model instead, if it did not conflict too sharply with the VAR, would be the more detailed story it provides about the causal chain connecting monetary policy to other variables in the system. If we eventually progress to the point where the large, main, policy model is a Bayesian DSGE, that model might play the role of base model, while smaller models that can be more quickly manipulated for making projections in real time, play the role of the less-well-fitting alternatives.

VI. CONCLUSION

It is a little discouraging that the biggest recent model reform, BEQM, which was worked on in part after the Smets and Wouters proof of concept, represents possibly the most complete turning away from probability-based inference of any large central bank model. On the other hand, there is apparently interest at the Bank of England in attempting a probability-based approach to BEQM or a variant of it, and there is active research toward building probability-based models at many central banks and other policy institutions. Developments in computational power and in

statistical and economic theory seem to be coming together to promise a period of rapid progress in policy modeling.

REFERENCES

- ALVAREZ-LOIS, P., R. HARRISON, L. PISCITELLI, AND A. SCOTT (2005): "Taking DSGE models to the policy environment," Discussion paper, Bank of England.
- DELNEGRO, M., AND F. SCHORFHEIDE (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.
- DOEPKE, M., AND M. SCHNEIDER (2004): "Real Effects of Inflation through the Redistribution of Nominal Wealth," Discussion paper, UCLA, NYU.
- EDGE, R. M., M. T. KILEY, AND J.-P. LAFORTE (2005): "An Estimated DSGE Model of the US Economy with an Application to Natural Rate Measures," Discussion paper, Board of Governors of the Federal Reserve.
- HAAVELMO, T. (1944): "The Probability Approach in Econometrics," *Econometrica*, 12(supplement), iii–vi+1–115.
- HARRISON, R., K. NIKOLOV, M. QUINN, G. RAMSAY, A. SCOTT, AND R. THOMAS (2005): *The Bank of England Quarterly Model*. Bank of England.
- SCHORFHEIDE, F. (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15(6), 645–670.
- SIMS, C. A. (1987): "A rational expectations framework for short-run policy analysis," in *New approaches to monetary economics*, ed. by W. A. Barnett, and K. J. Singleton, pp. 293–308. Cambridge University Press, Cambridge, England.

——— (2001): “Fiscal Consequences for Mexico of Adopting the Dollar,” *Journal of Money, Credit, and Banking*, 33(2,part2), 597–616.

——— (2002): “The Role of Models and Probabilities in the Monetary Policy Process,” *Brookings Papers on Economic Activity*, 2002(2), 1–62.

SIMS, C. A., D. F. WAGGONER, AND T. ZHA (2005): “General Methods for Restricted Markov-Switching Models with Independent State Variables,” Manuscript in Preparation for the *Econometric Review* special issue on Bayesian Methods in Dynamic Economies.

SMETS, F., AND R. WOUTERS (2003): “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area,” *Journal of the European Economic Association*, 1, 1123–1175.

DEPARTMENT OF ECONOMICS, PRINCETON UNIVERSITY

E-mail address: sims@princeton.edu