Commentary

Jon Faust

The Economic Policy conference at the Federal Reserve Bank of St. Louis has for several decades been one of the premier monetary policy conferences worldwide, and it is a great privilege to participate in this conference focusing on the measurement and forecasting of potential growth. I am particularly pleased to be discussing the paper by Christophe Cahn and Arthur Saint-Guilhem (2009), which is a beautiful example of a broad class of work that explores how traditional economic concepts and measures relate to similar concepts in the context of dynamic stochastic general equilibrium (DSGE) models that are rapidly coming into the policy process. This class of work is vitally important if policymakers are to meld successfully traditional methods and wisdom with the new models to improve the policy process. I will mainly attempt to explain this class of work, why it is important, and some techniques for improving it. While my points are fairly generic, the Cahn–Saint-Guilhem paper provides an excellent case study for illustrating the key issues.

DSGE MODELS AND A NEW CLASS OF RESEARCH

Around 1980, Lucas, Sims, and others issued devastating critiques of existing monetary policy models. One basis for these critiques was the claim that the existing methods were substantially ad hoc relative to the ideal to which the profession should aspire. While this critique was undeniably valid, the absence of better-founded alternatives meant that more-or-less traditional ad hoc approaches continued to be used and refined at central banks for the next 25 years or so. Meanwhile, the profession did the basic research required to create models with sounder foundations.

In the past few years, DSGE models have advanced to the point that they are coming into widespread use at central banks around the world. These models are still rife with ad hoc elements, but there is no doubt that there has been an order of magnitude advance in the interpretability of the predictions of the model in terms of well-articulated economic theory.

There is still considerable disagreement, however, over the degree to which the new models should supplant the traditional methods. I do not want to argue this point. Rather, I want to assert that these models have at least advanced to the point that they constitute interesting laboratories in which to explore various claims and principles that are important in the policy process. My focus is on how the models can best play this role.

Consider an analogy to medical research. In attempting to understand the toxicology of drugs in humans, we often use animal models. That is, we check if the drug kills the rat before we give it to humans. In any given pharmacological context, there is generally substantial disagreement on how literally we should take the model when extrapolating the results to humans. Despite this

Jon Faust is the Louis J. Maccini Professor of Economics at Johns Hopkins University.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 241-46.

disagreement, there is a broad consensus that the rat model is extremely valuable in formulating policy.

Similarly, I think we should all agree that DSGE models have at least attained something akin to rat, or at least fruit fly, status. Under this agreement, a wide range of work becomes valuable and important. In particular, I think we should aggressively explore basic macroeconomic propositions treating these model economies as interesting economic organisms.

Although I am not sure the authors view it this way, the Cahn–Saint-Guilhem paper can be viewed in this perspective. Some notion of potential output is often at the center of policy discussions. One traditional measure of potential is based on the production function approach (PFA) as clearly described in the paper. Analysis of optimal policy in DSGE models suggests that for some purposes we should focus on a concept of potential as measured by the *efficient* level of output, known as flexible price output (FPO). FPO potential measures what output would be if certain distortions were not present.[1]

If we are to smoothly and coherently bring DSGE models and the associated measures into the policy process, it is important to know how PFA and FPO potential relate in the real world. One very useful step in this process, I argue, is exploring how both concepts operate in the simpler context of the DSGE model. That is, first understand the concepts as fully as possible in the rat before moving to the human case.

This type of work is relatively straightforward conceptually. Broadly, we must specify how to compute a model-based analog of both PFA and FPO potential. Then we simulate a zillion samples from the model, calculate both measures on each, and then summarize apparent similarities and dissimilarities.[2] For example, we might ask whether our traditional interpretation of PFA potential is correct in the context of the DSGE model.

The paper focuses on a particular question of this type. Movements in PFA potential are, in practice, often attributed to medium-term "structural features" of the economy as opposed to transitory demand or supply features. Is the interpretation warranted in the DSGE model? The paper finds (see their Table 3) that it is not. That is, a large portion of the variance of PFA potential is attributable to factors we would not usually consider "structural" in the sense this term is used in these discussions. FPO potential looks more structural in this regard. The paper elaborates this key result in a number of useful ways. What I want to discuss, however, is what we should make of this general class of work and how we can we make it better.

Let me note that this sort of work is multiplying. For example, I have been involved in a long line of work regarding the reliability of structural inferences based on long-run identifying restrictions in vector autoregressions (Faust and Leeper, 1997). At a presentation of my work with Eric Leeper in the early 1990s, Bob King asked why I did not assess the practical importance of the points using a DSGE model. I did not see the full merits of this at that time, but Erceg, Guerrieri, and Gust (2005) and Chari, Kehoe, and McGrattan (2008) have now taken up this suggestion (illustrating far more points than raised in the Faust-Leeper work) and considerably advanced the debate.

I would go so far as to argue that this sort of analysis should be considered a necessary component of best practice. That is, if anyone proposes a macroeconomic claim or advocates an econometric technique that is well defined in the new class of DSGE models, assessing the merits of the claim in the DSGE context should be mandatory. If it is coherent to apply the idea in the rat, we should do so before advocating its use in humans.

The work I am advocating cannot, however, be seen as part of some necessary or sufficient conditions for drawing *reliable conclusions* about reality. The mere fact that a particular claim is warranted in the DSGE model is neither necessary nor sufficient for the claim to be useful in practice. Similarly, the mere fact that a drug does not kill the rat is neither necessary nor sufficient for the drug's safety in humans. Just as judgment is required to draw lessons from animal studies, judgment will be required to draw lessons from

[1] Of course, many important issues remain in assessing this counterfactual, but these issues will not be important in this discussion.

[2] I add one important conceptual step in the discussion below.

DSGE studies. I believe that the results can be valuable nonetheless. In the remainder of the discussion, I highlight three points that, I believe, can make work in this spirit much more useful.

DOING IT BETTER

Don't Confuse the Rat with the Human

In animal studies, there is very rarely any confusion about when the authors are talking about the rats and when they are talking about the humans. The core of the research paper rigorously assesses some feature of the toxicology in rats and is clearly about the rat. Whatever one believes about the usefulness of the rat model, the point of the body of the paper is to support claims about the rat. This portion of the paper can be rigorously assessed without getting into unresolved issues about the ultimate adequacy of the rat model.

After settling issues about the rat, there is an active discussion about how the rat model results should be extrapolated to the human context.[3] This process is illustrated in the conclusions of a joint working group of the U.S. Environmental Protection Agency and Health Canada regarding the human relevance of animal studies of tumor formation (Cohen et al., 2004). They summarized their proposed framework for policy in the following four steps:

(i) Is the weight of evidence sufficient to establish the mode of action (MOA) in animals?

(ii) Are key events in the animal MOA plausible in humans?

(iii) Taking into account kinetic and dynamic factors, is the animal MOA plausible in humans?

(iv) Conclusion: Statement of confidence, analysis, and implications. (p. 182)

In the first step, we clarify the result in the model. The remaining steps involve asking serious questions about whether the transmission mechanisms in the model—to borrow a monetary policy term—plausibly operate similarly in the relevant reality.

In contrast, it is customary in macroeconomics to discuss quantities computed in the context of a DSGE model in a way that leaves it ambiguous at best whether the authors are advocating (or hoping) that we take them as statements about reality. I suspect that researchers arrive at the practice of treating statements about the model and reality as more-or-less equivalent under the rubric of "taking the model seriously." This seems to presume that the best way to take the model seriously is to take it literally. In toxicology, there is no doubt that policymakers take animal models seriously, but this never seems to require equating rats and humans. In my view, we should not confuse rats and humans; neither should we confuse DSGE models and reality.

Conceptual Clarity Before Computation

Broadly speaking, the point of the Cahn–Saint-Guilhem paper is to compare and contrast the behavior of two measures of potential output using a computational exercise on a DSGE model. Because it is so conceptually simple to implement computational experiments of the sort described above, it is very tempting to jump straight to the computer. I think work of this type would be better clarified by starting with careful conceptual analysis of the measures *before* computation. We can clearly lay out the expected differences and then many aspects of the computational work become exercises in measuring the empirical magnitude of effects that have been clearly defined.

I think this is particularly important in the macro profession where we seem to have a penchant for reusing labels for concepts that are quite distinct. "PFA potential" and "FPO potential" illustrate this point. "PFA potential" is meant to measure the level of output that would be attained if the current capital stock were used at some notion of "full capacity." "FPO potential" is, roughly speaking, the level of output that would be obtained if inputs were used *efficiently* as opposed to *fully*. It is clear that these two concepts of potential need not even be closely correlated. In any model in which the *efficient* level of, say,

[3] See Faust (2009) for a more complete discussion of this issue.

labor fluctuates considerably around the *full employment* level of labor, the two measures may be quite different. Clearly laying out the conceptual differences can be an incredibly enlightening step in what ultimately becomes a computational exercise.

One minor critique of the Cahn–Saint-Guilhem paper in this regard is that the work refers to FPO potential as simply the DSGE measure. There are many concepts of "potential" that might be useful for different questions in a DSGE model, and indeed we can discuss many versions of FPO potential, depending on how we implement the counterfactual regarding "if prices were flexible." Specific labels and careful analysis of the associated concepts can be very helpful.

Used properly, the sort of computational exercises with DSGE models that I am advocating can be an important tool for clarifying important conceptual issues. It may, at times, be tempting to simply substitute the relatively straightforward computational step for the sometimes painful step of careful conceptual analysis. Giving in to this temptation would be to miss an important opportunity.

Better Lab Technique

While the computational exercises I am advocating are conceptually straightforward, there are myriad subtle issues that fall under the umbrella of "lab technique." The new DSGE models are complicated and not fully understood. The Bayesian techniques being developed to analyze these models are also complicated and not fully understood. What we know from experience to date with DSGE models, and with similar tools applied in other areas, is that we can very easily create misleading results. For example, Sims (2003) has discussed such issues at length.

Much of the profession has long experience with the use of frequentist statistics and has become familiar with the myriad ways that one might inadvertently mislead. We need to be mindful of the fact that the profession is very new at assessing the adequacy of the new DSGE models using Bayesian techniques.

John Geweke (2005, 2007) has been at the forefront in developing flexible Bayesian tools for assessing model adequacy in the context of models that are known to be incomplete or imperfect descriptions of the target of the modeling. Abhishek Gupta (a Johns Hopkins graduate student) and I have recently been exploring these methods as they apply to DSGE models intended for policy analysis (Faust 2008, 2009; Faust and Gupta, 2009; and Gupta, 2009). I present just a flavor of one result with possible bearing on the topic of the Cahn–Saint-Guilhem analysis. The example is from Faust (2009), which reports results for the RAMSES model, a version of which is used by the Swedish Riksbank in its policy process.[4]

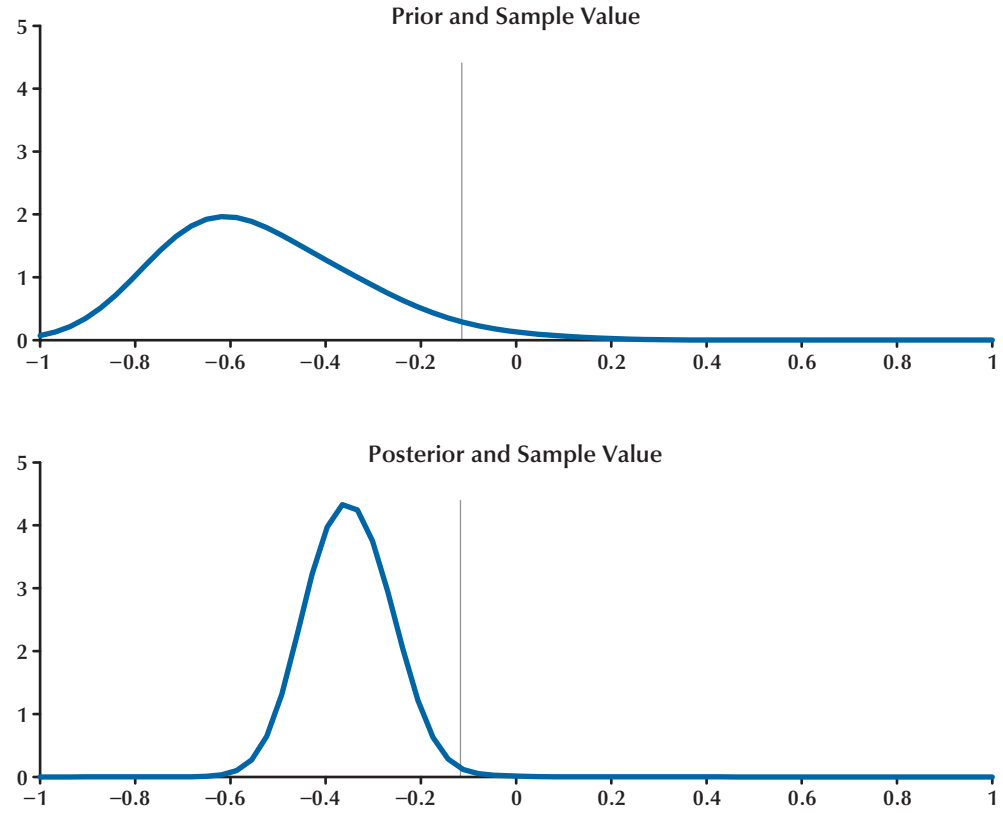
The simplest form of the idea is to take some feature of the data that is well defined outside the context of any particular macroeconomic model and about which we may have some prior beliefs. In the simplest form, we simply check whether the formal prior (which is largely arbitrary in current DSGE work) corresponds to our actual prior regarding this feature. Further, we check how both the formal prior and posterior compare with the data. A somewhat subtler version of this analysis instead considers prior and posterior predictive results for these features of interest.

As an example, consider the correlation between consumption growth and the short-term, nominally risk-free interest rate. Much evidence suggests that there is not a strong relation between short-term fluctuations in short-term rates and consumption growth. The upper panel of Figure 1 shows this marginal prior density implied by the prior over the structural parameters used in estimating RAMSES. The prior puts almost all mass on a fairly strong negative correlation, with the mode larger in magnitude than –0.5. The vertical line gives the value on the estimation sample of approximately zero. In short, the prior used in the analysis strongly favors the mechanism that higher interest rates raise saving and lower consumption. In the posterior (bottom panel), the mass is moved toward a negative correlation that is a bit smaller in magnitude, but the sample value is actually farther into the tail than it was in the prior.

[4] The developers of RAMSES (Riksbank Aggregated Macromodel for Studies of the Economy of Sweden) were exceedingly generous in helping me conduct this work.

Figure 1

Prior and Posterior Densities



NOTE: The figure shows the prior (upper panel) and posterior (lower panel) 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.

SOURCE: Author's calculations using computer code provided by Riksbank.

This result and related ones in the work cited above convince me that current DSGE models continue to have difficulty matching basic patterns in consumption and investment as mediated by the interest rate. If we were to use this model in policy, we might want to ask whether this is one feature—like differences between rats and humans—that we should explicitly adjust for in moving from model results to reality.

Of course, the forces driving short-run fluctuations in consumption are at the very center of the distinction between PFA potential and FPO potential. These results and others like them convince me that while the DSGE model provides an interesting lab, there is good reason to question how literal we should be in extrapolating these results to the real economy.

A more general lesson is that the methods just sketched can be applied to any data feature, including statistics like those reported, for example, in Table 4 in the Cahn–Saint-Guilhem paper. These techniques allow one to coherently take estimation and model uncertainty into account and to evaluate the importance of arbitrary aspects in the formal prior. I strongly urge the authors to move in this direction.

CONCLUSION

I commend the St. Louis Fed for holding a conference on this issue that is vital to the monetary policymaking process, and I commend Christophe and Arthur for their interesting work illuminating how two competing measures of potential output behave in the context of modern DSGE models. This line of work is extremely important. I have made three suggestions that I believe would improve any work of this type. I hope that these suggestions contribute to making work of this sort even more influential.

REFERENCES

Cahn, Christophe and Saint-Guilhem, Arthur. "Issues on Potential Growth Measurement and Comparison: How *Structural* Is the Production Function Approach?" Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4) pp. 221-40.

Chari, V.V.; Kehoe, Patrick J. and McGrattan, Ellen R. "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?" NBER Working Paper No. 14430, National Bureau of Economic Research, October 2008; www.nber.org/papers/w14430.pdf.

Cohen, Samuel M.; Klaunig, James; Meek, M. Elizabeth; Hill, Richard N.; Pastoor, Timothy; Lehman-McKeeman, Lois; Bucher, John; Longfellow, David G.; Seed, Jennifer; Dellarco, Vicki; Fenner-Crisp, Penelope and Patton, Dorothy. "Evaluating the Human Relevance of Chemically Induced Animal Tumors." *Toxicological Sciences*, April 2004, *78*(2), pp. 181-86.

Erceg, Christopher J.; Guerrieri, Luca and Gust, Christopher. "Can Long-Run Restrictions Identify Technology Shocks?" *Journal of the European Economic Association*, December 2005, *3*(6), pp. 1237-78.

Faust, Jon. "DSGE Models in a Second-Best World of Policy Analysis." Unpublished manuscript, Johns Hopkins University, 2008; http://e105.org/faustj/download/opolAll.pdf.

Faust, Jon. "The New Macro Models: Washing Our Hands and Watching for Icebergs." *Economic Review*, March 23, 2009, *1*, pp. 45-68.

Faust, Jon and Gupta, Abhishek. "Bayesian Evaluation of Incomplete DSGE Models." Unpublished manuscript, Johns Hopkins University, 2009.

Faust, Jon and Leeper, Eric M. "When Do Long-Run Identifying Restrictions Give Reliable Results?" *Journal Business and Economic Statistics*, July 1997, *15*(3), pp. 345-53.

Geweke, John. *Contemporary Bayesian Econometrics and Statistics*. Hoboken, NJ: Wiley, 2005.

Geweke, John. "Bayesian Model Comparison and Validation." Unpublished manuscript, University of Iowa, 2007; www.aeaweb.org/annual_mtg_papers/2007/0105_0800_0403.pdf.

Gupta, Abhishek. "A Forecasting Metric for DSGE Models." Unpublished manuscript, Johns Hopkins University, 2009.

Sims, Chris. "Probability Models for Monetary Policy Decisions." Unpublished manuscript, Princeton University, 2003.