DSGE Models: I Smell a Rat (and It Smells Good)*

Jon Faust
Johns Hopkins University and NBER

DSGE models have now reached a point where they can and do serve an important role in the monetary policy process. From the standpoint of real-world policymaking, however, there remain important areas of omission and coarse approximation in these models. I argue that macroeconomics should follow other fields such as toxicology in having a formal literature on how best to use models that are far from perfect as a basis for public policy.

JEL Codes: E2, E3, E6.

1. Introduction

I was very pleased to be invited to participate in the Third Annual Fall IJCB Conference on Monetary Policy Issues in Open Economics. When the IJCB was initially conceived, I was a Federal Reserve economist. From the outset of my years at the Federal Reserve, I observed and participated in much research that was, in my view, *serious* but that I understood was not publishable in a *serious* journal. I learned this fact of life early on from senior colleagues such as Dale Henderson and Dave Gordon. This never made much sense to me, however, and when the IJCB was founded, I hoped that it would become a place where a high-level conversation emerged about academic work and its relation to actual practice.

While the IJCB has become a very good journal, it has not, for the most part, played the role I had hoped for. But this was my fan-

*I would like to thank Abhishek Gupta, Eric Leeper, Andy Levin, Jonathan Wright, and all the participants at the IJCB conference for valuable comments. Author contact: Tel: (410) 510-7614; E-mail: faustj@jhu.edu; Web Site: http://e105.org/faustj.

tasy and not necessarily that of the journal management. Nonetheless, I will use my comments to make a case for a literature I believe we need and a literature where the IJCB could be in the vanguard.

The two outstanding papers in the session I was asked to discuss are "Is Exchange Rate Stabilization an Appropriate Cure for the Dutch Disease?" by Ruy Lama and Juan Pablo Medina of the International Monetary Fund and "The Performance of Monetary Policy Rules in a Multi-Sector Small Open Economy," by Carlos de Resende, Ali Dib, and Maral Kichian of the Bank of Canada.[1] These are each excellent examples of the use of modern dynamic stochastic general equilibrium (DSGE) models to answer important policy questions. The Dutch disease is a hearty perennial and has recently been an issue for the Bank of Canada. The key question for central banks is whether the blunt tool of monetary policy can productively be used to improve overall welfare when, following a commodity price shock, one portion of the economy is booming and another is suffering. Lama and Medina reach the very sensible conclusion that monetary policy is probably not the right tool.

The de Resende, Dib, and Kichian paper examines what measure(s) of prices should be the object of a central bank's price stability efforts. Our simple models give one appealing lesson. If stickiness of certain prices is the main source of inefficiency in the economy, then one would probably want to run policy so as to render the efficient path of the stickiest prices as smooth as possible—stickiness will be largely irrelevant if the first-best price does not move much. Even if we accept this insight, important practical questions remain. What would be the ideal price index to target? Given the need for public policy that can be explained and implemented in a clear manner, how close could we come to the ideal using some standard price index? The authors find that focusing on the CPI works well for Canada.

Both papers deal with important policy problems and, more importantly, problems that are shot through with dynamic and general equilibrium issues. It would be very difficult to analyze either question coherently without some comprehensive model.

As is the custom in papers of this sort, the authors draw concrete practical conclusions for central banks. Lama and Medina

[1]The paper by de Resende, Dib, and Kichian was presented at the conference but is not published in this issue.

(2010, p. 26) argue, "This paper provides theoretical support for the current policy of the Bank of Canada of not intervening in the foreign exchange market." De Resende, Dib, and Kichian similarly conclude (2011, p. i), "Our welfare analysis provides support for CPI-based targeting policies by the Canadian monetary authority." These statements trouble me.

Statements like these are the norm in policy-oriented macro, but since those early days at the Federal Reserve, I have found myself wondering if their authors are serious. I understand that the model in question may provide an *example* of an economy consistent with the conclusions. This is a necessary condition for supporting the claim, but it is not sufficient. Of course, I am not alone in being troubled in this regard. For example, Robert Solow (2010) recently argued:

> Especially when it comes to matters as important as macroeconomics, a mainstream economist like me insists that every proposition must pass the smell test: does this really make sense? I do not think that the currently popular DSGE models pass the smell test.

As my title hints, I too have been doing some smell testing in the DSGE arena. Like Solow, I smell a rat, but I think I differ from him in that I find that rat smell refreshing in some respects. But even more, I think we can do better than smell testing. I think we need a serious literature on questions such as, What warrants taking an academic result to the Board room? How can we best allow for the deficiencies in DSGE models in the policy process?

To emphasize that my discussion is not about criticizing the state of DSGE modeling, let us start with a brute assumption: for the foreseeable future, DSGE models will continue to have substantial areas of omission and coarse approximation that may be salient to the intended use in policymaking. I believe that this assumption is unassailable and will say no more about it.[2]

The point of this discussion, then, is to argue for a formal literature on best methods and practices for using materially flawed models in practical policymaking.

[2]I argue explicitly for this assumption in, e.g., Faust (2009) and Faust and Gupta (2011).

2. Lab Rats and DSGE Models

The necessity of making policy based on models that are far from ideal is not unique to macro. Policymakers regulating human exposure to potentially toxic chemicals face similar issues. Toxicology policy concerns a complicated dynamic general equilibrium system (the human) that is imperfectly understood. As with macro policy, experimentation with the actual target of policy is not considered acceptable. Policymakers have no better choice than to make policy based on results from models and whatever "natural experiments" on humans happen to be present in the data. In this arena, rats (and other mammals) are used as imperfect models of humans:

> Biological differences among species and the use of high experimental doses often make animal test results difficult to interpret with regard to human relevance. Despite the difficulties, animal studies have formed the cornerstone of toxicology and safety assessments. . . . Rodents are the most widely used animal models. (Lewis et al. 2008, p. 1)

The rat model is a fully microfounded, dynamic general equilibrium system. While the rat matches many features of the human for which it serves as a surrogate,[3] the list of omissions and coarse approximations could fill a book (e.g., Suckow, Weisbroth, and Franklin 2006).[4]

Despite these limitations, there is no doubt that the lab rat comprises an exquisite dynamic general equilibrium model. If we were to assess whether the lab rat or the standard macro model comes closer to the ideal of complete fidelity to important features of the target being modeled, I think the rat wins. Where the rat lacks, say, a gallbladder, the standard DSGE model lacks a housing sector and financial sector. On the relevant biological scale, the DSGE model

[3]As summarized by the National Human Genome Research Institute of the National Institutes of Health (2004), "[A]lmost all human genes known to be associated with diseases have counterparts in the rat genome and appear highly conserved through mammalian evolution, confirming that the rat is an excellent model for many areas of medical research."

[4]For example, rats have no gallbladder, and cannot vomit. As for coarse approximations, the rat's normal heart rate is about 300—even by macro standards a bad miss of the stylized fact in humans.

probably falls somewhere between the fruit fly and the rat, which makes DSGE models an impressive achievement.

Thus, when I apply Solow's smell test, I conclude that I am beginning to smell a rat. And it smells pretty good—at least relative to that fruit fly odor we had been living with.

The reason for raising the parallel with toxicology is to highlight a branch of the toxicological literature that has no equivalent in macro. In toxicology there is a vigorous literature regarding best methods and practices for formulating policy for humans using results regarding rats. This literature goes under names like "human relevance of animal studies" and "interspecies extrapolation."

For example, a joint working group of the U.S. Environmental Protection Agency and Health Canada conducted a detailed study of the human relevance of animal studies of tumor formation. Cohen et al. (2004, p. 182) summarize their framework:

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.

In the first step, we establish the result strictly in the model. This step the toxicologists share with macroeconomists. The remaining steps involve asking serious questions about whether the MOA in the model plausibly operates similarly in the model and the relevant reality. In monetary policymaking, we would probably say *transmission mechanism* rather than MOA, but the parallel is clear.

As Cohen et al. elaborate, in toxicology, there are both in vitro and in vivo studies, corresponding roughly to partial and general equilibrium studies, respectively.[5] There is a wealth of partial equilibrium (in vitro) studies regarding the effects of chemicals, say, on human tissue. But as macroeconomists know, partial equilibrium models are not enough:

[5]These terms are roughly taken to draw a distinction between lab work in the living being (in vivo) and other lab work.

> As our understanding of toxic mechanisms steadily increases, the role of in vitro methods in [toxicity testing] is obvious. However, a hazard assessment cannot easily be made without further knowledge of the compound's behaviour in the integrated system of an intact organism. Therefore, results obtained from in vitro studies in general, are often not directly applicable to the in vivo situation. (Vermeire et al. 2007, p. 257)

So toxicological policy advisers have a wealth of micro evidence but are unsure what this implies for general equilibrium. But the available general equilibrium models have important areas of omission and coarse approximation relative to humans. Finally, it is worth noting that there may be some epidemiological evidence on humans, which is the rough equivalent of aggregate time series work in macro.

In steps 2–4, the framework provides a guide to reconciling all these sources of information.

> The expanded framework takes into account both qualitative and quantitative aspects of the MOA, addresses the issue of the kinds of data that can be used in evaluating the MOA in humans, and provides a disciplined, transparent approach to comparing the key events of the MOA in laboratory animals to humans. (Cohen et al. 2007, p. 186)

A disciplined and transparent approach to assessing model relevance is just what I believe we should strive for in macro.

3. Real-World Relevance of DSGE Results

In macro, we generally perform step 1 of the framework, carefully establishing effects within the model. After doing this, the papers in the session follow the standard pattern of leaping straight to conclusions about much more complex organisms such as the functioning of the Bank of Canada in the world economy.

To take a more prominent example, Smets and Wouters performed the seminal work showing that modern DSGE models had reached the point of fitting certain data about as well as some well-respected benchmarks. This seems to have been a watershed event, marking the rapid adoption by central banks of models of this sort for use in policymaking. Smets and Wouters argued (2003, p. 1125):

> [Our results] suggest that the current generation of DSGE models with sticky prices and wages is sufficiently rich to capture the time-series properties of the data, as long as a sufficient number of structural shocks is considered. These models can therefore provide a useful tool for monetary policy analysis.

While the first sentence is arguably correct, the final sentence is a non sequitor. On what basis did achieving a reasonable fit to seven macro variables become a sufficient condition a tool to be "useful for policy analysis"? Of course, both Lucas and Sims rejected this view. Sims argues that we must think seriously about whether the MOA in the model is *actually* something functioning in reality:

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

Lucas argued that fit was irrelevant unless one gets *deep* aspects of the MOA right:

> More particularly, I shall argue that the features which lead to success in short-term forecasting are unrelated to quantitative policy evaluation. (1981, p. 105)

In conventional macroeconomics seminars, the practice of making the most naïve possible extrapolation from simple model to a much more complex reality is sometimes signaled by the statement that we are going to "take the model seriously." At this point, if one criticizes the realism of the model, one might hear the response "it takes a model to beat a model."

I wonder how this would go in toxicology. In my reading of the literature, there is never any doubt about taking the rat *seriously*, but no *serious* journal would accept naïve extrapolation in the name of *seriousness*. Similarly, in response to criticism about the adequacy of the rat models, I suspect one does not hear, "it takes a rodent to beat a rodent."

What I am advocating is that we firmly establish results in a model and then draw conclusions for reality in light of the DSGE model results, which help us understand general equilibrium effects

in a simple setting, and in light of all that is known from microeconomic studies (some of which we cannot yet reconcile in DSGE models), and in light of what is known from broad descriptive evidence regarding the economy in the wild. None of these alone is sufficient. This kind of work is difficult but, in my view, is an essential step in serious policymaking.

4. Related Arguments

The issues raised here overlap to a significant degree with the more generic argument about the "accept/reject" mentality that sometimes comes out of the simple-minded application of hypothesis testing. As Tiao and Xu (1993) argue, "[We should seek] development of diagnostic tools with a greater emphasis on assessing the usefulness of an assumed model for specific purposes at hand rather than on whether the model is true." In economics, Bruce Hansen (2005) makes a similar argument.

Riccardo Caballero (2010) has recently written an important critique of DSGE modeling in which he assails what he labels "the pretense of knowledge." My starting point—that we need to take seriously that there are important areas of omission and coarse approximation—is very much in line with Caballero's. Caballero then goes on to make thoughtful and wide-ranging comments on how to improve on our state of knowledge.[6]

My argument is more mundane. I take the current state of knowledge as given and propose we put some resources into how to make the most responsible use in policymaking of what we now know.

5. Strengths and Weaknesses of DSGE Models

Let me mention briefly some work I have been doing with a former graduate student, Abhishek Gupta (Gupta 2010; Faust and Gupta 2011). Our goal is to create better tools intended to reveal and highlight the strengths and weaknesses of DSGE models from the standpoint of practical policy analysis.

[6]On this count, I think there is much legitimate disagreement.

At central banks, if you ask why one needs a structural model as opposed to simply needing a reduced-form forecasting model, one often gets the practical response that structural models are needed to help "tell a story" about what is going on. An informative way to assess whether what is going on in DSGE models is a plausible account of actual economic fluctuations is to scrutinize the structural story that the DSGE model tells for the historical sample.

One version of this idea has long been standard in the DSGE literature. Researchers regularly present historical decompositions of observed output variation in terms of estimated structural shocks. For example, figure 1 is the historical decomposition of U.S. growth implied by the Smets-Wouters model as reported in Smets and Wouters (2007).

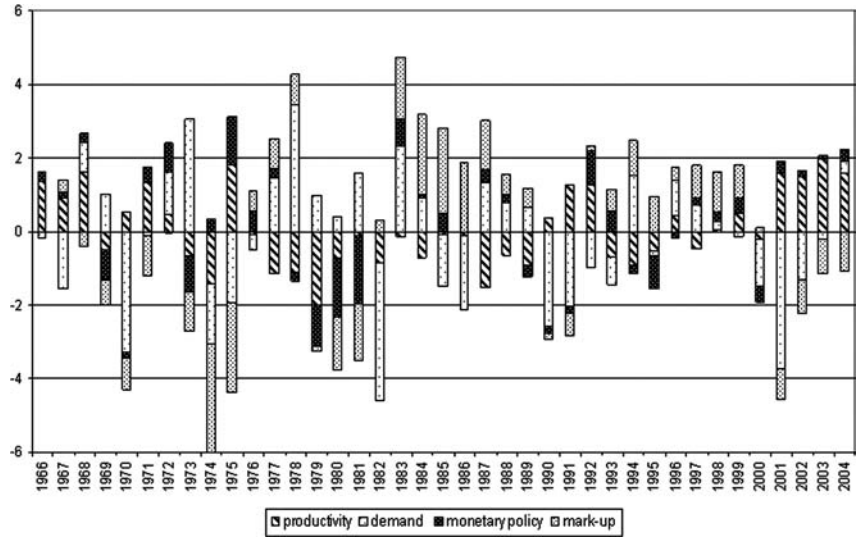
While these historical decompositions are a depiction of the causal story the models imply for the sample, they give no indication of whether this story is at all plausible—even from the standpoint of the model. It turns out that the estimated structural shocks in figure 1 have some very peculiar sample properties.[7] They consistently have a much larger standard deviation and a different pattern of correlation in recessions than in expansions. Of course, in population the shocks are uncorrelated and homoskedastic. The patterns the model deduces in the sample are, from the standpoint of the model, extremely freakish.

Overall, we find that if one takes the model *seriously*, the main story the model tells of the post-War recession is that recessions were repeated instances of the same perfect storm. For central bankers, the main lesson regarding recessions would be to prepare for something else, because the recessions experienced repeatedly in this sample will almost certainly never be seen again.

Gupta and I discuss a different interpretation we take more seriously. These results are teaching us about important differences between the MOA or causal channels in the model and those that operate in reality. If so, there could be important policy implications.

[7]Formally, we document results like this using a tool related to posterior predictive analysis (e.g., Geweke 2005). In particular, it is a version of discrepancy analysis as proposed by Gelman, Meng, and Stern (1996). While the decomposition in the figure is conditional on a point estimate of the model parameters, these methods take full account of uncertainty in the parameter estimates.

Figure 1. Historical Decomposition of U.S. Output Growth in Terms of the Structural Shocks in the Smets-Wouters Model



Source: Smets and Wouters (2007).
Notes: The seven shocks have been averaged over calendar years and summed across four categories. The productivity category includes the general and investment-specific technology shocks; the demand category includes the risk premium and government spending shocks; the markup category includes the price and wage markup shocks; and there is one monetary policy shock.

6. Conclusion

The profession and central banks can benefit greatly from papers like those of Lama and Medina and of de Resende, Dib, and Kichian. More generally, I believe that the current crop of DSGE models, like lab rats, are magnificent models. For the profession, they represent a titanic achievement—with all the images that adjective may bring to mind.

If macro is a serious applied science, then we should have a formal discussion on how best to use DSGE models in light of all the disparate sources of information we have and in light of the fact that our best general equilibrium models still have potentially

important areas of omission and coarse approximation. From this discussion should emerge an evolving standard of best practices in "model-to-world" extrapolation.

I started by alluding to the fact that central bank staffs have long conducted this kind of analysis but that this work has been almost entirely under cover from the standpoint of the academic literature. Both central banks and the macro field more generally would benefit from the open, analytic, and vigorous discussion that is the hallmark of a healthy academic literature. The IJCB is the ideal place for this literature to grow.

References

Caballero, R. J. 2010. "Macroeconomics after the Crisis: Time to Deal with the Pretense-of-Knowledge Syndrome." *Journal of Economic Perspectives* 24 (4): 85–102.

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, and D. Patton. 2004. "Evaluating the Human Relevance of Chemically Induced Animal Tumors." *Toxicological Sciences* 78 (2): 181–86.

de Resende, C., A. Dib, and M. Kichian. 2011. "The Performance of Monetary Policy Rules in a Multi-Sector Small Open Economy." Manuscript, Bank of Canada.

Faust, J. 2009. "The New Macro Models: Washing Our Hands and Watching for Icebergs." *Economic Review* (Sveriges Riksbank) 2009 (1): 45–68.

Faust, J., and A. Gupta. 2011. "Posterior Predictive Analysis for Evaluating DSGE Models." Manuscript, Johns Hopkins University.

Gelman, A., X. Meng, and H. Stern. 1996. "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies." *Statistica Sinica* 6 (4): 733–59.

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

Gupta, A. 2010. "A Forecasting Metric for Evaluating DSGE Models for Policy Analysis." Manuscript, Johns Hopkins University.

Hansen, B. 2005. "Challenges for Econometric Model Selection." *Econometric Theory* 21 (1): 60–68.

Lama, R., and J. P. Medina Guzman. 2010. "Is Exchange Rate Stabilization an Appropriate Cure for the Dutch Disease?" IMF Working Paper No. 10/182.

Lewis, A. S., L. A. Beyer, C. J. Langlois, C. J. Yu, and A. D. Wait. 2008. "Considerations in Toxicology Study Design and Interpretation: An Overview." Manuscript, IASC. Available at http://www.iasc.org/08_0814_IASC_Article_References.pdf.

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

National Human Genome Research Institute, National Institutes of Health. 2004. "Scientists Compare Rat Genome with Human, Mouse." Press Release, March 31. Available at http://www.genome.gov/11511308.

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

Smets, F., and R. Wouters. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1 (5): 1123–75.

———. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3): 586–606.

Solow, R. 2010. "Building a Science of Economics for the Real World." Prepared Statement, House Committee on Science and Technology, Subcommittee on Investigations and Oversight (July).

Suckow, M. A., S. H. Weisbroth, and C. L. Franklin, ed. 2006. *The Laboratory Rat*. American College of Laboratory Animal Medicine Series. Academic Press.

Tiao, G., and D. Xu. 1993. "Robustness of Maximum Likelihood Estimates for Multi-Step Predictions: The Exponential Smoothing Case." *Biometrika* 80 (3): 623–41.

Vermeire, T. G., A. J. Baars, J. G. M. Bessems, B. J. Blaauboer, W. Slob, and J. J. A. Muller. 2007. "Toxicity Testing for Human Health Risk Assessment." In *Risk Assessment of Chemicals: An Introduction*, ed. C. J. van Leeuwen and T. G. Vermeire. Springer Netherlands.