

THE STATE OF MACROECONOMIC POLICY MODELING: WHERE DO WE GO FROM HERE?

CHRISTOPHER A. SIMS
PRINCETON UNIVERSITY
SIMS@PRINCETON.EDU

1. MACROECONOMICS DONE RIGHT IS A SCIENCE

- a body of theory
- a flow of facts
- a collection of experts with more or less agreed upon principles for assessing the match of theory to data
- “match” can mean organize, explain, predict

2. A POTTED HISTORY OF MACROECONOMICS AS SCIENCE

- Tinbergen’s ambitious program of quantifying macroeconomic theory in a system of equations.
- Haavelmo’s recognition that, if macroeconomics were to behave like a science, its models would have to assert probability distributions for observable data, because that is the only way to allow objective assessment of the match of model to data when the model is inexact.
- Stagnation of the Cowles-style modeling agenda

3. STAGNATION

- Academic economists convinced themselves that the real-time policy analysis that is most of what goes on at central banks was pointless.
- Central bank staff economists realized that the Cowles probability theory did not actually generate usable probability models when applied at the scale needed for policy models, so they abandoned explicit use of the elegant theory that Haavelmo’s ideas had generated.

4. VAR’S

- They are credible probability models of the data, at least in their BVAR variant.
- Structural VAR’s are usable for policy analysis.
- They are used all the time. Everyone knows how to use them.
- Nowhere are they the main policy model. They did not end or counteract the stagnation.

Date: March 31, 2005.

5. WILL FITTED BAYESIAN NEO-KEYNESIAN DSGE'S TURN THE TIDE?: REASONS FOR HOPE

- They are — at least close to — credible probability models of the data.
- Because they “work”, people are motivated to understand the approach to inference that justifies them and thereby may be acquiring a language in which uncertainty about policy choices can be connected to the probabilities we attach to model inference.
- They are better than structural VAR's as tools for spinning elaborate stories about how the economy works, which may make them more saleable.
- Weather forecasting, which shares some forms of difficulty with macro policy modeling, has become very much better (shows more “skill” in the jargon of the atmospheric scientists) in recent decades. Why not us?

6. REASONS FOR PESSIMISM: POPULARITY FOR THE WRONG REASONS

- They are better than structural VAR's as tools for spinning elaborate stories about how the economy works, which may make them more saleable.
- We know these stories aren't true. There is no K , there is no P , there is no C . There are no infinitesimal monopolistically competitive firms. Individuals do not respond instantly, continuously and rationally to the macroeconomic state. And that these things are all myths matters for how the data behave.
- It may be an advantage of VAR's that they restrain our impulse to take our story-spinning too seriously. If Bayesian DSGE's displace methods that try to get by with weak identification and in the process reinforce the excess weight we give to story-spinning, they may set us back. This is, as I read it, a version of a point that is made also by Jon Faust (2005).

7. REASONS FOR PESSIMISM: THEY CAN'T DELIVER

- The models actually fit carefully so far are not yet at the scale needed for policy analysis. MCMC methods, powerful as they are, may not scale up. Cowles SE methods didn't. We've had about 35 years of trying to deliver on the idea of replacing Keynesian-style models with DSGE RE models. The recent signs of progress have to be weighed against the 35 years of stasis.
- Some people think that at most a minority of the population can ever actually learn to follow the inverse-probability reasoning of Bayesian inference, and that the set of people who can do so when more than two variables are involved is smaller still.

8. REASONS FOR PESSIMISM: INSTITUTIONS AND SOCIOLOGY: WEATHER AS A COMPARISON

- (I) There are several big government agencies devoted to forecasting, with a large teams of experts working on modeling.
- (II) There are multiple competing models whose structures are known to all and whose real time forecasting records are publicly available.
- (III) There are academics who work on the modeling methods and there is interaction between them and the modeling groups.
- (IV) The flow of data about forecasts and actual values is thick, and the history is regularly analyzed.
- (V) For weather forecasting (as opposed to long-term climate modeling) there is no call for projections conditional on policy actions.

9. WEATHER, CONTINUED

- (I) Central banks are in some ways like NOAA. Their scale in personnel, but not in computer hardware, may be similar in order of magnitude.
- (II) The models are more or less public in many cases. But the way those models forecast in real time is much less open. In the US, there is a five year delay imposed on the release of Federal Reserve model forecasts. In most countries, even where model forecasts are available, forecasts by the main policy model of policy variables are not available.
- (III) Few academic economists work on models or methods directly relevant to the policy-modeling needs of central banks.
- (IV) Instead of generating data on multiple variables at hundreds of locations every day that match to model output, as do weather data systems, macroeconomic data systems generate data at monthly or quarterly frequencies on a much shorter list of variables. And such data on historical model performance as do exist are much more infrequently analyzed.
- (V) Of course macro models are called on to make policy projections.

10. ANOTHER POINT OF COMPARISON

- Weather forecasts recently show a sharper version of something I noted in Sims (2002): that subjective forecasts improve on the best model forecasts mainly at short horizons, and apparently in large part because subjective forecasters can react to erratically arriving detailed current data more promptly than a large model (Baars and Mass, 2005).
- This seems to be true now, at least for some purposes, of the Fed's formal model, (though of course it can't be checked for the most recent 5 years!).
- It didn't used to be true. It used to be conventional wisdom among econometricians that formal models broke down for long term projections. That this result

emerged in a 2004-5 study of weather models was regarded as a change from previous patterns, and a sign of the improvement in the models.

11. ARE THERE ANY LESSONS HERE?

- We can't make the data flow faster, but we could do a lot better at systematically tracking and comparing model performance, in both unconditional and conditional projection.
- Why haven't we?

12. INSTITUTIONS AND SOCIOLOGY: MODELING IN CENTRAL BANKS AND ACADEMIA

- The interaction of macro models with policy is a reason they have been kept less open.
- The long run cost in slow progress has been high, and may remain so if the institutional structure stays as it is.
- The weak interaction with academia means that it is a common, though not universal, pattern that ambitious staff are reluctant to be labeled "model mechanics", and hence do not stay with models they have developed for more than a few years. Successors placed in charge of the model have even less incentive to devote time to maintain and improve it, if they are concerned about external recognition.
- Those in charge of forecasting and policy projection are not necessarily enthusiastic about systematically discussing past errors and their sources, or maintaining comparable records on the performance of multiple models or multiple variants on the same model. Ideally the role of critical historical analysis should be in different hands than the role of model maintainer and forecast-maker.
- In the absence of a systematic historical accounting of model and subjective projections, there is a tendency for models to be de-emphasized or abandoned after some period of below-average performance, or even in periods of above average performance when the projections are uncomfortable for the policy preferences of the senior staff. Or what is nominally the same model is modified too frequently and unsystematically.

13. ARE THERE INSTITUTIONAL CHANGES THAT COULD IMPROVE MATTERS?

- What about a macroeconomic version of NOAA?
- Not clear that there is a way to sustain the necessary political support for this. Forecasts and policy projections are likely to be disliked by one side or another of political disputes over policy.
- It is not an accident that much of the best policy-oriented macroeconomic research goes on inside central banks, particularly inside the wonderfully redundant US Federal Reserve system. Central banks can see the value of policy

modeling and have the seignorage to finance it independent of direct political scrutiny.

- If we tried to move modeling to an independent agency, the direct political scrutiny would probably begin, and either distort the research or undercut its funding. And modeling progress depends on demanding users. Outside the Fed there is no constituency of model users analogous to the many people and businesses who feel direct economic effects from the quality of weather forecasts.
- Probably the best we can hope for is better recognition by the central banks of the long-run costs of keeping policy models and their real-time projections obscure, and some corresponding incremental institutional changes.
- For example, if two or three regional Federal Reserve banks committed to regularly publishing forecasts from their own models, including forecasts of policy variables and conditional forecasts given a few alternate assumptions about policy, an academic literature evaluating these forecasts might arise, academic research efforts aimed at improving on these models might appear, and the prospect of external recognition for forecast excellence might help slow the turnover of skilled modelers.
- One could also imagine academic initiatives to generate funding for one or more research centers aimed at developing maintaining, and evaluating policy models. There used to be something like this at Michigan, and the Wharton model for a while also aimed in this direction. They were not able to sustain broad academic interest in this kind of project, which is another reason for pessimism.

14. LET'S TALK ABOUT SOMETHING LESS GLOOMY: ASSUME THE REASONS FOR OPTIMISM PREVAIL

In that case there are interesting research challenges and ideas ahead.

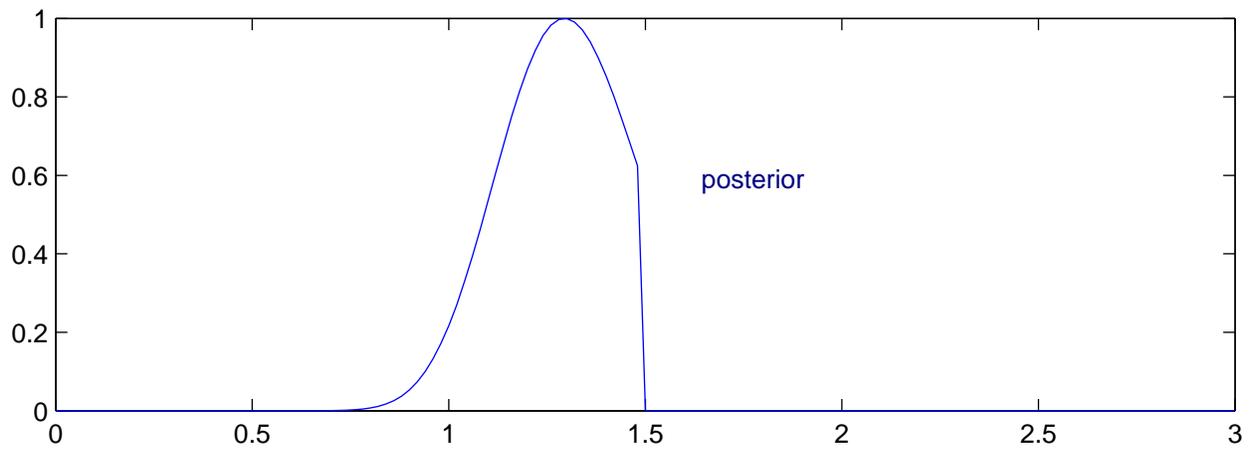
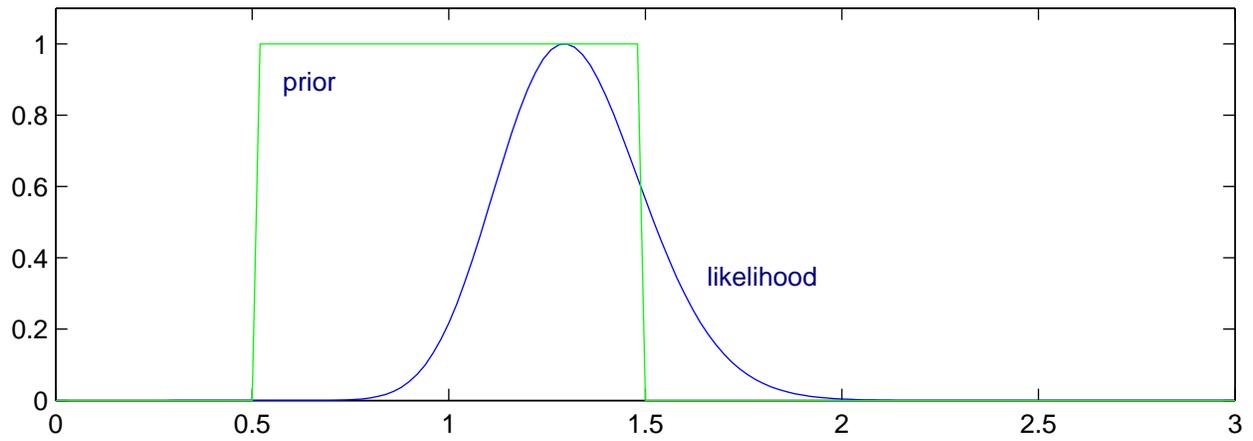
- Learning to use likelihood-based measures of fit in evaluating parameter uncertainty and (as a special case of that) model uncertainty.
- Learning to think about inference as at least implicitly Bayesian non-parametrics, with uncertainty spread over an infinite-dimensional space.
- Deep questions about limitations on the complexity of probability models that we and our computers can handle.
- Learning to deal with multiple models, or multiple versions of models, that differ little in fit but that have differing implications for policy.
- Getting beyond linearity and Gaussianity.
- Implementing intuitive, non-conjugate priors on large models.
- MOS for econometric policy models?

15. LIKELIHOOD-BASED MEASURES OF FIT

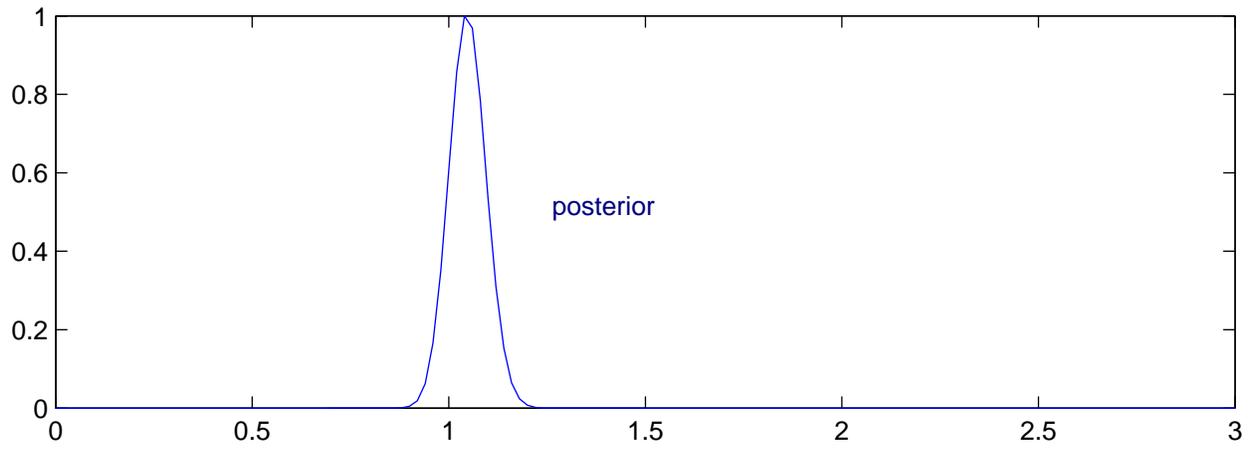
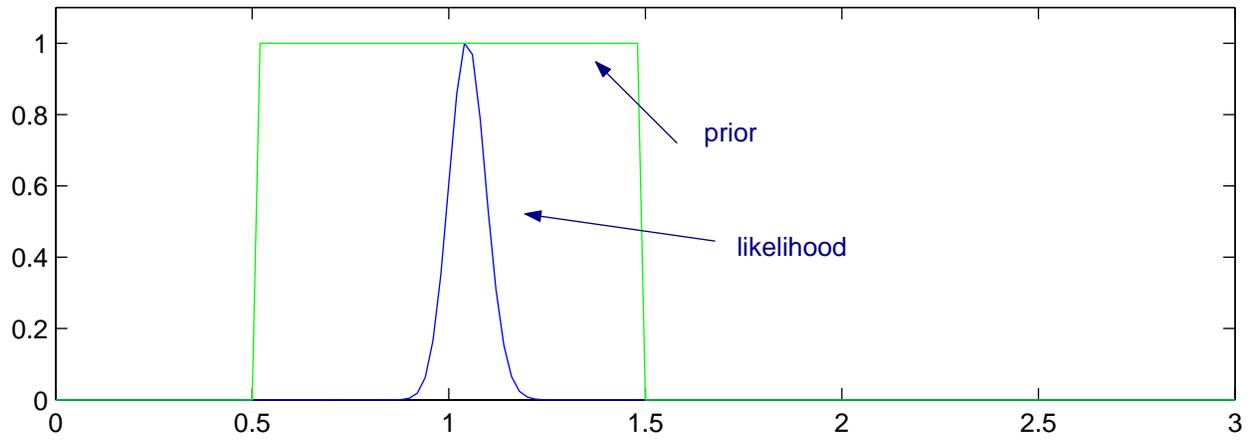
- The promise that Bayesian methods can provide an assessment of cross-model uncertainty was one source of initial enthusiasm for Bayesian DSGE's.

- It is even claimed in various places in the literature that Bayesian (or other likelihood-based) methods can be used to choose among or weight models that we know to be false, in the sense that there are other models that clearly fit better.
- The latter claim is just wrong, as Schorfheide (2000) showed.
- Nonetheless, my view is that it should no longer be respectable to use models for policy analysis that are known to fit badly. DSGE's can fit as well as VAR's, and we should not be satisfied to use models that don't.
- But integrating likelihoods by Monte Carlo methods is still an art, even more than estimation of continuous parameters. Convergence in high-dimensional models in my experience is fragile, slow, and hard to assess. This will undoubtedly get better as more people do it and experience accumulates.
- The phenomena grouped under the "Lindley paradox" heading are pervasive when we deal with discrete parameters. It is important to recognize that this does not mean that there is something wrong with the Bayesian calculation of weights on models and that we should therefore look to some other principle for picking weights.
- Model A may emerge with posterior probability $1 - 10^{-12}$ and another set of models we were interested in may therefore turn out to get negligible weight. We may not in fact be confident that model A is the truth, because we feel its parameterization is too restrictive to be believable. The right response is not to reweight the other models, but to relax the restrictions on A so we have more confidence in it, or to diagnose the problems with the other models so they can be modified to compete with A. The "shoot the messenger" approach of, for example, equal-weight averaging of the models despite their disparate fits, is unjustifiable.
- If computational difficulties and loss of nerve in the face of Lindley paradox phenomena lead to widespread use of "fitted" DSGE's with ad hoc fit criteria, much of the promise of Bayesian DSGE's will have been lost.

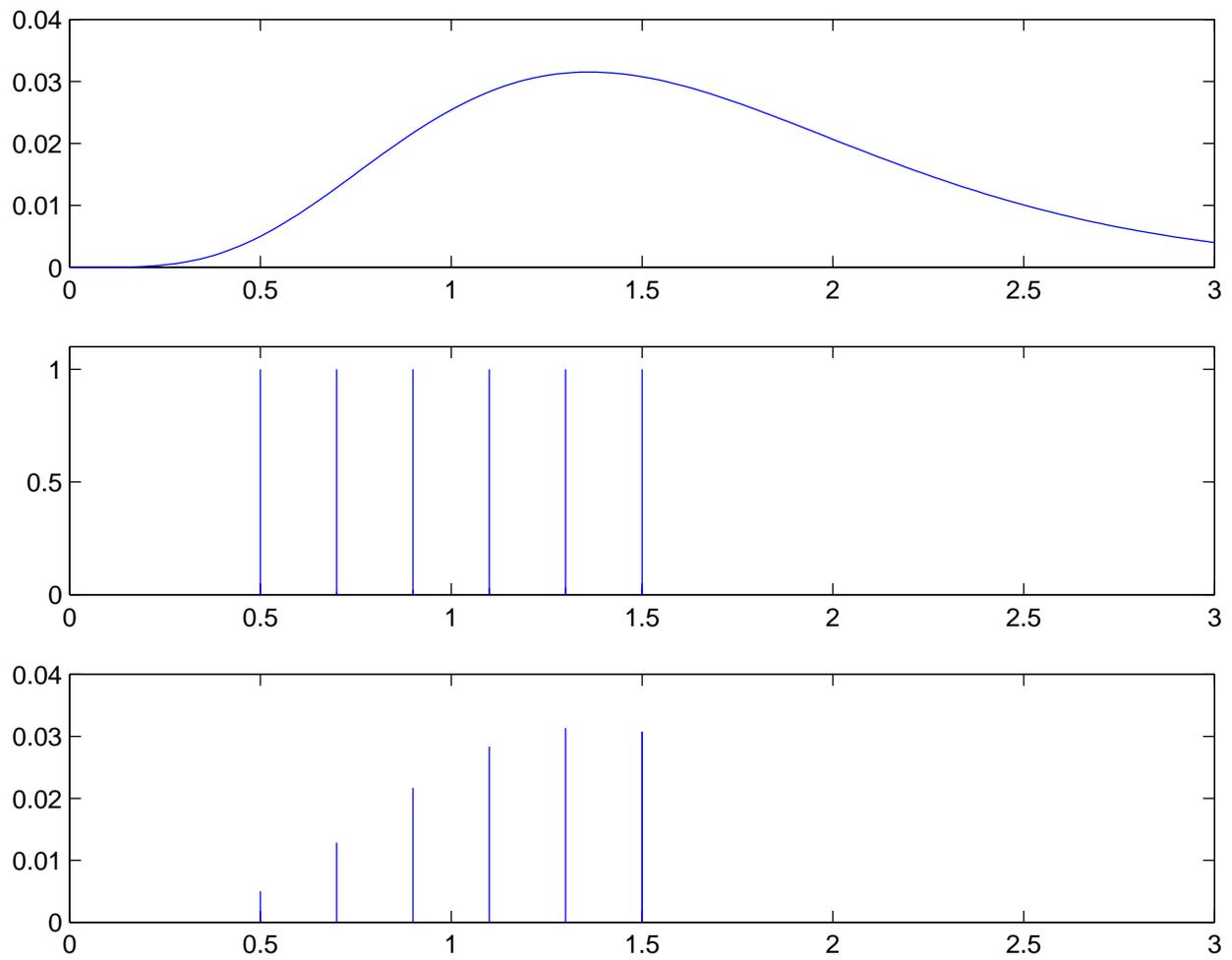
16.



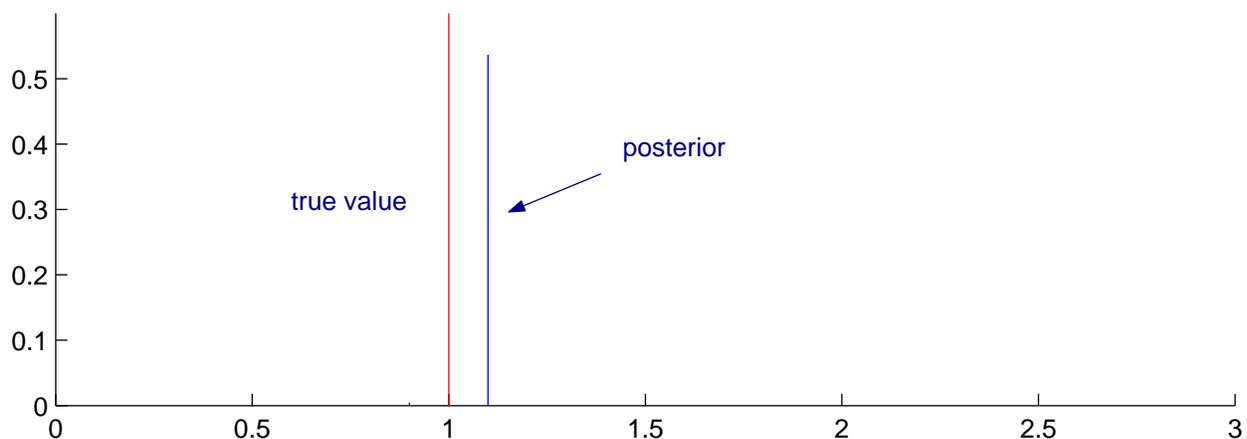
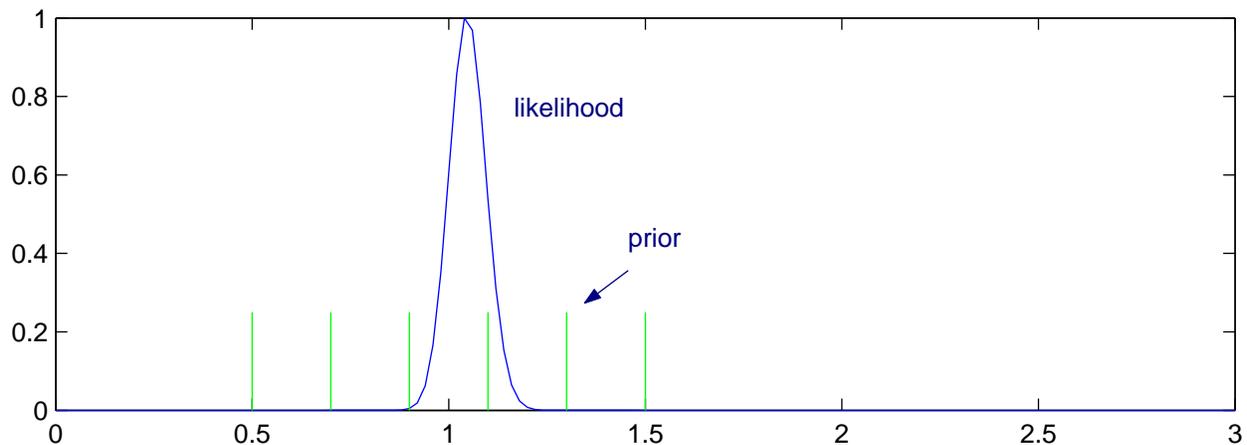
17.



18.



19.



20. LINEARITY AND GAUSSIANTY: ASSET PRICES

- Faust 2005 in his paper for this conference cites the uncovered interest parity anomaly as something that DSGE's still don't fit.
- Linearized DSGE's obviously strictly enforce UIP and thus contradict the data.
- Going to second order is not much help.
- Blake LeBaron (1998) shows that the distribution of returns from trading strategies that attempt to exploit deviation from UIP are extremely non-normal, with fat tails, and that the trading strategies do not produce as attractive returns as estimated regression equations might lead one to expect.
- If extremely non-normal returns is part of the explanation of deviations from UIP, linearized models won't help with it.
- This point applies also to term structure and to asset pricing in general, when multiple asset prices appear in a model.

21. LINEARITY AND GAUSSIANTY: INFERENCE

- Tao Zha and I have just finished a paper documenting how strong is the time variation in volatility of US macro data, and how strong are the effects on inference about other things of treating that time variation too inflexibly.
- I used to think that linearized DSGE's must produce accurate second-order approximations to likelihood, but now think this is true only in special cases.
- Nonlinear models will predict time variation in disturbance variances, and a linearized model will lose that aspect of model predictions.
- However the kind of time variation in volatility that Tao and I find does not look like the smooth and predictable variation that would emerge from a DSGE.
- So it seems unlikely that simply drawing out the higher-order implications of existing DSGE's will much improve fit — indeed it could well worsen fit, since no one is thinking about implications for time-varying residual variances when they specify or calibrate these models.
- This area seems to me a hornets nest just around the corner for macroeconometric DSGE policy models.

REFERENCES

- BAARS, J. A., AND C. F. MASS (2005): "Performance of National Weather Service Forecasts Compared to Operational, Consensus, and Weighted Model Output Statistics," Discussion paper, Department of Atmospheric Sciences, University of Washington.
- BLAKE LEBARON (1998): "A Dynamic Trading Strategy Approach to Deviations from Uncovered Interest Parity," Discussion paper, Brandeis University.
- JON FAUST (2005): "Is Applied Monetary Policy Analysis Hard?," Discussion paper, Federal Reserve Board, Washington, DC.
- SCHORFHEIDE, F. (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15(6), 645–670.
- SIMS, C. A. (2002): "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity*, 2002(2), 1–62.