Bayesian Analysis of DSGE Models – Rejoinder

Sungbae An
Singapore Management University

Frank Schorfheide*
University of Pennsylvania, CEPR and NBER

This Version: August 10, 2006

*An: School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903; email: *sungbae@smu.edu.sg.* Schorfheide: Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104; email: *schorf@ssc.upenn.edu.*

We would like to thank all the discussants for their stimulating comments. While our article to a large extent reviews current practice of Bayesian analysis of DSGE models the discussants provide many ideas to improve upon the current practice, thereby outlining a research agenda for the years to come. In our rejoinder we will briefly re-visit some of the issues that were raised.

Identification and Sensitivity Analysis. Fabio Canova as well as Fabio Milani and Dale Poirier raise the issue of identification in DSGE models. At a conceptual level "unidentifiability causes no real difficulty in the Bayesian approach." This quote is taken from D.V. Lindley and precedes Dale Poirier's (1998) article on revising beliefs in non-identified models. On the other hand, as Fabio Canova points out in his comments, Bayesian methods "may end up covering up identification problems if improperly used." At the least, it will be useful for the audience to know which of the structural parameters (or functions of parameters) are not identifiable or poorly identified by the available sample information. In principle, one could re-parameterize the DSGE models in terms of identifiable parameters ϕ and non-identifiable parameters ψ such that

$$\theta = F(\phi, \psi)$$

It is well understood that the prior distribution of $\psi|\phi$ is not updated and will affect the shape of the posterior of θ even in large samples. Hence, the audience might be particularly concerned about the sensitivity of the substantive conclusions drawn from the econometric analysis to changes in $p(\psi|\phi)$.

If the relationship between θ, ϕ, and ψ can be derived analytically, as for the slope of the Phillips curve κ, such a sensitivity analysis is fairly easily implementable. Unfortunately, in general it is very difficult to construct the function $F(\phi, \psi)$. Nevertheless, Fabio Canova's comment outlines some low-tech and high-tech approaches to detect identification problems that hopefully will be adopted by applied researchers in the future. At a minimum, we think that it will be useful to generate bivariate scatter plots as in Figure 1 of our article based on simulated data sets, possibly of different lengths. For the model studied in the paper, Figure 1 nicely illustrates that information about the parameters of the exogenous shock processes, accumulates much faster than information about the policy rule coefficients. This suggests that a sensitivity analysis should begin with changes in the prior distribution of the parameters that characterize the central bank behavior.

We followed the suggestion of Fabio Milani and Dale Poirier and repeated the analysis of Section 6.3 by placing a prior distribution on the structural parameter $1/\phi$ instead of the slope coefficient κ of the Phillips curve. We assumed that $1/\phi$ has a Gamma distribution with mean 0.012 and standard deviation of 0.006. The prior distribution of all other parameters is the same as in Table 3 of our paper. The functional relationship

$$\kappa = \tau \frac{1}{\phi} \frac{1-\nu}{\nu \pi^2}$$

induces a prior distribution for κ that has a mean of 0.29 and a standard deviation of 0.30. Figure 1 of this rejoinder depicts marginal densities of ν. As expected, the marginal distribution of ν is updated even if the model is solved with linear techniques. However, if we use a quadratic approximation to the model dynamics the likelihood function contains additional information about ν and the posterior is even more concentrated.

Truncation of Prior Distributions. Both Fabio Canova and John Geweke mention in their discussions potential pitfalls related to the truncation of prior distributions. Hence, we would like to provide a couple of clarifying remarks. With the exception of Lubik and Schorfheide (2004) researchers tend to restrict the parameter space to the subspace in which the linearized DSGE model has a unique rational expectations solutions. Let us denote this subspace by Θ_D and let $\mathcal{I}\{\theta \in \Theta_D\}$ be the indicator function that is one if θ is in the determinacy region and zero otherwise. The posterior is proportional to

$$p(Y|\theta) \propto p(Y|\theta)p(\theta)\mathcal{I}_{\theta \in \Theta_D}, \tag{1}$$

where $p(\theta)$ typically has non-zero density outside of the determinacy region. It certainly does for the $p(\theta)$ specified in Table 2 of our article. Two interpretations are possible: we could either define a modified likelihood function $\tilde{p}(Y|\theta) = p(Y|\theta)\mathcal{I}_{\theta \in \Theta_D}$ and write

$$p(Y|\theta) \propto \tilde{p}(Y|\theta)p(\theta) \tag{2}$$

or we could define a modified prior distribution

$$\tilde{p}(\theta) = \frac{1}{c} p(\theta)\mathcal{I}_{\theta \in \Theta_D}, \quad \text{where} \quad c = \int_{\theta \in \Theta_D} p(\theta) d\theta$$

and let

$$p(Y|\theta) \propto p(Y|\theta)\tilde{p}(\theta). \tag{3}$$

We actually follow the latter approach. In Table 2 of the article we reported $p(\theta)$, whereas we plotted draws from $\tilde{p}(\theta)$ in Figure 1. John Geweke correctly points out that we have to adjust our calculation of the log marginal data densities for the fact that the "raw" prior $p(\theta)$ does not integrate to one. We do so, but it turns out that the adjustment is small since our prior only assigns 2% probability to the indeterminacy region: $\ln 0.98 \approx$

-0.02. Fabio Canova points out that if one absorbs the indicator function into the likelihood function, then shape differences between prior and posterior could be due to the presence of the indicator function $\mathcal{I}\{\theta \in \Theta_D\}$ instead of the non-truncated likelihood $p(Y|\theta)$.

High-Quality Computations. John Geweke and Tao Zha provide many suggestions on how to ensure that the posterior computations are carried out in a reliable manner and how to diagnose convergence problems of MCMC calculations. The techniques mentioned should definitely become part of the toolbox for the Bayesian analysis of DSGE models.

In our exposition we focused mostly on informal methods, such as comparisons of recursive means of parameter draws from multiple chains in Figures 3 and 6 as well as a comparison of the output generated by the Random-Walk Metropolis Algorithm and the Importance Sampler. Tao Zha suggests to report potential reduction scale statistics for the DSGE model parameters based on a large number of parallel Markov chains. This is a good idea – in fact, these statistics are part of the default output of the software package DYNARE, which automates many of the estimation techniques for linearized DSGE models described in our article.

It is, however, important to keep in mind, that even more formal methods than the ones used in our paper do not necessarily guard against pitfalls as the analysis of the New Keynesian model with the output growth rule shows. If it turns out that the starting values for the parallel Markov Chains are chosen such that the chains move toward the vicinity of the low mode, neither the potential scale reduction statistic nor a statistical examination of the recursive parameter means will flag the computational inaccuracy. On the other hand, once an explorative analysis of the posterior surface has unveiled a multi-modal structure of the posterior, we can customize the diagnostics and explicitly design sampling algorithms that handle the complicated shape of the posterior.

As suggested by John Geweke, we now use a hybrid MCMC algorithm with transition mixture to deal with the bimodal posterior distribution that arises in the analysis of the output-growth-rule version of the DSGE model $\mathcal{M}_2(L)/\mathcal{D}_2(L)$. We briefly describe the algorithm.

A Hybrid MCMC Algorithm for a Bimodal Posterior

1. Using different starting values, apply a numerical optimization procedure to search for modes $\tilde{\theta}_{(j)}$, $j = 1, \ldots, J$ of the posterior density.

2. For each mode, compute the inverse of the Hessian, denoted by $\tilde{\Sigma}_{(j)}$.

3. Let $q_j(\theta)$ be the density of a multivariate t-distribution with mean $\tilde{\theta}_{(j)}$, scale matrix $c_j\tilde{\Sigma}_{(j)}$, and ν_j degrees of freedom.

4. Let π_j, $j = 1, \ldots, J$ a set of probabilities and define the transition mixture

$$q(\theta) = \sum_{j=1}^{J} \pi_j q_j(\theta)$$

5. Choose a starting value $\theta^{(0)}$ for instance by generating a draw from $q(\theta)$.

6. For $s = 1, \ldots, n_{sim}$, draw ϑ from the transition mixture $q(\vartheta)$ as follows. First, randomly select a mixture component j according to the mixture probabilities $\pi_1, \ldots, \pi_J$. Second, draw ϑ from the distribution with density $q_j(\vartheta)$.

   The jump from $\theta^{(s-1)}$ to ϑ is accepted ($\theta^{(s)} = \vartheta$) with probability $\min\{1, r_j(\theta^{(s-1)}, \vartheta)|Y\}$ and rejected ($\theta^{(s)} = \theta^{(s-1)}$) otherwise. Here

$$r_j(\theta^{(s-1)}, \vartheta|Y) = \frac{\mathcal{L}(\vartheta|Y)p(\vartheta)/q_j(\vartheta)}{\mathcal{L}(\theta^{(s-1)}|Y)p(\theta^{(s-1)})/q_j(\theta^{(s-1)})}.$$

In Section 4.1 of our article we found two modes of the posterior for $\mathcal{M}_2(L)/\mathcal{D}_2(L)$. In slight abuse of our previous notation, let $j = L$ correspond to the "low" mode, and $j = H$ to the "high" mode. We set $c_L = 5$, $\nu_L = 3$, $c_H = 4$, and $\nu_H = 2$. Thus, the proposal density associated with the high mode is more diffuse. We assign equal probabilities to the two components of the transition mixture $\pi_L = \pi_H = 1/2$.

Figure 2 of this rejoinder shows the acceptance ratios of the Metropolis chain conditional on the mixture component that has been selected (either q_L or q_H). The acceptance ratios are computed recursively. To generate the upper panel of Figure 2 we initialize the Markov chain at the low mode. It is apparent that the recursively computed acceptance rate of a candidate draw from q_L decreases as the chain progresses, while the acceptance rate from q_H increases. In fact, after 40,000 draws the acceptance ratio conditional on q_L is effectively zero, whereas the acceptance rate for q_H draws stabilizes around 25%. These results suggest that an exploration of the posterior in the neighborhood of the high mode potentially is able to provide a good approximation to the overall posterior distribution, at least for the combination of model and data set considered in our article. The second panel of Figure 2 depicts conditional acceptance ratios for a chain that is initialized at the high mode. The acceptance ratio for draws generated from q_L is essentially zero, which confirms our conclusion.

Most of the empirical literature has used the Random-Walk Metropolis Algorithm to generate draws from the posterior distribution of the DSGE model parameters, in part, because the relevant software is readily available over the internet. While the problems with multi-modal posteriors discussed above suggest that the Random-Walk Metropolis algorithm might not be the most efficient way to proceed, there is little research that tries to compare the performance of alternative sampling schemes and develop potentially better MCMC algorithms for DSGE models. The comments by Malin Adolfson, Jesper Linde, and Mattias Villani provide an important step in this direction and we hope that the authors will continue their line of research.

Improvements in MCMC methods for DSGE models are particularly important as central banks are starting to estimate large scale DSGE models with often more than 50 parameters using the techniques described in this survey paper. Bayesian statisticians have a lot of expertise in boosting the performance of MCMC computations by customizing the sampling algorithms. We hope that this survey article and the discussions will stimulate some interest in the statistics community to conduct research on more efficient algorithms for the Bayesian analysis of DSGE models.

The Interpretation of DSGE-VARs. We strongly disagree with some of the statements made by Tao Zha about the DSGE-VAR approach. His discussion of the DSGE-VAR framework gives the impression that the DSGE-VAR is a reduced-form specification instead of an identified VAR. To be clear: the DSGE-VAR is an identified VAR. The restrictions used to identify the DSGE-VAR differ from those that Tao Zha and his co-authors have used in their own empirical work but that does not make the approach *incoherent*, unless *incoherence* means "your priors are different from mine."

We start out from a structural VAR that can be written as

$$y_t = \Phi' y_{t-1} + \Sigma_{tr}(\Sigma)\Omega\epsilon_t, \tag{4}$$

where ϵ_t is a vector of orthogonal structural shocks with unit variance, Ω is an orthonormal matrix, and $\Sigma_{tr}(\Sigma)$ is the lower triangular matrix obtained from the Cholesky decomposition of Σ. It has the property $\Sigma_{tr}\Sigma_{tr}' = \Sigma$. In the article, we introduced the one-step ahead forecast errors $u_t = \Sigma_{tr}\Omega\epsilon_t$. It is straightforward to verify that the covariance matrix of u_t is given by Σ. It is well-known that the likelihood function of the VAR only depends on Φ and Σ, but not on Ω.

A *coherent* Bayesian approach specifies a joint prior distribution for Φ, Σ, and Ω and updates the prior using the likelihood function to obtain a posterior distribution. This is exactly what we do. Many priors in the literature are such that they concentrate on a subspace of the Φ - Σ - Ω parameter space such that

$$F(\Phi, \Sigma, \Omega) = 0$$

and $F(\cdot)$ can be inverted to uniquely determine Ω as a function of Φ and Σ.

The approach proposed by Del Negro and Schorfheide (2004) and reviewed in our survey article can be represented by a function that is indexed by the DSGE model parameters θ:

$$F(\Phi, \Sigma, \Omega|\theta) = 0.$$

The actual function $F(\cdot|\theta)$ used by Del Negro and Schorfheide (and in the survey article) has the property that $\Omega = \Omega^*(\theta)$. The overall prior distribution for the DSGE-VAR has the following structure:

$$p_\lambda(\Phi, \Sigma, \Omega, \theta) = p_\lambda(\Phi, \Sigma|\theta)p(\Omega|\theta)p(\theta), \tag{5}$$

where $p(\Omega|\theta)$ is essentially a point mass at $\Omega^*(\theta)$. This prior is combined with the VAR likelihood to calculate a posterior

$$p_\lambda(\Phi, \Sigma, \Omega, \theta|Y) \propto p(Y|\Phi, \Sigma)p_\lambda(\Phi, \Sigma|\theta)p(\Omega|\theta)p(\theta). \tag{6}$$

Since Ω does not enter the VAR likelihood function and $p(\Omega|\theta)$ is a point mass, the posterior computations are implemented in two steps. We first generate draws from the joint posterior distribution of Φ, Σ, θ, that is, we are estimating a reduced form DSGE-VAR. Second, for each draw (Φ, Σ, θ) we evaluate $\Omega^*(\theta)$.

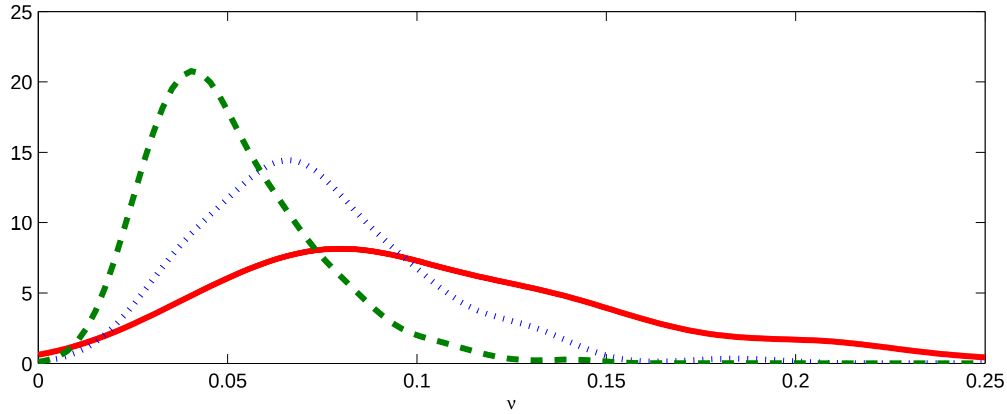
The impulse responses reported in our article are based on the evaluation of the joint posterior distribution of all parameters and not, as asserted in Tao Zha's discussion, by conditioning on the posterior mean estimate $\hat{\theta}$ when applying the mapping $F(\Phi, \Sigma, \Omega|\theta) = 0$ to recover Ω as a function of the reduced form parameters. Hence, there is nothing *incoherent* about the analysis, regardless of the value of the hyperparameter λ. In fact, our parameterization of the structural VAR in terms of Φ, Σ, and Ω has the advantage that it explicitly distinguishes between identifiable parameters, namely Φ and Σ, and non-identifiable parameters, namely Ω. It is clear from this parameterization that the conditional prior $p(\Omega|\Phi, \Sigma, \theta) = p(\Omega|\theta)$ will not be updated through the likelihood function.

A second question is whether our choice of $p(\Omega|\Phi, \Sigma, \theta)$ is economically reasonable. A twenty-five year old literature on structural VARs shows that identification assumptions are rarely uncontroversial. Our choice of Ω is to a large extent designed to mimic DSGE model impulse responses with the VAR. Hence, the impulse responses may or may not be similar to identification schemes obtained under more conventional identification schemes.

More traditional restrictions used to identify VARs can be incorporated in our analysis by modifying the DSGE model such that it satisfies these restrictions in the first place. In fact, since the premise of the DSGE-VAR analysis is that the DSGE model provides a good albeit not perfect approximation of reality, strong views about the identification of particular structural shocks can and should be directly incorporated into the underlying DSGE model. It is true that our identification scheme puts a lot of weight on matching DSGE and DSGE-VAR responses in the short-run. To the extent that the DSGE model misses some of the short-run adjustment dynamics it is worthwhile to extend the framework such that implicitly more weight is placed on medium-run responses.

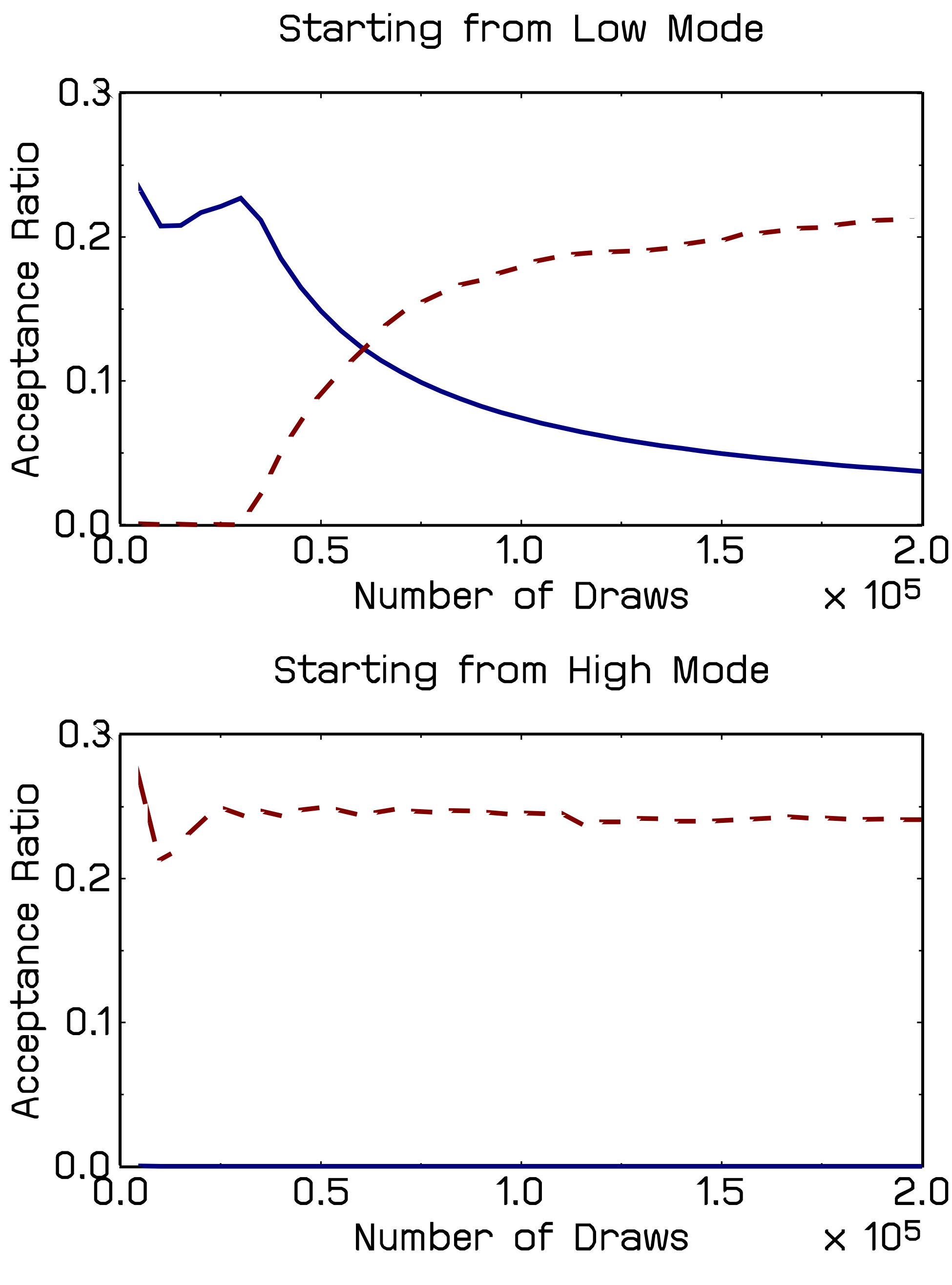
Better DSGE Models. Fabio Milani and Dale Poirier as well as Tao Zha pointed out that one should work with better DSGE model, that potentially incorporate learning mechanisms, regime switches, and time-varying shock volatilities. We agree with that general statement. The goal of the review paper was to survey tools for constant-coefficient homoskedastic DSGE models. One of the next steps in this research agenda will be to generalize these tools so that one can handle richer DSGE models. Some steps in this directions have been undertaken already. We hope that this article and the discussions will stimulate further research along these lines.

Figure 1: Linear versus Quadratic Approximation – Marginal Posterior of ν



Notes: Marginal densities of ν: prior (solid), linear/Kalman posterior (dotted), and quadratic/particle posterior (dashed).

Figure 2: Independence Metropolis Chain with Mixture Transition



Notes: Acceptance ratios are computed recursively and plotted as a function of the number of draws. The chain starts either from the low mode (the upper panel) or from the high mode (the lower panel). Each line corresponds to the acceptance ratio conditional on the candidate draw being generated from q_L (solid) or from q_H (dashed).