Comment on Marco Del Negro, Frank Schorfheide, Frank Smets, and Raf Wouters, 'On the Fit of New-Keynesian Models'*

Lawrence J. Christiano†

January 22, 2007

*I am particularly grateful for the extensive advice and expert research assistance of Cosmin Ilut. I am also grateful for conversations with Joshua Davis, Martin Eichenbaum and Giorgio Primiceri, and for the research assistance of Patrick Higgins. My comment has benefited from the reactions and advice of Marco Del Negro and Frank Schorfheide. I am grateful for the financial support of the National Science Foundation.

†Department of Economics, Northwestern University, and National Bureau of Economic Research.

1. Introduction

I am very grateful to have been given the opportunity to discuss this important and influential paper by Del Negro, Schorfheide, Smets and Wouters (DSSW). It represents a notable step forward in the ongoing enterprise of introducing Bayesian ideas into the analysis of macroeconomic time series (the closest antecedents are Ingram and Whiteman (1994) and Del Negro and Schorfheide (2006).) As dynamic, stochastic general equilibrium (DSGE) models become increasingly useful from an empirical standpoint, we need increasingly sophisticated methods to diagnose how well they fit. Because these developments in DSGE modeling are relatively recent, and have occurred rather suddenly, we are short on diagnostic methods. The contribution of this paper is to present and apply such a method, building on the work of Del Negro and Schorfheide (2006).

I begin with a brief review of the DSSW procedure. That procedure works with a 'hybrid model' that is a combination of an unrestricted vector autoregression (VAR) for the data and the VAR implied by the econometrician's DSGE model. The combination is indexed by a scalar parameter, λ, where the hybrid model reduces to the unrestricted VAR when λ is small and to the DSGE model as $\lambda \rightarrow \infty$. The best hybrid model is the one associated with $\hat{\lambda}$, the value of λ that results in the highest marginal likelihood for the data. If $\hat{\lambda}$ is large, then the DSGE model is a good one. If $\hat{\lambda}$ is sufficiently small, this is evidence that the researcher needs to go back to the drawing board to improve the DSGE model.

My comment focuses on two questions: 'what is the rationale for using the marginal likelihood to assess alternative values of λ?', and 'what should the cutoff values of $\hat{\lambda}$ be for deciding whether a DSGE model is good or bad?'. After addressing these questions, I ask whether there are other procedures for evaluating model fit. I turn to this question in the conclusion.

The two basic ingredients in the marginal likelihood are the likelihood of the data, which is assumed to be Normal, and the priors over model parameters. In practice, the choices made on both dimensions are controversial. Based on the skewness and kurtosis properties of residuals in an estimated VAR, I find strong evidence against the Normality assumption. Also, the choice of priors is as heavily influenced by computational tractability as by plausibility. The marginal likelihood is a compelling only to the extent that its two ingredients are compelling.

I report the results of computational experiments with simple examples which suggest that the magnitude of deviation from Normality, which is statistically very significant, is not large enough to distort the DSSW analysis. Regarding the choice of priors, in my comment I merely question the appropriateness of the DSSW priors. I suggest a way to construct an alternative set of priors that may better capture a researcher's actual priors over VAR parameters. However, it is beyond the scope of this comment to investigate whether a DSSW-style analysis is robust to such an alternative specification of priors.

Next, I turn to the question of how large is a 'large' and how small is 'small', in the case of $\hat{\lambda}$. I construct two Monte Carlo experiments in which artificial data are generated by a DSGE model, and the econometrician correctly specifies the model. This allows me to assess how small $\hat{\lambda}$ must be for the econometrician to conclude that there is something wrong with the DSGE model. Not surprisingly, I find that the answer depends on two things: (i) the details of the DSGE model and (ii) the number of free parameters in the unrestricted VAR relative to the number free parameters in the DSGE model. This suggests that the DSSW method could be made even more useful if explicit guidance could be provided to link the lower cutoff value of $\hat{\lambda}$ to the model used in the analysis and to the number of degrees of freedom. I also construct a Monte Carlo experiment in which the econometrician's DSGE model is misspecified. The DSSW method is shown to have power in that it discovers with very high probability that the DSGE model is a poor one. All the experiments suggest one simple improvement to the DSSW method, which would help it to better identify weaknesses in model fit. In addition to reporting $\hat{\lambda}$ itself, there should be an analysis of the rate at which the marginal likelihood declines for $\lambda > \hat{\lambda}$. The experiments suggest that such a measure would help to sharply differentiate between good and bad-fitting DSGE models. On the whole, the Monte Carlo experiments support DSSW's conclusion that there is information about model fit in their method.

DSSW argue that the best hybrid model which emerges from their analysis is of independent interest. The idea is that it can serve as a basis for thinking about how to improve the DSGE model in cases when $\hat{\lambda}$ is small. This is possible, though I am skeptical. As emphasized by DSSW, the marginal likelihood penalizes models with a large number of free parameters. In practice, the parameters of the hybrid model are those of the unrestricted VAR, modified to resemble those implied by the DSGE model. Such a hybrid model can lead to improvement in the marginal likelihood simply because the DSGE model has substantially fewer free parameters and not because the hybrid model is necessarily closer to the 'true' reduced form in a sense that is relevant to the economic analyst. Still, it is possible to evaluate the DSSW idea that the hybrid model is useful for identifying directions for improvement in the DSGE model by constructing the type of experiments analyzed in this comment. In such an experiment, the econometrician would be modeled as analyzing artificial data that are generated from a wrong model, and uses the hybrid model to identify directions for improvement.

The following section briefly reviews the DSSW procedure. Section 3 evaluates the marginal likelihood as a measure of model fit. Section 4 investigates how one should interpret the magnitude of $\hat{\lambda}$. The final section concludes.

2. The DSSW Procedure

Let the constants and the parameters on the lag coefficients in the VAR representation of the data, Y, be denoted by Φ. Let the variance-covariance matrix of the one-step-ahead forecast errors in this VAR be denoted by Σ. The mapping from the DSGE model parameters, θ, to the VAR representation of Y is denoted $\Phi(\theta)$ and $\Sigma(\theta)$. DSSW assume the data have a Normal distribution, so that the likelihood of the data is a function only of Φ and Σ :

$$L(Y|\Phi,\Sigma).$$

To evaluate the marginal likelihood requires integration over the model parameters. One possibility is to replace Φ and Σ by $\Phi(\theta)$ and $\Sigma(\theta)$, respectively, and specify a prior over θ. But, this does not serve DSSW's purpose because this presumes the DSGE model is true, and that the only thing not known about it is the values of θ. In their assessment of the fit of the DSGE model, DSSW want to be open to the possibility that the model does not fit well. To see how DSSW proceed, consider Figure 1. On the horizontal axis are the VAR parameters, reduced to a single dimension for the sake of the discussion. In the middle of the horizontal axis, I have placed the values of the VAR parameters implied by the DSGE model, with parameter values, θ. If the DSGE model were true, then the prior over Φ,Σ conditional on θ would be a single spike above zero on the horizontal axis. But, this would defeat a basic objective of DSSW, which is to evaluate the fit of the DSGE model and, in particular, to entertain the possibility that the fit of the DSGE model is poor. For this reason, DSSW construct a prior distribution on Φ,Σ which assigns a positive probability to the state of the world in which the DSGE model is false. Conditional on a value for θ, the prior has mode $\Phi(\theta), \Sigma(\theta)$ and is denoted

$$P(\Phi, \Sigma|\theta, \lambda),$$

where the value of the scalar, λ, controls how quickly the prior drops to zero (see Figure 1). As λ goes to infinity, the prior converges to a single spike over $\Phi(\theta), \Sigma(\theta)$, and corresponds to the case in which the DSGE model is believed to be true. With small values of λ, the prior becomes more and more diffuse and a sufficiently small value of λ captures the view that the DSGE model provides very little prior information Φ, Σ. The marginal likelihood of the data, conditional on the priors and on a value for λ is denoted $\mathcal{L}(Y, \lambda)$, which is defined as follows:

$$\mathcal{L}(Y, \lambda) = \int_\theta \int_{(\Phi,\Sigma)} L(Y|\Phi, \Sigma) P(\Phi, \Sigma|\theta, \lambda) P(\theta) d(\Phi, \Sigma) d\theta.$$

The DSSW procedure computes $\hat{\lambda}$ as the solution to

$$\hat{\lambda} = \arg\max_{\lambda \geq 0} \mathcal{L}(Y, \lambda). \tag{2.1}$$

In principle,evaluating $\mathcal{L}(Y, \lambda)$ requires solving a massive numerical integration problem. For example, suppose we had an $m = 10$ variable VAR with four lags and a constant term in each equation, so that $k = 41$. Then, the number of elements in Φ is 410 and the number of elements in Σ is 55. In this case, the parameters Φ, Σ alone contribute 465 dimensions to the integration problem, while the model parameters, θ, contribute another 30-40. Numerical integration in such a high-dimensional space, while not impossible, would be a major impediment to implementing the DSSW procedure. To avoid this, DSSW specify $P(\Phi, \Sigma|\theta, \lambda)$ to be conjugate with the normal likelihood. That is, $P(\Phi, \Sigma|\theta, \lambda)$ is the product of the inverse Wishart density for Σ and the Multivariate Normal density for Φ conditional on Σ. The scalar, λ, controls how tightly concentrated this density is about $\Phi(\theta), \Sigma(\theta)$. With this specification of the prior the product,

$$\int_{(\Phi,\Sigma)} L(Y|\Phi,\Sigma)\, P(\Phi,\Sigma|\theta,\lambda)\, d(\Phi,\Sigma),$$

can be evaluated analytically for given values of θ, λ. With this dramatic reduction in the dimension of the integration problem, evaluating $\mathcal{L}(Y,\lambda)$ is computationally feasible. The computational problem is nevertheless quite cumbersome, however, so that in practice the maximization problem in (2.1) must be limited to a coarse grid of λ's.

For a specific value of λ the mode of the posterior distribution of Φ, Σ is DSSW's hybrid VAR. So, a side-product of the calculations is a 'best' hybrid parameterization. If that parameterization is 'far' from the DSGE model (i.e., $\hat{\lambda}$ is small) this is indication that the DSGE model fits poorly. If the hybrid parameterization corresponds closely to that implied by the DSGE model (i.e., $\hat{\lambda}$ is large) this is an indicator of good fit.

3. The DSSW Strategy: A priori Considerations

In this section, I raise some questions about the a priori appeal of using the marginal likelihood to evaluate model fit. That the strategy in principle has some appeal is suggested by the fact that we even use it in ordinary day-to-day conversation. For example, in a discussion about the reason one's car will not start or about the solution to a murder mystery, the hypothesis that best explains the pertinent facts commands the most attention. However, although the marginal likelihood may seem in principle to be an attractive way to select among models, in practice one has to make severe compromises for the approach to be tractable. This is a shortcoming of the strategy, because the impact of these compromises on the outcome of the analysis is hard to judge. I now review these considerations by examining the two basic ingredients in constructing the marginal likelihood: the likelihood function and the priors.

DSSW follow convention in specifying the likelihood function of the data to be Normal. I investigated the plausibility of this specification by fitting a four-lag, seven variable VAR using monthly US data for the period 1955Q4-2006Q1. The variables used are

$$y_t = [\log C_t/Y_t,\ \ \log I_t/Y_t,\ \ \text{inflation}_t, \log Y_t/L_t - \log W_t/P_t, R_t, \log L_t, GDPgrowth_t].$$

Here, C_t denotes real, per capita non-durables and services consumption; Y_t denotes real, per capita GDP; L_t denotes per capita hours worked; inflation denotes inflation in the personal consumption expenditure index; W_t/P_t denotes real labor compensation for whole economy (e.g., farmers and government included); R_t denotes federal funds rate; I_t denotes real, per capita gross private domestic investment plus household purchases of durable goods. All real variables were obtained by deflating by the GDP chain price index. I computed skewness and kurtosis statistics for each of the seven VAR disturbances and I computed their $p-$value relative to the null hypothesis that the underlying disturbances are Normal. The $p-$values were computed by (i) simulating 1,000 artificial data sets using the fitted VAR, drawing the disturbances, u_t, from the multivariate Normal distribution with mean zero and variance-covariance matrix equal to its estimated sample analog and (ii) computing the percent of times that the empirically estimated statistic is exceeded by its analog computed across the artificial samples. The results are reported in Table 1. The kurtosis statistics are particularly large, and all kurtosis statistics but the one on the disturbance in the inflation equation have $p-$values that are less than one percent. The kurtosis statistic on the interest rate has a $p-$value that is less than 0.00 percent, after rounding. The kurtosis statistic on GDP growth is also particularly large, having a $p-$value of 0.02 percent. The skewness statistic on the interest rate and on inflation are both very large, having a $p-$value below one percent. I conclude that the evidence against the normality assumption is substantial.

Table 1: Empirical Skewness and Kurtosis Statistics, and Corresponding $p-$ values

Disturbance	Kurtosis		Skewness	
	statistic, $\mathbf{s}$	prob($s > \mathbf{s}$| normal)	statistic, $\mathbf{s}$	prob($s > \mathbf{s}$| normal)
$\log C_t/Y_t$	**1.40**	0.18%	-0.17	84.64%
$\log I_t/Y_t$	**1.59**	0.12%	0.07	33.04%
PCE inflation$_t$	**0.79**	2.00%	**0.41**	0.68%
$\log Y_t/l_t - \log W_t/P_t$	**1.04**	0.66%	**0.39**	1.22%
R_t	**11.01**	0.00%	**1.62**	0.00%
$\log l_t$	**1.27**	0.40%	0.11	26.46%
GDP growth$_t$	**1.85**	0.02%	-0.02	55.00%

Next, consider the priors over the VAR parameters. In practice, it is difficult to deviate from the usual Normal/Wishart assumption. This is because doing so results in a prior that is not conjugate with the Normal distribution, which then forces one to confront the formidable numerical integration problem discussed in the previous section. But, whether the DSSW strategy is compelling requires that one accept the Normal/Wishart specification as a reasonable specification of one's priors over VAR parameters. DSSW do take some steps in the direction of a defense. They show that under their assumptions, the height of the prior in Figure 1 is inversely proportional to how surprised one expects to be at VAR parameterizations different from $\Phi(\theta)$, $\Sigma(\theta)$. The surprise is measured by the expected drop in the likelihood, conditional on the DSGE model and on θ.

The DSSW defense of their specification of the prior over the VAR parameters may have some appeal. Perhaps it is consistent with the notion that in case one's most preferred DSGE model is wrong, the most likely alternative is somewhere nearby. However, for this type of argument to rationalize the type of Normal/Wishart prior distribution used here would seem to require an extraordinary coincidence. One could assess how well the Normal/Wishart represents a researcher's priors over VAR parameters by the following exercise. Assign a set of probabilities to a range of models, and a sub-probability distribution over the parameters of each model, conditional on that model being true. This specification of model priors induces a prior distribution over VAR parameters. If each DSGE model were not too similar, it seems safe to speculate that these priors over VAR parameters would have a very different shape - possibly with multiple local peaks - than what we see in Figure 1.

In sum, the use of the marginal likelihood to assess the plausibility of alternative hypotheses requires a number of detailed assumptions. The Normality assumption on the likelihood function seems outright inconsistent with the data. The primary advantage of the prior distribution over VAR parameters seems to lie with computational tractability. Use of the marginal likelihood to select models has some a priori appeal. But, this appeal rests on two propositions: that the right likelihood is used and that the prior distribution corresponds to the priors held by actual researchers. Evidence has been presented that the first proposition is false. The second proposition is yet to be established.

4. The DSSW Strategy: How it Works in Practice

The DSSW paper does not provide guidance on how exactly $\hat{\lambda}$ should be used to evaluate model fit. How big should $\hat{\lambda}$ be for one to feel comfortable about a DSGE model? How small does $\hat{\lambda}$ have to be to justify going back to the drawing board to redesign the model? In the DSSW paper, a value of $\hat{\lambda}$ in the region of 0.75 and 1.5 is presented. Does this mean the model fits well, or poorly, or something in between?

To shed light on these questions, I implemented Monte Carlo experiments using artificial data sets of length 200 observations generated using three DSGE models. One DSGE model, which I refer to as the RBC model, is the Long-Plosser real business cycle model. The other two DSGE models are different versions of the Clarida, Gali and Gertler (2000) (CGG) sticky price model (CGG1 and CGG2) (see also Gali, Lopez-Salido, and Valles (2003).) Three experiments were done. The first two use RBC and CGG1. In these experiments the econometrician computes $\hat{\lambda}$ knowing the true model, though not the values of some of its parameters. These experiments provide a sense of the sort of $\hat{\lambda}$'s to expect when the right model is in hand. In the third experiment, the true model is CGG2, but the econometrician mistakenly believes the true model is a version of the real business cycle model. This experiment provides a sense of the sort of $\hat{\lambda}$'s to expect when the econometrician's model is wrong. In the first subsection below I describe the technical aspects of the data generating mechanisms and experiments (this could be skipped in a quick read). In the second subsection I summarize the Monte Carlo results.

4.1. The Experiments

To help economize on computer time, each Monte Carlo experiment is designed so that the mapping from the DSGE model parameters estimated by the econometrician to a VAR representation for the data, $\Phi(\theta)$ and $\Sigma(\theta)$ is trivial. Multiple artificial data sets of length 200 observations are generated from a specific true DSGE model and the $\hat{\lambda}$ that solves (2.1) is computed in each data set. In each data set, the integral in $\mathcal{L}(Y, \lambda)$ is computed using Geweke's (1999) modified harmonic mean method and 100,000 MCMC trials. The procedure was tuned so that the acceptance rate in the MCMC trials averaged approximately 30 percent. Marco Del Negro and Frank Schorfheide kindly supplied me with their MATLAB code for the calculations.

As an additional step to keep the required computer time down, the maximization in (2.1) is restricted to the following set of possible values of λ:

$$0.11 \quad 0.25 \quad 0.43 \quad 0.67 \quad 1.00 \quad 1.5 \quad 2.33 \quad 4.00 \quad 9.00 \quad \infty \,.$$

When transformed into $\lambda/(1+\lambda)$, this corresponds to the ten equally spaced grid points, 0.1, 0.2, ..., 1.0. The quantity, $\lambda/(1+\lambda)$, is of interest because it corresponds to the relative weight assigned in the hybrid model to the DSGE model.

The RBC Model

The preferences, technology and shocks in the Long-Plosser model are as follows:

$$\begin{aligned} & E_0 \sum_{t=0}^{\infty} \beta^t \left[\log C_t - \frac{\exp(\tau_t)}{1+\psi} l_t^{1+\psi} \right], \\ C_t + K_{t+1} \quad \leq \quad & K_t^{\alpha} \left(\exp(z_t) \, l_t \right)^{(1-\alpha)} = Y_t, \\ & \tau_t, \; z_t : \text{ iid mean zero random variables, variance } \sigma_\tau^2, \; \sigma_z^2, \end{aligned}$$

where C_t denotes consumption, l_t denotes labor, K_{t+1} denotes capital, z_t is a technology shock, and τ_t is a preference shock. I set $\alpha = 1/3$, $\beta = 0.99$. The parameters whose values are estimated by the econometrician are $\theta = (\psi, \sigma_\tau^2, \sigma_z^2)$, whose true values are $(1, 0.02^2, 0.02^2)$. The econometrician's prior distribution for each parameter is inverted gamma, with mode equal to the corresponding true value. Specifically, denote the inverted gamma density for the random variable, x, by $f(x)$. Then,

$$f(x) = \frac{\zeta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\zeta}{x}\right),$$

where Γ is a gamma function. I assume $\alpha = 10$ and ζ is determined by the assumption on the mode of the distribution.

Artificial data sets on $k_{t+1} = \log(K_{t+1})$ and $\log(Y_t/l_t)$ are generated by the RBC model and provided to the econometrician, whose mapping from θ to a VAR representation is defined by

$$\begin{pmatrix} k_{t+1} \\ \log(Y_t/l_t) \end{pmatrix} = \begin{pmatrix} \gamma_k \\ \gamma_a \end{pmatrix} + \begin{bmatrix} \alpha & 0 \\ \alpha & 0 \end{bmatrix} \begin{pmatrix} k_t \\ \log(Y_{t-1}/l_{t-1}) \end{pmatrix} + \begin{pmatrix} (1-\alpha)\left(z_t - \frac{1}{1+\psi}\tau_t\right) \\ (1-\alpha) z_t + \frac{\alpha}{1+\psi}\tau_t \end{pmatrix},$$

where

$$\begin{aligned} \gamma_k &= \frac{1-\alpha}{1+\psi} \log\left(\frac{1-\alpha}{1-\beta\alpha}\right) + \log \beta\alpha \\ \gamma_a &= -\frac{\alpha}{1+\psi} \log\left(\frac{1-\alpha}{1-\beta\alpha}\right). \end{aligned}$$

This VAR representation can be derived from the wel-known fact that the solution to this model is given by $K_{t+1} = \beta\alpha Y_t$.

The CGG1 Model

In the CGG model the equilibrium allocations under a specific monetary policy rule (the 'equilibrium') are expressed as a deviation from the best equilibrium achievable when the monetary policy rule is dropped (the 'Ramsey equilibrium', or the 'natural equilibrium'). In the Ramsey equilibrium inflation, π_t, is always zero and the nominal rate of interest is given by

$$rr_t^* = \log \frac{1}{\beta} + \rho \Delta a_t + \frac{1-\zeta}{1+\varphi} \tau_t, \tag{4.1}$$

where β is the discount rate of the representative household (see Appendix A for the specification of preferences and technology underlying this economy).The equilibrium conditions are:

$$\beta E_t \pi_{t+1} + \kappa x_t - \pi_t = 0 \text{ (Calvo pricing equation)} \tag{4.2}$$

$$-\left[r_t - E_t \pi_{t+1} - rr_t^*\right] + E_t x_{t+1} - x_t = 0 \text{ (intertemporal Euler equation)} \tag{4.3}$$

$$u_t + \phi_\pi \pi_t - r_t = 0 \text{ (monetary policy rule),} \tag{4.4}$$

where r_t is the equilibrium nominal rate of interest and $x_t \equiv y_t - y_t^*$ is the deviation of equilibrium output, y_t, from Ramsey equilibrium output, y_t^*. Also,

$$\kappa = \frac{\left(1-\xi_p\right)\left(1-\beta\xi_p\right)(1+\varphi)}{\xi_p},$$

where ξ_p is the probability that an intermediate goods producer is not able to reoptimize its price in any given period. The monetary policy shock, u_t, the growth rate of technology, Δa_t, and a labor preference shock, τ_t, are assumed to have the following scalar first order autoregressive representations, respectively:

$$\begin{aligned} u_t &= \delta u_{t-1} + \eta_t \\ \Delta a_t &= \rho \Delta a_{t-1} + \varepsilon_t \\ \tau_t &= \zeta \tau_{t-1} + \varepsilon_t^\tau . \end{aligned}$$

In these expressions, the innovations are iid and have variances, σ_η^2, σ_ε^2, $\sigma_{\varepsilon^\tau}^2$, respectively. I adopt the following model parameterization:

$$\begin{aligned} \phi_\pi &= 1.5,\ \beta = 0.99,\ \varphi = 1,\ \rho = 0.2,\ \xi_p = 0.75,\ \delta = 0.2, \\ \zeta &= 0.5,\ \lambda_f = 1.25,\ \sigma_\eta = 0.005,\ \sigma_\varepsilon = 0.01,\ \sigma_{\varepsilon^\tau} = 0.006. \end{aligned}$$

To solve this model, first write the four equilibrium conditions, (4.1), (4.2), (4.3), and (4.4), in matrix form:

$$E_t \left[\alpha_0 z_{t+1} + \alpha_1 z_t + \beta_0 s_{t+1} + \beta_1 s_t \right] = 0,$$

where

$$z_t = \begin{pmatrix} \pi_t \\ x_t \\ r_t - \log(1/\beta) \\ rr_t^* - \log(1/\beta) \end{pmatrix}, \; s_t = \begin{pmatrix} \Delta a_t \\ u_t \\ \tau_t \end{pmatrix}, \; s_t = P s_{t-1} + \epsilon_t,$$

and α_0, α_1, β_0, β_1, P are functions of the model parameters. Then, the solution is

$$z_t = B s_t,$$

were the 4×3 matrix B uniquely solves the following linear system of equations:

$$(\beta_0 + \alpha_0 B) P + \beta_1 + \alpha_1 B = 0.$$

The vector of DSGE model parameters, θ, estimated by the econometrician has 8 dimensions, and is composed of φ, ξ_p and the two parameters associated with each of the three shocks. In the case of each element of θ, the prior is an inverted gamma distribution with mode equal to the true value of the parameter, and standard deviation equal to the true value of the parameter, divided by 2. The econometrician estimates a three variable VAR using data on x_t, π_t, and r_t. In Appendix B it is verified that with this DSGE, the data satisfy a first order VAR.

The CGG2 Model

In this version of the CGG model I replaced the technology shock process by the following stationary representation:

$$a_t = \rho a_{t-1} + \varepsilon_t,$$

where $\rho = 0.95$ and all other parameters are as in the CGG1 model. With this change, the equilibrium condition for the Ramsey rate of interest, (4.1), is replaced by

$$rr_t^* = \log \frac{1}{\beta} + (1-\rho)\, a_t + \frac{1-\zeta}{1+\varphi}\tau_t,$$

and log of the Ramsey level of output satisfies

$$y_t^* = \gamma + a_t - \frac{1}{1+\varphi}\tau_t,\ \gamma = -\frac{\log \lambda_f}{1+\varphi}. \tag{4.5}$$

Here, λ_f $(= 1.25)$ is a parameter that controls the elasticity of demand for intermediate goods and corresponds to the markup earned in steady state by monopolists in the model. Equilibrium output, Y_t, is obtained as follows:

$$\log Y_t = x_t + y_t^*,$$

and equilibrium employment, l_t, is obtained from:

$$\begin{aligned} \log l_t &= \log Y_t - a_t \\ &= x_t + \gamma - \frac{1}{1+\varphi}\tau_t. \end{aligned}$$

Artificial data on $\log Y_t$ and $\log l_t$ are generated using CGG2 and provided to the econometrician, who mistakenly assumes the data were generated by a version of the RBC model. In this version the preference shock has a first order autocorrelation structure

$$\tau_t = \rho_\tau \tau_{t-1} + \varepsilon_{\tau,t},$$

where $\rho_\tau = 0.9$ and $E\varepsilon_{\tau,t}^2 = \sigma_{\varepsilon_\tau}^2$. The first order, bivariate VAR representation that the econometrician (falsely) deduces for the data is

$$\begin{pmatrix} \log Y_t \\ \log l_t \end{pmatrix} = \begin{pmatrix} \gamma_Y \\ \gamma_l \end{pmatrix} + \begin{bmatrix} \alpha & (1-\alpha)\,\rho_\tau \\ 0 & \rho_\tau \end{bmatrix} \begin{pmatrix} \log Y_{t-1} \\ \log l_{t-1} \end{pmatrix} + \begin{pmatrix} (1-\alpha)\left(z_t - \frac{1}{1+\psi}\varepsilon_{\tau,t}\right) \\ -\frac{1}{1+\psi}\varepsilon_{\tau,t} \end{pmatrix},$$

where

$$\gamma_Y = \alpha \log(\alpha\beta) + (1-\alpha)\frac{1-\rho_\tau}{1+\psi}\log\left(\frac{1-\alpha}{1-\beta\alpha}\right),$$
$$\gamma_l = \frac{1-\rho_\tau}{1+\psi}\log\left(\frac{1-\alpha}{1-\beta\alpha}\right).$$

To derive the first row of this representation, note first that the log of the production function is

$$\log Y_t = \frac{1}{3}k_t + \frac{2}{3}\left(z_t + \log(l_t)\right).$$

Then, use the solution of the model to express k_t as a function of $\log Y_{t-1}$ and to express $\log(l_t)$ as a function ε_t^τ and $\log(l_{t-1})$.

The DSGE model parameters estimated by the econometrician are $\theta = (\psi, \sigma_z, \sigma_{\varepsilon_\tau})$. The econometrician is assumed to use an inverted gamma prior on each of the three parameters with mode $(1, 0.02, 0.02)$, respectively, and standard deviation equal to the mode, divided 7.

4.2. The Results

In this section, I present the results of the Monte Carlo experiments described above. Consider first the results when the data generating mechanism is RBC and the DSSW method is implemented with a bivariate, unrestricted VAR that has 1 lag (see RBC (lag 1) in Table 2). In this case, the unrestricted model has 9 free parameters (e.g., six parameters associated Φ and three with Σ), while the econometrician's DSGE model has 3 free parameters. Note that even though the econometrician has the true model in hand, a substantial fraction (6 percent) of artificial data sets result in a $\hat{\lambda}$ that assigns a relative weight of 1/2 or less to the DSGE model. The reason $\hat{\lambda}$ sometimes provides evidence against the DSGE model when it is true is that there is a positive probability of data sets in which the unrestricted VAR fits better than the true VAR. To see this, consider the likelihood ratio statistic formed from twice the difference of the log-likelihood associated with the estimated unrestricted VAR and the log-likelihood associated with the true VAR implied by the DSGE. Asymptotically, this is a realization from a Chi-square distribution with 9 degrees of freedom. The average value of this statistic over all data sets with the indicated value of $\hat{\lambda}$ is reported in the row beneath the results for RBC (lag 1). Note how these likelihood ratio statistics tend to be higher in data sets associated with a low $\hat{\lambda}$. With some exceptions, this general pattern is also a feature of the CGG1 experiment. The exceptions may reflect Monte Carlo sampling uncertainty.

There are two features of the log marginal likelihood in the RBC (lag 1) results that are worth emphasizing. One is illustrated in Figure 2. This shows a type of shape that occurs a nontrivial fraction of times in data generated by RBC (lag 1). In these cases the log, marginal likelihood is concave over most values of λ, and then rises sharply for λ near ∞. The abrupt change in the behavior of the log marginal likelihood for large values of λ seems puzzling. A consequence of this shape is that results are sensitive to the specification of the set of λ's over which the maximization in (2.1) is done. For example, if the upper bound is $\lambda = 200$ rather than $\lambda = \infty$, then the percent of artificial data sets with $\hat{\lambda}/\left(1+\hat{\lambda}\right) \leq 0.9$ is 64, rather than the 26.5 reported in Table 1. In the calculations for the oral presentation of this comment, the upper bound on λ was $\lambda = 5$ and the reported frequency of small $\hat{\lambda}$'s was even greater. To see this, note from Figure 2 that $\hat{\lambda}$ is 0.43 if the upper bound on the λ's considered in the maximization in (2.1) is 9, and $\hat{\lambda} = \infty$ if the upper bound on is ∞. The puzzling shape of the log, marginal likelihood in the case of RBC (lag 1) was not observed in the other Monte Carlo experiments.

A second notable feature of the log, marginal likelihood corresponding to RBC (lag 1) is that it exhibits very little variation across different values of λ. For example, the average difference between the log marginal likelihood at the smallest value of λ and the one associated with $\hat{\lambda}$ is only 7.7, which is considerably smaller than what is reported in the empirical example presented in DSSW. This finding is illustrated in Figure 3, which displays the log, marginal likelihood associated with a different artificial data set from the one underlying Figure 2. It bears emphasis that the log, likelihoods in Figures 2 and 3 are chosen for illustrative purposes. Both are atypical, in that they imply very low values of $\hat{\lambda}$.

The lack of variation in the log marginal likelihood motivated me to consider an unrestricted VAR with additional lags. The results are reported in Table 2, in the row labeled 'RBC (lag 4)'. In this case, $\hat{\lambda} = \infty$ in almost all the artificial datasets. Also, the mean difference of the log marginal likelihood at the lowest value of λ and at $\hat{\lambda}$ is now 32.4. This degree of variation in the log marginal likelihood is similar to what is reported in DSSW. In RBC (lag 4) the number of free parameters in the unrestricted VAR jumps to 21, versus the 3 free parameters in the econometrician's DSGE model. As emphasized by DSSW, the log marginal likelihood assigns a substantial penalty to free parameters and this is manifest here in the form of a sharp preference in favor of the DSGE model over the unrestricted VAR.

Consider now the CGG1 experiment with 1 lag. In this case, the VAR has three variables, and so the number of unrestricted parameters is 18, compared with the 8 free parameters of the econometrician's DSGE model. In this experiment small $\hat{\lambda}$'s are still possible, though this is less likely than it is in the case of RBC (lag 1). Thus, in 4 percent of the data sets, the relative weight assigned to the DSGE model is 60 percent or less in the case of CGG1 (lag 1), versus 10 percent for RBC (lag 1). Presumably, the improved performance of the DSSW method reflects the greater number of parameters in the unrestricted VAR in the CGG (lag 1) experiment than in the RBC (lag 1) experiment.

In CGG1 (lag 1), the average difference between the log marginal likelihood at the lowest value of λ and at $\hat{\lambda}$ is 16.6, which is smaller than the value reported for the empirical example reported by DSSW. This led me to consider the case in which there are four lags in the unrestricted VAR, raising the number of free parameters from 18 to 45, while the number of free parameters in the DSGE model remains at 8. As in the case of the RBC model, the frequency of low values of $\hat{\lambda}$ declines, though less dramatically than we saw in RBC (lag 4). In particular, 9.5 percent of the $\hat{\lambda}$'s assign a weight of 70 percent or less to the DSGE model in CGG1 (lag 4), versus zero percent for RBC (lag 4). In CGG1 (lag 4), the average difference between the log marginal likelihood at the lowest value of λ and at $\hat{\lambda}$ is 35.4, which is closer to the empirical example in DSSW.

In the lag 4 versions of both RBC and CGG1, there is a substantial difference between the log marginal likelihood at the lowest value of λ and at $\lambda = \hat{\lambda}$, as in DSSW. However, in my examples there is relatively less difference between the log marginal likelihood at $\lambda = \hat{\lambda}$ and at $\lambda = \infty$. For example, in CGG2 the mean decline in the log marginal likelihood from $\lambda = \hat{\lambda}$ to $\lambda = \infty$ when $\hat{\lambda} < \infty$ is a little less than unity. This is substantially smaller than the sharp drop in the log marginal likelihood reported in DSSW.

My next experiment, CGG2, investigates the behavior of $\hat{\lambda}$ and the log marginal likelihood when the econometrician's DSGE model is by construction false. In this case the data are generated by a version of the CGG model, but the econometrician's DSGE model is a version of the RBC model. The unrestricted VAR is estimated with 1 lag. Because two variables are included in the analysis, the unrestricted VAR has 9 free parameters. The econometrician's DSGE model has 3 free parameters. The results for this experiment are dramatic. In each artificial data set, $\hat{\lambda}$ is the lowest value of λ.Thus, the DSSW method correctly reveals, with probability one, that the

DSGE model is misspecified. Moreover, the slope of the log marginal likelihood is very steep, with the difference between the log marginal likelihood at the lowest and highest values of λ being on the order of 300-500. I presume that the finding that the unrestricted VAR is always the best model in this experiment is an artifact of the specification that it has only one lag. If more lags had been permitted in the unrestricted VAR, then $\hat{\lambda}$ would have exceeded the lowest value of λ at least occasionally. The feature of this example that I expect to be robust is that $\hat{\lambda}$ is substantially less than infinity, and that the log marginal likelihood declines steeply for $\lambda > \hat{\lambda}$.

Of the examples considered, the only one that can replicate DSSW's finding that the log marginal likelihood declines steeply for $\lambda > \hat{\lambda}$ is the one in which the econometrician's model is false. The two examples in which the econometrician's DSGE model is true do occasionally produce a $\hat{\lambda}$ substantially less than infinity. However, it is rare for the slope of the log marginal likelihood to be steeply negative for $\lambda > \hat{\lambda}$. These findings suggest that the DSSW method would have even greater power to identify evidence against DSGE models if the method formally integrated the slope of the log marginal likelihood for $\lambda > \hat{\lambda}$ into the diagnostic procedure.

Finally, recall the empirical evidence of leptokurtosis reported in the previous section. In principle, the use of the Normal likelihood in Bayesian analysis entails specification error. To investigate whether this error distorts the DSSW analysis, I redid the RBC experiment using disturbances that exhibit the amount of kurtosis observed in the data. My results were essentially unchanged from what is reported in Table 2, consistent with the proposition that the amount of leptokurtosis observed in the data is not enough to distort Bayesian analyses that use the Normal likelihood. Of course, these findings (as all other findings in Table 2!) are only indicative and need to be substantiated by similar additional experiments.

Table 2: Cumulative Distribution of $\hat{\lambda}$										
Experiment	$\frac{\hat{\lambda}}{1+\hat{\lambda}}$									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
RBC (lag1)	0	0	0.5	2.5	6.0	10.0	15.0	20.5	26.5	100
likelihood ratio	NA	NA	18	21.4	17.3	16.2	14.4	13.3	13.0	8.8
RBC (lag4)	0	0	0	0	0	0	0	0	0.5	100
likelihood ratio	NA	NA	NA	NA	NA	NA	NA	NA	27.5	22.5
CGG1 (lag1)	0.5	1.0	1.0	1.0	1.5	4.0	11.0	24.5	53.0	100
likelihood ratio	14.3	12.5	12.5	12.5	15.2	20.7	21.7	18.9	16.9	15.3
CGG1 (lag4)	0	0	0	0	0	2.0	9.5	23.5	50.5	100
likelihood ratio	NA	NA	NA	NA	NA	70.5	59.4	54.3	49.1	45.1
CGG2	100	100	100	100	100	100	100	100	100	100

Notes: (i) Entries indicate percent, out of 200 simulations, that $\hat{\lambda}$ is less than or equal to value indicated in column heading; (ii) (lag x): x indicates the number of lags in the unrestricted VAR; (iii) likelihood ratio - average likelihood ratio (LR) statistic over all artificial data sets having $\hat{\lambda}$ indicated in column heading, where LR is twice the difference of log-likelihood of unrestricted VAR versus log likelihood of true VAR; (iv) NA indicates 'Not Applicable' because there were no $\hat{\lambda}$'s in this entry.

5. Conclusion

DSSW have provided a valuable service in describing and implementing a measure of fit for DSGE models. The Monte Carlo evidence presented in this comment suggests four ways in which the DSSW measure of fit could be made even more useful:

- it would be useful if a lower cutoff value of $\hat{\lambda}$ were provided, such that for smaller $\hat{\lambda}$, the researcher knows there is with high probability a problem with the DSGE model. The Monte Carlo experiments in my comment suggest that such a cutoff would be a function of, among other things, the difference between the number of free parameters in the unrestricted VAR and in the DSGE model.

- the rate at which the marginal likelihood declines for $\lambda > \hat{\lambda}$ should be integrated formally into the DSSW procedure. The Monte Carlo experiments suggest that a steep rate of decline is a reliable signal that the econometrician's DSGE model fits poorly. There are various ways this rate of decline could be measured. One way would be to report Bayesian probability intervals for λ.

- in the absence of a stronger defense for the priors used in the DSSW analysis, it would be useful to have evidence that results based on the DSSW priors are robust to plausible alternatives. A practical impediment to evaluating robustness is that priors which deviate from DSSW's are unlikely to have convenient conjugacy properties. As a result, the numerical integration problem in (2.1) would be computationally very burdensome in practical situations. Still, robustness could be studied in the type of simple examples considered in my comment, where computational limitations are less binding.

- I have provided evidence that the DSSW results are robust to the kind of evidence against Normality one observes in the data. A more systematic investigation of robustness would be useful.

DSSW compare their procedure with alternative measures of model fit based on out-of-sample forecasting performance. Further comparisons of this type would be of interest. Measures of out-of-sample forecasting performance appear to offer at least four advantages over the DSSW method:

1. the computational burden is minimal by comparison with the substantial resources required to evaluate (2.1).

2. computational tractability limits the range of model comparisons that can be done with the DSSW method, while there is no limit to the models one can compare under out-of-sample forecasting criteria. For example, one can compare the forecasting performance of the DSGE model with that of a Bayesian VAR, as in Ingram and Whiteman (1994). In practice, Bayesian VAR's are more useful at forecasting than unrestricted VAR's because of parameter parsimony. Alternatively, one could compare DSSW's hybrid model under alternative specification of λ with a Bayesian VAR under the out-of-sample forecasting criterion.

3. classical sampling theory offers some assistance in determining whether differences in the out-of-sample root mean square error performance of alternative models are statistically significant (see, e.g., Christiano 1990, Appendix D). This contrasts with the DSSW method, in which a small $\hat{\lambda}$ suggests the presence of evidence against a DSGE model, but there is no guidance, as yet, on how small such a $\hat{\lambda}$ must be (see the first bullet point above).

4. the out-of-sample forecast performance criterion is transparent and is of obvious interest to everyone. By contrast, for the marginal likelihood to be compelling, one must first confront several difficult - in some cases, possibly unresolvable - questions. What likelihood is appropriate for the data? Are the researcher's priors faithfully captured by the choice of prior distribution? Implicit in the

> DSSW procedure is the assumption that one prior is suitable for everyone. But, why should researchers with different backgrounds and experiences use the same prior?

DSSW show that an out-of-sample forecast root mean square error criterion produces fit results similar to what the DSSW procedure produces. In view of the advantages of the out-of-sample forecasting approach, DSSW's findings would appear to be a powerful argument in its favor.

A. Preferences and Technology Underlying Clarida-Gali-Gertler Model.

Although the preferences and technology underlying the Clarida-Gali-Gertler model are well known, we include them here for completeness. In particular, the representative household's preferences are:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(\log C_t - \exp(\tau_t) \frac{l_t^{1+\varphi}}{1+\varphi} \right), \ \varphi > 0,$$

where C_t and l_t denote consumption and employment, respectively, and τ_t is a labor supply shock. A budget constraint allows the household to finance consumption by participating in a competitive labor market and by participating in a loan market in which the log of the gross nominal rate of interest is r_t. In equilibrium the loan market must clear with zero trade.

Final output is produced by competitive firms using intermediate intputs, $Y_t(i)$, $i \in (0,1)$ using the following technology:

$$Y_t = \left(\int_0^1 Y_t(i)^{\frac{1}{\lambda_f}} di \right)^{\lambda_f}, \ \lambda_f \geq 1.$$

The technology for producing $Y_t(i)$ is

$$Y_t(i) = A_t l_t(i), \ a_t = \log(A_t),$$

where $l_t(i)$ is employment by the i^{th} intermediate good producer. This producer, subject to Calvo-sticky price frictions, is able to reoptimize its price with probability $1-\xi_p$. With the complementary probability, the intermediate good producer cannot change its price. In steady state, equilibrium inflation is zero, $Y_t = C_t$, and $l_t = \int_0^1 l_t(i)\, di$.. In the text, Δa_t denotes the first difference of a_t.

B. First Order VAR Representation of CGG1 Model

Let $\tilde{z}_t$ denote the 3×1 vector composed of the first three elements of z_t and let $\tilde{B}$ denote the first three rows of B, so that $\tilde{B}$ is a square matrix. I found that B is invertible in the numerical examples I considered. The solution for $\tilde{z}_t$ is written $\tilde{z}_t = \tilde{B}s_t$ or,

$$\tilde{B}^{-1}\tilde{z}_t = s_t.$$

Then, multiply on the left by the matrix lag operator, $I - PL$, to obtain:

$$(I - PL)\,\tilde{B}^{-1}\tilde{z}_t = \epsilon_t,$$

or,

$$\tilde{B}^{-1}\tilde{z}_t = P\tilde{B}^{-1}\tilde{z}_{t-1} + \epsilon_t.$$

Multiply on the left by $\tilde{B}$ to obtain the first order VAR representation for $\tilde{z}_t$:

$$\tilde{z}_t = \tilde{B}P\tilde{B}^{-1}\tilde{z}_{t-1} + \tilde{B}\epsilon_t.$$

Then, $\Phi\,(\theta)$ is constructed from $\tilde{B}P\tilde{B}^{-1}$ and $\Sigma\,(\theta)$ corresponds to $\tilde{B}V\tilde{B}'$, where V is the variance-covariance matrix of ϵ_t. This establishes that the variables in the CGG1 model has a first order, VAR representation.

References

[1] Christiano, Lawrence J., 1990, 'P^* : Not the Inflation Forecaster's Holy Grail,' Federal Reserve Bank of Minneapolis, *Quarterly Review*.

[2] Clarida, Richard, Jordi Gali and Mark Gertler, 2000, 'Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory,' *Quarterly Journal of Economics*, 115, 147-180.

[3] Del Negro, Marco, and Frank Schorfheide, 2006, 'Priors from General Equilibrium Models for VARs,' *International Economic Review*, 45, 643-673.

[4] Gali, Jordi, J. David Lopez-Salido, and Javier Valles, 2003, "Technology Shocks and Monetary Policy: Assessing the Fed's Performance." *Journal of Monetary Economics*. 50 (May): 723 - 743.

[5] Geweke, John, 1999, 'Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication,' *Econometric Reviews*, 18, 1-126.

[6] Ingram, Beth F. and Charles H. Whiteman, 1994, "Supplanting the 'Minnesota' prior Forecasting macroeconomic time series using real business cycle model priors", *Journal of Monetary Economics*, 34, 497-510.

Figure 1: Priors Over VAR Parameters

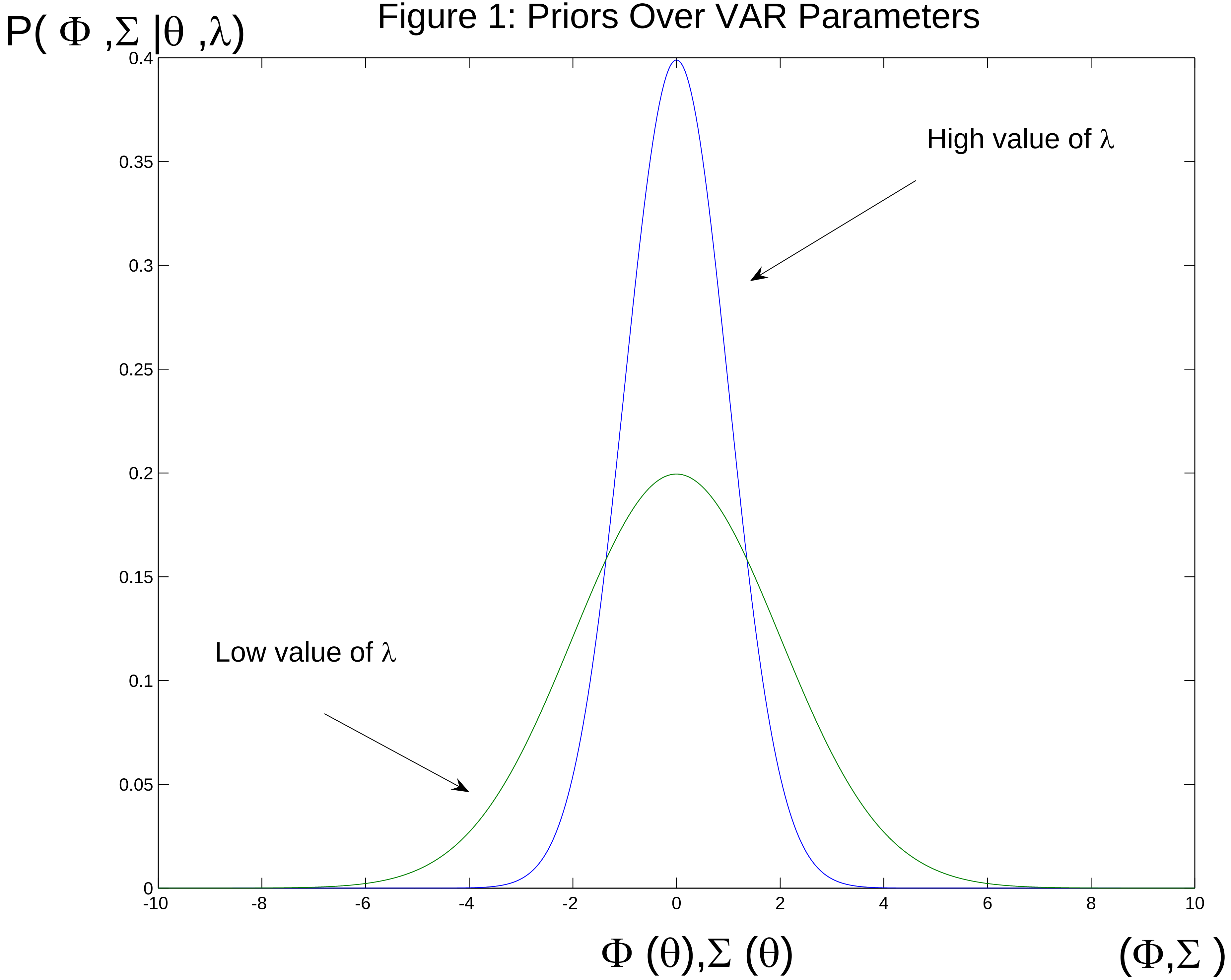

Figure 2: Realization of log, Marginal Likelihood, RBC (lag 1)

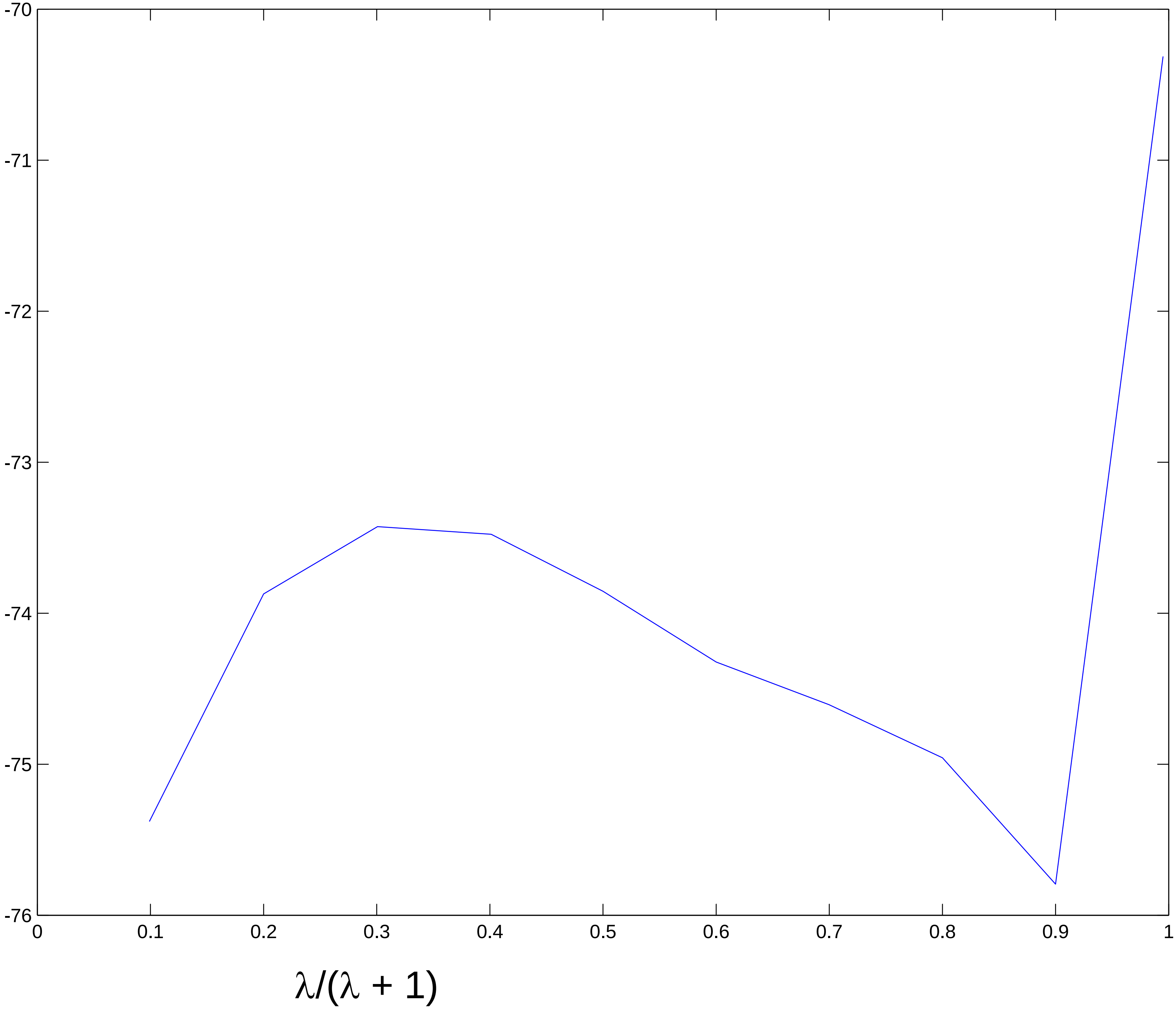

Figure 3: Realization of log, Marginal Likelihood, RBC (lag1)