

DSGE MODELS WITH STUDENT- t ERRORS

Siddhartha Chib¹ and Srikanth Ramamurthy²

¹*Olin Business School, Washington University in St. Louis, St. Louis, Missouri, USA*

²*Sellinger School of Business, Loyola University Maryland, Baltimore, Maryland, USA*

□ *This paper deals with Dynamic Stochastic General Equilibrium (DSGE) models under a multivariate student- t distribution for the structural shocks. Based on the solution algorithm of Klein (2000) and the gamma-normal representation of the t -distribution, the TaRB-MH algorithm of Chib and Ramamurthy (2010) is used to estimate the model. A technique for estimating the marginal likelihood of the DSGE student- t model is also provided. The methodologies are illustrated first with simulated data and then with the DSGE model of Ireland (2004) where the results support the t -error model in relation to the Gaussian model.*

Keywords Bayesian inference; Marginal likelihood; MCMC; Metropolis-Hastings; Particle-filtering; State-space model; Multivariate-student- t distribution.

JEL Classification C01; C11; C15; C22; C32; E27; E37; E47.

1. INTRODUCTION

Dynamic Stochastic General Equilibrium (DSGE) models are a useful and popular framework for examining macroeconomic dynamics (Christiano et al., 2005). In these models, households and firms are modeled in an optimizing, stochastic, multiperiod setting, and the behavior of these agents is combined with those of the government and the monetary authorities. The first order conditions of this structural model are log-linearized and the reduced form dynamics of the endogenous variables, such as consumption, output, inflation and interest rates, are obtained through a solve step using a method specifically designed for linear dynamic systems that contain conditional expectations of future endogenous variables. The solve step produces reduced form dynamics that can be expressed in a linear state space model (SSM)

Address correspondence to Siddhartha Chib, Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Bookings Drive, St. Louis, MO 63130, USA; E-mail: chib@wustl.edu

but with the complication that the parameters that define the SSM are analytically intractable functions of the structural parameters. The stochastic component of this SSM arises from shocks that affect, for example, the utility and production functions in the structural model. These structural shocks are conventionally modeled as a vector autoregressive process with Gaussian innovations. The assumption of Gaussianity is particularly convenient in the context of a SSM. Analysis of such models is now primarily implemented by Bayesian methods on account of the developments recorded in Lubik and Schorfheide (2004), Smets and Wouters (2003, 2007), Justiniano and Primiceri (2008), Fernandez-Villaverde and Rubio-Ramirez (2005), Sims et al. (2008), and Chib and Ramamurthy (2010).

Our goal in this paper is to expand on this literature by providing a framework for estimating and comparing DSGE models under a multivariate student- t distribution for the structural shocks. We consider this alternative assumption because the distributional assumption concerning the innovations is not intrinsically a part of the structural model. We focus on the student- t alternative because (as we show) it maintains the tractability of the Gaussian model and also allows one to encompass the Gaussian case by varying the degrees of freedom of the t distribution. The choice of this distribution is also particularly meaningful in a special issue in honor of Arnold Zellner since the use of the student- t as an alternative to the Gaussian error was pioneered by Zellner (1976) in the regression context. That early work in the regression setting was further extended by Chib et al. (1988), and to various other models by Kitagawa (1987), Albert and Chib (1993), and Carlin et al. (1992), to name a few. In the context of DSGE models, Curdia et al. (2012), in independent work, also consider fat tailed innovations within a DSGE setup. The modeling approach in that paper is somewhat different from that pursued here in that they allow for low frequency changes in the volatility of the structural shocks. This also translates into a slightly different estimation strategy.

In this paper, we show that by exploiting the well know representation of a t distribution as a gamma scale mixture of normals (for example Carlin et al., 1992), it is possible to develop a complete Bayesian approach for the DSGE- t model. The approach we develop relies crucially on the method of Klein (2000) to solve the structural model conditioned on the gamma variables. The Klein method is especially useful for the t -model because the original process for the innovations is preserved in the reduced form SSM. The approach we discuss is much simpler than (say) a particle filtering based summary of the posterior distribution that does not involve the scale mixture representation of the t -distribution. We conclude our inferential treatment of the model by showing how the model marginal likelihood can be found by the method of Chib (1995). In this connection,

we calculate the likelihood ordinate by a particle filter method based on Chib et al. (2002) and the posterior ordinate by the method of Chib and Jeliazkov (2001).

The rest of the paper is organized as follows. In the following section we describe the model setup and provide details of the methodology for estimation and model comparison. Section 3 illustrates these techniques with the aid of two examples. The first deals with a SSM that is designed to mimic the structure of the reduced form of DSGE models. The next example is the DSGE model in Ireland (2004) where we document support for the student- t model. Section 4 concludes.

2. THE SETUP

Consider the class of structural DSGE models in which the first order conditions are log-linearized and the reduced form dynamics of the endogenous variables $x_t : k \times 1$, say, are obtained by the solution method of Blanchard and Kahn (1980), as generalized in Klein (2000). As mentioned earlier, the parameters in this reduced form are the output of a highly nonlinear, analytically intractable map from the set of structural parameters ψ . Partition $x_t = (k_t, d_t)$ into the “forward looking variables” k_t and the “backward-looking variables”¹ d_t , and let $z_t : q \times 1$ denote the innovations in the model whose evolution is defined by the stationary VAR process

$$z_{t+1} = Pz_t + v_{t+1}, \quad (1)$$

where, for simplicity, P is assumed to be a diagonal matrix, $v_{t+1} \stackrel{i.i.d.}{\sim} t_q(0, \Sigma, \nu)$, a multivariate- t distribution with diagonal dispersion matrix $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_q^2)$, and ν degrees of freedom. Combining the reduced form dynamics of the endogenous variables from the Klein method with the innovation process, one can write

$$\begin{pmatrix} k_t \\ d_{t+1} \\ z_{t+1} \end{pmatrix} = \begin{bmatrix} 0 & G_{kd} & G_{kz} \\ 0 & G_{dd} & G_{dz} \\ 0 & 0 & P \end{bmatrix} \begin{pmatrix} k_{t-1} \\ d_t \\ z_t \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix} v_{t+1}, \quad (2)$$

where the matrices G_{ij} , delivered by the solution, are functions of ψ . We denote the latter system compactly by

$$\alpha_{t+1} = G\alpha_t + Fv_{t+1}$$

¹As defined in Blanchard and Kahn (1980), backward and forward looking refer to ‘predetermined’ and ‘non-predetermined’ variables, respectively. Klein (2000) provides a generalization of these definitions. For instance, a backward-looking variable is one that has an exogenously specified initial value and prediction error.

The goal, therefore, is to numerically summarize this posterior based on a sample of draws $\{\theta^g\}_{g=1}^N$ from it by a suitable MCMC procedure. Before we describe the MCMC procedure, note that, unlike in the Gaussian case, it is not easy to calculate the likelihood function marginalized of the states. It requires the use of the computationally expensive particle filter. Conditioned on $\lambda = \{\lambda_{jt}\}$, however, the quantity $f(y|\theta, \lambda)$ is available from the Kalman filter. Hence, the conditional posterior $\pi(\theta|y, \lambda)$ can be sampled by the Tailored Randomized Block-Metropolis Hastings (TaRB-MH) method of Chib and Ramamurthy (2010). In turn, given the conditionally Gaussian structure of the system, $\{\alpha_t\}|\lambda, \theta$ and $\lambda|\{\alpha_t\}, \theta$ can be sampled directly. This is the approach that we take in our MCMC sampling, which, in effect, involves drawing from the joint posterior of θ , λ and the states $\{\alpha_t\}$. This approach is summarized in the following steps.

Algorithm: MCMC Algorithm for DSGE Models with Student- t Errors.

1. Set $g = 1$ and initialize the parameters θ^0 and λ^0 .
2. Sample $\theta^g | y, \lambda^{g-1}$ by TaRB-MH.
3. Sample $\{\alpha_t\}^g, \lambda^g | y, \theta^g, t \leq n$ by drawing.
 - (a) $\{\alpha_t\}^g$ from $\{\alpha_t\} | y, \theta^g, \lambda^{g-1}$.
 - (b) λ^g from $\lambda_{jt} | y, \theta^g, \{\alpha_t\}^g \sim \text{gamma}\left(\frac{v_j+1}{2}, \frac{v_j+\sigma_j^{-2g}(\alpha_{jt+1}^g - G_j \alpha_t^g)^2}{2}\right), j \leq q, t \leq n,$
and G_j is the j th row of G in (4).
4. Increment g to $g + 1$ and go to 2.
5. Return the sample $\{\theta^g\}_{g=1}^N$.

Further details related to these steps now follow.

2.1.1. Sampling θ

Although the general procedure for sampling θ is straightforward, generating an efficient sample in step 2 above is challenging. The source of this problem is the solution step required to construct the matrices in the SSM. Consequently, to calculate $f(y|\theta, \lambda)$ for a given value of θ , one first needs to solve the model. As mentioned earlier, the solution scrambles the parameters such they cannot be delineated even in a small scale problem that is of economic interest. As a result, the likelihood function is a complicated object in these models, displaying erratic behavior for small changes in parameter values. This is a well documented fact in the DSGE literature (see, for instance, the survey paper by An and Schorfheide, 2007). For this reason, “off the rack” sampling schemes, like the single block Random Walk M-H (RW-MH), do not generate draws that mix well

in the parameter space. This issue was raised in Chib and Ramamurthy (2010) and is the key motivation for the TaRB-MH algorithm. Extensive details of the algorithm and comparison to the RW-MH in the context of several examples are available in that paper. Here we focus on the practical implementation of the procedure in the context of the current setup.

In the g th iteration, begin by creating a random partition of blocks $(\theta_1, \dots, \theta_B)$ from a permuted sequence of θ . The goal is to sample each of these B blocks in turn by a tailored M-H scheme (introduced originally in Chib and Greenberg, 1994). Consider the l th block, $l = 1, \dots, B$, and suppose that blocks $1, \dots, l - 1$ denoted by Ψ_{l-1} have been updated with current values Ψ_{l-1}^g . Also, let Ψ^l be the blocks $\theta_l, \dots, \theta_B$ with current values $\Psi^{l,g-1}$ from the previous iteration, and likewise, let λ^{g-1} be the most recent value of λ . To update the l th block.

1. Calculate

$$\hat{\theta}_l = \arg \max_{\theta_l} \log f(y | \Psi_{l-1}^g, \theta_l, \Psi^{l+1,g-1}, \lambda^{g-1}) \pi(\theta_l)$$

by simulated annealing (or another suitable numerical optimization procedure).

2. Calculate the inverse of the negative Hessian evaluated at $\hat{\theta}_l$

$$\hat{V}_l = \left(- \frac{\partial^2 \log f(y | \Psi_{l-1}^g, \theta_l, \Psi^{l+1,g-1}, \lambda^{g-1}) \pi(\theta_l)}{\partial \theta_l \theta_l'} \right)^{-1} \Bigg|_{\theta_l = \hat{\theta}_l}.$$

3. Simulate

$$\theta_l^g \sim t(\hat{\theta}_l, \hat{V}_l, \nu_l),$$

where the degrees of freedom ν_l in the student- t proposal density is set to a low value, say, 15.

4. Accept θ_l^g as the updated value of θ_l with probability α given by

$$\alpha = \min \left\{ \frac{f(y | \Psi_{l-1}^g, \theta_l^g, \Psi^{l+1,g-1}, \lambda^{g-1}) \pi(\theta_l^g)}{f(y | \Psi_{l-1}^g, \theta_l^{g-1}, \Psi^{l+1,g-1}, \lambda^{g-1}) \pi(\theta_l^{g-1})} \frac{t(\theta_l^{g-1} | \hat{\theta}_l, \hat{V}_l, \nu_l)}{t(\theta_l^g | \hat{\theta}_l, \hat{V}_l, \nu_l)}, 1 \right\}.$$

The g th iteration is complete when all B blocks have been updated as above.

2.1.2. Sampling α

The general principle for sampling the states by calculating the smoothed moments follows Carter and Kohn (1994). This is commonly referred to as the “Filter-Forward-Sample-Backward” scheme. The filtering

procedure is described in almost any graduate level textbook of time series. For instance, following Hamilton (1994), one first iterates for $t = 1, \dots, n$ the Kalman filter

$$\begin{aligned} e_t &= y_t - a - B\hat{\alpha}_{t|t-1}, & f_t &= B\hat{R}_{t|t-1}B', & K_t &= G\hat{R}_{t|t-1}B'f_t^{-1}, \\ \hat{\alpha}_{t|t} &= \hat{\alpha}_{t|t-1} + \hat{R}_{t|t-1}B'f_t^{-1}e_t, & \hat{\alpha}_{t+1|t} &= G\hat{\alpha}_{t|t}, \\ \hat{R}_{t|t} &= \hat{R}_{t|t-1} - \hat{R}_{t|t-1}B'f_t^{-1}B\hat{R}_{t|t-1}, & \hat{R}_{t+1|t} &= G\hat{R}_{t|t}G' + F\Sigma_tF', \end{aligned}$$

where $\Sigma_t = \text{diag}(\sigma_j^2/\lambda_{jt})$, storing for $t = 1, \dots, n$, $\hat{\alpha}_{t|t}$ and $\hat{R}_{t|t}$, and for $t = 0, \dots, n-1$, $\hat{R}_{t+1|t}$ with initializations $\hat{\alpha}_{1|0} = 0$ and $\text{vec}(\hat{R}_{1|0}) = (I - G \otimes G)^{-1} \text{vec}(F\Sigma_1F')$.

Then in the backward simulation one draws for $t = n, \dots, 1$

$$\alpha_t \sim \mathcal{N}(\alpha_{t|n}, R_{t|n}),$$

where the moments of the normal distribution are calculated as, for $t = n$, $\hat{\alpha}_{n|n}$ and $\hat{R}_{n|n}$ obtained from the terminal values in the forward iteration, and for $t = n-1, \dots, 1$

$$\alpha_{t|n} = \hat{\alpha}_{t|t} + J_t(\alpha_{t+1} - G\hat{\alpha}_{t|t}), \quad R_{t|n} = \hat{R}_{t|t} + J_t(\hat{R}_{t+1|n} - \hat{R}_{t+1|t})J_t',$$

where $J_t = \hat{R}_{t|t}G'\hat{R}_{t+1|t}^{-1}$.

We make the following implementational remark concerning the the lack of measurement errors and the evolution errors Fv_t . Typically, in empirical DSGE models, the number of measurements m is less than the number of states p . Thus, despite the lack of measurement errors, only certain states are inferred without errors. Further, as noted earlier, because only z_t is truly latent, only these need sampling in the backward step. Consequently, when calculating $\hat{R}_{t+1|t}^{-1}$, only the inverse of the lower $q \times q$ submatrix needs to be calculated. Then given z_t , the remaining states are automatically updated by the evolution process.

2.1.3. Sampling λ

Sampling λ follows the standard gamma update (see, for instance, Albert and Chib, 1993). Because of the structure of the state transition equation emanating from the solution to the model, sampling $\lambda_{jt} | y, \theta, \alpha_t$ is equivalent to sampling from $\lambda_{jt} | \theta, z_t$. Furthermore, because of the independence of the innovations, sampling λ further simplifies to drawing

$$\lambda_{jt} | \theta, z_t \sim \text{gamma} \left(\frac{v_j + 1}{2}, \frac{v_j + \sigma_j^{-2}(z_{jt+1} - \rho_j z_{jt})^2}{2} \right),$$

where ρ_j is the (j, j) th element of P .

2.2. Model Comparison

The advantage of the student- t model for the structural shocks is that it allows for different error specifications by simply varying the degrees of freedom v_j in the t -density. Subsequently, it might be of independent interest to the researcher to compare the fit of these various models to the data. For instance, the objective could be to compare the fit between the Gaussian error model and another with fatter tails. This can be easily accomplished by setting a large v_j , say 200, for the former and a small v_j , say 10, for the latter.

To facilitate model comparison by marginal likelihood, we follow the approach of Chib (1995) and Chib and Jeliazkov (2001) adapted to the TaRB-MH scheme by Chib and Ramamurthy (2010). The starting point of the Chib method is the basic marginal likelihood identity

$$m(y) = \frac{f(y | \theta)\pi(\theta)}{\pi(\theta | y)}, \tag{6}$$

which is evaluated at a value θ^* , usually taken to be a high density point in the support of the posterior. Since the ordinate of the prior density at θ^* is readily available, we now describe how the remaining two terms are estimated. To estimate the denominator, we first decompose the posterior ordinate as

$$\pi(\theta^* | y) = \pi(\theta_1^* | y)\pi(\theta_2^* | y, \theta_1^*) \dots \pi(\theta_B^* | y, \theta_1^*, \dots, \theta_{B-1}^*), \tag{7}$$

where $\theta_1, \dots, \theta_B$ is a (random) partition of θ into B blocks, and estimate each of the components on the right from the output of the main TaRB-MH algorithm described above and the output from subsequent reduced MCMC runs. As shown in Chib and Jeliazkov (2001), a simulation consistent estimate of the typical l th ordinate $\pi(\theta_l^* | y, \theta_1^*, \dots, \theta_{l-1}^*)$, $l = 1, \dots, B$, is obtained as

$$\begin{aligned} &\hat{\pi}(\theta_l^* | y, \theta_1^*, \dots, \theta_{l-1}^*) \\ &= \frac{n_1^{-1} \sum_{g=1}^{n_1} \alpha(\theta_l^g, \theta_l^* | y, \Psi_{l-1}^*, \Psi^{l+1,g}, \lambda^g) q_l(\theta_l^* | \Psi_{l-1}^*, \Psi^{l+1,g}, y)}{n_1^{-1} \sum_{j=1}^{n_1} \alpha(\theta_l^*, \theta_l^j | y, \Psi_{l-1}^*, \Psi^{l+1,j}, \lambda^g)}, \tag{8} \end{aligned}$$

where Ψ_{l-1} and Ψ^{l+1} are as defined earlier, the numerator α is the probability of moving from θ_l^g to θ_l^* , the denominator α is the probability of moving from θ_l^* to θ_l^j and q is the proposal density evaluated at θ_l^* . Practically, therefore, one iteratively draws N samples $\{(\theta_l^g, \theta_{l+1}^g, \dots, \theta_B^g), \lambda^{(g)}\}_{g=1}^N$ by

TaRB-MH, randomizing only over the parameters in Ψ^{l+1} . Then for each draw one calculates the numerator α as

$$\min \left\{ \frac{f(y | \Psi_{l-1}^*, \theta_l^*, \Psi^{l+1,g}, \lambda^g) \pi(\theta_l^*) t(\theta_l^g | \hat{\theta}_l, V_l, v_l)}{f(y | \Psi_{l-1}^*, \theta_l^g, \Psi^{l+1,g}, \lambda^g) \pi(\theta_l^g) t(\theta_l^g | \hat{\theta}_l, V_l, v_l)}, 1 \right\}$$

and q_l as $t(\theta_l^* | \hat{\theta}_l, V_l, v_l)$. It is important to note the conditioning on Ψ_{l-1}^* above when calculating the moments of the proposal density. The denominator is more straightforward. In each iteration, one simply supplements the numerator draws by an additional draw of θ_l from $t(\theta_l | \hat{\theta}_l, V_l, v_l)$, taking care to respect the conditioning on the remaining parameters when calculating $\hat{\theta}_l$ and V_l , to evaluate the denominator α .

We make the following remarks.

1. Calculating the posterior ordinate as above requires minimal programming requirements beyond the estimation step.
2. For $l = 1$, the draws $\{(\theta_1^g, \theta_2^g, \dots, \theta_B^g), \lambda^g\}_{g=1}^N$ is a sample from $\pi(\theta | y)$. These draws can be conveniently recycled from the draws in the estimation step.

We now describe the computation of the likelihood ordinate $f(y | \theta^*)$. Note that this is the likelihood value evaluated at θ^* marginalized over λ . Because marginalized over λ the SSM is no longer Gaussian, the likelihood function cannot be calculated by the Kalman filter. Instead, the one step ahead density

$$f(y_t | Y_{t-1}, \theta^*) = \int f(y_t | \alpha_t, \theta^*) f(\alpha_t | Y_{t-1}, \theta^*) d\alpha_t, \quad (9)$$

where $Y_{t-1} = (y_1, \dots, y_{t-1})$ and

$$f(\alpha_t | Y_{t-1}, \theta^*) = \int f(\alpha_t | \alpha_{t-1}, \theta^*) f(\alpha_{t-1} | Y_{t-1}, \theta^*) d\alpha_{t-1}, \quad (10)$$

can be estimated from the output of a particle filter. It may be noted that the density $f(y_t | \alpha_t, \theta^*)$ in DSGE models is degenerate. In this case, one needs to assume a small Gaussian measurement error to define this density. We discuss this further in the context of the simulated data example. With this simplification, the particles from $\{f(\alpha_t | Y_{t-1}, \theta^*)\}_{t=1}^n$, starting with particles from $\alpha_0 | Y_0, \theta^*$, can be used to form the likelihood estimate. In our work, we adapt the version of the Pitt and Shephard (1999) auxiliary particle filter that was given in Chib et al. (2002). It is defined through the following steps.

Algorithm: Particle Filter for Calculating $f(y | \theta^*)$.

1. Initialize $t = 1$ and draw a sample $\{\alpha_{0|0}^j = (s_{0|0}^j, z_{0|0}^j)\}_{j=1}^M$ from $f(\alpha_0 | Y_0, \theta^*)$.
2. Simulate

$$z_t^j \sim t(Pz_{t-1|t-1}^j, \Sigma, v),$$

and calculate $s_t^j = G_s \alpha_{t-1|t-1}^j$.

3. Estimate the one step ahead density as

$$\hat{f}(y_t | Y_{t-1}, \theta^*) = \frac{1}{M} \sum_{j=1}^M \mathcal{N}_m\{y_t | a + B\alpha_t^j, \Omega\},$$

where $\alpha_t^j = [s_t^j, z_t^j]$ and $\Omega : m \times m$ is a diagonal matrix with ω_{ii} small such as 10^{-3} or 10^{-5} .

4. Calculate $\hat{\alpha}_{t|t-1}^j = G \alpha_{t-1|t-1}^j$.
5. Calculate $p^j = \mathcal{N}_m\{y_t | a + B\hat{\alpha}_{t|t-1}^j, \Omega\}$ and $w^j = \frac{p^j}{\sum p^j}$.
6. Draw R values $\hat{\alpha}_{t|t-1}^r$ and $\alpha_{t-1|t-1}^r, r = 1, \dots, R$, with replacement from $\{\hat{\alpha}_{t|t-1}^j\}$ and $\{\alpha_{t-1|t-1}^j\}$ with respective probability w^j , where R is at least five times larger than M .
7. For each $\hat{\alpha}_{t|t-1}^r$, simulate

$$z_t^r \sim t(Pz_{t-1|t-1}^r, \Sigma, v),$$

and calculate $s_t^r = G_s \alpha_{t-1|t-1}^r$.

8. Calculate $p_n^r = \mathcal{N}_m\{y_t | a + B\alpha_t^r, \Omega\}, p_d^r = \mathcal{N}_m\{y_t | a + B\hat{\alpha}_{t|t-1}^r, \Omega\}, w_i^r = \frac{p_i^r}{\sum p_i^r}, i = n, d$, and $w^r = \frac{w_n^r}{w_d^r}$.
9. Draw M values from $\{\alpha_t^r\}$ with probabilities w^r to produce M values from $f(\alpha_t | Y_t, \theta^*)$.
10. Increment t to $t + 1$ and goto 2.

Finally, calculate the log likelihood ordinate

$$\log f(y | \theta^*) = \sum_{t=1}^n \log \hat{f}(y_t | Y_{t-1}, \theta^*). \tag{11}$$

3. EXAMPLES

We now apply the methods discussed above to two examples. First we consider a small state space model with simulated data that provides a testbed for both the estimation and model comparison procedures discussed above. Next we fit the DSGE model in Ireland (2004) to real

data. For each example, we consider four different choices of v ranging from 5 to 30 and compare the resulting models via marginal likelihood.

3.1. Small State Space Model

For illustrative purposes consider the following state space model that mimics the general structure in (3)–(4)

$$y_t = [1 \ 0] \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{bmatrix} \quad (12)$$

$$\begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{bmatrix} \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v_{t+1}, \quad (13)$$

where $v_t \sim t(0, \sigma_v^2, v)$. We generate 100 observations of y_t by setting $g_{11} = g_{12} = g_{22} = 0.6$, $\sigma_v^2 = 1$, and $v = 5$.

Interestingly, g_{12} is not identified through the likelihood. To illustrate this point, we plot the likelihood ordinate over a range of values for each of the four parameters holding the other three fixed at their true values in Fig. 1. As one can see in the plot for g_{12} , the likelihood function is

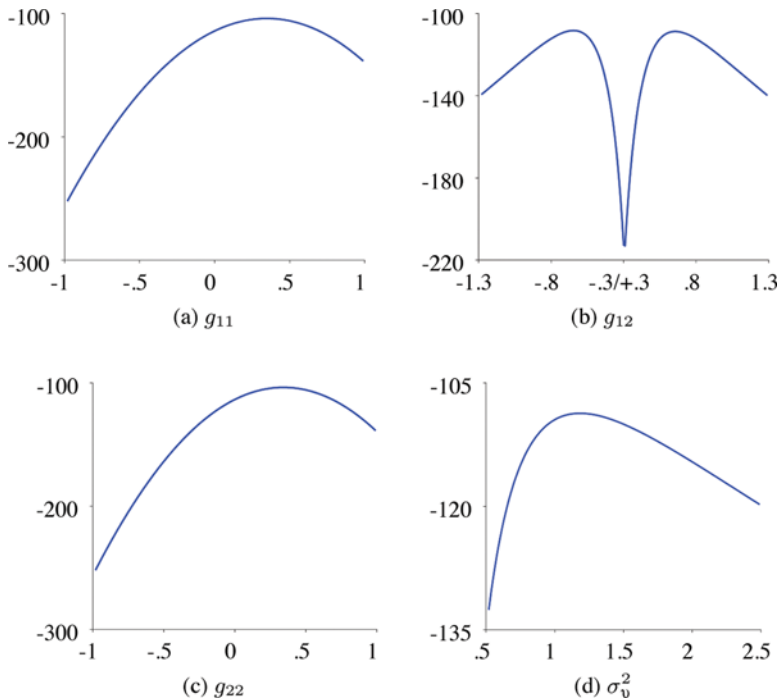


FIGURE 1 Log-likelihood for a range of values of the given parameter holding others fixed at their true values. (Figure available in color online.)

symmetric around $g_{12} = 0$, with peaks around 0.6 and -0.6 . One might be able to overcome this identification problem by specifying an informative prior on g_{12} . However, since we are not really interested in parameter estimation in this problem, we fix g_{12} at its true value of 0.6.

The main focus in this example is model comparison. Because the data is generated with $v = 5$, this should reflect in the estimate of the marginal likelihood. To this end, we compare four models $\mathcal{M}_1 - \mathcal{M}_4$ for each of $v = 5, 10, 15$, and 30, respectively. Note that there is no measurement error in this setup. Consequently, as mentioned in the particle filter section, conditioned on the states, the data density is degenerate. To resolve this issue, we introduce a small measurement error $u_t \sim \mathcal{N}(0, \sigma_u^2)$ with $\sigma_u^2 = 10^{-3}$ fixed. The variance of the measurement error was chosen to be small enough so that the particle filter was stable. It is important to note that while the magnitude of the error affects the marginal likelihood estimates, it does not alter the ranking of the models. To illustrate this point, as well as for sensitivity analysis, we also report the results for $\sigma_u^2 = 10^{-2}$. For further robustness checks, we revisit the same model with a real measurement error, taking σ_u^2 as an additional parameter. In the latter model we set $\sigma_u^2 = 10^{-3}$ and $v = 10$ in the DGP. The remaining parameters are set to the same values as before.

Table 1 reports the log marginal likelihood estimates for each of the four models $\mathcal{M}_1 - \mathcal{M}_4$ with and without measurement errors. These results are based on 5,000 draws (beyond an initial burn-in of 500

TABLE 1 Marginal likelihood estimates for the SSM examples

	Log(marginal likelihood)	
	w/o measurement error	w/measurement error
\mathcal{M}_1	-121.74 (0.0102)	-109.08 (0.0240)
\mathcal{M}_2	-125.04 (0.0226)	-105.95 (0.0215)
\mathcal{M}_3	-129.13 (0.0163)	-106.24 (0.0167)
\mathcal{M}_4	-402.20 (0.0142)	-107.66 (0.0154)

Notes: 1.) The numerical standard errors are reported in parentheses. 2.) These estimates are based on the following prior specification. Each of g_{11} and g_{22} are assumed to be apriori normally distributed with a mean of 0.1 and variance of 2.0. For the example without the measurement error, σ_v^2 is assumed to follow a lognormal distribution with mean 0.6 and variance 0.95. For the model with the measurement error the prior on both σ_v^2 and σ_u^2 is lognormal with mean 0.88 and variance 4.46. 3.) The estimates reported for the model without a measurement error are based on $u_t \sim \mathcal{N}(0, 10^{-3})$. The marginal likelihood estimates for $u_t \sim \mathcal{N}(0, 10^{-2})$ for models $\mathcal{M}_1 - \mathcal{M}_4$ are -119.41 , -121.84 , -123.68 , and -155.52 , respectively.

draws) following the algorithms introduced in Section 2. The draws are constrained to the stationarity restriction on G , which in these examples requires that $|g_{ii}| < 1$, $i = 1, 2$. As can be seen from the results, in both cases the suggested method picks the correct model based on the DGP. For the case without a measurement error, model \mathcal{M}_1 is clearly preferred by the data over any of the other three competing models. In the case with a measurement error, there is substantial evidence in favor of model \mathcal{M}_2 compared to models \mathcal{M}_1 and \mathcal{M}_4 , with \mathcal{M}_3 being a close second. Also worth noting are the small numerical standard errors² (in parenthesis) of these estimates that point to their robustness.

We conclude this example with the following remark. Suppose that v was considered an additional parameter with a priori discrete support $\{5, 10, 15, 30\}$. Then, under a discrete uniform prior on v , the posterior distribution of v is also readily available from this exercise.³ This is due to the fact that the marginal posterior probability that $v = j \in \{5, 10, 15, 30\}$ is proportional to the marginal likelihood $m(y | v = j)$.

3.2. Ireland (2004) Model

The log-linearized DSGE model in Ireland (2004) is summarized by the following system of equations:

$$\begin{aligned}
 \hat{x}_t &= \alpha_x \hat{x}_{t-1} + (1 - \alpha_x) \mathbb{E}_t \hat{x}_{t+1} - (\hat{r}_t - \mathbb{E}_t \hat{\pi}_{t+1}) + (1 - \omega)(1 - \rho_a) \hat{a}_t, \\
 \hat{\pi}_t &= \beta \alpha_\pi \hat{\pi}_{t-1} + \beta(1 - \alpha_\pi) \mathbb{E}_t \hat{\pi}_{t+1} + \psi \hat{x}_t - \hat{e}_t, \\
 \hat{g}_t &= \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t, \\
 \hat{x}_t &= \hat{y}_t - \omega \hat{a}_t, \\
 \hat{r}_t &= \rho_r \hat{r}_{t-1} + \rho_\pi \hat{\pi}_t + \rho_g \hat{g}_t + \rho_x \hat{x}_t + \varepsilon_{r,t},
 \end{aligned} \tag{14}$$

where x_t , g_t , π_t , r_t , and y_t denote output gap, output growth, inflation, nominal interest rate, and stochastically detrended output, respectively, and the hats denote log-deviation of the variables from their steady state or average values; \mathbb{E}_t denotes the expectation of the relevant variables by the agents in this economy, formed under rational expectations; finally, \hat{a}_t , \hat{e}_t , and \hat{z}_t capture exogenous shifts in preferences, costs of production, and technology, respectively. These equations, in ascending order, represent a forward looking IS curve, a new Keynesian Phillips curve, growth rate of output, growth rate of output gap, and the modified Taylor rule (1993).

²Computed as in Chib (1995).

³Alternatively, one could also explicitly derive the posterior distribution of v as shown in Albert and Chib (1993) at a negligible cost.

Further, it is assumed that the exogenous driving processes \hat{a}_t , \hat{e}_t , and \hat{z}_t evolve independently of one another as

$$\begin{aligned}\hat{a}_t &= \rho_a \hat{a}_{t-1} + \varepsilon_{a,t} \\ \hat{e}_t &= \rho_e \hat{e}_{t-1} + \varepsilon_{e,t} \\ \hat{z}_t &= \varepsilon_{z,t}.\end{aligned}\tag{15}$$

Whereas in Ireland (2004) the shocks $\varepsilon_t = [\varepsilon_{a,t}, \varepsilon_{e,t}, \varepsilon_{z,t}, \varepsilon_{R,t}]'$ are assumed to be distributed as multivariate normal, we assume here that each of these shocks follows an independent student- t distribution, which, following the discussion in Section 2, we express as

$$\varepsilon_{jt} \mid \lambda_{jt} \sim \mathcal{N}(\mathbf{0}, \sigma_j^2 / \lambda_{jt})$$

There are two other parameters z and π in the non-linear model that determine the steady-state values of output growth and inflation, respectively. In addition, β determines the steady-state value of the short-term nominal interest rate through the relation $\bar{r} = z\pi/\beta$. Following Ireland, the values of z , π and β are set to the average levels of output growth, inflation and interest rates in the data. Also, ψ and ρ_r are fixed at 0.10 and 1.00, respectively. We collect the remaining 12 parameters of interest in the vector θ

$$\theta = (\omega, \alpha_x, \alpha_\pi, \rho_\pi, \rho_g, \rho_x, \rho_a, \rho_e, \sigma_a, \sigma_e, \sigma_z, \sigma_r).$$

The structural definition of these parameters impose the following bounds. Each of ω , α_x , α_π , ρ_a , and ρ_e are between 0 and 1, ρ_π , ρ_g , and ρ_x are all greater than 0, and the scale parameters σ_i^2 lie in the region \mathcal{S}_Σ that satisfy the usual positivity and positive definiteness constraints. Lastly, we deal with only the determinacy region of the solution space.⁴

The data for the fitting are the series of demeaned log-deviations of output growth \hat{g}_t , inflation $\hat{\pi}_t$, and the short-term nominal interest rate \hat{r}_t from their steady-states or average values for the period 1980:I to 2003:I, with all three observables measured in decimal units.⁵ In the following section we turn to the specification of the prior distribution $\pi(\theta)$ for the parameters in this model.

3.2.1. Prior Distribution

As widely noted in the literature (again refer to An and Schorfheide, 2007), an informative prior distribution can play an critical role in the

⁴Detailed derivations of the solution step and state space representation are available in the notes accompanying the original paper at Peter Ireland's website.

⁵The data is available from the Federal Reserve Bank of St. Louis (FRED) website.

estimation of DSGE models by adding mass to a priori important regions of the parameter space that are otherwise not captured by the likelihood function. It is therefore important to construct a prior that is meaningful. Our preferred method for constructing such a prior is to reason in terms of the implied distribution of the data (Chib and Ergashev, 2009; Del Negro and Schorfheide, 2008).

To construct the prior distribution following this approach, start with a prior specification for the parameters. Then draw parameter values from this prior a large number of times (say 10,000) ensuring that the draws satisfy all the constraints mentioned above. For each draw of the parameters, solve the model and generate the time series of the outcomes, namely \hat{g}_t , $\hat{\pi}_t$, and \hat{r}_t . Finally, check whether this simulated data is reasonable. Specifically, we inspect the median, 5% and 95% quantiles of the implied data. If the low and high quantiles look unreasonably wide or too narrow for any time period, we repeat the process after adjusting some or all of the marginal components of the prior. Note that this procedure of prior elicitation does not involve the real data in any way.

We now summarize a sample prior constructed in this fashion for our illustrative model.

- (1) We assume a beta distribution for the parameters constrained to the $(0, 1)$ interval. Thus, ω , α_x , α_π , ρ_a , and ρ_e are each assumed to be distributed as

$$Beta(\alpha_i, \beta_i),$$

where the (α_i, β_i) are equal to $(3, 12)$, $(3.5, 31.5)$, $(3.5, 31.5)$, $(10, 1.8)$, and $(10, 1.8)$ in the order of the parameters listed above. These values roughly correspond to a mean and standard deviation of $(0.20, 0.10)$, $(0.10, 0.05)$, $(0.10, 0.05)$, $(0.85, 0.10)$, and $(0.85, 0.10)$, respectively.

- (2) Parameters ρ_π , ρ_g , and ρ_x are each assumed to follow a Gamma distribution

$$\rho_i \sim G(\alpha_i, \beta_i),$$

with (α_i, β_i) being $(9, 30)$, $(9, 30)$, and $(16, 64)$ for $i = \pi, g$, and x , respectively. This translates to means and standard deviations of $(0.30, 0.10)$, $(0.30, 0.10)$, and $(0.25, 0.0625)$.

- (3) Finally, we suppose a standard Inverse Gamma prior on 10^4 times the variance parameters

$$\sigma_i^2 \times 10^4 \sim IG(\alpha_i, \beta_i)$$

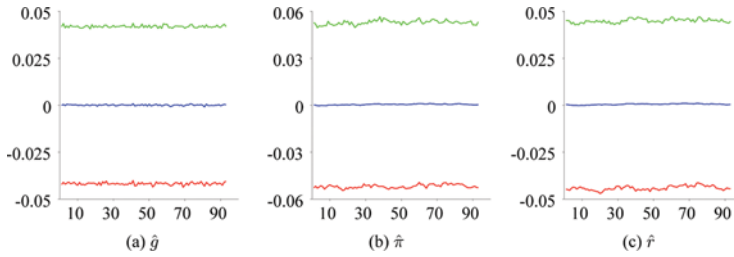


FIGURE 2 Implied \hat{g} , $\hat{\pi}$, and \hat{r} from the prior distribution of the parameters in Ireland (2004). See Table 4 for a summary of the prior distribution. (Figure available in color online.)

$(\alpha_i, \beta_i) = (3, 60), (2, 0.08), (2.1, 5.5), (2.06, 0.53)$. This transformation is helpful in specifying an *IG* prior given the extremely small values of these parameters (relative to the others).

The implied data for each of the variables are shown in Fig. 2. As one can see from the figure, this prior implies a maximum (minimum) quarterly deviation of roughly 4.5% (−4.5%) for output and the rate of interest, and around 5% (−5%) in the case of inflation. This is understandably a substantial change in the variables when measured on a quarterly basis. From our perspective, though, this indicates a reasonably well dispersed prior distribution on the parameters.

3.2.2. Results

As in the previous example we estimate and compare four different models for the same four values of ν . The principal goal of our proposed method, as in any MCMC scheme, is to explore the posterior distribution globally. We would also like the sampler to be efficient, making large moves across the support of the target, while at the same time generating reasonable acceptance rates. On both dimensions the TaRB-MH performed well with inefficiency factors (the inverse of the numerical efficiency measure in Geweke, 1992) below 20 and acceptance rates in the 40–50% range for all the parameters. To provide the specifics of these results, the sampler was initialized at the prior mean and run for 1,000 iterations beyond a burn-in of 1,000 iterations. Also the degrees of freedom in the student- t proposal density was set to 15.

The extent (or lack thereof) of the serial correlation among the draws can also be observed through a plot of the autocorrelation function of the sampled draws. We plot these in Fig. 3 for the MCMC sample from model \mathcal{M}_1 . These plots reveal that the serial correlations among the draws decay quickly to zero for all the parameters. This figure also includes the kernel smoothed histogram of the marginal posterior superimposed on the prior distribution of the parameters. As can be observed, for most parameters,

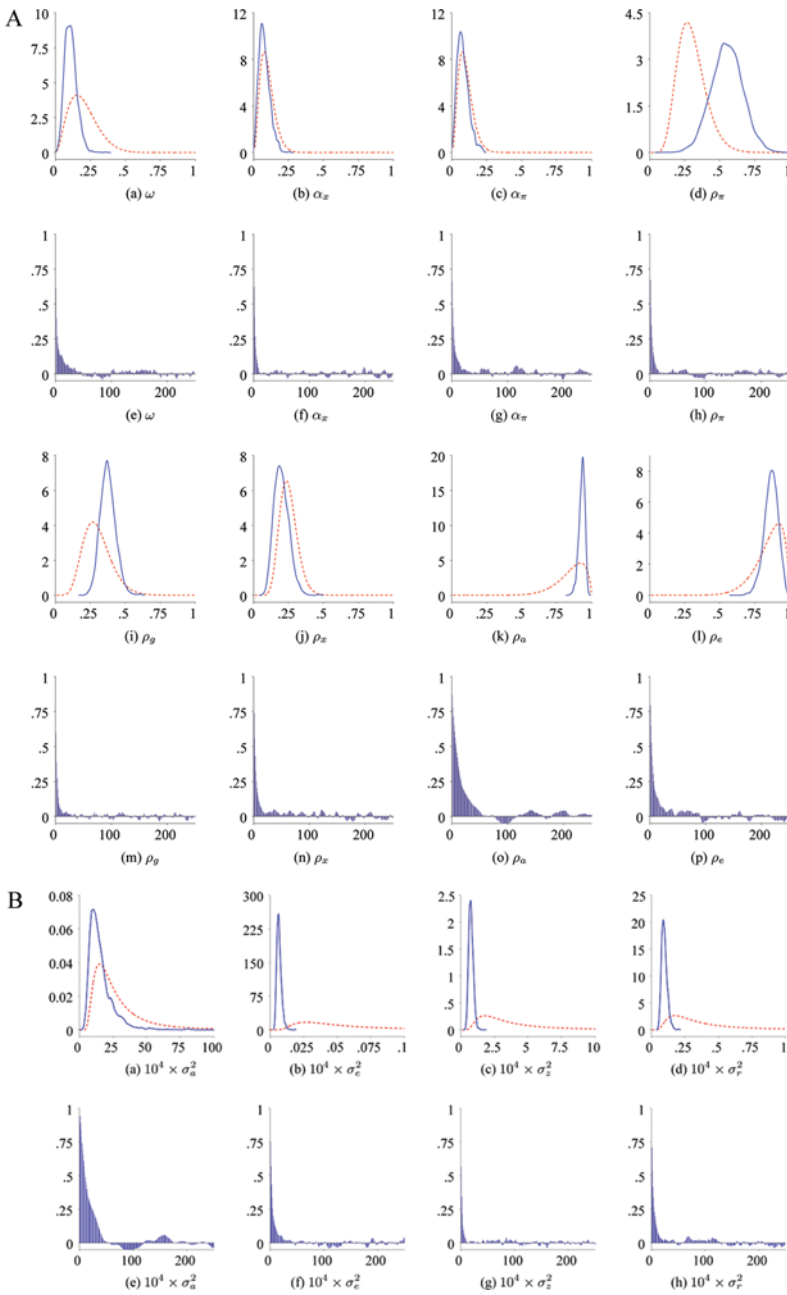


FIGURE 3 Sampling results for the Ireland (2004) model using the TaRB-MH algorithm: A) Marginal prior-posterior plots and autocorrelation functions for the structural and autoregressive parameters, B) Marginal prior-posterior plots and autocorrelation functions for the variances of the shock parameters. (Figure available in color online.)

TABLE 2 Posterior summary and log marginal likelihood for the Ireland (2004) model

Parameter	Prior		Posterior mean			
	Mean	Std. Dev.	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
ω	0.20	0.10	0.0713	0.0893	0.0930	0.1003
α_x	0.10	0.05	0.0392	0.0527	0.0691	0.0750
α_π	0.10	0.05	0.0316	0.0627	0.0801	0.0513
ρ_π	0.30	0.10	0.6563	0.6677	0.5876	0.6510
ρ_g	0.30	0.10	0.2960	0.4215	0.4021	0.3435
ρ_x	0.25	0.0625	0.0885	0.1261	0.1646	0.1002
ρ_a	0.85	0.10	0.9580	0.9544	0.9464	0.9351
ρ_e	0.85	0.10	0.9467	0.8990	0.8689	0.9428
$\sigma_a^2 \times 10^4$	30.00	30.00	14.928	17.569	17.500	14.623
$\sigma_e^2 \times 10^4$	0.08	1.00	0.0155	0.0118	0.0067	0.0210
$\sigma_z^2 \times 10^4$	5.00	15.00	0.5169	0.5723	0.6267	0.6375
$\sigma_r^2 \times 10^4$	0.50	2.00	0.0526	0.0902	0.0789	0.0740
log-marginal likelihood			1090.02 (0.1158)	1083.92 (0.1069)	1078.64 (0.0782)	1072.52 (0.0449)

Remark: The results reported in this table are based on prior mean as the starting value. However, the results are insensitive to this choice.

the likelihood function carries information beyond that contained in the prior.

Table 2 provides a summary of the sampled draws for all four models along with the log marginal likelihood estimates for each model (numerical standard errors in parentheses). These estimates are based on an MCMC sample size of $N = 10,000$ after discarding the initial 1,000 draws. The variance of each of the three measurement errors was set to 10^{-5} to enable comparison to the standard Gaussian error model, for which the estimated log-marginal likelihood is roughly 1,070. These results present strong evidence in favor of the t error model with $\nu = 5$ degrees of freedom. We note that, while the estimated values of marginal likelihood differ for different specifications of the measurement error variance, the ranking of the models in terms of their fit to the data remains unchanged.

4. CONCLUSION

This paper develops a Bayesian framework for analyzing log-linearized DSGE models with student- t errors. By exploiting the well known gamma-normal representation of the t distribution and the TaRB-MH algorithm in Chib and Ramamurthy (2010), the estimation methodology provided here is both efficient and scalable. A complementary technique for estimating the marginal likelihood is also discussed. These methodologies rely on the solution method of Klein (2000). The methodology is first tested with a

simulated data example in the context of a simple state-space model. This example is followed by the DSGE model in Ireland (2004) that is fit to real data. The results from the latter support the t -error model in comparison to the Gaussian model. Thus, the framework and the results are important for the next generation of DSGE models.

REFERENCES

- Albert, J. H., Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88(422):669–679.
- An, S., Schorfheide, F. (2007). Bayesian analysis of DSGE models. *Econometric Reviews* 26(2):113–172.
- Blanchard, O. J., Kahn, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica* 48(5):1305–1312.
- Carlin, B. P., Polson, N. G., Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association* 87(418):493–500.
- Carter, C., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81:541–553.
- Chib, S. (1995). Marginal likelihood from the Gibbs Output. *Journal of the American Statistical Association* 90:1313–1321.
- Chib, S., Ergashev, B. (2009). Analysis of multi-factor affine yield curve models. *Journal of the American Statistical Association* 104(488):1324–1337.
- Chib, S., Greenberg, E. (1994). Bayes inference in regression models with ARMA (p,q) errors. *Journal of Econometrics* 64:183–206.
- Chib, S., Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician* 49:327–35.
- Chib, S., Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association* 96:270–281.
- Chib, S., Nardari, F., Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108:281–316.
- Chib, S., Ramamurthy, S. (2010). Tailored randomized block MCMC methods with application to DSGE models. *Journal of Econometrics* 108:19–38.
- Chib, S., Tiwari, R. C., Jammalamadaka, S. R. (1988). Bayes prediction in regressors with elliptical errors. *Journal of Econometrics* 38:349–360.
- Christiano, L. J., Eichenbaum, M., Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113(1):1–45.
- Curdia, V., Del Negro, M., Greenwald, D. (2012). Rare Shocks, Great Recessions. *Federal Reserve Bank of New York Staff Reports*, No. 585.
- Del Negro, M., Schorfheide, F. (2008). Forming priors for DSGE models (And How it Affects the Assessment of Nominal Rigidities). *Journal of Monetary Economics* 55(7):1191–1208.
- Geweke, J. F. (1992). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computing Science and Statistics: Proceedings of the Twenty-third Symposium*, pp. 571–578.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- Fernandez-Villaverde, J., Rubio-Ramirez, J. F. (2005). Estimating dynamic equilibrium economics: Linear versus nonlinear likelihood. *Journal of Applied Econometrics* 20:891–910.
- Ireland, P. N. (2004). Technology shocks in the new keynesian model. *Review of Economics and Statistics* 86(4):923–936.
- Kitagawa, G. (1987). Non-Gaussian State-Space modeling of nonstationary time series. *Journal of the American Statistical Association* 82(400):1032–1041.
- Klein, P. (1999). Using the generalized schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control* 24(2000):1405–1423.
- Pitt, M., Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association* 94(446):590–599.
- Justiniano, A., Primiceri, G. (2008). The time varying volatility of macroeconomic fluctuations. *American Economic Review* 98(3):604–641.

- Sims, C. A. (2002). Solving linear rational expectations models. *Computational Economics* 20(1–2): 1–20.
- Sims, C. A., Waggoner, D. F., Zha, T. (2008). Methods for inference in large multiple-equation Markov-switching models. *Journal of Econometrics* 146(2):255–274.
- Smets, F., Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of European Economic Association* 97(3):586–606.
- Smets, F., Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* 97(3):586–606.
- Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy* 39:195–214.
- Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *Journal of the American Statistical Association* 71:400–405.