Tailored Multiple-block MCMC Methods for Analysis of DSGE Models

Siddhartha Chib Srikanth Ramamurthy

September 2008

Abstract

In this paper we develop new Markov chain Monte Carlo schemes for Bayesian estimation of DSGE models. The motivation for our work arises from some of the shortcomings of the single block random walk Metropolis Hastings (M-H) algorithm (RW-MH), the sampling method that has been used to date in this context. In our basic (TaB-MH) algorithm, the parameters of the model are randomly clustered at every iteration into an arbitrary number of blocks. Then each block is sequentially updated through an M-H step. Furthermore, the proposal density for each block is tailored to the location and curvature of the target density based on the output of a suitably formulated version of simulated annealing, following Chib and Greenberg (1994, 1995) and Chib and Ergashev (2008). We also provide an extension of this algorithm for sampling multi-modal distributions. In this version, which we refer to as the TaBMJ-MH algorithm, at a pre-specified move jumping iteration (say every 100th), a single-block proposal is generated from one of the modal regions using a mixture proposal density, and this proposal is then accepted according to an M-H probability of move. At the non-mode jumping iterations, the draws are obtained by applying the Tab-MH algorithm. The methodological developments are completed by showing how the approach in Chib (1995) and Chib and Jeliazkov (2001) can be adapted to these sampling schemes for estimating the model marginal likelihood. We illustrate our methods with the aid of stylized problems and two DSGE models that have appeared in the literature. The first is the model in Ireland (2004) where we show that the Tab-MH algorithm is more reliable and efficient than the RW-MH algorithm. Our second example is the model in An and Schorfheide (2007). The posterior distribution in this model is more challenging to simulate on account of multiple modes. As shown by these authors, the RW-MH algorithm is unable to jump from the low modal region to the high modal region, and vice-versa. The TaBMJ-MH method, on the other hand, does not suffer from this problem and moves between the two modal regions and explores the posterior distribution globally in an efficient manner.

*The authors gratefully acknowledge the support from “Center for Research in Economics and Strategy (CRES), in the Olin Business School, Washington University in St. Louis.”

†Chib: Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Bookings Drive, St. Louis, MO 63130 (e-mail: chib@wustl.edu); Ramamurthy: Department of Economics, Washington University in St. Louis, Campus Box 1208, 1 Bookings Drive, St. Louis, MO 63130 (e-mail: sramamur@artsci.wustl.edu).
Key words: Dynamic Stochastic General Equilibrium models; Markov chain Monte Carlo; Metropolis-Hastings algorithm; marginal likelihood; randomized blocks; tailored proposal densities; multi modal densities; simulated annealing.
1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are arguably the dominant framework for modeling and analyzing macroeconomic phenomena. Over the past few years there has been substantial interest in the fitting of these models from a Bayesian perspective implemented by Markov chain Monte Carlo (MCMC) methods (for example, Fernandez-Villaverde and Rubio-Ramirez (2004), Lubik and Schorfheide (2004), Smets and Wouters (2003, 2007), An and Schorfheide (2007), Canova (2007), DeJong and Dave (2007)). One reason for the interest in the Bayesian approach is that it provides an avenue (namely through the prior distribution) for incorporating substantive information about the parameters. This proves important because the parameters in these models are often ill-determined (or unreasonable) when fit by maximum likelihood on account of the stylized nature of the models and the relatively small data sets that can be utilized for the fitting. The attractiveness of Bayesian methods has also grown due to the availability of powerful MCMC simulation methods that provide the technology for sampling the posterior distribution of the parameters (Chib and Greenberg (1995), Chib (2001)).

One feature of the current generation of Bayesian DSGE fitting methods is that they sample the posterior distribution by what is formally known as a single block random-walk Metropolis-Hastings (M-H) algorithm (RW-MH). In the RW-MH algorithm the parameters are sampled in a single block by drawing a proposal from a random walk process. This proposal value is then accepted as the next draw according to the corresponding M-H probability of move (which in this case is essentially the ratio of the posterior density at the proposed value and the posterior density at the current value); if the proposed value is rejected, the current value is retained as the new value of the Markov chain.

The main appeal of this sampling method is its speed and relative ease of implementation. It has been noticed, however, that in the realm of DSGE models, where the posterior distributions can be irregular, the tuning of the RW-MH algorithm is not straightforward. Substantial pre-run tuning effort is often necessary, which can add considerably to the overall time of
implementation. For one, it turns out that the effect of the starting value of the Markov chain can wear off slowly, thus necessitating large simulation sample sizes of the order of many hundreds of thousands. In some situations, for instance in a multi-modal problem, the effect of the starting value may not wear off in realistic sampling time. This problem was demonstrated by An and Schorfheide (2007) who noted that the algorithm was unable to escape the region surrounding one of the modes. We revisit this example later in the paper. Another problem is that the variance of the increment in the random walk proposal can be difficult to set, especially in higher-dimensional problems, and the sampler performance can be severely comprised by a poor choice of it. With too small a variance the search process can be extremely slow, whereas with a large variance there can be many rejections and the same value can be repeated many times in the chain. As a result, the sampled sequence tends to exhibit high serial correlations and slow convergence to the posterior distribution.

Our goal in this paper is to suggest new MCMC schemes for fitting DSGE models that are more reliable than the existing approach. We achieve this by combining the efficiency of tailored adaptive proposals (Chib and Greenberg (1994)) with a flexible blocking strategy that virtually eliminates pre-run tuning. In our basic approach, which we refer to as the TaB-MH algorithm, the parameters of the model are randomly clustered at every iteration into a random number of blocks. Then each block is sequentially updated through an M-H step in which the proposal density is tailored to closely mimic the target density of that block. Since the posterior surface can be irregular in DSGE models, we construct our tailored proposal densities from the output of a suitably formulated version of simulated annealing, following Chib and Ergashev (2008). In effect, the TaB-MH algorithm has the features of a general black box approach that requires little user involvement. We also provide an extension of this algorithm for sampling multi-modal distributions. In this version, which we refer to as the TaBMJ-MH algorithm, at a pre-specified move jumping iteration (say every 100th), a single-block proposal is generated from one of the modal regions using a mixture proposal density, and this proposal is then accepted according to an M-H probability of move. At the non-
mode jumping iterations, the draws are obtained by applying the Tab-MH algorithm. It should be noted that while these methods are developed with the DSGE applications in mind, the methods are general and can be applied to any Bayesian problem.

We complete our methodological developments by presenting an approach for calculating the model marginal likelihood, which is needed for the comparison of alternative DSGE models. The method we present is based on the framework of Chib (1995), and its M-H version in Chib and Jeliazkov (2001), with suitable modifications to accommodate the randomized block approach proposed in this paper.

We illustrate the performance of our methods with the aid of two DSGE models. The first is the model in Ireland (2004). In this example we provide a comparative analysis of the RW-MH and TaB-MH algorithms. Specifically, we pick three different starting values - the prior mean, a local mode and the dominant mode. We also specify different variance-covariance matrices for the RW proposal. We then run three independent RW chains initiated at each of the starting values for 250,000 iterations following a burn-in of 50,000 iterations. Our results indicate that the chains initialized at the first two values do not fully navigate the posterior distribution. In some instances, this problem is not overcome even when the simulation sample size is increased to one million. In contrast, the sampled draws from the TaB-MH algorithm quickly move to the same regions of the parameter space, regardless of the starting value. The TaB-MH output is also efficient with the sample autocorrelation functions (ACF) of the draws decaying to zero within 15-20 lags for most parameters. As a result, a simulation sample size of, say, 10,000 with a burn-in of a 1000 iterations, is sufficient to adequately explore the posterior distribution in this problem.

Our second example is the model in An and Schorfheide (2007). The posterior distribution in this model is more challenging to simulate on account of possibly two modes. As shown by the authors, the RW sampler was simply unable to escape the attraction of the individual modes even when the sampler was run for a million iterations following a burn-in of 100,000 iterations. What is especially interesting is that, despite a difference of approximately 8 on the
log scale between the unnormalized posterior ordinates at the two modes, the RW chain did not cross over from the low modal region to the high modal region. The TaBMJ-MH method, on the other hand, jumps to the high mode almost instantaneously and explores the posterior distribution globally in an efficient manner.

The rest of the paper is organized as follows. To motivate the discussion of our methods, as well as for completeness of the paper, we begin with a brief discussion of the general setup of linearized DSGE models in the following section. In Section 3 we present the details of our approach for sampling the posterior distribution and discuss a suitable way of calculating the marginal likelihood. Section 4 deals with the application of our methods to the two DSGE models mentioned above. Concluding remarks are provided in Section 5.

2 The setup

In this paper our focal application is the class of macroeconomic models commonly referred to as DSGE models. These are structural microfounded models set in a general equilibrium framework. Thus, the outcome of the economic variables is determined by the collective actions of the agents in the economy being modeled. Further, in the face of uncertainty, the agents take into account the future expected values of the variables when making their decisions. In the ensuing discussion, as well as in the applications in Section 3, this decision making process is assumed to be based on rational expectations (expectations based on the full information set available up to the current period). For both theoretical and inferential purposes, one is usually interested in analyzing the behavior of the economy around its deterministic steady state. Since these models are typically highly nonlinear, one often works with first order approximations. This reduces the model to a system of stochastic expectational difference equations, popularly known as a linear rational expectations model. For the rest of this paper, this is our structural model of interest.1

As discussed in An-Schorfheide (2007), Sargent (1989), Smets-Wouters (2003) and others,

---

1These stylized assumptions are inconsequential from the point of view of this paper as the MCMC methods are applicable to a wide variety of problems.
the process of transforming a structural model to its reduced form involves two steps. The first step is to solve the model. In the solution step, the expectational terms are integrated out to obtain a stochastic law of motion for the endogenous variables of the model. Here, one generally restricts attention to the subset of parameter space, $S_D$, that generates a unique stable solution to the model (the determinate solution). Several solution methods have been proposed in the literature to solve linear rational expectations models (for a concise comparative discussion of these methods see Anderson (2008)). We adopt the method in Sims (2002) to obtain the model solution. Following this approach the variables $s_t$ in the model are written as a function of their past values (the vector autoregressive component), the innovations $\varepsilon_t$ (assumed to be a zero mean Gaussian random vector) and an endogenous rational expectations error $\eta_t$

$$G_0(\theta)s_t = G_1(\theta)s_{t-1} + G_2(\theta)\varepsilon_t + G_3(\theta)\eta_t. \quad (1)$$

Here, $G_{j=0}^3$ are matrices of appropriate dimensions involving the parameters $\theta$ of the model. Henceforth, we let $S_L$ denote the linear restrictions on the parameters imposed by the underlying theoretical model.

The details of the solution algorithm are beyond the scope of this paper. We simply note that the solution proceeds by applying the generalized Schur decomposition to the pair of matrices $(G_0, G_1)$ from which the set of generalized eigenvalues of the preceding matrix pair can be computed. The general idea behind this exercise is to isolate the stable and unstable blocks of the system. Loosely speaking, for a unique solution to exist, we require that the number of unstable roots of the system equal the number of non-predicted variables in the model. The solution is then constructed by solving ‘forward’ (‘backward’) the unstable (stable) block of the system. From an empirical standpoint, the convenience of a linear framework is that the output of the solution is a vector Markov process for the endogenous variables of the model

$$s_t = D(\theta)s_{t-1} + F(\theta)\varepsilon_t \quad (2)$$

where the matrices $D(\theta)$ and $F(\theta)$ are awkward implicit functions of the model parameters, obtained from the solution.
For purposes of conducting inference on the parameters, the preceding vector Markov process can be viewed as the state transition equation in a linear Gaussian state space model (SSM). Accordingly, the SSM formulation is completed by defining a measurement equation that relates the states \( s_t \) to a vector of observable data \( y_t \).

\[
y_t = a(\theta) + B(\theta)s_t
\]  

(3)

Depending on the model and representation of the data, the vector \( a \) and the matrix \( B \) may be constants or involve the parameters of the model.

Let \( \Omega(\theta) \) be the variance-covariance matrix of \( \varepsilon \). It is assumed that the parameters in \( \Omega \) satisfy the usual positivity and positive definiteness constraints, \( S_\Omega \). Given a sample of data \( y_{n \times T} = \{ y_t \}, t = 1, \ldots, T \), the joint density of the data (the likelihood function) given the parameters can be calculated as

\[
f(y|\theta) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2}\left|\Sigma_{t|t-1}\right|^{1/2}} \exp \left\{ -\frac{1}{2} \eta_{t|t-1}'\Sigma_{t|t-1}^{-1}\eta_{t|t-1} \right\} I_{S_L}(\theta)I_{S_\Omega}(\theta)I_{S_D}(\theta)
\]  

(4)

where \( \eta_{t|t-1} = y_t - B\hat{s}_{t|t-1} \) is the one-step ahead prediction error, \( \hat{s}_{t|t-1} \) is the one-step ahead estimate of the state and \( \Sigma_{t|t-1} \) is the conditional variance of the prediction error, each given information up to time \( (t - 1) \). These quantities are obtained from the usual Kalman filter recursions (see for instance, Harvey, 1990).

It is important to recognize that in most practical models of interest, the solve procedure can only be performed numerically for a given value of the parameters. This means that the model has to be solved numerically for every value of the parameters at which the likelihood function is to be evaluated. Also notice that the likelihood function is restricted to the parameter space \( \theta \in S_L \cap S_\Omega \cap S_D \). Of these, the determinacy constraint is the most tedious to verify as it requires the model to be solved for the particular value of the parameters. Therefore, to verify if a particular value of the parameters satisfies the restrictions, it is efficient to check the constraints in the order listed above. Alternatively, when possible, one could also work with transformed parameters that are unrestricted, requiring only the determinacy check.
In the Bayesian context, the empirical state space model is completed with a prior distribution $\pi(\theta)$ on the parameters. The posterior distribution of the parameters $\pi(\theta|y)$ is calculated (up to the normalizing constant) as the product of the likelihood function and the prior distribution

$$\pi(\theta|y) \propto f(y|\theta) \times \pi(\theta)$$  \hspace{1cm} (5)

From the Bayesian point of view, a summary of the posterior (for instance, the mean or the median) serves as an estimate of the parameters. However, as the posterior distribution is typically intractable, such summaries are not obtained easily. Instead, one relies on MCMC methods to draw sample variates from the posterior distribution. These sampled draws are then used to find the posterior mean, posterior quantiles and other summaries of the distribution. In the following section we discuss how this general approach can effectively operationalized.

### 3 Posterior sampling and marginal likelihood estimation

#### 3.1 Basic sampling approach

In MCMC methods, one samples the posterior density by simulating a Markov chain whose invariant density is the specified target density. The sampled draws, beyond a suitable burn-in phase, are then taken as a (correlated) sample from the posterior density. While the general idea is clear, designing a sampler that mixes well (has low serial correlations in the sampled draws), and converges quickly to the invariant distribution, is challenging, especially in the context of DSGE models.

The sampling schemes that we focus on belong to the class of tailored multiple block MH algorithms as discussed in Chib and Greenberg (1994, 1995). In a multiple-block MH algorithm, the parameters are grouped into several distinct blocks, and then each block of parameters is updated in sequence by a MH step, conditioned on the most current value of the parameters in the remaining blocks. It is noteworthy that a multiple-block M-H algorithm has not been implemented before in the setting of DSGE models. It is our aim to show, however, that versions of the multiple-block M-H algorithm are capable of effectively sampling the
complex posterior distributions that arise from DSGE models.

A key question in the multiple block sampling method is the number and composition of the blocks. One principle is to form groups so that parameters in different blocks are not strongly correlated whereas those within a block are. Unlike usual statistical models, however, where location and scale parameters (for instance) may be approximately orthogonal and hence placed in separate groups, in a DSGE model such apriori knowledge is rarely available. In fact, because of the solve process, which converts the structural model into a state-space model as discussed above, the parameters become scrambled in an unspecified way, thus creating correlations between parameters that cannot be analytically isolated.

Another complication is that the DSGE model posteriors tend to be irregular so that the “grouping by correlation” criterion is not necessarily the ideal. To see why, suppose that a bivariate posterior distribution has positively sloped contours for some portion of the support but has vertical contours for another portion of the support. In that case, it is meaningful in the positively sloped regions to sample the parameters in a group but to sample the parameters one at a time to escape into the other portion of the support. In other words, in irregular problems, a permanent grouping of parameters in one block is not necessarily optimal.

Our novel idea to blocking, which works extremely well and has never been tried before in the MCMC literature, is to randomize the formation of blocks in every iteration. This means that the number of blocks and its components are randomized. One clear virtue of this randomized blocking scheme is that it avoids the pitfalls from a poor choice of a priori blocks. Another is that it allows for the groupings to change, which is important in irregular problems (as in the preceding illustration). We note that (if necessary) such a scheme could also be organized to allow for some parameters in \( \theta_i \) to form a fixed block, and the remaining parameters to form random blocks. In this sense, this randomized blocking strategy is flexible enough to accommodate different model specific blocking needs.

The second component of our approach is that the proposal density in the M-H step is constructed to adapt to both the location and curvature of the posterior density of that block.
While this notion of local tailoring dates back to Chib and Greenberg (1994, 1995), we further enhance this approach by basing the tailored densities on the output of the simulated annealing optimization method (following Chib and Ergashev (2008)). This enhancement is particularly helpful in dealing with irregular distributions, and also for localizing the proposal density far enough away from the current region (depending on where the global mode for that block of parameters is located) to permit large moves.

Our multiple block tailored M-H algorithm can be summarized as follows.

**Algorithm: TaB-MH algorithm**

**Step 1** Initialize $\theta^{(0)} \in S_L \cap S_\Omega \cap S_D$ and fix $n_0$ (the burn-in) and $M$ (the MCMC sample size)

**Step 2** In each iteration $j, j = 1, \ldots, n_0 + M$, randomly generate blocks $(\theta_{j,1}, \theta_{j,2}, \ldots, \theta_{j,p_j})$

**Step 3** Within each iteration, sample each block $\theta_{j,l}, l = 1, \ldots, p_j$, by an M-H step with a tailored proposal density as described below.

**Step 4** Repeat steps 2-3 $n_0 + M$ times, discard the draws from the first $n_0$ iterations and save the subsequent $M$ draws $\theta^{(n_0+1)}, \ldots, \theta^{(n_0+M)}$

Generating random blocks in Step 2 above is, of course, straightforward and does not require comment. We therefore focus on an explanation of Step 3. Suppose that at the end of the $(j-1)$st iteration of the MCMC iteration, there are $p_j$ randomly constructed blocks $(\theta_{j,1}, \ldots, \theta_{j,p_j})$. Let $\theta_{j,-l}$ denote the most current value of all the blocks except the $l$th. Then to construct the tailored proposal density for $\theta_{j,l}$ we find

$$\hat{\theta}_{j,l} = \arg \max_{\theta_{j,l}} \log \left\{ f(y | \theta_{j,l}, \theta_{j,-l}) \times \pi(\theta) \right\}$$

from a suitably formulated version of simulated annealing.

In simulated annealing one proposes a random modification to the current guess of the maximum which is then accepted or rejected probabilistically. Moves that lower the function value can sometimes be accepted. The probability of accepting such downhill moves
declines over iterations according to a “cooling schedule,” thus allowing the method to converge. Specifically, starting from an initial temperature \( t_0 \) and an initial guess for the maximum \( \theta_{j,l}^{(0)} \), such that \( (\theta_{j,1}, \ldots, \theta_{j,l}^{(0)}, \ldots, \theta_{j,p}) \in S_L \cap S_{\Omega} \cap S_D \), repeated proposals are generated for a randomly chosen element \( \theta_m \) of \( \theta_{j,l} \) using a random walk process

\[
\theta'_m = \theta_m + s \mathcal{N}(0, 1)
\]

where \( s > 0 \) is a suitable scale factor. As in the M-H step we restrict the proposals to the constrained set. Perturbations resulting in a higher function value are always accepted, whereas those resulting in a lower function evaluation are accepted with probability

\[
p = \exp\{\Delta[\log(f(y|\theta_{j,l}, \theta_{j,-l}) \times \pi(\theta))] / t\} < 1
\]

where \( \Delta[\log(f(y|\theta_{j,l}, \theta_{j,-l}) \times \pi(\theta))] \) is the change in the log of the objective function, computed as the log of the objective function at the perturbed value of the parameters minus the log of the objective function at the existing value of the parameters. As one can verify, this probability is a decreasing function of the temperature, with convergence achieved as \( t \to 0 \).

Our implementation of this algorithm is as follows. We first divide the search process into various stages, denoted by \( k, k = 1, 2, \ldots, K \), with the length of each stage \( l_k \) given by \( b + l_{k-1} \), where \( b \in \mathbb{N} \) is the stage expansion factor. We then specify the initial temperature \( t_0 \) which is held constant in each stage but reduced across stages according to the linear cooling schedule \( t_k = a t_{k-1} \), where \( 0 < a < 1 \) is the cooling constant. Within each stage, the algorithm searches for the maximum, proposing values from a random walk process as mentioned above.

Once \( \hat{\theta}_{j,l} \) has been so found, we calculate the curvature of the target posterior distribution of that block as the negative inverse of the Hessian:

\[
V_{j,l} = \left(-\frac{\partial^2 \log\{f(y|\theta_{j,l}, \theta_{j,-l}) \times \pi(\theta)\}}{\partial \theta_{j,l} \partial \theta'_{j,l}}\right)^{-1} \bigg|_{\theta_{j,l}=\hat{\theta}_{j,l}}
\]

Our proposal density \( q_l(\theta_{j,l} | \theta_{j,-l}, y) \) of \( \theta_{j,l} \) is then given by

\[
q_l(\theta_{j,l} | \theta_{j,-l}, y) = t(\theta_{j,l} | \hat{\theta}_{j,l}, V_{j,l}, \nu),
\]
a multivariate student-\( t \) density with \( \nu > 2 \) degrees of freedom. Note that this proposal density depends on the current values of the remaining parameters (blocks) and changes from one iteration to the next as the constituents of the blocks change.

We now draw a proposal value \( \theta_{j,l}^\dagger \) from the latter proposal density. If the proposal value violates any of the constraints it is rejected immediately. Otherwise, the proposed value is taken as the new value of the block with the M-H probability of move given by

\[
\alpha_l(\theta_{j,l}, \theta_{j,l}^\dagger | \theta_{j,-l}, y) = \min \left\{ \frac{f(y | \theta_{j,l}^\dagger, \theta_{j,-l}) \pi(\theta_{j,l}^\dagger, \theta_{j,-l}) t(\theta_{j,l} | \hat{\theta}_{j,l}, V_{j,l}, \nu)}{f(y | \theta_{j,l}, \theta_{j,-l}) \pi(\theta_{j,l}, \theta_{j,-l}) t(\theta_{j,l} | \hat{\theta}_{j,l}, V_{j,l}, \nu)}, 1 \right\}
\]

(7)

If the proposed value is rejected, the current value of the block is retained as the new value of that block. Step 3 of the algorithm is completed by repeating this process for each block.

As an instructive example, consider the following arbitrary SSM

\[
\begin{bmatrix}
y_{1,t} \\
y_{2,t} \\
s_{1,t} \\
s_{2,t}
\end{bmatrix}
= \begin{bmatrix} a_1 \\ a_2 \\ d_{11} \\ d_{21} \end{bmatrix}
+ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\begin{bmatrix}
s_{1,t} \\
s_{2,t}
\end{bmatrix}
+ \begin{bmatrix} d_{12} \\ d_{22} \end{bmatrix}
\begin{bmatrix}
s_{1,t-1} \\
s_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

where \( \varepsilon_t \sim \mathcal{N}(0, \Omega) \) and \( \Omega = \text{diag}(\sigma_1^2, \sigma_2^2) \). The parameters of interest are

\[\theta = [a_1, a_2, d_{11}, d_{12}, d_{21}, d_{22}, \sigma_1^2, \sigma_2^2]'\]

We generate 200 observations from the data generating process

\[\theta_{DGP} = [0.2, 1.4, -0.45, 0.30, 1.50, 0.10, 2.00, 0.75]'\]

The likelihood function \( f(y | \theta) \) can now be calculated through the Kalman filter as outlined in Section 2. The model is completed by specifying a prior distribution \( \pi(\theta) \) of the parameters. For simplicity, assume that the parameters are a priori independent. We suppose that the parameters in \( \mathbf{a} \) and \( \mathbf{D} \) are each normally distributed with mean 0 and standard deviation 1, and that each of the variance parameters in \( \Omega \) have an inverse gamma distribution with
parameters $\alpha = 3, \beta = 2$ (which corresponds to a mean, as well as standard deviation, of 1). Thus, the posterior distribution (up to the normalizing constant) can be written as

$$\pi(\theta|y) \propto f(y|\theta) \times \prod_{1}^{6} \mathcal{N}(0, 1) \prod_{1}^{2} \mathcal{IG}(3, 2).$$

This is the target function that we would like to explore. In addition to the positive definiteness constraint on $\Omega$, which in this case translates to $\sigma_i^2 > 0$, we impose the stationarity constraint, that is, the eigenvalues of $D$ are all less than one in absolute value.

Following step 1 in the TaB-MH algorithm, we initialize the sampler at $\theta^{(0)} = [0.50, 0.50, 0.50, 0.00, 0.00, 0.50, 1.00, 1.00]'$. Then steps 2 and 3 are iterated 10,000 times as follows. In every iteration the parameters are first randomly grouped into various blocks. In our scheme this is accomplished by randomly permuting the index of the parameters at the beginning of an iteration. Then block 1 is initialized at the first element of this shuffled index. A new block is started at every subsequent element with probability 0.2 (or in other words, with probability 0.8 that element is included in the current block). Once the blocks have been so formed in a given iteration, we proceed to step 3, setting the parameters of the simulated annealing algorithm at $t_0 = 2, a = 0.4, K = 5, b = 5$ and $s = 0.1$. For each block, we now find the mode and curvature of the current target density to construct a student-$t$ proposal density with 15 degrees of freedom. We then apply the M-H step, revise the value of that block according to the probability given in (7), and repeat this process for the next block.

Figure 1 below shows the kernel smoothed histogram of the marginal posterior density for 4 select parameters. Notice that the marginal posterior densities are centered around the data generating process as one might expect in this problem. With 200 observations, the role of the prior distribution is diminished since the likelihood function carries substantial information about the parameters. The figure also includes plots of the autocorrelation function. As seen in these plots, there is virtually no sign of serial correlation as it drops to zero within 5 to 6 iterations for all the parameters. This sort of mixing is understandably difficult to achieve in more realistic problems.
3.2 Extension to multi-modal problems

In this section we propose a simple extension to our algorithm, still in the same class of tailored MH methods, that can be used to efficiently sample multi-modal distributions. Before that, consider once again the RW-MH algorithm. As mentioned earlier, one strategy for the RW-MH algorithm is to thoroughly scan the parameter space for the various modes and then start the RW chain at the highest one. The problems with this approach are twofold. One is that the mass in the various modal regions is not known apriori. The other, more pronounced issue, concerns truly multi-modal distributions. In this regard, a suitable version of the RW-MH algorithm that is often used is an intuitive modification of the regular RW-MH algorithm. The idea is to propose occasional large jumps that breaks the attraction to the current mode. This method, however, has limited applicability and often works well only in small dimensional problems.

To describe our modification of the TaB-MH algorithm, consider for simplicity the case where there are only two well defined modes. We assume that the modal values have been
found by initial optimization. Let the location of the two modes be $\mu_1$ and $\mu_2$. Also, let $V_1$ and $V_2$ denote the inverse of the negative Hessian matrices at the two modes. In addition to the proposal densities that are calculated within the TaB-MH algorithm, define the following mixture proposal

$$q(\theta|y) = p t(\theta|\mu_1, V_1, \nu_1) + (1 - p) t(\theta|\mu_2, V_2, \nu_2),$$

where $p$ is the probability of drawing a value from a $t$ density centered at $\mu_1$ and variance given by $V_1$. We now incorporate this mixture proposal in our regular TaB-MH algorithm in the following way. Once every few (say a 100) iterations we update all the parameters in one block using this mixture proposal. That is, we first draw a value $\theta^\dagger$ from the first modal region with probability $p$ (from the second modal region with probability $1 - p$). This value is subsequently accepted as the next draw with the probability of move $\alpha_{M,J}(\theta, \theta^\dagger|y)$ given by

$$\alpha_{M,J}(\theta, \theta^\dagger|y) = \min \left\{ \frac{f(y|\theta^\dagger)\pi(\theta^\dagger)}{f(y|\theta)\pi(\theta)} \frac{p t(\theta|\mu_1, V_1, \nu_1) + (1 - p) t(\theta|\mu_2, V_2, \nu_2)}{p t(\theta^\dagger|\mu_1, V_1, \nu_1) + (1 - p) t(\theta^\dagger|\mu_2, V_2, \nu_2)}, 1 \right\}. \quad (8)$$

The next (say) 100 draws of the chain are obtained from the TaB-MH algorithm. At that point one returns to the mode-jumping step and the whole process is repeated.

In summary, the TaBMJ M-H algorithm has the following form.

**Algorithm: TaBMJ-MH algorithm**

**Step 1** Initialize $\theta^{(0)} \in \mathcal{S}_L \cap \mathcal{S}_\Omega \cap \mathcal{S}_D$ and fix $n_0$ (the burn-in) and $M$ (the MCMC sample size)

**Step 2a** If not a mode-jumping iteration, then

**Step i** Randomly generate blocks $(\theta_{j,1}, \theta_{j,2}, \ldots, \theta_{j,p_j})$

**Step ii** For $l = 1, \ldots, p_j$, sample $\theta^\dagger_{j,l}$, from the tailored proposal density $t(\theta_{j,l}|\hat{\theta}_{j,l}, V_{j,l}, \nu)$ and accept this proposal with the probability of move $\alpha_l(\theta_{j,l}, \theta^\dagger_{j,l}|\theta_{j,\sim l}, y)$

**Step 2b** If a mode-jumping iteration, then
with probability $p$ draw $\theta^\dagger$ from the proposal density $t(\theta|\mu_1, V_1, \nu_1)$ and with probability $(1 - p)$ draw $\theta^\dagger$ from the proposal density $t(\theta|\mu_2, V_2, \nu_2)$ and accept proposal with the probability of move $\alpha_{M,J}(\theta, \theta|^\dagger|y)$

**Step 4** Repeat Steps 2a-2b $n_0 + M$ times, discard the draws from the first $n_0$ iterations and save the subsequent $M$ draws $\theta^{(n_0+1)}, \ldots, \theta^{(n_0+M)}$

The generalization of this method to more than two modes is straightforward. Suppose there are $M$ distinct modes, denoted by $\mu_k$, $k = 1, 2, \ldots, M$. The idea is to construct a proposal density that encompasses all $M$ modes, that is

$$q(\theta|y) = \sum_{k=1}^{M} p_k t(\theta|\mu_k, V_k, \nu_k); \quad \sum_k p_k = 1.$$ 

Now suppose that at the end of the $j$th iteration the current state of the sampler is $\theta$ located in one of the $M$ modal regions, say the $m$th mode ($m \leq M$). Then in the mode jumping step, the proposal $\theta^\dagger$ can be drawn from any one of the $M$ modes. The weights $p_k$ can be selected in any manner. However, given that one does not know a priori the relative probability mass in each of the modal regions, we suggest fixing $p_k$ at $1/M$ for all $k$. This way the decision to jump modes is left to the sampler, which makes the necessary move across the modes according to the relative probability content in them. After all, if the proposal density is well designed, the transition probability of the Markov chain to a particular modal region should equal the probability content of the region. The corresponding $\alpha$ is calculated as

$$\alpha_{M,J}(\theta, \theta|^\dagger|y) = \min \left\{ \frac{f(y|\theta^\dagger)\pi(\theta^\dagger)}{f(y|\theta)\pi(\theta)} \frac{\sum_{k=1}^{M} p_k t(\theta|\mu_k, V_k, \nu_k)}{\sum_{k=1}^{M} p_k t(\theta^\dagger|\mu_k, V_k, \nu_k)}, 1 \right\}. \quad (9)$$

As an illustration, consider the following bivariate contaminated normal distribution. The target function is defined as

$$\pi(\theta|y) = c\mathcal{N}(\mu_1, \Sigma_1) + (1 - c)\mathcal{N}(\mu_2, \Sigma_2) \quad (10)$$

where $\mu_1 = [1, -1]'$, $\mu_2 = 6\mu_1$, $\Sigma_1 = \text{diag}(1.3, 1.3)$, $\Sigma_2 = \text{diag}(0.05, 0.05)$ and $c = 0.99$. The motivation for the specific choice of $c$ comes from the posterior distribution in the An
and Schorfheide model. The significance of $c$ is that the difference in the ordinates at the two modes is roughly 1.34 in the log scale. Although this is a simple problem in two dimensions, the challenge arises because of the tiny second mode in the tails of an otherwise unimodal distribution. For this problem, we initialize the TaBMJ-MH algorithm at the low mode. The mode jumping step is proposed every 100 iterations with $p = 0.5$. The simulation results from 20000 draws are shown in Figure 2. The proposal densities used for this experiment are both student t-distributions, with 5 degrees of freedom for the regular TaB-MH algorithm and 30 degrees of freedom for the mode jumping step. For the sake of illustration we plotted all 20000 draws without any burn-ins to show when the initial jump from the low to the high mode occurred. The top row plots the time series of the draws for the two random variables. Note that the sampler rightly revisits the low mode roughly 1% of the time. The bottom row includes a bivariate scatter plot of the draws (left) along with a 3 dimensional surface plot of the actual density.

For a second example we construct a two component mixture of six dimensional normals. We pick the values for the first mode $\mu_1$ to equal the low modal values of six of the parameters that differ widely across the two modes in the An-Schorfheide (2007) model (discussed in Section 4). To make the problem interesting we center the second mode $\mu_2$ at 15 times $\mu_1$. Thus, the two modal regions are widely separated.

\[
\mu_1 : (1.41, 0.81, 0.49, 0.80, 1.07, 0.30)
\]

\[
\mu_2 : (21.15, 12.15, 7.35, 12.00, 16.05, 4.50)
\]

We also match the variance-covariance matrices $\Sigma_1$ and $\Sigma_2$ to the reduced variance-covariance matrices at the two modes in the aforementioned model for the relevant parameters. The specific values are given by

\[
\Sigma_1 = \begin{bmatrix}
0.0885 & -0.0023 & 0.0077 & 0.0041 & -0.0229 & -0.0025 \\
-0.0023 & 0.0055 & 0.0015 & 0.0028 & 0.0013 & 0.0001 \\
0.0077 & 0.0015 & 0.0031 & 0.0018 & -0.0011 & -0.0002 \\
0.0041 & 0.0028 & 0.0018 & 0.0029 & 0.0004 & -0.0012 \\
-0.0229 & 0.0013 & -0.0011 & 0.0004 & 0.0169 & 0.0004 \\
-0.0025 & 0.0001 & -0.0002 & -0.0012 & 0.0004 & 0.0024
\end{bmatrix}
\]
Figure 2: Sampling from a bivariate contaminated normal using the TaBMJ-MH algorithm:
The top panel plots the actual draws (without any burn-ins) against time for the two variables.
The bottom left panel shows a bivariate scatter plot of the draws. The actual target function is
the 3-dimensional surface plot in the bottom right panel.

Finally, we assign 20% probability to the first mode. Our target distribution takes the form

\[ \pi(\theta | y) = 0.2 \mathcal{N}(\mu_1, \Sigma_1) + 0.8 \mathcal{N}(\mu_2, \Sigma_2) \]

The results from 25000 iterations of the TaB-MH algorithm are summarized in Figure 3.
Rows 1 and 3 plot the time series of the actual draws for the first three parameters (the results
are identical for all six parameters). Notice the fluid movement of the sampler across the two
modes. We also plot the kernel smoothed histogram of the draws in rows 2 and 4. These plots
Figure 3: Sampling from a two component mixture of 6-variate normals using the TaBMJ-MH algorithm: The top row shows the sampled draws against time (again without any burn-ins). The simulated marginal density (kernel smoothed histogram) plots for the variables are shown in the bottom row.

suggest an accurate replication of the target density.

3.3 Marginal likelihood computation

One of the advantages of the Bayesian framework is that it is possible to compare competing models in a formal way through marginal likelihoods and Bayes factors. In this section we show how the framework of Chib (1995), and its M-H version in Chib and Jeliazkov (2001), can be modified to accommodate the randomized block approach proposed in this paper.

In the Chib framework, one starts by noting that the marginal likelihood (which is integral
of the sampling density with respect to the prior density) can be expressed in the form

\[
m(y) = \frac{f(y|\theta)\pi(\theta)}{\pi(\theta|y)},
\]

by a rearrangement of Bayes theorem. Importantly, this is an identity in \( \theta \). Thus, the marginal likelihood can be obtained by evaluating this identity at a value \( \theta^* \) (say), which is usually taken to be a high density point in the support of the posterior. Clearly, the first two components in the latter expression are readily computed. As for the third component, we follow the approach developed in Chib (1995) and decompose the posterior ordinate into marginal and conditional components each of which is then estimated in turn.

In order to apply this approach we fix the number of blocks in the posterior ordinate estimation step to the average number of blocks (say \( B \)) that are realized in the Tab-MH run. Then, we construct the blocks \( \theta_1, \ldots, \theta_B \) with each block consisting of randomly chosen components from \( \theta \) and write

\[
\pi(\theta^*|y) = \pi(\theta_1^*|y)\pi(\theta_2^*|y, \theta_1^*) \ldots \pi(\theta_B^*|y, \theta_1^*, \ldots, \theta_{B-1}^*).
\]

In this decomposition, the typical ordinate is \( \pi(\theta_l^*|y, \theta_1^*, \ldots, \theta_{l-1}^*) \), \( l = 1, \ldots, B \), which can now be estimated in a manner akin to that in Chib (1995) by utilizing the output from appropriate reduced runs in which the parameters in \( \theta_l \) form a fixed block, and the remaining free parameters form random blocks.

Specifically, following Chib and Jeliazkov (2001), the simulation-consistent estimate of the \( l \)th ordinate is obtained as

\[
\hat{\pi}(\theta_l^*|y, \theta_1^*, \ldots, \theta_{l-1}^*) = \frac{n_1^{-1} \sum_{g=1}^{n_1} \alpha(\theta_l^{(g)}, \theta_l^*|y, \Psi_{l-1}, \Psi^{l+1,(g)})q_l(\theta_l^*|\Psi_{l-1}, \Psi^{l+1,(g)}, y)}{n_1^{-1} \sum_{j=1}^{n_1} \alpha(\theta_l^*, \theta_l^{(j)}|y, \Psi_{l-1}, \Psi^{l+1,(j)})},
\]

where \( \Psi_{l-1} = (\theta_1, \ldots, \theta_{l-1}), \Psi^{l+1} = (\theta_{l+1}, \ldots, \theta_B) \), and \( \alpha(\theta_l, \theta_l'|y, \Psi_{l-1}, \Psi^{l+1}) \) and \( q_l(\theta_l|y, \Psi_{l-1}, \Psi^{l+1}) \) are as defined in (6)-(7).

In the preceding expression, the average in the numerator is with respect to draws from the conditional distribution \( \pi(\theta_l^{(g)}, \Psi^{l+1}|y, \Psi_{l-1}^*) \). Accordingly, the \( n_1 \) draws \( \{\theta_l^{(g)}, \Psi^{l+1,(g)}\} \) in
the numerator are obtained from the $l$th stage reduced MCMC run, in which the parameters in the preceding blocks, collected in $Ψ_{l-1}$, are held fixed at $Ψ_{l-1}^*$. Furthermore, the sampling in each of these reduced runs is implemented by the TaB-MH algorithm, with the parameters in $θ_l$ forming a fixed block, and the parameters in $Ψ^{l+1}$ forming random blocks. As in the main run, this randomization serves to increase the efficiency of the sampling in the reduced runs.

On the other hand, the average in the denominator is taken with respect to the conditional product measure $π(Ψ^{l+1}|y, Ψ^*_l)$. Conveniently, draws from $π(Ψ^{l+1}|y, Ψ^*_l)$ are available from the calculation of the $(l+1)$st stage numerator which are supplemented with a drawing of $θ_l$ from $q_l(θ_l|y, Ψ^*_l, Ψ^{l+1})$. In the final stage, we draw $n_1$ values of $θ_B$ from $q_B(θ_B|y, Ψ^*_B)$ and compute the denominator average based on these draws.

Having estimated $π(θ^*|y)$, the log of the marginal likelihood is then available as

$$\log \hat{m}(y) = \log f(y|\theta^*) + \log π(\theta^*) - \sum_{l=1}^{B} \log \hat{π}(\theta_l^*|y, θ_1^*, \ldots, θ_{l-1}^*).$$  \hspace{1cm} (14)

We summarize this readily implementable calculation in algorithmic form as follows.

**Algorithm: Marginal likelihood**

**Step 1** Partition $θ$ into $B$ blocks, each block consisting of randomly chosen elements from $θ$

**Step 2** In the $l$th stage, $l = 1, \ldots, B$, repeat $n_1$ times:

(a) Generate $θ^{(g)}_l$, $Ψ^{l+1,(g)}_l$ from $π(θ_l, Ψ^{l+1}|y, Ψ^*_l)$ by the TaB-MH algorithm, randomizing only over the parameters in $Ψ^{l+1}$, and calculate the $l$th stage numerator summand in (13)

(b) Supplement the preceding draw with a draw $θ^{(g)}_{l-1}$ from $q_{l-1}(θ_{l-1}|y, Ψ^*_{l-2}, Ψ^{l,(g)})$ and calculate the $(l - 1)$st stage denominator summand in (13)

(c) Store these values

**Step 3** For the final stage, draw $n_1$ values $\{θ_B\}$ from $q_B(θ_B|y, Ψ^*_B)$ and compute the $l$th stage denominator in (13)
Step 4 Calculate the log of the marginal likelihood as in (14)

We complete the calculation of the marginal likelihood by finding the numerical standard error of the estimate which we get from the approach given in Chib (1995). As we show in our applications, the numerical standard error of the marginal likelihood estimate is typically low which indicates, that in conjunction with our sampling scheme, this method of estimating the marginal likelihood is reliable and efficient.

4 Applications

In this section we apply the methods discussed above to two DSGE models. In the first example (Ireland, 2004), we provide a comparative analysis of the TaB-MH and RW-MH algorithms. Subsequently, we illustrate the TaBMJ-MH algorithm in the context of the model in An and Schorfheide (2007) where the posterior distribution is multi-modal.

4.1 Ireland (2004) model

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

\[
\begin{align*}
\hat{x}_t &= \alpha_x \hat{x}_{t-1} + (1 - \alpha_x)E_t \hat{x}_{t+1} - (\hat{r}_t - E_t \hat{r}_{t+1}) + (1 - \omega)(1 - \rho_a)\hat{a}_t \\
\hat{r}_t &= \beta \alpha_r \hat{r}_{t-1} + \beta(1 - \alpha_r)E_t \hat{r}_{t+1} + \psi \hat{x}_t - \hat{e}_t \\
\hat{y}_t &= \hat{y}_t - \hat{y}_{t-1} + \hat{\varepsilon}_t \\
\hat{a}_t &= \hat{y}_t - \omega \hat{a}_t \\
\hat{r}_t &= \rho_r \hat{r}_{t-1} + \rho_\pi \hat{\pi}_t + \rho_g \hat{y}_t + \rho_x \hat{x}_t + \varepsilon_{r,t}
\end{align*}
\]  

(15)

where \( x_t, y_t, \pi_t, r_t \) and \( \pi_t \) 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; \( E_t \) 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 \( \hat{\varepsilon}_t \) 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). Further, it is assumed that the exogenous driving processes \( \hat{a}_t, \hat{e}_t, \hat{z}_t \) evolve independently of one another as

\[
\begin{align*}
\hat{a}_t &= \rho_a \hat{a}_{t-1} + \varepsilon_{a,t} \\
\hat{e}_t &= \rho_e \hat{e}_{t-1} + \varepsilon_{e,t} \\
\hat{z}_t &= \varepsilon_{z,t}
\end{align*}
\]

The innovations in the model \( \varepsilon_t = [\varepsilon_{a,t}, \varepsilon_{e,t}, \varepsilon_{z,t}, \varepsilon_{R,t}]' \) are assumed to be distributed as multivariate normal

\[
\varepsilon_t \sim \mathcal{N}_4(0, \Omega)
\]

where \( \Omega = \text{diag}(\sigma_a^2, \sigma_e^2, \sigma_z^2, \sigma_R^2) \). These equations are derived by log-linearizing the original microfounded model around its deterministic steady state, details of which are available in Ireland (2004).

There are two other parameters \( z \) and \( \pi \) in the non-linear model that determine the steady-state values of output growth and inflation, respectively. In addition, \( \beta \) determines the steady-state value of the short-term nominal interest rate through the relation \( \bar{r} = z\pi/\beta \). Following Ireland (2004), the values of \( z, \pi \) and \( \beta \) are set to the average levels of output growth, inflation and interest rates in the data. Also, \( \psi \) and \( \rho_r \) are fixed at 0.10 and 1.00, respectively. We collect the remaining 12 parameters of interest in the vector \( \theta \)

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

The parameters \( (\omega, \alpha_x, \alpha_\pi, \rho_\pi, \rho_g, \rho_x) \), where \( \omega, \alpha_x, \alpha_\pi \) are each assumed to be between 0 and 1, and each of \( \rho_\pi, \rho_g, \) and \( \rho_x \) are greater than 0, may be called the structural parameters. It is also assumed that \( (\rho_a, \rho_e) \) each lie between 0 and 1. Recalling the notations introduced in Section 2, we let \( S_L \) denote the subset of \( R^{12} \) satisfying these linear constraints. Finally, the variance parameters \( \sigma_i^2 \) lie in the region \( S_\Omega \) that satisfy the usual positivity and positive definiteness constraints.
To solve this model using Sims’ (2002) method mentioned in Section 2, we define

\[ s_t = [\hat{y}_t, \hat{r}_t, \hat{\pi}_t, \hat{\pi}_t, \hat{x}_t, \hat{z}_t, \hat{\pi}_{t+1}, \hat{x}_{t+1}]' \]

\[ \eta_t = [\hat{\pi}_t - \hat{\pi}_{t-1}, \hat{x}_t - \hat{x}_{t-1}]' \]

and \( \varepsilon_t \) is as defined above. Using these notations the set of equations (15)-(16) can be written succinctly as in (1).

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. The resulting measurement equation has the straightforward form

\[
\begin{bmatrix}
\hat{g}_t \\
\hat{\pi}_t \\
\hat{r}_t \\
y_t
\end{bmatrix} =
\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
s_t \\
Bs_t(17)
\end{bmatrix}
\]

In the following section we turn to the specification of the prior distribution \( \pi(\theta) \) for the parameters in this model.

4.1.1 Prior distribution

As mentioned in Section 1, an informative prior distribution can play an important role in the 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, thus smoothing the likelihood function to a certain extent. 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 (2008), Del Negro and Schorfheide (2008)).

Continuing our discussion of the Ireland model, we start by assuming that the parameters are apriori independent. Thus, the joint prior distribution is simply the product of the marginal distribution of each parameter. We now sample parameters from this prior 10,000 times, ensuring that the draws satisfy all the constraints mentioned above. For each draw of the parameters, we solve the model and generate the time series of the outcomes, namely \( \hat{g}_t \),
\( \hat{\pi}_t \) and \( \hat{\tau}_t \). We now 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.

- We assume a beta distribution for the parameters constrained to the \((0, 1)\) interval. Thus, \( \omega, \alpha_x, \alpha_\pi, \rho_a \) and \( \rho_e \) are each assumed to be distributed as

\[
\text{Beta}(\alpha_i, \beta_i)
\]

where the \((\alpha_i, \beta_i)\) 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.

- Parameters \( \rho_\pi, \rho_g \) and \( \rho_x \) are each assumed to follow a Gamma distribution

\[
\rho_i \sim G(\alpha_i, \beta_i)
\]

with \((\alpha_i, \beta_i)\) 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 \((0.25, 0.0625)\).

- Finally, we suppose a standard Inverse Gamma prior on 10,000 times the variance parameters:

\[
10000 \times \sigma_i^2 \sim IG(\alpha_i, \beta_i)
\]

\((\alpha_i, \beta_i) = (3, 60), (2, 0.08), (2.1, 5.5), (2.06, 0.53)\). This transformation is helpful in specifying an \( IG \) prior given the extremely small values of these parameters (relative to the others) needed to generate plausible implied data. It is perhaps interesting to note that we started with a mean of 0.3 and standard deviation equal to 0.15 in the
untransformed scale, but this translated to growth rates of more than 100% in the log scale.

Remark: Alternatively, one could convert the data into percentage units rather than working with decimal units and avoid rescaling the σ’s.

The implied data for each of the variables are shown in Figure 4. 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.

![Figure 4](image_url)

Figure 4: Implied $\hat{g}$, $\hat{\pi}$ and $\hat{r}$ from the prior distribution of the parameters in Ireland (2004). See Table 1 for a summary of the prior distribution.

4.1.2 Posterior and MCMC sampling

Before presenting the results from the TaB-MH algorithm, we focus on the implementation of the RW-MH algorithm for this problem. To illustrate the importance of the starting value of the chain and the variance of the random walk proposal we start three different RW chains, one at the prior mean, one at a local mode and the third at the dominant mode. For the first chain we specify the variance of the random walk proposal to be $k \times I_{12}$, where $I_{12}$ is the twelve

---

2These modes are found by a standard deterministic optimizer based on the method of steepest decent and through experimentation with the starting values.
dimensional identity matrix, whereas for the other two chains we let it be $k$ times the variance at the respective modes (calculated as the negative inverse of the Hessian). For the second and third chains it was reasonably easy to find a suitable $k$ (that generated roughly 50% acceptance rates) within a few tries. For the first chain, though, this task proved extremely tedious and time consuming. We started with $k = 0.5$ and finally stopped at $k = 0.0001$. Even with such a small variance the acceptance rate was only a disappointing 8.5%, which points to the complex correlation structure between the parameters. A summary of the results from the three chains, each run for 250,000 iterations following a burn-in of 50,000 iterations, are reported in Table 1 along with their respective starting values.

Table 1–Summary of results from three RW-MH chains for the parameters in Ireland (2004) model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Chain I</th>
<th>Chain II</th>
<th>Chain III</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.2000</td>
<td>0.0701</td>
<td>0.1058</td>
</tr>
<tr>
<td>$\alpha_x$</td>
<td>0.1000</td>
<td>0.0651</td>
<td>0.0629</td>
</tr>
<tr>
<td>$\alpha_y$</td>
<td>0.1000</td>
<td>0.0825</td>
<td>0.0605</td>
</tr>
<tr>
<td>$\rho_{\pi}$</td>
<td>0.3000</td>
<td>0.6079</td>
<td>0.5515</td>
</tr>
<tr>
<td>$\rho_{\pi}$</td>
<td>0.3000</td>
<td>0.4022</td>
<td>0.3593</td>
</tr>
<tr>
<td>$\rho_x$</td>
<td>0.2500</td>
<td>0.1825</td>
<td>0.1760</td>
</tr>
<tr>
<td>$\rho_m$</td>
<td>0.8500</td>
<td>0.9583</td>
<td>0.9334</td>
</tr>
<tr>
<td>$\rho_{a}$</td>
<td>0.7500</td>
<td>0.3201</td>
<td>0.5284</td>
</tr>
<tr>
<td>$\rho_{e}$</td>
<td>0.8500</td>
<td>0.8843</td>
<td>0.8874</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>10000</td>
<td>30.000</td>
<td>16.2229</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>10000</td>
<td>5.0000</td>
<td>0.6977</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>10000</td>
<td>0.5000</td>
<td>0.0857</td>
</tr>
</tbody>
</table>

Note: The columns start and post mean report the starting value and the mean of the sampled values for each of the three chains. Chain I is started at the prior mean with a variance of $0.0001 \times I_{12}$. Chain II is initiated at a local mode at which the ordinate of the (unnormalized) log posterior is roughly 992.68. The starting value for Chain III is the dominant mode, reaching a height of 1197.30 in the log scale.

A quick glance at the table reveals that sufficient exploration of the target distribution is not achieved with 250,000 iterations of chains I and II. In fact, further sampling shows that a million iterations also may not be enough. On the other hand, chain III, that was initiated at the dominant mode, performed quite well. The performance of these chains is better assessed in Figure 5 where we plot the sampled draws over time along with the corresponding sample
ACF for a couple of parameters from each of the three chains. As the top row shows, the first and second chains essentially did not converge, especially for $\sigma_a$. Not surprisingly, the sampled draws from these chains are highly persistent. This is in fact true even for chain III, where the ACFs retain significant mass for certain parameters even at lags of 250 and higher.

As might be evident from this exercise, unless one is able to exploit in some way the correlation among the various parameters, designing a sampler that makes large moves across the support of the distribution and quickly navigates to the meaningful regions of the posterior is not straightforward. The process can require substantial pre-run tuning and even so one is not assured of getting reliable results.

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 TaB-MH performed exceedingly well as can be seen in Table 2. To elaborate on the specifics of these results, we initialized the TaB-MH chain at the prior mean. Notice that unlike the RW-MH algorithm this does not require any sort of pre-run tuning. We also set the initial temperature $t_0$ in simulated annealing to 2, the rate of temperature decline $a$ to 0.4, the total number of stages $K$ to 6, the stage length increment $b$ to 10. We ran the chain for 10,000 iterations beyond a burn-in of 1000 iterations. We also specified a student-t distribution with $\nu = 15$ degrees of freedom for the proposal density.

The first point to note is that, unlike in the RW-MH algorithm, the sampling results in the case of the TaB-MH algorithm are virtually identical regardless of where the chain is initialized. Second, the sampler navigates to the same region of the parameter space within a few iterations. Specifically, a TaB-MH chain initialized at the prior mean climbs to the main regions of the posterior distribution within 30 iterations, whereas one initialized at the local mode corresponding to chain II in the RW-MH algorithm takes less than 10 iterations with the same settings for the SA parameters. This is not surprising given that information on both the

---

3The results reported here are virtually identical regardless of where the chain is initialized
location of the mode for a given block, as well as the correlation among the parameters in that block, is utilized in the sampling. Consequently, the autocorrelation among the draws decays
Table 2—Posterior sampling results using the TaB-MH algorithm for the Ireland (2004) model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Mean</th>
<th>Standard deviation</th>
<th>Posterior Mean</th>
<th>Numerical S.E.</th>
<th>90 percent interval</th>
<th>Inefficiency factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.20</td>
<td>0.10</td>
<td>0.1089</td>
<td>0.0010</td>
<td>[0.0381,0.2036]</td>
<td>5.2791</td>
</tr>
<tr>
<td>$\alpha_x$</td>
<td>0.10</td>
<td>0.05</td>
<td>0.0778</td>
<td>0.0006</td>
<td>[0.0186,0.1669]</td>
<td>2.7625</td>
</tr>
<tr>
<td>$\alpha_\pi$</td>
<td>0.10</td>
<td>0.05</td>
<td>0.0807</td>
<td>0.0009</td>
<td>[0.0184,0.1819]</td>
<td>4.9731</td>
</tr>
<tr>
<td>$\rho_\pi$</td>
<td>0.30</td>
<td>0.10</td>
<td>0.5522</td>
<td>0.0023</td>
<td>[0.3341,0.7767]</td>
<td>4.1913</td>
</tr>
<tr>
<td>$\rho_g$</td>
<td>0.30</td>
<td>0.10</td>
<td>0.3747</td>
<td>0.0011</td>
<td>[0.2751,0.4867]</td>
<td>3.9146</td>
</tr>
<tr>
<td>$\rho_x$</td>
<td>0.25</td>
<td>0.0625</td>
<td>0.2001</td>
<td>0.0016</td>
<td>[0.1108,0.3134]</td>
<td>9.2058</td>
</tr>
<tr>
<td>$\rho_\pi$</td>
<td>0.85</td>
<td>0.10</td>
<td>0.9310</td>
<td>0.0008</td>
<td>[0.8814,0.9662]</td>
<td>15.013</td>
</tr>
<tr>
<td>$\rho_e$</td>
<td>0.85</td>
<td>0.10</td>
<td>0.8674</td>
<td>0.0016</td>
<td>[0.7582,0.9555]</td>
<td>9.7198</td>
</tr>
<tr>
<td>$10000\sigma_a^2$</td>
<td>30.00</td>
<td>30.00</td>
<td>15.7994</td>
<td>0.3784</td>
<td>[6.0171,38.228]</td>
<td>15.814</td>
</tr>
<tr>
<td>$10000\sigma_e^2$</td>
<td>0.08</td>
<td>1.00</td>
<td>0.0068</td>
<td>0.0000</td>
<td>[0.0041,0.0107]</td>
<td>6.2913</td>
</tr>
<tr>
<td>$10000\sigma_z^2$</td>
<td>5.00</td>
<td>15.00</td>
<td>0.7633</td>
<td>0.0030</td>
<td>[0.4785,1.1145]</td>
<td>3.1988</td>
</tr>
<tr>
<td>$10000\sigma_r^2$</td>
<td>0.50</td>
<td>2.00</td>
<td>0.0969</td>
<td>0.0005</td>
<td>[0.0635,0.1443]</td>
<td>6.3380</td>
</tr>
</tbody>
</table>

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

quickly to zero. This translates into an efficient sampling scheme, one in which the Markov chain can be expected to converge quickly to its stationary distribution.

The serial correlations among the sampled draws can be summarized in terms of the so-called inefficiency factors. The inefficiency factors approximate the ratio of the numerical variance of the posterior mean from the MCMC chain relative to that from hypothetical iid draws. For a given sequence of draws the inefficiency factor is computed as

$$1 + 2 \sum_{l=1}^{L} \rho_k(l)$$

where $\rho_k(l)$ is the autocorrelation at lag $l$ for the $k$th sequence, and $L$ is the value at which the autocorrelation function tapers off (the higher order autocorrelations are also downweighted by a windowing procedure, but we ignore this aspect for simplicity). Whereas in the best case scenario of the RW-MH algorithm (refer to chain III in Table 1) the inefficiency factors range from 60 and 230, those for the TaB-MH algorithm are between 2.5 and 15 $^4$. Thus, in some cases the TaB-MH algorithm performs close to an independence sampler.

$^4$The inefficiency factors are computed for a maximum lag of $L = 250$.  

31
The extent (or lack thereof) of the serial correlation can also be observed through a plot of the autocorrelation functions of the sampled draws. We plot these functions in Figure 6.
Figure 6: Sampling results for the Ireland (2004) model using the TaB-MH algorithm (cont’d): Marginal prior-posterior plots and autocorrelation functions for the variances of the shock parameters.

These plots show that the serial correlations among the draws decay quickly to zero for all the parameters. In contrast, the serial correlations from the best performing RW-MH chain III, are considerably more persistent. This indicates that the mixing properties of our sampler are markedly better than those of the RW-MH algorithm. 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, the likelihood function carries information beyond that contained in the prior.

4.1.3 Marginal likelihood

In Section 3.2 we showed how the marginal likelihood can be estimated within the Chib (1995) framework when sampling is done with our tailored randomized block method. In this section, we apply that method to the Ireland model. Recall that the marginal likelihood computation requires decomposing the posterior ordinate into marginal-conditional densities. Unlike the fixed block case discussed in Chib and Jeliazkov (2001), there is some freedom here in select-
ing the number of blocks (stages), as well as the components of these blocks. Therefore, it is worthwhile to consider the marginal likelihood estimate, and the numerical standard errors that result, as the number of stages (and the sample size \( n_1 \) in the reduced MCMC runs) are varied. Accordingly, we consider both two and three stage schemes for calculating the posterior ordinate. We consider these many stages because the average number of blocks in the full MCMC run for this model is around 3. We also compare our results to the estimate of the marginal likelihood when sampling is done with the help of the RWMH algorithm.

The results are summarized in Table 3. For the two stage and three stage computations under the TaB-MH algorithm, we calculate the marginal likelihood from 5,000, 10,000 and 15,000 draws in the reduced runs. Under the single block RWMH scheme, the computations are carried out in 1-stage, with the draws in the numerator of equation (13) being those from the posterior sampling and those in the denominator being those from \( \mathcal{N}(\theta^*, c\Sigma) \), where \( \Sigma \) is the variance matrix computed at the posterior mode. In this calculation we use the output from our most efficient implementation of the RWMH algorithm (chain III in the preceding section) and illustrate the results for samples of sizes 75,000, 150,000 and 250,000.

Table 3–Log marginal likelihood estimates (with numerical standard errors) for the Ireland (2004) model based on the outputs from the TaB-MH and RWMH algorithms

<table>
<thead>
<tr>
<th>( n_1 )</th>
<th>TaB-MH</th>
<th>RWMH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 stage</td>
<td>3 stage</td>
</tr>
<tr>
<td>5,000</td>
<td>1170.08</td>
<td>1170.26</td>
</tr>
<tr>
<td></td>
<td>(0.0324)</td>
<td>(0.0400)</td>
</tr>
<tr>
<td>10,000</td>
<td>1170.18</td>
<td>1170.29</td>
</tr>
<tr>
<td></td>
<td>(0.0268)</td>
<td>(0.0302)</td>
</tr>
<tr>
<td>15,000</td>
<td>1170.15</td>
<td>1170.33</td>
</tr>
<tr>
<td></td>
<td>(0.0216)</td>
<td>(0.0250)</td>
</tr>
</tbody>
</table>

It is noteworthy that the numerical standard error of the marginal likelihood estimate (which are reported in parentheses in the table) are lower under the more efficient sampling scheme. This is in keeping with the discussion in Chib and Jeliazkov (2001). As one can see from the table, the numerical standard errors under the less efficient RWMH algorithm are between 15 and 20 times higher than those under the TaB-MH algorithm, for any sample
length. Furthermore, in this case, the estimates (for different reduced MCMC sample sizes) from the the 2-stage run are marginally more efficient than those from the 3-stage run.

4.2 Application to An and Schorfheide (2007)

While in many problems the posterior is relatively well defined in that there is a unique dominant mode, there can be instances where the posterior is (at least seemingly) multi-modal. In the realm of DSGE models, one such interesting example is demonstrated by An and Schorfheide (2007) in a 13-parameter model. In this model, the two distinct modes that show up during optimization appear to be separated by a deep valley, posing a severe impediment for the RW-MH algorithm. In fact, the authors show that the RW sampler was simply unable to escape the attraction of the individual modes even when the sampler was run for a million iterations following a burn-in of 100,000 iterations. What is especially interesting is that, despite a difference of approximately 8 in the log scale between the unnormalized posterior ordinates at the two modes, the RW chain does not cross over from the low modal region to the high modal region. Since the probability content in each of these separated regions is not known apriori, it is not possible to approximate the posterior distribution globally by running multiple chains of the RW-MH algorithm.

Before presenting the model in An and Schorfheide (2007), it is worth mentioning that the TaB-MH algorithm proposed in the previous section is far less susceptible to such problems. The first point to note is that a purely randomized block version of the TaB-MH sampler quickly navigates to the high modal region when initialized at the prior mean. In contrast, a RW-MH chain initiated at the same value converges to the low modal region. More importantly, even when the TaB-MH algorithm is started at the low mode, it does not get trapped in the region. The simple strategy employed here exploits the fact that six of the parameters differ significantly across the two modes. Thus, by grouping these six parameters in one block, and the remaining seven parameters in another, one is able to move across the two modes quite easily. Nonetheless, the TaBMJ-MH method is more efficient and easier to implement in the
case of multi-modal distributions. For the rest of this section we briefly present the model in An and Schorfheide (2007), henceforth referred to as the AS07 model.

The linearized DSGE model in AS07 is given as follows.

\[
\hat{y}_t = E_t \hat{y}_{t+1} + \hat{g}_t - E_t \hat{g}_{t+1} - \frac{1}{\tau} (\hat{r}_t - E_t \hat{r}_{t+1} - E_t \hat{z}_{t+1}) \tag{19}
\]

\[
\hat{\pi}_t = \beta E_t \hat{\pi}_{t+1} + \kappa (\hat{y}_t - \hat{g}_t) \tag{20}
\]

\[
\hat{c}_t = \hat{y}_t - \hat{g}_t \tag{21}
\]

\[
\hat{r}_t = \rho r \hat{r}_{t-1} + (1 - \rho_r) \psi_1 \hat{\pi}_t + (1 - \rho_r) \psi_2 (\Delta \hat{y}_t + \hat{z}_t) + \varepsilon_{r,t} \tag{22}
\]

where \( y_t, \pi_t, r_t \) and \( c_t \) denote output, inflation, nominal interest rate and consumption, respectively, and the hats denote log-deviation of the variables from their steady state or average values. In addition to the interest rate shock \( \varepsilon_{r,t} \) in the Taylor rule above, the system is driven by two other shocks, namely, a consumption shock and a technology shock. The evolution of these exogenous processes is governed by the independent AR(1) processes

\[
\hat{g}_t = \rho_g \hat{g}_{t-1} + \varepsilon_{g,t} \tag{23}
\]

\[
\hat{z}_t = \rho_z \hat{z}_{t-1} + \varepsilon_{z,t} \tag{24}
\]

where \( \varepsilon_{i,t} \) is normally distributed with mean 0 and standard deviation \( \sigma_i \). As before, we collect the shocks in the vector \( \varepsilon_t = [\varepsilon_{r,t}, \varepsilon_{g,t}, \varepsilon_{z,t}]' \). Then \( \varepsilon_t \sim N_3(0, \Omega) \), where \( \Omega = \text{diag}[\sigma_r^2, \sigma_g^2, \sigma_z^2] \).

As in the previous section, define the terms

\[
s_t = [\hat{\pi}_t, \hat{y}_t, \hat{r}_t, \hat{c}_t, \hat{g}_t, \hat{z}_t, E_t \hat{\pi}_{t+1}, E_t \hat{y}_{t+1}]'
\]

and

\[
\eta_t = [\hat{r}_t - E_{t-1} \hat{r}_t, \hat{g}_t - E_{t-1} \hat{g}_t]'
\]

to express the model compactly as in (1). The model is solved, subject to the determinacy constraint, to obtain a stochastic law of motion for the 7-dimensional state vector \( s_t \).

The simulated data used in the fitting are assumed to be the series of quarterly observations on per capita GDP growth rates \( \hat{Y}_t = \gamma^Q + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t) \), annualized inflation rates
π_t = π^A + 400\hat{\pi}_t, and annualized nominal interest rate r_t = π^A + r^A + 4γQ + 400\hat{r}_t, where γQ, r^A, and π^A are related to the steady states of the relevant variables. Table 4 contains the parameters of the data generating process, along with the prior distribution, both of which are taken from AS07. It is assumed that the sample size is 80. To account for the lagged output in the data, the state vector is augmented with \hat{y}_{t-1}. The redefined state vector takes the form

\[ s_t = [\hat{\pi}_t, \hat{\gamma}_t, \hat{r}_t, \hat{\psi}_t, \hat{\gamma}_Q, \hat{\pi}_A, \hat{r}_A, \hat{\psi}_A, \hat{\gamma}_A, \hat{y}_{t-1}]' \]

and the resulting SSM is then written as

\[
\begin{bmatrix}
\hat{Y}_t \\
π_t \\
r_t \\
y_t
\end{bmatrix} =
\begin{bmatrix}
γ(Q) \\
π^A \\
\pi^A + r^A
\end{bmatrix} +
\begin{bmatrix}
0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 4 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\gamma \\
π^A \\
r^A
\end{bmatrix}
\begin{bmatrix}
\theta
\end{bmatrix}
\]

\[ s_t = D(θ)s_{t-1} + F(θ)\varepsilon_t. \] (26)

### Table 4–DGP and prior distribution for the model parameters in An and Schorfheide (2007)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DGP</th>
<th>Density</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ</td>
<td>2.00</td>
<td>Gamma</td>
<td>2.00</td>
<td>0.50</td>
</tr>
<tr>
<td>κ</td>
<td>0.15</td>
<td>Gamma</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>ψ_1</td>
<td>1.50</td>
<td>Gamma</td>
<td>1.50</td>
<td>0.25</td>
</tr>
<tr>
<td>ψ_2</td>
<td>1.00</td>
<td>Gamma</td>
<td>0.50</td>
<td>0.25</td>
</tr>
<tr>
<td>ρ_r</td>
<td>0.60</td>
<td>Beta</td>
<td>0.50</td>
<td>0.20</td>
</tr>
<tr>
<td>ρ_g</td>
<td>0.95</td>
<td>Beta</td>
<td>0.80</td>
<td>0.10</td>
</tr>
<tr>
<td>ρ_z</td>
<td>0.65</td>
<td>Beta</td>
<td>0.66</td>
<td>0.15</td>
</tr>
<tr>
<td>r^A</td>
<td>0.40</td>
<td>Gamma</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>π^A</td>
<td>4.00</td>
<td>Gamma</td>
<td>7.00</td>
<td>2.00</td>
</tr>
<tr>
<td>γ_Q</td>
<td>0.50</td>
<td>Normal</td>
<td>0.40</td>
<td>0.20</td>
</tr>
<tr>
<td>σ_r</td>
<td>0.20</td>
<td>Inverse Gamma</td>
<td>0.50</td>
<td>0.26</td>
</tr>
<tr>
<td>σ_g</td>
<td>0.80</td>
<td>Inverse Gamma</td>
<td>1.25</td>
<td>0.65</td>
</tr>
<tr>
<td>σ_z</td>
<td>0.45</td>
<td>Inverse Gamma</td>
<td>0.63</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Remark: The IG(\tilde{α}, \tilde{β}) distribution, sometimes referred to as an Inverse Gamma type-I distribution, on σ is the derived distribution from the assumption that σ^2 \sim IG(α, β).
4.2.1 Sampling results from the TaBMJ-MH algorithm

From the point of view of this paper, an interesting aspect of this model is the occurrence of possibly multiple modes. Specifically, by maximization of the posterior distribution with respect to the parameters one can find two distinct optima with the optima differing primarily in terms of the components \((\tau, \psi_2, \rho_r, \rho_z, \sigma_g, \sigma_z)\). The modal values are presented in Table 4. While it is not a simple task to ascertain the importance of the two modal regions in the posterior distribution, one point to note is that the difference in the (unnormalized) posterior ordinates at the two modes is around 8.3 in the log scale. We can understand the implication of this difference on the relative probability mass in the two regions by the following simple exercise.

Denote the high mode by \(\mu_1\) and the low mode by \(\mu_2\). Assume that the underlying posterior distribution can be approximated by a two component mixture of 13-dimensional normals as follows

\[
p \mathcal{N}(\theta|\mu_1, \Sigma_1) + (1 - p) \mathcal{N}(\theta|\mu_2, \Sigma_2)
\]

where the \(\Sigma\)s are the negative inverse hessians at the respective modes. Under this working assumption, \(p\) is the only unknown. From the knowledge of the ordinates at the two modes one can write

\[
\frac{p \mathcal{N}(\mu_1|\mu_1, \Sigma_1) + (1 - p) \mathcal{N}(\mu_1|\mu_2, \Sigma_2)}{p \mathcal{N}(\mu_2|\mu_1, \Sigma_1) + (1 - p) \mathcal{N}(\mu_2|\mu_2, \Sigma_2)} = \exp\{8\}
\]

which yields \(p = 0.99995\). This suggests that the mass in the second mode is around 0.005%, which is small enough to be ignorable. Arguably, this calculation is only suggestive of the actual probability of the low modal region. Nonetheless, it leads to the conjecture that the mass in the second modal region is insignificant as one might intuitively expect from the disparity in the heights of the two modes.

For sampling the posterior distribution in this model we initialized the TaBMJ-MH sampler at the low mode. For the usual TaB-MH iterations with fully randomized blocks, the parameters for the simulated annealing algorithm were chosen as follows. We set the number of stages \(K\) to 10, the increment of stage length \(b\) to 15, the initial temperature to 75 and the
Table 5–Posterior sampling results using the TaBMJ-MH algorithm for the An-Schorfheide (2007) model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode 1</th>
<th>Mode 2</th>
<th>Mean</th>
<th>S.E.</th>
<th>90 percent interval</th>
<th>Inefficiency factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau )</td>
<td>2.05</td>
<td>1.41</td>
<td>2.12</td>
<td>0.0392</td>
<td>[1.04,3.74]</td>
<td>29.91</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.16</td>
<td>0.18</td>
<td>0.17</td>
<td>0.0059</td>
<td>[0.03,0.42]</td>
<td>34.54</td>
</tr>
<tr>
<td>( \psi_1 )</td>
<td>1.55</td>
<td>1.57</td>
<td>1.66</td>
<td>0.0149</td>
<td>[1.12,2.42]</td>
<td>19.88</td>
</tr>
<tr>
<td>( \psi_2 )</td>
<td>0.96</td>
<td>0.81</td>
<td>1.00</td>
<td>0.0101</td>
<td>[0.70,1.35]</td>
<td>36.38</td>
</tr>
<tr>
<td>( \rho_r )</td>
<td>0.59</td>
<td>0.49</td>
<td>0.59</td>
<td>0.0054</td>
<td>[0.41,0.72]</td>
<td>49.38</td>
</tr>
<tr>
<td>( \rho_g )</td>
<td>0.94</td>
<td>0.97</td>
<td>0.92</td>
<td>0.0033</td>
<td>[0.79,0.98]</td>
<td>41.57</td>
</tr>
<tr>
<td>( \rho_z )</td>
<td>0.58</td>
<td>0.80</td>
<td>0.54</td>
<td>0.0094</td>
<td>[0.21,0.83]</td>
<td>31.57</td>
</tr>
<tr>
<td>( r^A )</td>
<td>0.64</td>
<td>0.62</td>
<td>0.68</td>
<td>0.0083</td>
<td>[0.07,1.43]</td>
<td>5.54</td>
</tr>
<tr>
<td>( \pi^A )</td>
<td>4.06</td>
<td>4.00</td>
<td>4.16</td>
<td>0.0212</td>
<td>[3.28,5.52]</td>
<td>14.48</td>
</tr>
<tr>
<td>( \gamma^Q )</td>
<td>0.50</td>
<td>0.54</td>
<td>0.48</td>
<td>0.0050</td>
<td>[0.11,0.80]</td>
<td>8.655</td>
</tr>
<tr>
<td>( \sigma_r )</td>
<td>0.22</td>
<td>0.24</td>
<td>0.23</td>
<td>0.0012</td>
<td>[0.18,0.32]</td>
<td>11.61</td>
</tr>
<tr>
<td>( \sigma_g )</td>
<td>0.76</td>
<td>1.07</td>
<td>0.76</td>
<td>0.0120</td>
<td>[0.45,1.33]</td>
<td>28.57</td>
</tr>
<tr>
<td>( \sigma_z )</td>
<td>0.54</td>
<td>0.30</td>
<td>0.61</td>
<td>0.0111</td>
<td>[0.30,1.01]</td>
<td>38.86</td>
</tr>
</tbody>
</table>

Remark: The results reported in this table are based on the low mode as the starting value. The TaBMJ-MH algorithm was run for 10,000 iterations, with the mode jumping step using the mixture distribution proposed every 100th iteration.

As expected, the transition from the low mode to the high mode was almost instantaneous (in that the first time a value from the high modal region was proposed it was accepted). This jump occurred at the 200th iteration. However, the reverse transition through the mode jumping step occurred only once in the 300th iteration. Beyond this, any value proposed from the low modal region in the mode jumping step got rejected. This is of course not surprising given the difference in the ordinates at the two modes. As one can verify from (8), the probability of accepting a draw from the low mode is virtually zero when the current state is in the high modal region. In the non-mode jumping steps, the usual TaB-MH sampler
otherwise moved across the support of the posterior distribution as in the Ireland model, with occasional visits to the low mode. This points to the potential of this algorithm in the global exploration of the posterior distribution.

The results from this sampler are summarized in Table 5. This table also contains the modal information of the parameters. These results are based on an acceptance rate of 45%. Particularly noteworthy in this table is the movement of the sampler across the two modal regions as indicated by the 90 percent intervals (column 6 in table). Also worth recognizing is the fine mixing of the sampler as indicated by the inefficiency factors. The highest inefficiency factor is around 50, with the average inefficiency factor across all the parameters being 27.\(^5\) These factors combined together suggest an efficient, well mixing sampler. The movement of the sampler across the support of the posterior is better illustrated in Figure 7. The plots show the sampled draws from the posterior distribution for the six parameters that differ across the

\(^5\)These inefficiency factors are low given the irregularity of the posterior surface in this problem and in relation to what can be achieved without recourse to our sampling scheme.
modes. As one can see, the sampler visited the low modal region a few times. However, the proportion of draws from this region are not substantial enough to define a distinct mode. Overall, this further strengthens our belief that the low modal region has negligible probability content.

We conclude this section by calculating the marginal likelihood. As in the Ireland example, we present the marginal likelihood estimates from both the two and three stage implementations of the Chib method and different MCMC sample sizes. For comparison, we also include the marginal likelihood estimate based on the RWMH output. The estimates are summarized in Table 6. Just as in the previous example, one can see from the numerical standard errors (which are given in parentheses) that the RWMH based estimate is relatively inefficient. It is evident from the results that 75,000 iterations of the RWMH algorithm are not sufficient to estimate the marginal likelihood. In contrast, a stable and efficient estimate can be obtained with a sample size of 5000 from the Tab-MH algorithm, with either 2 or 3 stages.

Table 6–Log marginal likelihood estimates (with numerical standard errors) for the An and Schorfheide (2007) model based on the outputs from the TaB-MH and RWMH algorithms

<table>
<thead>
<tr>
<th></th>
<th>TaB-MH</th>
<th>RWMH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2 stage</td>
</tr>
<tr>
<td>5,000</td>
<td>-206.64</td>
<td>-206.18</td>
</tr>
<tr>
<td></td>
<td>(0.1009)</td>
<td>(0.0944)</td>
</tr>
<tr>
<td>10,000</td>
<td>-206.58</td>
<td>-206.29</td>
</tr>
<tr>
<td></td>
<td>(0.0729)</td>
<td>(0.0671)</td>
</tr>
<tr>
<td>15,000</td>
<td>-206.55</td>
<td>-206.34</td>
</tr>
<tr>
<td></td>
<td>(0.0586)</td>
<td>(0.0544)</td>
</tr>
</tbody>
</table>

5 Conclusion

In this paper we provide new MCMC approaches for estimating DSGE models. The organizing principle is the sampling of the posterior density by a specially constructed version of a (randomized) multiple-block M-H algorithm that is designed to deal with the sort of considerations that arise in the context of these models. Specifically, the blocks and its constituents
are constructed randomly and the proposal density in the M-H sampling is found from the output of simulated annealing. The combined thrust of these innovations proves extremely effective as we show in the context of our two examples. We also provide a straightforward extension of our general approach that can be used for sampling distributions in which one suspects multiple modes. On both dimensions, this improves significantly on what has been achieved so far. Finally, we show how the model marginal likelihood, which is needed in the comparison of alternative DSGE models, can be computed efficiently from the output of our tailored randomized block sampling schemes. In sum, given that the methods proposed here are reliable and efficient, they have the potential for broadening the appeal and practicality of Bayesian methods in the fitting of DSGE models.

References


