Waggoner, Daniel F.; Wu, Hongwei; Zha, Tao

Working Paper

The dynamic striated Metropolis-Hastings sampler for high-dimensional models

Working Paper, No. 2014-21

Provided in Cooperation with:
Federal Reserve Bank of Atlanta

Suggested Citation: Waggoner, Daniel F.; Wu, Hongwei; Zha, Tao (2014) : The dynamic striated Metropolis-Hastings sampler for high-dimensional models, Working Paper, No. 2014-21, Federal Reserve Bank of Atlanta, Atlanta, GA

This Version is available at:
https://hdl.handle.net/10419/114478

FEDERAL RESERVE BANK *of* ATLANTA **WORKING PAPER SERIES**

The Dynamic Striated Metropolis-Hastings Sampler for High-Dimensional Models

Daniel F. Waggoner, Hongwei Wu, and Tao Zha

Working Paper 2014-21
November 2014

Abstract: Having efficient and accurate samplers for simulating the posterior distribution is crucial for Bayesian analysis. We develop a generic posterior simulator called the "dynamic striated Metropolis-Hastings (DSMH)" sampler. Grounded in the Metropolis-Hastings algorithm, it draws its strengths from both the equi-energy sampler and the sequential Monte Carlo sampler by avoiding the weaknesses of the straight Metropolis-Hastings algorithm as well as those of importance sampling. In particular, the DSMH sampler possesses the capacity to cope with incredibly irregular distributions that are full of winding ridges and multiple peaks and has the flexibility to take full advantage of parallelism on either desktop computers or clusters. The high-dimensional application studied in this paper provides a natural platform to put to the test generic samplers such as the DSMH sampler.

JEL classification: C32, C63, E17

Key words: dynamic striation adjustments, simultaneous equations, Phillips curve, winding ridges, multiple peaks, independent striated draws, irregular posterior distribution, importance weights, tempered posterior density, effective sample size

---

This research is supported in part by the National Science Foundation Grant SES 1127665. Simulations are programmed in C++, tailored to parallel and grid computations. The views expressed herein are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or the National Bureau of Economic Research.

Please address questions regarding content to Daniel F. Waggoner, Federal Reserve Bank of Atlanta and Emory University, Research Department, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8278, dwaggoner@frbatlanta.org; Hongwei Wu, Federal Reserve Bank of Atlanta and Emory University, Research Department, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8413, hongwei.wu@atl.frb.org; or Tao Zha, Federal Reserve Bank of Atlanta and Emory University, Research Department, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8353, zmail@tzha.net.

I. Introduction

We develop a new posterior simulation method that allows researchers to estimate high-dimensional economic and statistical models that have irregular likelihoods with multiple peaks and complicated winding ridges. We undertake this research mainly because, in recent years, the Bayesian estimation and evaluation of multivariate dynamic models have played a central role in assessing how well the model fits to the data and in selecting the best-fit model for forecasting and policy analysis (Geweke, 1999; Christiano, Eichenbaum, and Evans, 1999, 2005; An and Schorfheide, 2007; Smets and Wouters, 2007).

Standard Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings algorithm, work well for estimating models with likelihoods or posterior distributions that have smooth Gaussian shapes. For high-dimensional economic and statistical models, however, the likelihood or the posterior distribution can be non-Gaussian with highly irregular shapes and multiple peaks. These problems can severely compromise the accuracy of previous MCMC samplers for Bayesian inference.

To tackle such problems we develop a new posterior simulation method, called the dynamic striated Metropolis-Hastings (DSMH) sampler. It draws the strengths of two recently developed samplers: equi-energy (EE) (Kou, Zhou, and Wong, 2006) and sequential Monte Carlo (SMC) (Chopin, 2004; Durham and Geweke, 2012; Herbst and Schorfheide, 2014) simulators. The basic idea behind these two techniques is to start with a tractable initial distribution one can sample from and then to transform this initial distribution gradually to the desired posterior distribution through a sequence of stages. At each stage the sample from the previous stage is used to form a new sample for the current stage. At the final stage the sample comes from the desired posterior distribution. The differences between these techniques is how the initial stage is formed and how information from the previous stage is used for the current stage.

The contribution of this paper has several dimensions. First, the DSMH sampler is designed to take advantage of the strengths of Metropolis-Hastings sampling on the one hand and overcome its well-known weaknesses on the other hand. We divide the target distribution into various levels and define a striation as a parameter space in which the posterior kernel is between the two adjacent levels. Striations are *dynamically* adjusted when we move from one stage to another. This dynamic adjustment ensures that each striation remains fully populated by *independent* striated draws within the striation. Among other things, it marks a major departure from the EE sampler and plays an indispensable role in the DSMH sampler. In theory we show that convergence holds for our sampler with dynamically adjusted striations. In practice this dynamic feature is the key for the DSMH sampler to avoid

getting stuck in a local parameter region with highly correlated draws. Consequently the DSMH sampler is capable of exploring the entire posterior distribution.

Second, like the SMC sampler, the DSMH sampler takes advantage of parallel computing and approximates marginal data densities or marginal likelihoods as a by-product. The SMC sampler relies on importance sampling to reweight the sample of random draws (particles) at each stage. The problem inherent in importance sampling is that particles tend to collapse so that only a small fraction of the sample receives most weights. The remedy, called "the mutation," is to use the Metropolis-Hastings algorithm to resample new particles when the importance sampler begins to collapse. By contrast, because the DSMH sampler is grounded in the Metropolis-Hastings algorithm and utilizes importance weights *only for an initial draw* at each stage, it does not suffer the degenerate problem characteristic of the SMC sampling. As long as striations are dynamically adjusted across stages to maintain a sufficient probability for each striation, independent sampling within the striation enables the DSMH sampler to traverse the entire parameter space.

Third, we apply the DSMH sampler to structural vector autoregressions (SVAR) models. This application is relevant and important for several reasons. Many multivariate dynamic models such as dynamic stochastic general equilibrium (DSGE) models are closely connected to VAR models (Ingram and Whiteman, 1994; Del Negro and Schorfheide, 2004). Understanding how the DSMH sampler works for SVAR models provides a first step toward extending application to other multivariate dynamic models. We show that the exact Gibbs sampler exists at every stage of our sampler. This powerful result allows us to obtain accurate posterior draws from the Gibbs sampler at each stage and compare this "true" distribution to the distribution simulated from the DSMH sampler. Because the sampling quality is sensitive to the values of tuning parameters (as in any Monte Carlo simulation technique), these values that work for our SVAR application serve as an informative benchmark for other applications in which the exact Gibbs sampler may no longer be available.

Because the posterior distributions for reduced-form VAR models or SVAR models with recursive identification are well behaved, a successful application of the DSMH sampler to these models does not necessarily mean that the sampler is capable of exploring irregular high-dimensional distribution. The SVAR models studied in this paper, however, are far more challenging than those models. The identification is non-recursive and involves simultaneous equations in the spirit of Sims and Zha (2006). Moreover, we choose to work on the *unnormalized* SVAR model so as to make the posterior distribution populated with at least as many as 2^n isolated peaks, where n is the number of equations. A combination of simultaneity and un-normalization makes the posterior distribution incredibly complex, full of thin winding ridges between peaks. Thus the posterior distribution in our application is

complicated enough to serve as a serious testing ground for any generic Monte Carlo sampler.[1] By generating independent draws from the Gibbs sampler, we are able to evaluate how well the DSMH sampler works against the underlying distribution.

The rest of the paper is organized as follows. Section II develops the DSMH sampler with theoretical justifications. Section III addresses a number of practical issues that are relevant to the end user. Section IV discusses two major difficulties that lie at the heart of estimation of multivariate dynamic models. Section V presents two challenging SVAR models that put the generic DSMH sampler to the full test. Both models have highly irregular posterior distributions. SectionVI offers concluding remarks.

II. The Dynamic Striated Metropolis-Hastings Sampler

In this section we deliver a detailed description of the DSMH sampler and discuss conditions under which convergence holds. Because the DSMH sampler combines the strengths of both the EE and SMC samplers, we contrast it with these other samplers throughout the section to facilitate understanding of our new sampler.

II.1. **Generic Algorithm.** Let $Y_T = (y_1, \ldots, y_T)$ denote the observable variables, where T is the total number of observations and y_t denotes an $n \times 1$ vector of variables observed at time t. The likelihood function is denoted by $p(Y_T|\theta)$, where $\theta \subset \Theta \subset \mathbb{R}^m$ is a vector of parameters. Combining the likelihood and the prior probability density $\pi(\theta)$, we obtain the posterior kernel $p(Y_T|\theta)\pi(\theta)$.[2] To simplify notation, we denote the posterior kernel by $p(\theta)$ with the understanding that $p(\theta)$ depends on the data Y_T. The DSMH sampler proceeds through a series of stages, each associated with a target probability distribution on Θ. The initial stage's target distribution must be tractable, i.e. one must be able to sample independently from the distribution and be able to compute its probability density, not just its kernel. This initial distribution is gradually transformed until, at the final stage, the target is the posterior distribution.[3] At each stage one obtains a sample from the target distribution using the sample obtained from the previous stage. The sample at the final stage is from the posterior distribution. The DSMH sampler is designed to ensure

[1] It is beyond the scope of this paper to test other generic samplers. How well a particular sampler works depends on fine tuning as well as many factors, and generally requires an extraordinary amount of time to deal with a specific application at hand. See, for example, Bognanni and Herbst (2014).

[2] A probability kernel is non-negative and integrates to a positive finite number that may not be one, while a probability density is nonnegative and integrates to exactly one.

[3] Although the basic idea is the same for the DSMH, EE, and SMC samplers, but how this idea is implemented differs substantially across these samplers. Even within the SMC sampler, for instance, the implementation differs considerably between the method of Durham and Geweke (2012) and Herbst and Schorfheide (2014).

the sample is representative, even if the posterior distribution has a complicated shape in a high-dimensional space.

II.1.1. *Stages.* We transform the posterior distribution by tempering the probability kernel $p(\theta)$. For any real number λ satisfying $0 < \lambda \leq 1$, define

$$f_\lambda(\theta) = p(\theta)^\lambda.$$

The value λ controls the degree of tempering.[4] When $\lambda = 1$, $f_1(\theta)$ is the posterior kernel; as λ tends to zero, $f_\lambda(\theta)$ tends to one for all θ with positive posterior probability.

To see how tempered posterior kernels behave, consider a simple one-dimensional example in which $ay_t = \varepsilon_t$, where ε_t is a standard normal random variable. If the prior on a is Gaussian and centered at the origin, then the tempered posterior is proportional to

$$|a|^{\lambda T} \exp\left(-\frac{\lambda T a^2}{2\sigma^2}\right),$$

where σ is a function of the sample variance of the data and the variance of the prior. Figure 1 plots tempered posterior kernels of a for λ varying from 0.001 to 1.0 with $T = 20$ and $\sigma = 1$.[5] This simple example embodies the two essential features of the DSMH sampler. First, the most tempered posterior kernel is very flat (the thick solid line in the figure) and similar to a Gaussian distribution with a large standard deviation. Second, as λ tends to one, the tempered posterior kernels gradually converge to the posterior distribution (the thick dashed line in the figure). More important is the fact that the two peaks of the posterior kernel grow from the much flatter peaks of the tempered posteriors. This is one of the main reasons for the DSMH sampler to temper the posterior kernel, not just the likelihood function, as is done in the SMC sampler literature.

To define stages we choose λ_i, for $1 \leq i \leq H$, such that $0 < \lambda_1 < \cdots < \lambda_{H-1} < \lambda_H = 1$. For $1 \leq i \leq H$, the target distribution for the i^{th} stage is $f_{\lambda_i}(\theta)$.[6] Note that the final target distribution $f_{\lambda_H}(\theta)$ is the posterior kernel as required. Recommendations for how to choose λ_i are given in Section III.2.

A tractable distribution for the initial (0^{th}) stage must be specified. We recommend a student-t distribution with ν degrees of freedom, mean μ_0, and variance Ω_0. In Section III.3 we discuss how to choose ν, μ_0, and Σ_0. We denote the target distribution at the initial stage by $f_0(\theta)$, but the reader should keep in mind that this distribution does not come from

[4]The EE literature refers to $1/\lambda$ as temperature and uses it to measure the degree of tempering. We prefer λ since it is this term that directly appears in all formulae.

[5]These kernels are unnormalized and thus have multiple peaks. We use unnormalized kernels to illustrate how the sampler handles irregular posterior kernels.

[6]We use $f_{\lambda_i}(\theta)$ to denote both the probability kernel and the actual distribution itself.

the family of tempered posterior kernels and that $f_0(\theta)$ must be a density function, not just a kernel.

For each stage i we obtain a sample from $f_{\lambda_i}(\theta)$, which we denote by $\left\{\theta^{(i,\ell)}\right\}_{\ell=1}^{NG}$, where NG is the total number of simulations. The sample $\left\{\theta^{(0,\ell)}\right\}_{\ell=1}^{NG}$ consists of independent random draws from the initial distribution and the sample $\left\{\theta^{(H,\ell)}\right\}_{\ell=1}^{NG}$ comes from the posterior distribution. In general, the sample $\left\{\theta^{(i,\ell)}\right\}_{\ell=1}^{NG}$ depends on the sample $\left\{\theta^{(i-1,\ell)}\right\}_{\ell=1}^{NG}$. This dependence allows the DSMH sampler to take full advantage of the sample at previous stages and the nature of this dependence marks a major departure from the SMC sampler as discussed in the following two sections.

II.1.2. *Striations.* Most of the time the DSMH sampler at the i^{th} stage functions as the standard random walk Metropolis algorithm with the target $f_{\lambda_i}(\theta)$. But occasionally a proposal draw comes from the sample at the previous stage. Random draws are accepted or rejected with an appropriate Metropolis-Hastings acceptance criterion. How to simulate random draws from the previous stage is crucial to the efficiency of the sampler. When simulating from the tempered distribution at the previous stage, we would like those draws to be similar to the current draw in terms of the level (height) of the posterior kernel. As a result, those draws from the previous stage are likely to be accepted, and because they are simulated *independently*, the sampler moves efficiently among the values of θ that have similar posterior values. On the other hand, the random walk component of the sampler allows movement among the values of θ that have different posterior values. Put differently, proposal draws from the tempered distribution at the previous stage move independently *within* the same level set while serially-correlated random walk proposal draws move *between* level sets.

We call a "striation" the set of all values of θ that have similar posterior values. Striations at the i^{th} stage are defined by a sequence of $M+1$ levels, denoted by $L_{i,k}$, satisfying $0 = L_{i,0} < L_{i,1} < \cdots < L_{i,M-1} < L_{i,M} = \infty$. For $1 \leq k \leq M$, the k^{th} striation is the set

$$S_{i,k} = \{\theta \in \Theta \mid L_{i,k-1} \leq p(\theta) \leq L_{i,k}\}. \tag{1}$$

We choose the levels so that the probability that $\theta \in S_{i,k}$ is equal to $1/M$. This probability is with respect to the distribution at the *previous* stage. If

$$I_{i-1} = \int_{\theta \in \Theta} f_{\lambda_{i-1}}(\theta) d\theta, \tag{2}$$

the levels are chosen to satisfy

$$\frac{1}{M} = \int_{\theta \in S_{i,k}} \frac{f_{\lambda_{i-1}}(\theta)}{I_{i-1}} d\theta.$$

It is generally impossible to find analytic expressions for setting the levels. One can, however, use the sample from the previous stage to set the levels by simply choosing $L_{i,k}$ so that an

equal number of draws lie in each striation. We find that this simple rule works well in practice for determining the levels.[7] There are tradeoffs in determining the number of levels. On the one hand, one would like to have striations as small as possible to allow the random walk Metropolis algorithm to move between striations more efficiently. This argues for a larger M. On the other hand, we need the sample in each striation to be representative, which argues for a smaller M. We find that setting M to 20, so that each striation contains 5% of the draws, is a good generic value. In Section III.4 we discuss this choice and its consequences.

In general, levels chosen at stage $i-1$ differ from those chosen at stage i. This is an important departure from the EE sampler developed by Kou, Zhou, and Wong (2006). In that setup, the levels are the same for all stages so that $L_{i,k} = L_k$ for all i. Furthermore, the levels follow a geometric progression such that $L_{k+1} = \gamma L_k$ with $\gamma > 1$ being a key tuning parameter. And with the requirement $\gamma L_{M-1} = \sup_\theta\{p(\theta)\}$, the sequence is completely specified by γ and has no room for flexibility. By allowing levels differ across stages, our approach has two substantive advantages. First, it is unnecessary to maximize the posterior, which is a hard problem in and out of itself. Indeed, optimization proceeds best if one can sample first and then use sampled draws as starting points for the maximization routine. Second, because the striations are fixed by the EE sampler, we find that most of the striations contain no draws in the later stages of the sampler. This phenomenon negates one of the main advantages of utilizing striations. Indeed, the term "dynamic" in DSHM refers to the important fact that levels can change from stage to stage. Such a *dynamic* adjustment is critical because it ensures that each striation remains fully populated so that all of the information in the previous sample is efficiently exploited.

II.1.3. *Metropolis-Hastings.* We now turn to the details of the Metropolis-Hastings proposal distribution. The proposal distribution is a mixture of a Gaussian and $f_{\lambda_{i-1}}(\theta)$, the distribution at the previous stage. If $\theta^* \equiv \theta^{(i,\ell)}$ is the most recent draw from the Metropolis-Hastings sampler at the i^{th} stage, the proposal density of θ given θ^* is

$$g_i(\theta, \theta^*) = (1-p)\phi_{\mathfrak{c}_i\Omega_i}(\theta - \theta^*) + p\chi(\theta, \theta^*)\frac{M f_{\lambda_{i-1}}(\theta)}{I_{i-1}},$$

where $\phi_{\mathfrak{c}_i\Omega_i}(\cdot)$ is the density of the mean-zero Gaussian distribution with variance $\mathfrak{c}_i\Omega_i$ and $\chi(\theta, \theta^*)$ is the indicator function that returns one if θ and θ^* are in the same striation and zero otherwise.[8] The mixture form of $g_i(\theta, \theta^*)$ dictates that with probability $1-p$, θ is drawn

[7] In principle one could also allow an unequal number of draws to lie in each striation as long as there are enough draws in each striation.

[8] Given Ω_i, $\mathfrak{c}_i$ is chosen so that the acceptance rate of the Metropolis-Hastings algorithm is approximately equal to some target rate α, which is taken to be 30% in our application. A practical operation for determining $\mathfrak{c}_i$ is to run the sampler for a short period of time, compute the acceptance rate, and then adjust $\mathfrak{c}_i$ accordingly

from the Gaussian distribution centered at θ^* and with probability p, θ is drawn from the distribution at the previous stage but from the same striation that contains θ^*. Draws from the previous stage that lie in a particular striation can be easily obtained by selecting, with equal probability, any of the previously obtained draws $\left\{\theta^{(i-1,\ell)}\right\}_{\ell=1}^{NG}$ that *lie in that striation*. A draw θ from the proposal distribution is accepted with probability

$$\min\left\{1, \frac{f_{\lambda_i}(\theta)}{f_{\lambda_i}(\theta^*)} \frac{g_i(\theta^*, \theta)}{g_i(\theta, \theta^*)}\right\}. \tag{3}$$

If the draw is accepted, then $\theta^{(i,\ell+1)} = \theta$ and if it is rejected, then $\theta^{(i,\ell+1)} = \theta^*$.

To evaluate the proposal density $g_i(\theta, \theta^*)$, one must be able to evaluate the integral I_{i-1}.[9] One can, however, approximate

$$\frac{g_i(\theta^*, \theta)}{g_i(\theta, \theta^*)}$$

without the knowledge of this quantity. When θ and θ^* are close to each other, $\phi_{\mathfrak{c}_i\Omega_i}(\theta - \theta^*)$ is much larger than either $f_{\lambda_{i-1}}(\theta)$ or $f_{\lambda_{i-1}}(\theta^*)$. When θ and θ^* are far apart from each other, both $f_{\lambda_{i-1}}(\theta)$ and $f_{\lambda_{i-1}}(\theta^*)$ are much larger than $\phi_{\mathfrak{c}\Omega_i}(\theta - \theta^*)$. Thus, when θ is sampled from the Gaussian distribution, $g_i(\theta^*, \theta)/g_i(\theta, \theta^*) \approx 1$; when θ is sampled from the distribution at the previous stage, $g_i(\theta^*, \theta)/g_i(\theta, \theta^*) \approx f_{\lambda_{i-1}}(\theta^*)/f_{\lambda_{i-1}}(\theta)$. Evidence from our SVAR model suggests that if we use this approximation, we would incorrectly accept or reject a proposal draw approximately once in a million draws. Thus, the DSMH sampler remains efficient even in situations where I_i may not be well estimated. This is one of the key features that make the DSMH sampler attractive and is highlighted by the application in Section V.3.

II.2. **Theoretical Foundation.** In this section we give necessary and sufficient conditions for convergence of the DSMH sampler.

Assumption 1. $\int_{\theta\in\Theta} p(\theta)^{\lambda_1} d\theta < \infty$.

For $\lambda_1 \leq \lambda_i \leq 1$, note that $p(\theta)^{\lambda_i} \leq p(\theta)^{\lambda_1}$ when $p(\theta) \leq 1$ and $p(\theta)^{\lambda_i} \leq p(\theta)$ when $1 \leq p(\theta)$. Since integrals of both $p(\theta)^{\lambda_1}$ and $p(\theta)$ are finite, the integral of $p(\theta)^{\lambda_i}$ is also finite. Under Assumption 1, therefore, $p(\theta)^{\lambda_i}$ is in fact a probability kernel. Assumption 1 is a very weak assumption and certainly much less restrictive than those in the SMC literature as it does not impose uniform boundedness on the posterior density. And Assumption 1 is all we need for applying the DSHM sampler as shown next.

Proposition 1. *If* $f(\theta) = \sum_j |q_j(\theta)| \exp(-\theta'\Sigma_j^{-1}\theta)$, *where* $q_j(\theta)$ *is a polynomial function and* Σ_j *is a symmetric and positive definite matrix, then* $\int_{\theta\in\Theta} f(\theta)^{\lambda} d\theta < \infty$ *for* $0 < \lambda \leq 1$.

until the acceptance rate is close to α. The entire procedure is then repeated with a slightly longer running time to ensure that the acceptance rate is indeed close to α.

[9] In Section III we discuss techniques for estimating this quantity, which is of independent interest, as well as how to choose Ω_i and p.

Proof. Since $\left(\sum_j |q_j(\theta)| \exp(-\theta'\Sigma_j^{-1}\theta)\right)^\lambda \leq \sum_j \left(|q_j(\theta)| \exp(-\theta'\Sigma_j^{-1}\theta)\right)^\lambda$, it suffices to show the result for $f(\theta) = |q(\theta)| \exp(-\theta'\Sigma^{-1}\theta)$. Let $\Sigma = TT'$ be the Cholesky decomposition of Σ and define $x \in \mathbb{R}^m$ by $x = T^{-1}\theta$. Since $q(Tx)$ is a polynomial, there exist positive numbers B and b_k such that $|q(Tx)| \leq \prod_{k=1}^m |x_k^{b_k}|$ for $B < \|x\|$. Thus, there exists a number C such that $\int_{\theta\in\Theta} f(\theta)^\lambda d\theta \leq C + |\det T| \prod_{k=1}^m \int_{B<|x_k|} |x^{\lambda b_k}| \exp(-\lambda x_k^2) dx_k < \infty$. □

Since a large class of economic and statistical models, including the examples in this paper, have posterior distributions of the form stated in Proposition 1, this proposition establishes the applicability of the DSMH sampler to such models. Indeed, the following proposition shows that Assumption 1 is sufficient to ensure convergence of the DSHM sampler.

Proposition 2. *Under Assumption 1, the DSMH sampler at the i^{th} stage converges to $f_{\lambda_i}(\theta)$. In particular, at the final stage, the DSMH sampler converges to the posterior distribution.*

Proof. If the Metropolis-Hastings transition kernel is aperiodic and irreducible with respect to $f_{\lambda_i}(\theta)$, then the Metropolis-Hastings algorithm at the i^{th} stage converges to the distribution $f_{\lambda_i}(\theta)$. See Tierney (1994), Theorem 1, for a discussion of these concepts and this result. Because the support of our proposal distribution is all of $\mathbb{R}^m$, the transition kernel is aperiodic. Because a subset of $\mathbb{R}^m$ is of positive probability with respect to $f_{\lambda_i}(\theta)$ if and only if it is of positive probability with respect to $f_{\lambda_{i-1}}(\theta)$, the transition kernel is irreducible with respect to f_{λ_i}. □

III. Practical Issues

In this section we discuss tuning parameters that the user can set and other tuning decisions that are made automatically by our implementation. Table 1 provides an overview of the tuning parameters available to the user, the recommended settings, and the speed-reliability tradeoffs. The following sections give detailed discussions, all of which are relevant to Table 1.

III.1. **Importance Weights.** To make the DSMH sampler operational, we need information about the distribution of $f_{\lambda_i}(\theta)$ before we form the sample $\left\{\theta^{(i,\ell)}\right\}_{\ell=1}^{NG}$ but after we have obtained the sample $\left\{\theta^{(i-1,\ell)}\right\}_{\ell=1}^{NG}$. In particular, we use this information to generate a small number of starting values for the Metropolis-Hastings step and to approximate the mean and variance of $f_{\lambda_i}(\theta)$ for constructing the Metropolis-Hastings proposal density. To accomplish these tasks we calculate importance weights on the sample from the previous stage. The unnormalized importance weight of $\theta^{(i-1,\ell)}$ is $\tilde{w}_\ell = f_{\lambda_i}(\theta^{(i-1,\ell)})/f_{\lambda_{i-1}}(\theta^{(i-1,\ell)})$. The normalized importance weight is $w_\ell = \tilde{w}_\ell/(\sum_{k=1}^{NG} \tilde{w}_k)$. Estimates of mean μ_i and variance Ω_i of $f_{\lambda_i}(\theta)$

are given by

$$\mu_i = \sum_{\ell=1}^{NG} w_\ell \theta^{(i-1,\ell)} \text{ and } \Omega_i = \sum_{\ell=1}^{NG} w_\ell (\theta^{(i-1,\ell)})(\theta^{(i-1,\ell)})' - \mu_i \mu_i'.$$

Starting values for the Metropolis-Hastings step are taken independently from an importance-weighted sample of $\left\{\theta^{(i-1,\ell)}\right\}_{\ell=1}^{NG}$.

The notion of effective sample size (ESS) based on importance weights is a useful diagnostic for evaluating the quality of the aforementioned estimates and starting values. It is defined as

$$\text{ESS}^{(\text{IW})} = \frac{\left(\sum_{\ell=1}^{NG} \tilde{w}_\ell\right)^2}{\sum_{\ell=1}^{NG} \tilde{w}_\ell^2} = \frac{1}{\sum_{\ell=1}^{NG} w_\ell^2},$$

where the superscript "IW" standards for importance weights. This value is between 1 and NG, with a larger number indicating a better estimate. We use $\text{ESS}^{(\text{IW})}$ to discipline the choice of λ_i and the choice of the initial distribution $f_0(\theta)$ for the DSHM sampler. These choices are made to guarantee that $\text{ESS}^{(\text{IW})}$ is always equal to or greater than $NG \times \text{ess}_{\text{min}}^{(\text{IW})}$, where $\text{ess}_{\text{min}}^{(\text{IW})}$ is between zero and one. The smaller $\text{ess}_{\text{min}}^{(\text{IW})}$ is, the faster the sampler can finish. But too small a value would lead to unreliable samples. We recommend to set $\text{ess}_{\text{min}}^{(\text{IW})} = 0.10$ so that the effective sample size based on importance weights is always at least 10% of the the total sample. We note that this lower bound is often smaller than what is recommended for the SMC sampler (Durham and Geweke, 2012; Herbst and Schorfheide, 2014). The reason we can afford to relax the required minimum effective sample size is because importance sampling plays only a peripheral role in the DSMH sampler, even though it is central to the SMC sampler. The SMC sampler relies on importance sampling to reweight the sample of particles (draws) at each stage. The problem embedded in importance sampling is that particles tend to collapse and as a result only a small fraction of the sample receives most weights. The remedy, called "the mutation ", is to use the Metropolis-Hastings algorithm to resample new particles when the importance sampler begins to collapse. The mutation is an essential part of the SMC sampler and its resampling scheme brings to the surface the well-known weakness of importance sampling that lies at the heart of Bayesian inference. By contrast the DSMH sampler is grounded in the Metropolis-Hastings algorithm and uses importance sampling only for an initial draw when the algorithm moves from the previous stage to the current stage. Thus, there is no collapsing problem as long as striations are dynamically adjusted across stages to maintain a sufficient probability for each striation. Once one is able to sample efficiently from $f_{\lambda_0}(\theta)$ at the initial stage (a requirement also for the EE sampler and the SMC sampler), independent sampling within the striation is designed to make the DSMH sampler travel across the entire parameter space according to the tempered posterior distribution at each stage.

III.2. **Choosing λ_i.** There are two considerations in choosing the value of λ_i: how small should λ_1 be (the first stage) and how quickly should λ_i converge to 1 (the last stage)? To answer the first question we consider the model in the form of $g_t(y_t, \theta) = \varepsilon_t$, a common form for dynamic multivariate models, where ε_t is an $n \times 1$ vector of exogenous shocks for $1 \leq t \leq T$. If the probability density of ε_t is $p_\varepsilon(\varepsilon_t, \theta)$, the likelihood function can be expressed as

$$\prod_{t=1}^{T} \left| \det \left[\frac{\partial g_t}{\partial \theta} \right] \right| p_\varepsilon(g_t(y_t), \theta).$$

Whatever complexity the possibly nonlinear function $g_t(y_t, \theta)$ might bring to our problem, the product of the determinants adds another dimension of complexity: a high-order polynomial of degree nT. As illustrated in Figure 1, the tempered kernel $f_{\lambda_1}(\theta)$ must be a diffuse mound-shaped distribution and to achieve this diffuseness λ_1 needs to be small enough to wash away the effects of the high-order polynomial. For this reason, λ_1 should be at most $1/(nT)$ and we recommend to set it equal to $1/(10nT)$.

Given the value of λ_1, we use $\text{ESS}^{(\text{IW})}$ to determine how fast the value of λ_i for $i > 1$ should converge to one. Since $\text{ESS}^{(\text{IW})}$ is equal to

$$\frac{\left(\sum_{\ell=1}^{NG} p(\theta^{(i-1,\ell)})^{\lambda_i - \lambda_{i-1}} \right)^2}{\sum_{\ell=1}^{NG} p(\theta^{(i-1,\ell)})^{2(\lambda_i - \lambda_{i-1})}},$$

one can see that as λ_i decreases to λ_{i-1}, $\text{ESS}^{(\text{IW})}$ increases to NG. Thus, there always exists a value of λ_i such that $\text{ESS}^{(\text{IW})} \geq NG \times \text{ess}_{\min}^{(\text{IW})}$. We set λ_i to be the largest value in $(\lambda_{i-1}, 1]$ such that $\text{ESS}^{(\text{IW})} = NG \times \text{ess}_{\min}^{(\text{IW})}$ for $1 < i < H$ and $\text{ESS}^{(\text{IW})} \geq NG \times \text{ess}_{\min}^{(\text{IW})}$ for $i = H$. The required value of λ_i can be found numerically by the bisection method.

By choosing λ_i this way, we do not know in advance how many stages are required or how long the sampler would run. We find that λ_i chosen in this manner converges geometrically to one. In practice, therefore, we are able to estimate how many stages are required and how long the sample will run after observing only the first few λ_i's.

III.3. **Initial Distribution.** The initial distribution, $f_0(\theta)$, is of student-t with $\nu > 2$ degrees of freedom, mean μ_0, and variance Ω_0. It is straightforward to obtain independent samples from this distribution and the distribution has a known density function. For the examples discussed in this paper, the exogenous shocks are Gaussian and thus setting the degrees of freedom to 30 works remarkably well. In general, the distributions may have fatter tails and a smaller number of degrees of freedom might be more appropriate.

The mean and variance, μ_0 and Ω_0, are taken to be the sample mean and variance of the stage-one distribution $f_{\lambda_1}(\theta)$. Because λ_1 was chosen to make $f_{\lambda_1}(\theta)$ a diffuse mound-shaped distribution, the standard random-walk Metropolis algorithm is effective for simulating a

large sample from $f_{\lambda_1}(\theta)$, which in turn is used to form μ_0 and Ω_0. We recommend the following procedure.

(1) Set μ_0 to zero and Ω_0 to the identity matrix.
(2) Obtain a sample of length NG from $f_{\lambda_1}(\theta)$, using the random-walk Metropolis algorithm with variance $\mathfrak{c}\Omega_0$, where $\mathfrak{c}$ is tuned to the acceptance rate α.
(3) Set μ_0 to the sample mean and Ω_0 to the sample variance.
(4) Obtain an independent sample of length NG from a student-t distribution with mean μ_0 and variance Ω_0.
(5) Compute $\text{ESS}^{(\text{IW})}$ with respect to $f_{\lambda_1}(\theta)$ from this sample. If $\text{ESS}^{(\text{IW})} > NG \times \text{ess}_{\text{min}}^{(\text{IW})}$, stop; otherwise, return to step (2).

The procedure usually succeeds after one or two iterations. If it fails to find $\text{ESS}^{(\text{IW})}$ greater than $NG \times \text{ess}_{\text{min}}^{(\text{IW})}$ after five iterations, either the degrees of freedom ν is badly chosen or λ_1 is not small enough. We recommend to begin with $\nu = 30$ so that $f_0(\theta)$ is approximately Gaussian, and then decrease the value of ν as needed. If no such ν can be found, λ_1 should be lowered. For the SVAR models studied in this paper and our other experiments, the procedure finds a large value of $\text{ESS}^{(\text{IW})}$ after two iterations with the recommended values $\nu = 30$ and $\lambda_1 = 1/(10nT)$.

III.4. **Thinning and Probability of Striated Proposal.** It is well known that the standard random walk Metropolis-Hastings algorithm produces serially correlated draws that meander only in one local region, especially for high-dimensional problems. While the acceptance of a proposal draw from the sample obtained at the previous stage breaks this dependence, one could still have a long sequence of serially correlated draws. For this reason, we recommend that each draw be saved with probability p_s. To save NG draws, therefore, one must simulate approximately NG/p_s draws. The thinning probability p_s is chosen by the user, but we recommend to set it equal to $\alpha\bar{\mathfrak{c}}$, where $\bar{\mathfrak{c}}$ is the average value of the scaling factor $\mathfrak{c}_i$. We interpret $1/\alpha\mathfrak{c}_i$ as the number of random-walk Metropolis draws required to move a distance of one standard deviation. Of course, p_s must be chosen before we know $\bar{\mathfrak{c}}$, but we find that $\mathfrak{c}_i$ does not vary much from stage to stage. After a few stages are completed, one can evaluate how stable the chosen thinning probability is and restart the sampler if necessary. In the application studied in this paper, the scaling factor $\mathfrak{c}_i$ is generally between 0.1 and 0.2. As the number of parameters increases, however, one should expect $\bar{\mathfrak{c}}$ to decrease.

The probability of making a striated proposal draws, p, is related to p_s in two ways. The random-walk Metropolis algorithm needs time to explore the distribution locally, before a striated proposal draw is accepted and the sampler moves to another region of the parameter space. Since we save only a portion of these draws, it must be that $p < p_s$. And since striated

proposal draws come from the sample obtained at the previous stage, we need to avoid over-sampling from the previous tempered distribution. If we set $p = 0.1p_s$, then the number of striated proposals is equal to 10% of the total number of draws simulated at the previous stage. Heuristically this value is reasonable enough to safeguard against over-sampling.

III.5. **Parallelism.** The DSMH sampler is designed to utilize the parallelism efficiently. We have G independent groups and within each group we simulate N draws. To exploit parallel computing, the size of G should be a multiple of the number of computing cores available on the hardware device at hand. On a cluster, where there are often a hundred or more cores available, we recommend to set G to the actual number of computing cores available. On an average desktop computer, where there are only 4 to 16 cores, we recommend to set G to a multiple of the actual number of computing cores with G being at least 32. Since each independent group has its own starting value that is chosen from the importance-weighted sample at the previous stage, this recommended value of G simply avoids an unfortunate draw of the starting value that may be in the extremely low-probability region. Since the random draws from independent groups can be used to form within-group and between-group sample variances for a useful diagnostic of how effective the sampler is, a large value of G is recommended. At each stage a total of NG draws are stored for use in the next stage. Because it is the total number of draws that matters, G can be tailored to the computing environment while N is adjusted to target NG at the desired level.

III.6. **Marginal Likelihood.** The marginal likelihood, often called the marginal data density (MDD) in the macroeconomics literature, is

$$p(Y_T) = \int_{\Theta} p(Y_T \mid \theta)p(\theta)d\theta. \tag{4}$$

Computing the MDD is necessary for calculating the Bayes factor or the posterior odds ratio when preforming model comparison. Estimation of the MDD is a by-product of the SMC sampler and can be obtained with no extra computational costs.[10] Such a by-product is also true of the DSMH sampler. Because $f_0(\theta)$ is a probability density, $I_0 = \int_{\theta\in\Theta} f_0(\theta)d\theta = 1$. For $1 \leq i \leq H$,

$$I_i = I_{i-1} \int_{\theta\in\Theta} \frac{f_{\lambda_i}(\theta)}{f_{\lambda_{i-1}}(\theta)} \frac{f_{\lambda_{i-1}}(\theta)}{I_{i-1}} d\theta.$$

If $\hat{I}_{i-1}$ is an estimate of I_{i-1}, therefore, I_i can be estimated from the sample $\{\theta^{i-1,\ell}\}_{\ell=1}^{NG}$ by

$$\hat{I}_i = \frac{\hat{I}_{i-1}}{NG} \sum_{\ell=1}^{NG} \frac{f_{\lambda_i}(\theta^{i-1,\ell})}{f_{\lambda_{i-1}}(\theta^{i-1,\ell})} = \frac{\hat{I}_{i-1}}{NG} \sum_{\ell=1}^{NG} \tilde{w}_\ell. \tag{5}$$

[10]For the EE sampler of Kou, Zhou, and Wong (2006), there is no discussion of how to estimate the MDD.

From (2) and (4) one can see that $I_H = p(Y_T)$. Hence the MDD can be approximated by $\hat{I}_H$. The estimates $\hat{I}_i$ are extremely fast to compute, but is inaccurate if the weights become unbalanced. As discussed at the end of Section II.1.3, the simulated sample can still be representative even if the importance weights are unbalanced. In this case, one can use the posterior draws to estimate the MDD through other means such as the bridge-sampling method (Meng and Wong, 1996) and the Mueller method described in Liu, Waggoner, and Zha (2011). This alternative estimate serves as cross-verification of the quality of the estimated MDD through updated importance weights. In Section V.4 we provide a revealing example that highlights this problem—a problem that cannot be detected simply by numerical standard errors (NSEs)—and propose a diagnostic tool for detecting and alleviating such a problem.

III.7. **Within and Between Sample Variances.** One of the tuning parameters, as discussed in Section III.1, is the *minimal* ESS for importance weights . With G independent groups of posterior draws, we can estimate the *actual* ESS for the DSMH draws as follows. Let ϑ be any scalar element of the parameter vector θ. Denote $\bar{\vartheta}_j = \frac{1}{N}\sum_{\ell=1}^{N}\vartheta_{\ell,j}$, $\bar{\vartheta} = \frac{1}{G}\sum_{j=1}^{G}\bar{\vartheta}_j$, and $Var(\vartheta) = \sigma_\vartheta^2$. The between-group sample variance of ϑ is $B_\vartheta = \frac{N}{G}\sum_{j=1}^{G}\left(\bar{\vartheta}_j - \bar{\vartheta}\right)^2$. The within-group sample variance of ϑ is $W_\vartheta = \frac{1}{G}\sum_{j=1}^{G}\left[\frac{1}{N}\sum_{\ell=1}^{N}(\bar{\vartheta}_{\ell,j} - \bar{\vartheta}_j)^2\right]$. If $\vartheta_{\ell,j}$ is sampled independently, it follows that with a reasonable number of simulations (e.g., $N = 500$ and $G = 50$), $B_\vartheta \sim \sigma_\vartheta^2$, $W_\vartheta \sim \sigma_\vartheta^2$, and thus $B_\vartheta/W_\vartheta \sim 1$. The ESS for a random sample of ϑ generated by the DSMH sampler is approximated as

$$\mathrm{ESS}_\vartheta^{(\mathrm{DSMH})} = \frac{NG}{B_\vartheta/W_\vartheta}. \tag{6}$$

IV. Specific Difficulties for High-Dimensional Models

Consider economic or statistical models of the general form

$$Y_T = \mathcal{M}(\mathcal{E}_T; \theta, \psi), \tag{7}$$

where $\mathcal{E}_T = (\varepsilon_1, \cdots, \varepsilon_T)$ are unobserved exogenous shocks and ψ is a vector of nuisance parameters (e.g., unobserved regimes in Markov-switching models). This general form includes VAR and DSGE models as a special case. For illustrative purposes, consider $\mathcal{M}(\cdot)$ in the following parametric form:

$$A(\theta, \psi)Y_T = c(\theta, \psi) + \mathcal{E}_T, \tag{8}$$

where $A(\cdot)$ is an $nT \times nT$ matrix and $c(\cdot)$ is an nT-dimensional vector. Note that any linear state-space model can be expressed in the form of (8), where $A(\cdot)$ and $c(\cdot)$ are often complicated non-linear functions of θ and ψ. We assume, in our application, that y_t and

ε_t have the same dimension of n and $\mathcal{E}_T$ has the standard Gaussian distribution.[11] The likelihood function for model (8) becomes

$$p(Y_T|\theta,\psi) = |\det A(\theta,\psi)| \exp\left(-\frac{1}{2}\left(A(\theta,\psi)Y_T - c(\theta,\psi)\right)'\left(A(\theta,\psi)Y_T - c(\theta,\psi)\right)\right). \quad (9)$$

The posterior kernel is $p(Y_T|\theta,\psi)p(\theta,\psi)$, where $p(\theta,\psi)$ is the prior probability density. From expression (9) one sees two major difficulties for estimating this kind of models:

(1) The determinant function is a multivariate polynomial of degree nT.

(2) $A(\cdot)$ and $c(\cdot)$ may be complicated nonlinear functions of the parameters.

To deal with the first difficulty, we consider SVAR models in which there are no nuisance parameters, both $A(\theta)$ and $c(\theta)$ are linear in θ, and the exponent in the exponential term of (9) is quadratic. If the determinant term were not present in (9), the likelihood function would be a Gaussian probability density. The determinant term, however, induces multiple peaks and complicated ridges into the likelihood function as well as the posterior kernel. Because the determinant term is prevalent in multivariate dynamic models (including DSGE models), SVAR models provide a natural benchmark for testing the DSMH simulator.

For SVAR models, Waggoner and Zha (2003a) develop an efficient Gibbs sampler. Since it is the exact Gibbs sampler, one can apply the method developed by Chib (1995) to accurately compute the MDD (see also Fuentes-Albero and Melosi (2013)). For recursive SVAR models, the Gibbs sampler produces independent draws. If the identification is non-recursive, the draws are serially correlated but convergence is so rapid that it is feasible to draw a large number of starting values independently and then apply the Gibbs sampler to obtain independent draws. Thus we can use the posterior draws generated by the Gibbs sampler as the "truth" to gauge the accuracy of the DSMH sampler and help develop diagnostic tools for more general models.

One class of more general models we study in this paper are Markov-switching SVARs (MSSVARs) proposed by Sims and Zha (2006). This class involves nuisance parameters, namely the hidden Markov states. In addition to the first difficulty discussed above, we now encounter the second difficulty: $A(\cdot)$ and $c(\cdot)$ are much more complicated functions of the underlying parameters. Because of this additional difficulty, posterior simulations become more challenging. Sims, Waggoner, and Zha (2008) use the Metropolis-within-Gibbs algorithm to make posterior draws, but the specific Gibbs design depends on a particular model specification and is prone to analytical and programming errors when the model specification changes. The DSMH sampler is generic. In Section V.5 we use the turning parameters and the diagnostic tool gained from our experiments with the benchmark SVAR model to show how the DSMH sampler works for this complicated example.

[11] All these special assumptions can be relaxed, for the DSMH sampler applies to model (7) as long as its posterior kernel satisfies Assumption 1.

V. Application

In this section we present two simultaneous-equation high-dimensional models for the purpose of testing the DSMH sampler: a SVAR model and a Markov-switching SVAR model. Both models pose significant challenges for any sampler.

V.1. **Benchmark SVAR.** Structural vector autoregressions have the following representation:

$$y_t' A_0 = C' + \sum_{h=1}^{l} y_{t-h}' A_h + \varepsilon_t', \text{ for } 1 \leq t \leq T, \tag{10}$$

where

- l is the lag length;
- ε_t is an n-dimensional column vector of unobserved random i.i.d. standard Gaussian shocks at time t;
- A_0 is an invertible $n \times n$ matrix and A_h is an $n \times n$ matrix for $1 \leq h \leq l$;
- C is an $n \times 1$ vector of constant terms.

The initial conditions $y_0, \cdots, y_{1-l}$ are taken as given. In our notation, the parameter vector θ is the collection of all the parameters in model (10). The prior distribution takes the form suggested by Sims and Zha (1998), which expands on the original Minnesota prior Litterman (1986).

With the prior of Sims and Zha (1998), the posterior probability density function is proportional to

$$|A_0|^T \prod_{k=1}^{n} \exp\left(-\frac{T}{2}\theta'\Sigma\theta\right), \tag{11}$$

where Σ is a symmetric and positive definite matrix that depends on the data and the prior. Consider exclusion restrictions placed on some of the parameters. Waggoner and Zha (2003a) show that the form (11) of the posterior kernel remains the same but θ consists of only parameters that are not excluded; furthermore, that paper derives a Gibbs sampler for any posterior kernel of form (11) that is efficient and converges very rapidly.

If we raise expression (11) to any positive power, the function form remains the same as (11). Thus, the Gibbs sampler of Waggoner and Zha (2003a) applies to $f_\lambda(\theta)$ for $0 < \lambda \leq 1$. We record this result in the following proposition.

Proposition 3. *For model* (10) *with exclusion restrictions and the prior of Sims and Zha (1998), there exists a Gibbs sampler for* $f_\lambda(\theta)$ *for* $0 < \lambda \leq 1$.

V.2. **Monthly Empirical Model.** To show how the DSMH sampler handles the first difficulty highlighted in Section IV, we apply the DSMH sampler to a three-variable monthly

SVAR model with thirteen lags.[12] The three variables are those commonly used by monetary DSGE models: log output gap (x_t), GDP-deflator inflation (π_t), and the federal funds rate (R_t). The U.S. data are monthly from 1988:1 to 2014:6, covering the post-Volcker period of U.S. history. Output gap is measured by the difference between actual real GDP and potential real GDP published by the Congressional Budget Office. Both actual and potential GDP series as well as GDP deflator are interpolated to monthly frequency using the methodology suggested by (Leeper, Sims, and Zha, 1996; Bernanke, Gertler, and Watson, 1997). Federal funds rates are monthly average effective rates and annualized.

A majority of applications in the SVAR literature concern restrictions imposed on A_0 (Sims and Zha, 2006; Rubio-Ramírez, Waggoner, and Zha, 2010). We follow this approach in our application. The identification for this three-variable SVAR is

$$A_0 = \begin{bmatrix} a_{0,11} & a_{0,12} & 0 \\ a_{0,21} & a_{0,22} & 0 \\ a_{0,31} & 0 & a_{0,33} \end{bmatrix}. \tag{12}$$

The identification (12) is non-recursive but the model is *globally identified* (Rubio-Ramírez, Waggoner, and Zha, 2010). The identifying restrictions are consistent with new-Keynesian models but with fewer restrictions than the stylized model of (Rudebusch and Svensson, 1999) to maintain the fit of the model. The first equation (the first column) characterizes the aggregate demand behavior in which output gap responds to both inflation and the interest rate.[13] The second equation (the second column) is consistent with the Phillips-curve relationship in which inflation reacts to output gap. The last equation (the third column) characterizes the monetary policy behavior that responds to output gap and inflation with only one-month delay (this assumption is reasonable given the fact that the monetary authority has no information about GDP and its price deflator within the month). The hyperparmaters for the prior, in the notation of (Sims and Zha, 1998), is $\lambda_1 = 0.7$, $\lambda_2 = 0.5$, $\lambda_3 = 0.1$, $\lambda_4 = 1.2$, $\mu_5 = 1.0$, and $\mu_6 = 1.0$. Our empirical results are not sensitive to this prior setting.

There are substantive reasons we choose model (10) to test the DSMH sampler for high-dimensional problems. First, SVARs have served as a benchmark for other multivariate dynamic models such as DSGE models. Second, model (10) with a long lag length or a large number of variables presents a challenging high-dimensional problem. Even for our "small-scale" three-variable SVAR model, the number of parameters is well over a hundred, 126 to

[12]Thirteen lags are used to remove possible residual seasonality, even though the data we use are seasonally adjusted.

[13]One could make a further restriction such that the coefficient of inflation and that of the interest rate have the same magnitude but with opposite signs. This restriction means that output gap responds to the current (realized) real interest rate.

be exact. Third, the posterior involves a term of the form $|\det A_0|^T$. This is a polynomial of degree nT, which for our monthly SVAR model is a polynomial of degree 915! Since our identification is non-recursive, this high degree means that the likelihood function contains many complicated winding ridges. Fourth, because we chose to work with the unnormalized posterior kernel, there are 8 peaks for our model.[14] A combination of un-normalization and simultaneity makes the likelihood function as well as the posterior kernel unusually complicated. For all these reasons, our SVAR model provides a complex and challenging environment for testing generic samplers.

According to Proposition 3, one is able to generate posterior draws very efficiently at each stage i using the Gibbs sampler. If identification is recursive, the Gibbs sampler produces independent draws. If the identification is non-recursive as in our case, the random draws are serially correlated but as shown in Waggoner and Zha (2003a) convergence is so rapid that it is feasible to draw independently a large number of starting values and apply the Gibbs sampler to obtain independent draws. Furthermore, one can apply the method developed by Chib (1995) to accurately compute the MDD. Thus we use the posterior draws generated by the Gibbs sampler as the "truth" to gauge the quality of the DSMH sampler and offer an invaluable tool for improving the efficiency of this sampler with different values of tuning parameters.

V.3. **Estimation Results.** The tuning parameters for the DSMH sampler are set as $N = 2000$, $G = 100$, and $H = 42$. It takes about 15 minutes on our 50-core cluster and 20 minutes on our 16-core desktop computer.[15] The processors on our desktop are more upgraded and faster than those on the cluster. A large number of G takes full advantage of parallel computation. As a comparison, we apply the straight Metropolis algorithm with the same value of N and G. The number of posterior draws from the Gibbs sampler is 20000, much less than that for the DSMH sampler. The Gibbs draws are independent and the calculated ESS using the within-group and between-group sample variances is about 19800 on average.

[14]Given the posterior form (11), changing the sign of an equation in model (10) does not change the posterior. Since there are 2^n possible ways to change signs, there are 2^n peaks in any unnormalized SVAR model. For detailed discussions, see Waggoner and Zha (2003b); Hamilton, Waggoner, and Zha (2007). As a scientific-reporting procedure, it is best to store the unnormalized posterior draws. If a researcher chooses a particular normalization rule for specific purposes, the normalization can be applied to the stored unnormalized draws without any extra computational costs.

[15]We stretch the problem to the extent that has unusual complications for the purpose of testing the DSMH sampler. The posterior distributions for many economic and statistical problems are likely to be better shaped than the example studied in this section and thus take less computational time to achieve an adequate number of effective simulated draws.

Four parameters determine the *simultaneous relationship* between output gap and inflation: $a_{0,11}$, $a_{0,21}$, $a_{0,12}$, and $a_{0,22}$. For clear illustration, we concentrate on the posterior distribution of these four parameters. Because it is impossible to display the four-dimensional distribution, we display a two-dimensional distribution at a time to get a glimpse of the complexity we are dealing with.

Figure 2 displays the two-dimensional marginal posterior distribution of $a_{0,12}$ and $a_{0,22}$, formed from the posterior draws. By "marginal" we mean the joint posterior probability of $a_{0,12}$ and $a_{0,22}$ after integrating out all other parameters. The bottom panel of Figure 2, used as the "truth," gives a partial view of the very complicated posterior distribution. A comparison between the top and bottom panels of Figure 2 reveals the inability of the straight random-walk Metropolis algorithm to trace out the distribution. As a result, a large region of the parameter space is completely missed by the Metropolis sampler and this phenomenon does not improve no matter how many more simulations are added. Figure 2 displays winding ridges and multiple peaks, but does not come close to reflecting the extent of the complexity because it is a two-dimensional marginal distribution by integrating out all the other parameters.

In a higher-dimensional parameter space, winding ridges and multiple peaks are much worse than what is displayed in Figure 2. To get a better idea of such a complicated posterior distribution, we display the two-dimensional distribution of $a_{0,11}$ and $a_{0,22}$ in Figure 3 and the two-dimensional distribution of $a_{0,21}$ and $a_{0,12}$ in Figure 4. In the bottom panel of Figure 3, there are multiple, isolated local peaks with long shallow ridges connecting these peaks. In the bottom panel of Figure 4, there is a large cross with thin ridges and peaks at the center as well as the four ends of the cross. Again straight random-walk Metropolis sampling (the top panels in both figures) not only misses many local peaks and winding ridges but also fails to cover a good portion of the area around the peak covered by the Metropolis draws themselves. The reason that the straight Metropolis algorithm cannot even cover the region near the peak is due to the higher-dimensional problem not revealed by multiple two-dimensional distributions. If one were able to display a four-dimensional probability distribution of $a_{0,11}$, $a_{0,12}$, $a_{0,21}$, and $a_{0,22}$ by combining Figures 2, 3, and 4, the graph would look much more complicated than each of the three figures. The joint probability distribution of all parameters would be beyond our visualization and imagination.

From all three figures one can see that the posterior draws generated by the DSMH sampler are representative of the underlying irregular posterior distribution. The empirical probability densities displayed in the middle panels capture those in the bottom panels remarkably well. Clearly, it is feasible for the DSMH sampler to trace out this highly complicated posterior distribution of 126 parameters.

There is another important lesson from this experiment: one should first simulate the unnormalized version of the model and then apply one's preferred normalization rule to the unnormalized posterior draws. Consider the slope coefficient of the Phillips curve $\rho_0 = -a_{0,12}/a_{0,22}$.[16] The slope coefficient ρ_0 is immune to sign normalization because the effect of a sign change is cancelled by both denominator and numerator. Since many complicated winding ridges connect the symmetric local peaks introduced by changing signs of equations, these ridges may not be explored if the posterior draws are normalized as simulations trudge along. The draws generated by the straight random-walk Metropolis sampler behave like normalized draws because they meander around one local peak as displayed in Figures 2, 3, and 4. Nonetheless these draws fail to even explore the ridges around this local peak. Consequently, even if the object of interest, such as the slope of the Phillips curve ρ_0, is immune to normalization, obtaining the unnormalized posterior draws is crucial to achieving accurate inference.

Figure 5 clearly illustrates how statistical inferences of ρ_0 are affected by how well the posterior distribution is fully explored. The top panel of Figure 5 indicates that the Phillips-curve slope is negative with more than 85% probability, implying a perverse relationship between output gap and inflation that contradicts what economists believe. But an accurate inference, according to the bottom panel, reveals so wide a dispersion that this estimate is statistically insignificant. The inference drawn from the DSMH sampler is the same as that from the Gibbs sampler: there is no tradeoff between inflation and output gap for the post-1987 period.

We demonstrate that the DSMH sampler has the remarkable capacity to trace out the complicated distribution in the high-dimensional parameter space. For the complicated posterior distribution discussed in this section, it is able to attain accurate statistical inferences that would be otherwise prohibitively costly to achieve with any sampler that cannot explore the entire distribution efficiently.

V.4. **Marginal Likelihood and Effective Sample Size.** As discussed in Section III.6, the MDD or marginal likelihood is a by-product of the DSMH sampler. The quality of this by-product, however, depends on the quality of importance weights. We record the estimated integral constant according to (5) at each stage and report the results in Table 2. The column with the heading "Truth" reports the integral constants calculated from the efficient algorithm of Chib (1995) based on independent Gibbs draws. The estimated integral

[16] We have also experimented with the long-run Phillips curve coefficient $\overline{\rho} = \left[\sum_{h=1}^{l} a_{h,12}/a_{0,22}\right] - a_{0,12}/a_{0,22}$ and the results are similar.

constants from the Mueller method based on DSMH simulations are almost identical to those listed in the "Truth" column *at each stage*.[17]

The estimated integral constants using importance weights according to (5) are upward biased in large magnitude due to the unbalanced weights. The numerical standard error for the estimated integral constant from importance weights is reported in the last column. It is calculated as

$$\text{NSE}^{(\text{IW})} \equiv \text{NSE}\left(\log \hat{I}_i\right) = \sqrt{\frac{1}{G}\sum_{j=1}^{G}\left[\log \hat{I}_i^{(j)} - \frac{1}{G}\sum_{k=1}^{G}\log \hat{I}_i^{(j)}\right]^2},$$

where the superscript (j) stands for the j^{th} group of posterior draws. These NSEs are *unrealistically* small relative to the actual error of the estimated integral constant from importance weights. If we had not known the accurate estimate of the integral constant in the "Truth" column, we would not have had the sense of the magnitude of the error. Thus, the NSE does not measure how biased the MDD estimate is.

By definition the integral constant at the final stage (42^{th} in our application) is the MDD. For this particular set of runs for Table 2, the integral constant calculated by the Mueller method is slightly larger than that calculated from the Gibbs samples. For other runs, the value is slightly lower (not reported here). The NSE for the estimated MDD from the Mueller method applied to the DSMH samples is 0.72. Since the NSE may not be a reliable indicator of how well the underlying sampler is, we compute the corresponding $\text{ESS}^{(\text{DSMH})}$ for each parameter according to (6). For all 126 parameters in our monthly model, the average $\text{ESS}^{(\text{DSMH})}$ is approximately $15,000$ and the minimum $\text{ESS}^{(\text{DSMH})}$ is around 8500. With these many effective posterior draws, the sample generated by the DSMH sampler is clearly representative.

V.5. **Markov-switching SVARs.** Because the DSMH sampler is generic, choosing appropriate values of its tuning parameters that work for high-dimensional problems is an extremely challenging task. We find those values that work well for the benchmark SVAR model discussed in the previous sections. These values serve as a very useful benchmark when one applies the DSMH sampler to other high-dimensional problems. In this section we apply it to a Markov-switching SVAR in the form of

$$y_t' A_0 = C + \sum_{h=1}^{l} y_{t-h}' A_h + \varepsilon_t' \Xi_{s_t}^{-1}, \text{ for } 1 \le t \le T, \tag{13}$$

[17]We use Sims, Waggoner, and Zha (2008)'s elliptic probability density as a proposal density for the Mueller method. Other efficient methods, including the bridge-sampling method (Meng and Wong, 1996), give almost identical results.

where $\Xi_{s_t}^{-1}$ is a diagonal matrix with the diagonal elements depending on a Markov process represented by s_t with the $\kappa \times \kappa$ transition matrix $Q = [q_{i,j}]$ such that $q_{i,j} = \text{Prob}(s_t = i|s_{t-1} = j)$ for $i, j = 1, \ldots, \kappa$. Markov-switching SVAR models have been effective in addressing relevant issues related to the 2008 financial crisis (Hubrich and Tetlow, Forthcoming). Time-varying volatility such as changing uncertainty has been a prominent feature in the data (Cogley and Sargent, 2005; Justiniano and Primiceri, 2008; Bloom, 2009; Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez, and Uribe, 2011). Sims, Waggoner, and Zha (2008) show that one tractable way to model time-varying volatility is to allow shock variances to follow a Markov process. If the identifying restrictions on A_0 are non-recursive, we no longer have the exact Gibbs sampler to generate independent draws.

To put the DSMH sampler to the test, we make the problem unusually demanding by estimating again the *unnormalized* version of model (13) with a two-state Markov process for s_t. The nonlinearity increases the complexity of the problem considerably by expanding the magnitude of the first difficulty and introducing the second difficulty discussed in Section IV. Moreover, evaluation of the posterior density function requires the Hamilton filter (Hamilton, 1989) and such a filter is highly nonlinear. Consequently, we increase the number of stages to 48 to maintain a representative sample for each striation (most of added stages occur when when λ_i tends to one). On our 50-core cluster, the Markov-switching SVAR model takes 50 minutes for each stage as compared to 15 minutes per stage for the benchmark SVAR model. The amount of time includes the time spent on tuning for the acceptance ratio, simulating draws, compute the log likelihood through the Hamilton filter, and calculating the log value of the integral constant using the Mueller method (as part of the diagnostic analysis). For the rest of tuning parameters, the values are set as those for the benchmark SVAR model.

Table 3 reports log values of the integral constants estimated by the Mueller method, those estimated by importance weights ($\log \hat{I}_i$), and the values of $\text{NSE}^{(\text{IW})}$ at various stages. In comparison to Table 2, there is no "Truth" column in Table 3 because the exact Gibbs sampler is unavailable for this simultaneous-equation Markov-switching SVAR model. As in the benchmark case, the estimate of the integral constant through importance weights with a small NSE is biased in comparison to the estimate by the Mueller method applied to the draws generated by the DSMH sampler.

There are several reasons for our confidence in the estimates by the Mueller method. First, in the benchmark case, Table 2 shows that the MDD estimate by the Mueller method is the same as the estimate by using the Gibbs sampler. Second, in the Markov-switching case, the Mueller method and the bridge-sampling technique deliver the same MDD estimate.[18] Third, we have a large effective sample size from the DSMH sampler for the Markov-switching

[18]The NSE for the estimate is 0.68.

model. The average value of $\text{ESS}^{(\text{DSMH})}$ is $24,461$ and out of the 131 parameters the minimum $\text{ESS}^{(\text{DSMH})}$ is $12,287$. As a snapshot of this complicated high-dimensional posterior distribution, Figure 6 plots the two-dimensional probability density of $a_{0,12}$ (x-axis) and $a_{0,22}$ from the posterior draws generated by the DSMH sampler. As one can see, it is able to generate the multiple peaks with winding ridges. These results indicate that the DSMH sampler is capable of estimating a nonlinear high-dimensional model with a highly irregular posterior distribution.

VI. Conclusion

We have shown that the DSMH sampler developed in this paper is capable of simulating incredibly irregular posterior distributions full of complicated winding ridges and multiple peaks in the high-dimensional parameter space. We have set the bar high by estimating two *unnormalized* monthly SVAR models with simultaneous equations and over a hundred parameters. To illustrate how hard the problem is, we have displayed part of the complexity inherent in this high-dimensional posterior distribution. The generic DSMH sampler has proven to a remarkably efficient posterior simulator that is capable of exploring such complexity.

As common in any posterior simulator, technical details such as the appropriate range of values for tuning parameters become indispensable. This task is challenging due to the nature of high dimension and unusual irregularity imbedded in the posterior distribution. The values of tuning parameters that work for our benchmark SVAR model provide a benchmark for estimating other dynamic structural models for which the exact Gibbs sampler may not be available. Indeed, we have applied the DSMH sampler to a Markov-switching SVAR model and shown that the sampler remains efficient.

We exclusively focus on SVAR models not only because they serve as a benchmark for other multivariate dynamic models but also because they provide a natural experiment for testing any simulator against the "truth" (the independent draws generated by the Gibbs sampler). If the simulator fails to trace out the labyrinthine shape of the posterior distribution as graphically displayed in the paper, it sends a strong signal about its capability of estimating other multivariate dynamic models when there is no known "truth" about the underlying posterior distribution. It is our hope that the DSMH sampler, thoroughly tested against high-dimensional SVAR models, will prove to be as powerful in other applications as in our application.

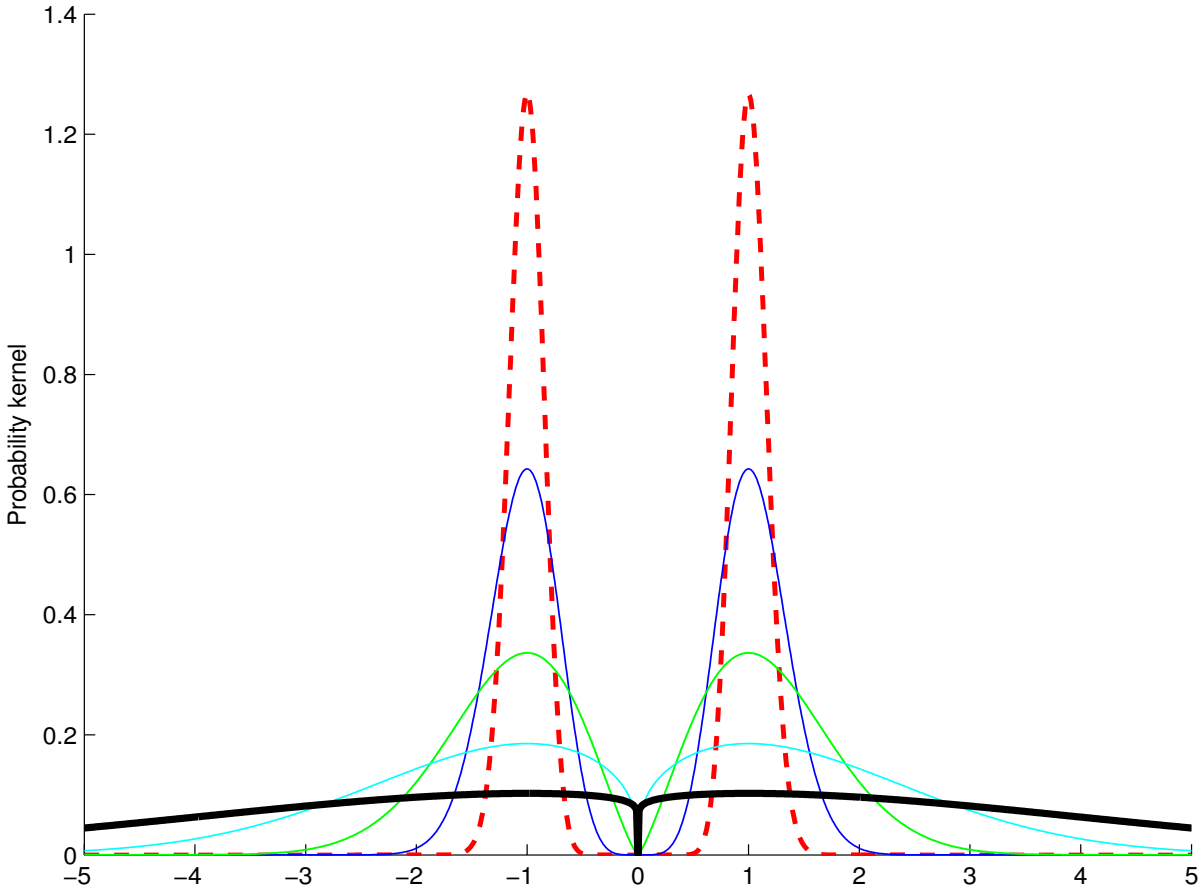

FIGURE 1. An illustrative example for tempered posterior density kernels. The thick dashed line is the posterior kernel (the most peaked) and the thick solid line is the most tempered kernel (the flattest).

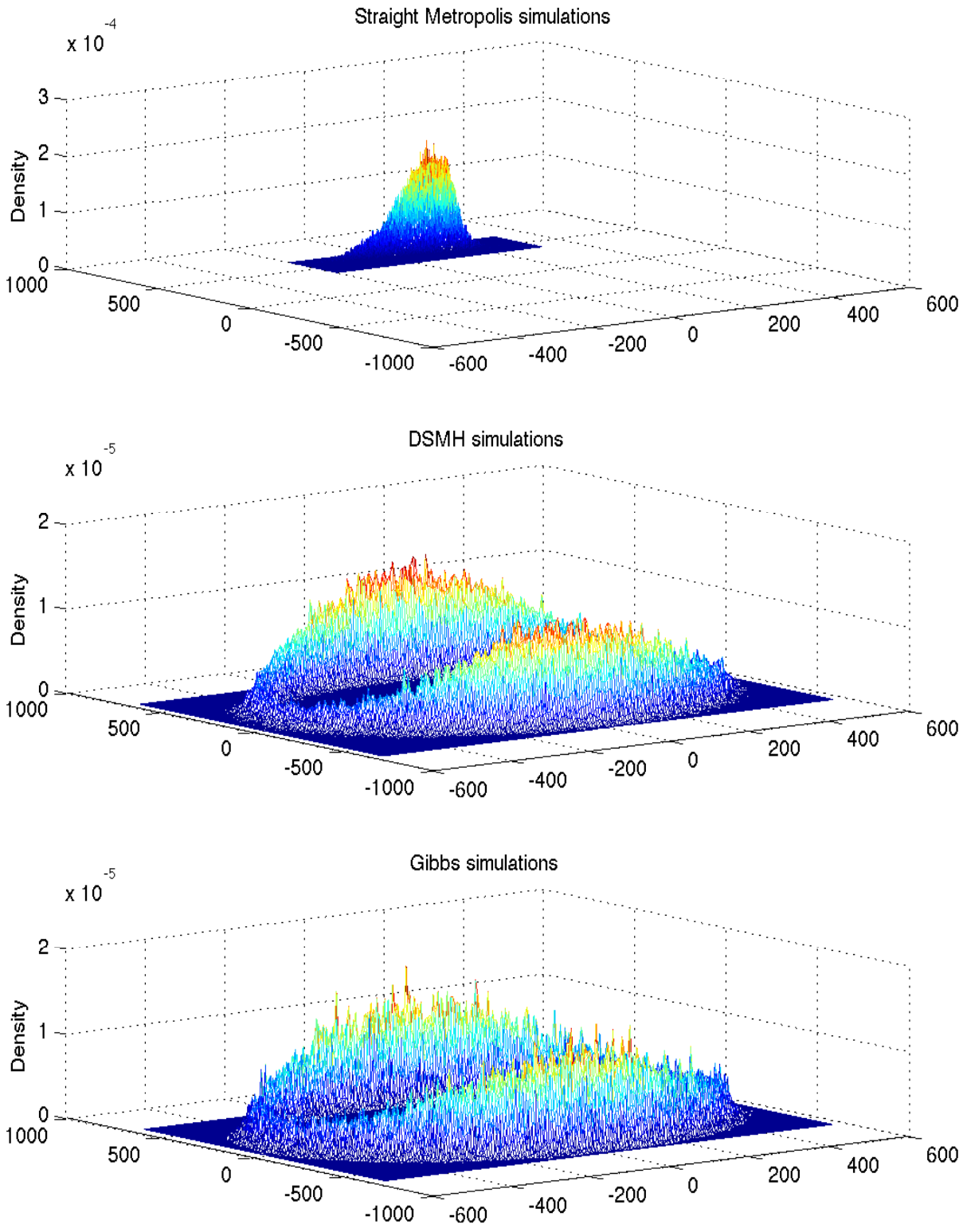

FIGURE 2. The two-dimensional probability density of $a_{0,12}$ (x-axis) and $a_{0,22}$ (y-axis) (after integrating out all other parameters). The probability density is formed empirically from the posterior draws generated by three methods: the straight random walk Metropolis algorithm, the DSMH sampler, and the Gibbs sampler.

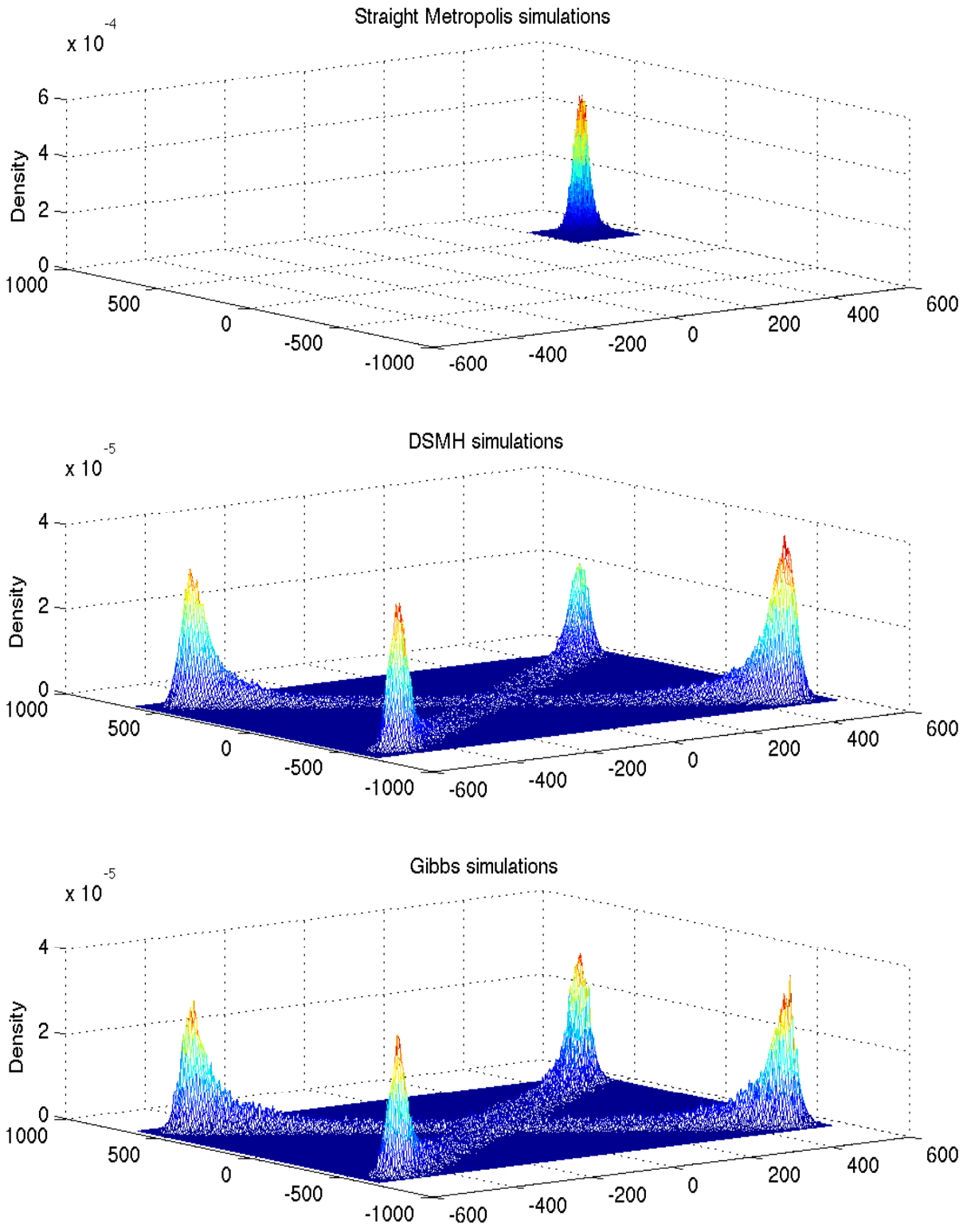

FIGURE 3. The two-dimensional probability density of $a_{0,11}$ (x-axis) and $a_{0,22}$ (y-axis) (after integrating out all other parameters). The probability density is formed empirically from the posterior draws generated by three methods: the straight random walk Metropolis algorithm, the DSMH sampler, and the Gibbs sampler.

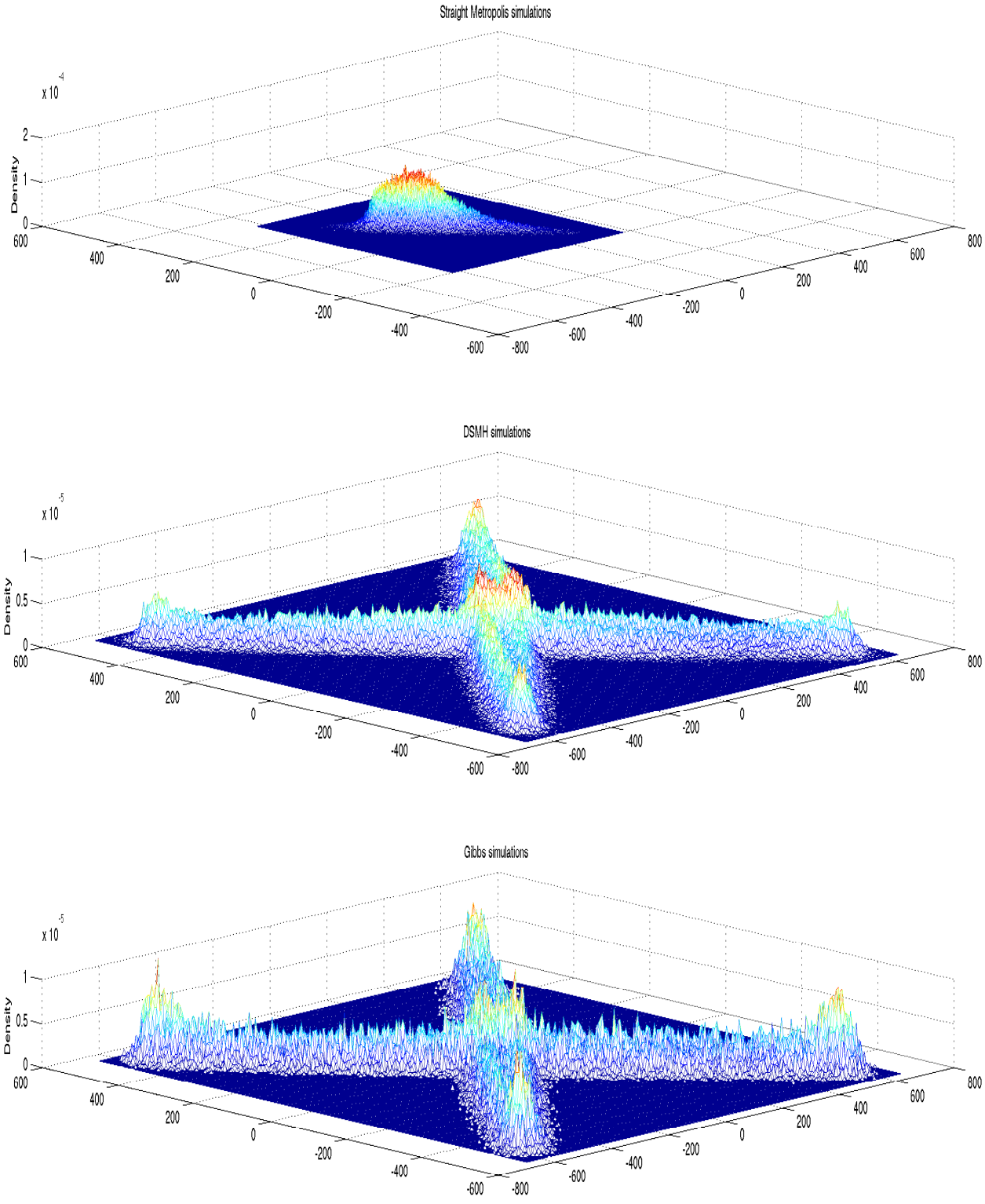

FIGURE 4. The two-dimensional probability density of $a_{0,21}$ (x-axis) and $a_{0,12}$ (y-axis) (after integrating out all other parameters). The probability density is formed empirically from the posterior draws generated by three methods: the straight random walk Metropolis algorithm, the DSMH sampler, and the Gibbs sampler.

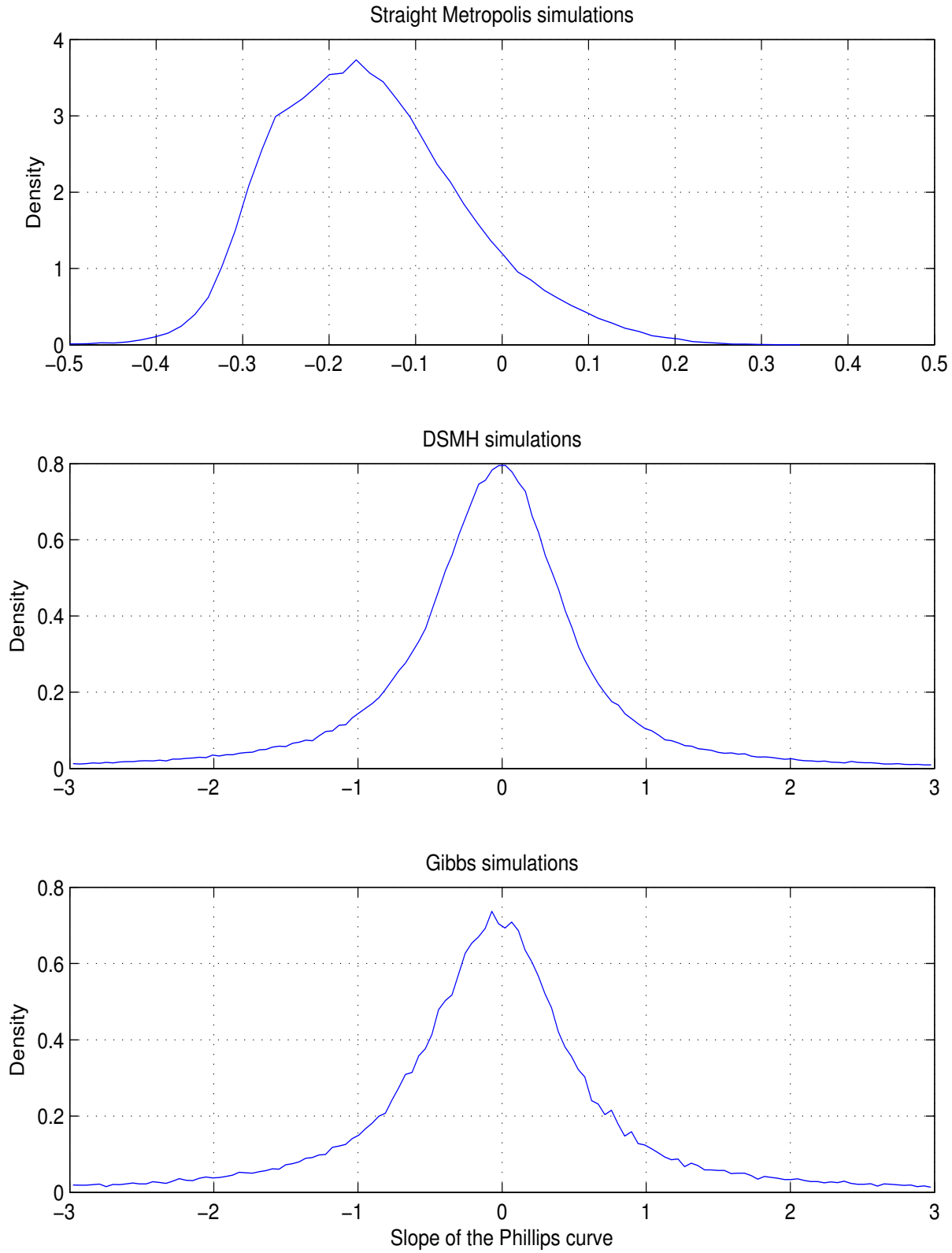

FIGURE 5. The slope coefficient of the Phillips curve. It is estimated empirically from the posterior draws generated by three methods: the straight random walk Metropolis algorithm, the DSMH sampler, and the Gibbs sampler.

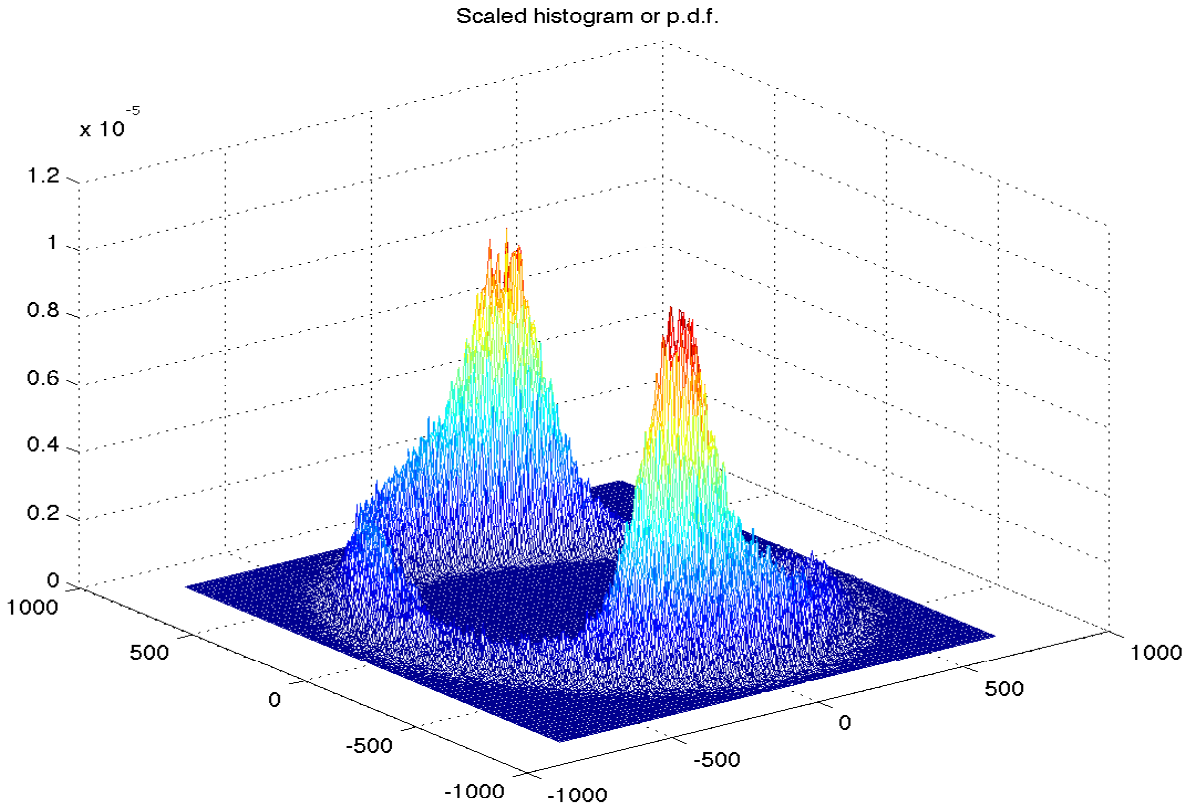

FIGURE 6. The two-dimensional probability density of $a_{0,12}$ (x-axis) and $a_{0,22}$ (y-axis) (after integrating out all other parameters). The probability density is formed empirically from the posterior draws generated by the DSMH sampler.

TABLE 1. Recommended values of tuning parameters

Parameter	Recommended Value	Faster Run Time	More Reliable Sample
NG	Problem specific	Smaller	Larger
$\text{ess}_{\text{min}}^{(\text{IW})}$	0.10	Smaller	Larger
λ_1	$1/(10nT)$	Larger	Smaller
p_s	$\alpha\bar{\mathfrak{c}}$	Larger	Smaller
p	$0.10p_s$	-	-
M	20	-	-
α	0.30	-	-
ν	30	-	-

Note: G should be a multiple of the number of computing cores and N is adjusted to target NG at its desired level.

TABLE 2. Estimated log integral constants ($\log I_i$) and log marginal data densities (at the final stage)

Stages	"Truth"	DSMH	IW	$\text{NSE}^{(\text{IW})}$
15	993.04	992.84	1059.77	0.28
16	978.07	977.83	1045.33	0.21
17	963.20	963.00	1030.54	0.16
18	948.45	948.32	1015.64	0.14
19	933.84	933.69	1000.85	0.15
20	919.42	919.30	986.25	0.15
21	905.23	905.11	971.89	0.13
22	891.35	891.21	957.86	0.17
23	877.84	877.71	944.20	0.15
24	864.82	864.70	931.00	0.16
25	852.44	852.29	918.44	0.14
26	840.84	840.72	906.68	0.16
27	830.30	830.16	895.93	0.17
28	821.09	820.94	886.55	0.18
29	813.66	813.52	878.96	0.18
30	808.53	808.38	873.69	0.17
31	806.50	806.33	871.47	0.18
32	808.57	808.40	873.37	0.20
33	816.29	816.06	880.88	0.18
34	831.73	831.52	896.17	0.21
35	858.23	858.02	922.46	0.20
36	900.88	900.63	964.93	0.24
37	968.34	968.11	1032.19	0.27
38	1076.13	1075.74	1139.70	0.32
39	1255.27	1254.95	1318.74	0.46
40	1578.83	1578.28	1642.04	0.60
41	2266.42	2265.48	2328.81	1.23
42	4421.96	4421.17	4479.37	2.23

Note: "Truth" represents integral constants calculated from independent Gibbs sampling, "DSMH" represents the estimates obtained by applying the Mueller method to DSMH draws, "IW" stands for "importance weights" that deliver the estimates according to (5), and "$\text{NSE}^{(\text{IW})}$" represents the numerical standard error for the estimate in the previous column.

TABLE 3. Estimated log integral constants ($\log I_i$) and log marginal data densities (at the final stage)

Stages	DSMH	IW	$\text{NSE}^{(\text{IW})}$
19	905.87	949.32	0.29
20	903.40	940.90	0.23
21	898.30	932.07	0.22
22	890.21	923.11	0.12
23	881.70	914.24	0.11
24	873.19	905.73	0.10
25	865.17	897.69	0.12
26	857.77	890.33	0.12
27	851.20	883.76	0.11
28	845.63	878.22	0.11
29	841.39	873.90	0.12
30	838.70	871.23	0.12
31	838.03	870.56	0.13
32	839.93	872.43	0.12
33	845.10	877.62	0.16
34	854.61	887.13	0.13
35	869.78	902.32	0.17
36	892.73	925.26	0.19
37	926.53	959.11	0.22
38	976.06	1008.74	0.26
39	1049.53	1082.32	0.25
40	1148.17	1180.94	0.28
41	1280.56	1313.20	0.31
42	1456.38	1489.19	0.49
43	1692.13	1724.67	0.78
44	2003.65	2035.96	0.41
45	2409.17	2442.03	0.38
46	2936.05	2969.70	0.32
47	3617.15	3653.50	0.36
48	4501.04	4537.84	0.42

Note: "DSMH" represents integral constants estimated by applying the Mueller method to DSMH draws, "IW" stands for "importance weights" that deliver the estimates according to (5), and "$\text{NSE}^{(\text{IW})}$" represents the numerical standard error for the estimate in the previous column.

References

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26(2–4), 113–172.

Bernanke, B. S., M. Gertler, and M. W. Watson (1997): "Systematic Monetary Policy and the Effects of Oil Price Shocks," *Brookings Papers on Economic Activity*, 1, 91–142.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77(3), 623–685.

Bognanni, M., and E. Herbst (2014): "Estimating (Markov-Switching) VAR Models without Gibbs Sampling: A Sequential Monte Carlo Approach," Unpublished Manuscript.

Chib, S. (1995): "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

Chopin, N. (2004): "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *Annals of Statistics*, 32, 2385–2411.

Christiano, L. J., M. S. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and To What End?," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1A, pp. 65–148. North-Holland, Amsterdam, Holland.

——— (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

Cogley, T., and T. J. Sargent (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics*, 8, 262–302.

Del Negro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.

Durham, G., and J. Geweke (2012): "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," Unpublished Manuscript.

Fernández-Villaverde, J., P. Guerrón-Quintana, J. F. Rubio-Ramírez, and M. Uribe (2011): "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review*, 101(6), 2530–2561.

Fuentes-Albero, C., and L. Melosi (2013): "Methods for Computing Marginal Data Densities from the Gibbs Output," *Journal of Econometrics*, 175(2), 132–141.

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–73.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

Hamilton, J. D., D. F. Waggoner, and T. Zha (2007): "Normalization in Econometrics," *Econometric Reviews*, 26(2-4), 221–252.

Herbst, E., and F. Schorfheide (2014): "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*.

Hubrich, K., and R. J. Tetlow (Forthcoming): "Financial Stress and Economic Dynamics: the Transmission of Crises," *Journal of Monetary Economics*.

Ingram, B. F., and C. H. Whiteman (1994): "Supplanting the "Minnesota" Prior: Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors," *Journal of Monetary Economics*, 34(3), 497–510.

Justiniano, A., and G. E. Primiceri (2008): "The Time Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98(3), 604–641.

Kou, S. C., Q. Zhou, and W. H. Wong (2006): "Equi-energy sampler with applications in statistical inference and statistical mechanics," *Annals of Statistics*, 34(4), 1581–1619.

Leeper, E. M., C. A. Sims, and T. Zha (1996): "What Does Monetary Policy Do?," *Brookings Papers on Economic Activity*, 2, 1–78.

Litterman, R. B. (1986): "Forecasting with Bayesian Vector Autoregressions — Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38.

Liu, Z., D. F. Waggoner, and T. Zha (2011): "Sources of Macroeconomic Fluctuations: A Regime-Switching DSGE Approach," *Quantitative Economics*, 2, 251–301.

Meng, X.-L., and W. H. Wong (1996): "Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration," *Statistica Sinica*, 6, 831–860.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77, 665–696.

Rudebusch, G. D., and L. E. O. Svensson (1999): "Policy Roles for Inflation Targeting," in *Monetary Policy Rules*, ed. by J. B. Taylor, chap. 5, pp. 203–262. University of Chicago Press, Chicago and London.

Sims, C. A., D. F. Waggoner, and T. Zha (2008): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*, 146(2), 255–274.

Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–968.

——— (2006): "Were There Regime Switches in US Monetary Policy?," *American Economic Review*, 96, 54–81.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Tierney, L. (1994): "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics*, 22(4), 1701–1728.

Waggoner, D. F., and T. Zha (2003a): "A Gibbs Sampler for Structural Vector Autoregressions," *Journal of Economic Dynamics and Control*, 28(2), 349–366.

——— (2003b): "Likelihood Preserving Normalization in Multiple Equation Models," *Journal of Econometrics*, 114(2), 329–347.

Federal Reserve Bank of Atlanta, Federal Reserve Bank of Atlanta, Federal Reserve Bank of Atlanta, Emory University, and NBER