Tailored Multiple-block MCMC Methods for Analysis of DSGE and Other Models

Siddhartha Chib Srikanth Ramamurthy

Washington University in St. Louis

March 2009

Introduction

- DSGE models are arguably the dominant framework for dealing with macro-economic dynamics
- From our perspective, provide a perfect setting for exploring a number of different MCMC issues because
  - highly nonlinear in the parameters
  - parameters subject to several linear and nonlinear constraints
  - intensive computation required to calculate the likelihood
  - occurrence of multiple modes
- Maximum likelihood estimates can be unreasonable
- Has led to an interest in Bayesian fitting (for example, Lubik and Schorfheide (2004), Fernandez-Villaverde and Rubio-Ramirez (2004), Smets and Wouters (2007))
- Based on single-block MCMC sampling: neither efficient nor scaleable

- We discuss new approaches for simulating the posterior
- Address the blocking problem that is central to this and other MCMC problems
- Approach to handle multiple modes
- Illustrate the ideas in the context of non-DSGE and DSGE models

Outline of presentation

- Setup
- Proposed methods
- Various applications
- Conclusion

Setup

- To understand the context, consider the model in Ireland (2004)
- This model after linearization around the steady state has the form

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

- Parameters of interest are

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

which are subject to

- the linear constraints $\mathcal{S}_L$: $\{\omega, \alpha_x, \alpha_\pi\} \in (0,1)$, $\{\rho_\pi, \rho_g, \rho_x\} \in (0,\infty)$, $\{\rho_a, \rho_e\} \in (0,1)$
- non linear constraints 1 $\mathcal{S}_\Omega$: σ_i^2 lie in the region that satisfy the usual positivity and positive definiteness constraints
- the determinacy constraint $\mathcal{S}_D$ (of a unique stable solution)

- To proceed, it is first necessary to solve the model for the endogenous variables and the expectational variables
- This requires that one first express the given model in the form

$$\mathbf{G}_0(\boldsymbol{\theta})\mathbf{s}_t = \mathbf{G}_1(\boldsymbol{\theta})\mathbf{s}_{t-1} + \mathbf{G}_2(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t + \mathbf{G}_3(\boldsymbol{\theta})\boldsymbol{\eta}_t$$

where $\mathbf{G}_{j=0}^{3}$ are matrices of appropriate dimensions involving the parameters $\boldsymbol{\theta}$ of the model

- For instance in the Ireland model if we define

$$\mathbf{s}_t = [\hat{y}_t, \hat{r}_t, \hat{\pi}_t, \hat{g}_t, \hat{x}_t, \hat{a}_t, \hat{e}_t, \hat{z}_t, \mathbb{E}_t\hat{\pi}_{t+1}, \mathbb{E}_t\hat{x}_{t+1}]$$
$$\boldsymbol{\varepsilon}_t = [\varepsilon_{a,t}, \varepsilon_{e,t}, \varepsilon_{z,t}, \varepsilon_{R,t}]'$$
$$\boldsymbol{\eta}_t = [\hat{\pi}_t - \mathbb{E}_{t-1}\hat{\pi}_t, \hat{x}_t - \mathbb{E}_{t-1}\hat{x}_t]$$

then it can be written as

$$\mathbf{G}_0(\boldsymbol{\theta})\mathbf{s}_t = \mathbf{G}_1(\boldsymbol{\theta})\mathbf{s}_{t-1} + \mathbf{G}_2(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t + \mathbf{G}_3(\boldsymbol{\theta})\boldsymbol{\eta}_t$$

for suitable choices of matrices $\mathbf{G}_{j=0}^{3}$

- One can then solve the model (by a series of computationally intensive steps) to produce a SSM for the state variables of the form

$$\begin{aligned} \mathbf{y}_t &= \mathbf{a}(\boldsymbol{\theta}) + \mathbf{B}(\boldsymbol{\theta})\mathbf{s}_t \\ \mathbf{s}_t &= \mathbf{D}(\boldsymbol{\theta})\mathbf{s}_{t-1} + \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t \end{aligned}$$

where $\mathbf{D}(\boldsymbol{\theta})$ and $\mathbf{F}(\boldsymbol{\theta})$ are awkward implicit functions of the model parameters, obtained from the solution, and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, $\boldsymbol{\Omega}$ p.d.

- For example, in the Ireland model the SSM has the form

$$\underbrace{\begin{bmatrix} \hat{g}_t \\ \hat{\pi}_t \\ \hat{r}_t \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{B}} \mathbf{s}_t$$

$$\mathbf{s}_t = \mathbf{D}(\boldsymbol{\theta})\mathbf{s}_{t-1} + \mathbf{F}(\boldsymbol{\theta})\varepsilon_t$$

where $\varepsilon_t \sim \mathcal{N}_4(\mathbf{0}, \boldsymbol{\Omega})$

- The modeling, the canonical representation, the solve step and the SSM form, can involve a large number of variables and parameters
- For example, the model of Smets and Wouter (2007) comprises 14 equations in 14 endogenous variables and 7 exogenous driving processes and 36 parameters
- Setting this up in canonical form for the solve step requires a 53 dimensional state vector

Inference

- Given the SSM, the joint density of the data $\mathbf{y}_{n\times T} = \{\mathbf{y}_t\}$, $t = 1, \ldots, T$, is, of course calculable as

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{t=1}^{T} \left[\frac{1}{(2\pi)^{n/2}|\Sigma_{t|t-1}|^{1/2}} \times \exp\left\{ -\frac{1}{2} y'_{t|t-1} \Sigma^{-1}_{t|t-1} y_{t|t-1} \right\} \right] \mathbb{I}_{\mathcal{S}_L}(\boldsymbol{\theta}) \mathbb{I}_{\mathcal{S}_\Omega}(\boldsymbol{\theta}) \mathbb{I}_{\mathcal{S}_D}(\boldsymbol{\theta})$$

- In the Bayesian context, we focus on the posterior density of $\boldsymbol{\theta}$

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta}) \times \pi(\boldsymbol{\theta})$$

and the question is how this complex distribution should be efficiently summarized

Tailored Block-MH (TaB-MH) algorithm

- Because of the intensity of the solve step, existing methods have relied on a single block RW-MH sampling approach
- Not efficient or scaleable
- We examine an alternative in which parameters are updated within each MCMC cycle by a sequence of M-H steps over (randomly) constructed blocks
- Blocks are constructed randomly because there is no natural blocking scheme in these models
- In addition, tailored proposal densities are used (following Chib and Greenberg (1994, 1995))
- To account for possible irregularities, tailoring is done by simulated annealing (Chib and Ergashev (2008))

Algorithm: TaB-MH

1. Initialize $\boldsymbol{\theta}^{(0)} \in \mathcal{S}_L \cap \mathcal{S}_{\boldsymbol{\Omega}} \cap \mathcal{S}_D$ and fix n_0 (the burn-in) and M (the MCMC sample size)
2. Randomly generate blocks $(\boldsymbol{\theta}_{j,1}, \boldsymbol{\theta}_{j,2}, \ldots, \boldsymbol{\theta}_{j,p_j})$
3. Sample each block $\boldsymbol{\theta}_{j,l}$, $l = 1, \ldots, p_j$, by a M-H step with a tailored proposal density
4. Repeat Steps 2-3 $n_0 + M$ times, discard the draws from the first n_0 iterations and save the subsequent M draws $\boldsymbol{\theta}^{(n_0+1)}, \ldots, \boldsymbol{\theta}^{(n_0+M)}$

Marginal Likelihood

- Interestingly, the framework of Chib (1995), and its M-H version in Chib and Jeliazkov (2001), can be applied in this setting, with suitable modifications to accommodate randomized blocking
- Starting point is the identity

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^*|\mathbf{y})},$$

  where the terms in the numerator are readily available
- Posterior ordinate estimated from a marginal-conditional decomposition

Example: Dynamic Factor Model

- Suppose

$$\begin{aligned} \mathbf{y}_t &= \mathbf{a} + \mathbf{B}\mathbf{s}_t + \mathbf{u}_t \\ \mathbf{s}_t &= \mathbf{G}\mathbf{s}_{t-1} + \varepsilon_t \end{aligned}$$

where $\mathbf{y}_t$ is a 10×1 vector of observables at time t, $\mathbf{s}_t$ is a 5×1 vector of time-t unobserved (latent) states, $\mathbf{a}$, $\mathbf{B}$ and $\mathbf{G}$ are matrices of appropriate dimensions, $\mathbf{u}_t \sim \mathcal{N}_{10}(\mathbf{0}, \mathbf{\Sigma})$ and $\varepsilon_t \sim \mathcal{N}_5(\mathbf{0}, \mathbf{\Omega})$.

Identification restrictions and parameter constraints

- $\mathbf{G}$ is diagonal
- $\mathbf{B}_{i,i} = 1$ for $i = 1, \ldots, 5$, $\mathbf{B}_{i,j} = 0$ for $i, j = 1, \ldots, 5$, $j > i$,
- $\mathbf{\Sigma} = \text{diag}\{\sigma_i^2\}_{i=1}^{10}$
- $\mathbf{\Omega} = \mathbf{I}_5$.
- reparameterize $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma}^* = \text{diag}\{\sigma_i^{2*}\}_{i=1}^{10}; \qquad \sigma_i^{2*} \in \mathcal{R}$$

  where $\sigma_i^2 = \exp\left(\sigma_i^{2*}\right)$
- stationarity restriction: $\mathbf{\Theta}_{\mathcal{S}} = \{\boldsymbol{\theta} : \text{abs}(\text{eig}(\mathbf{G})) < 1\}$
- This leads to 60 unknown parameters that we collect in the vector $\boldsymbol{\theta}$

Table 1–DGP for parameters in SSM example

Param.	DGP				
$\mathbf{G}_{1,1}, \ldots, \mathbf{G}_{5,5}$	0.80	0.20	0.75	0.60	0.10
$a_1, \ldots, a_5$	0.20	1.40	1.80	0.10	0.90
$a_6, \ldots, a_{10}$	1.00	2.00	0.10	2.20	1.50
$\mathbf{B}$	1.00	—	—	—	—
	0.50	1.00	—	—	—
	0.60	0.00	1.00	—	—
	0.00	0.20	-0.10	1.00	—
	-0.20	0.00	-0.70	0.00	1.00
	0.00	0.00	-0.40	-0.50	0.00
	0.30	0.20	0.00	0.00	-0.30
	-0.50	0.00	0.00	0.60	0.00
	0.00	-0.50	0.30	-0.10	0.00
	0.00	0.00	0.20	0.00	-0.40
$\sigma_1^2, \ldots, \sigma_5^2$	1.00	0.30	1.00	0.20	0.60
$\sigma_6^2, \ldots, \sigma_{10}^2$	0.50	1.00	1.00	0.75	0.60

Note: Prior variance in paranthesis.

- For notational convenience, denote

$$\begin{aligned}
\boldsymbol{\theta}_1 &= \text{vecr}(\{\mathbf{G}_{i,i}\}) \\
\boldsymbol{\theta}_2 &= \mathbf{a} \\
\boldsymbol{\theta}_3 &= \text{vecr}(\{\mathbf{B}_{i,j}\}), i, j = 2, \ldots, 5, j < i \\
\boldsymbol{\theta}_4 &= \text{vecr}(\{\mathbf{B}_{i,j}\}), i = 6, \ldots, 10, j = 1, \ldots, 5 \\
\boldsymbol{\theta}_5 &= \{\sigma_i^{2*}\}_{i=1}^{10}
\end{aligned}$$

and let

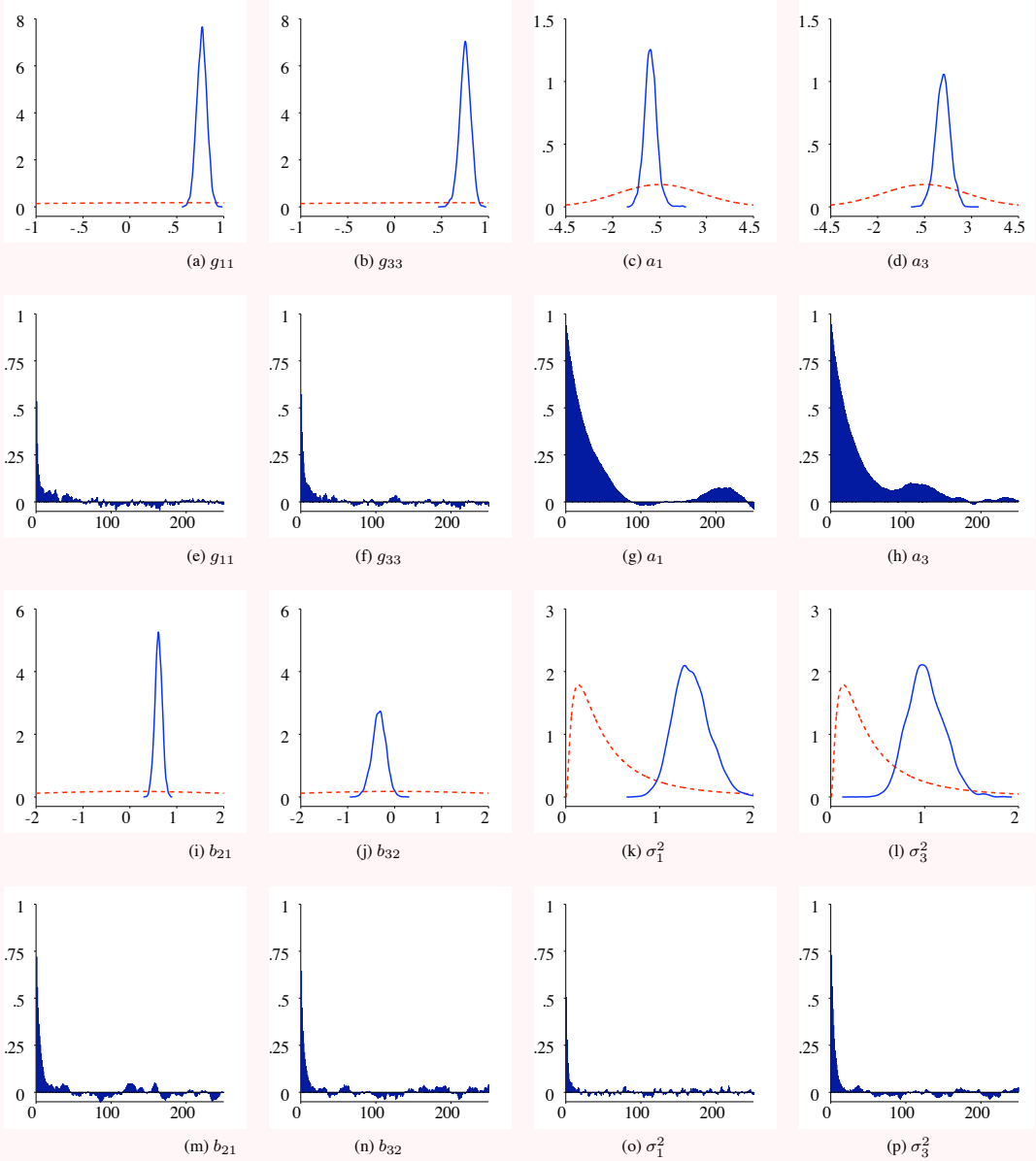
$$\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}_1|\mathbf{g}_0, \mathbf{V}_g)\mathcal{N}(\boldsymbol{\theta}_2|\mathbf{a}_0, \mathbf{V}_a)\mathcal{N}(\boldsymbol{\theta}_3|\boldsymbol{\theta}_{30}, \mathbf{V}_{\boldsymbol{\theta}_3})\times \\
\mathcal{N}(\boldsymbol{\theta}_4|\boldsymbol{\theta}_{40}, \mathbf{V}_{\boldsymbol{\theta}_4})\mathcal{N}(\boldsymbol{\theta}_5|\boldsymbol{\sigma}_0^*, \mathbf{V}_{\boldsymbol{\sigma}^*})\mathbf{I}_{\Theta_{\mathcal{S}}}.
\end{aligned}$$

Three MCMC schemes compared

- TaB-MH: fully randomized blocks run for 1,000 + 10,000 iterations
- FBTab-MH: fixed blocks with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \boldsymbol{\theta}_3', \boldsymbol{\theta}_4', \boldsymbol{\theta}_5')'$ run for 1,000 + 10,000 iterations
- RW-MH: single block sampler tuned to generate an acceptance rate of 30% run for 250,000 + 1 million iterations

Summary of results: inefficiency factors

- TaB-MH: inefficiency factors between 2 and 42, average of 10
- FBTaB-MH: inefficiency factors between 3 and 168, average of 38
- RW-MH: inefficiency factors between 632 and 5000, average of 2130



(a) g_{11} (b) g_{33} (c) a_1 (d) a_3

(e) g_{11} (f) g_{33} (g) a_1 (h) a_3

(i) b_{21} (j) b_{32} (k) σ_1^2 (l) σ_3^2

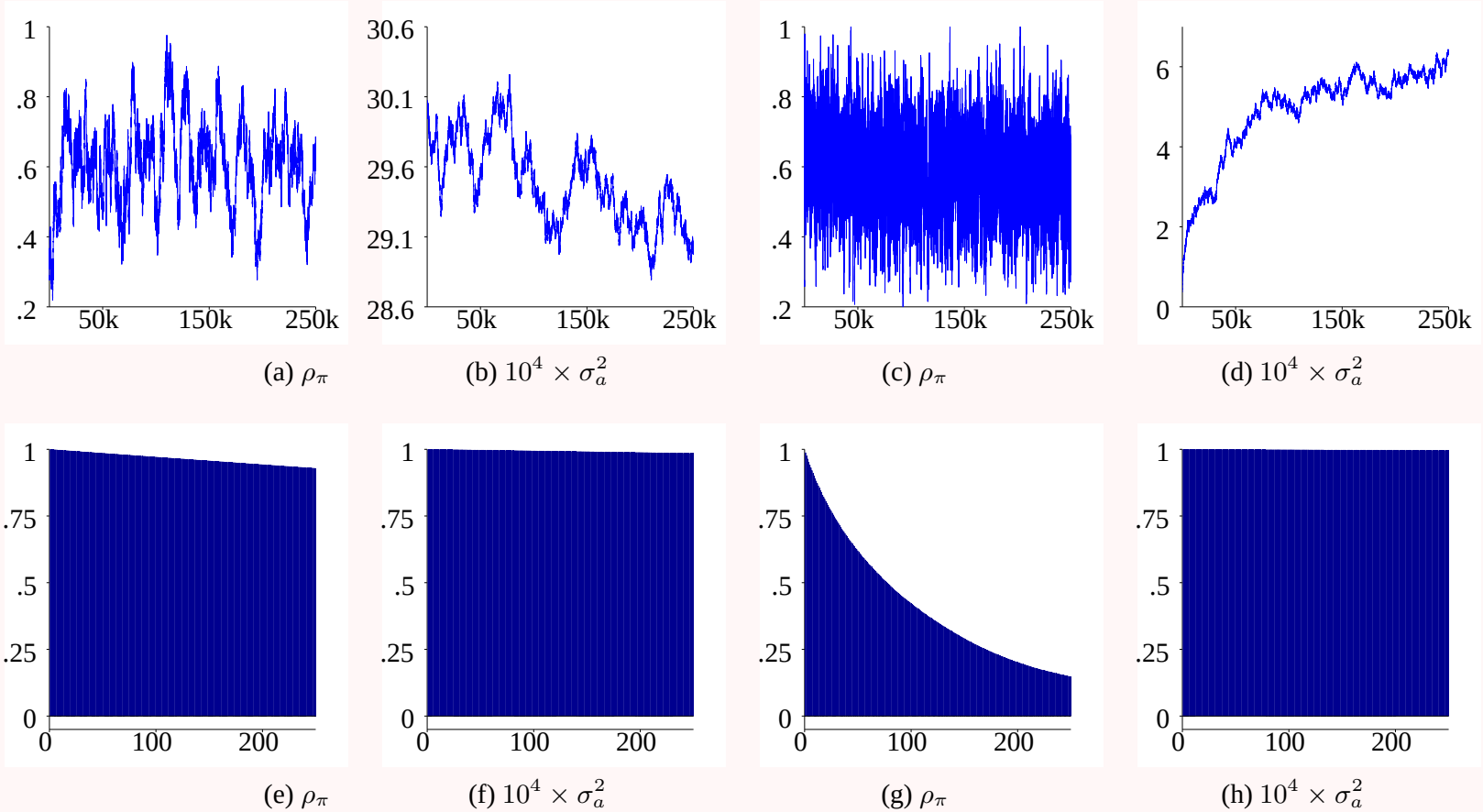
(m) b_{21} (n) b_{32} (o) σ_1^2 (p) σ_3^2

Example: Ireland04

- First, we start three different RW chains, each of length 250,000 following a burn-in of 50,000 iterations
- One at the prior mean, one at a local mode and the third at the dominant mode
- Also three different variance-covariance matrices for the proposal - $k \times \mathbf{I}_{12}$ for chain I and k times the variance at the modes for chains II and III.

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

	Chain I		Chain II		Chain III	
Parameter	start	post mean	start	post mean	start	post mean
ω	0.2000	0.0701	0.0589	0.1498	0.1058	0.1046
α_x	0.1000	0.0651	0.0612	0.0823	0.0629	0.0792
α_π	0.1000	0.0825	0.0443	0.0809	0.0605	0.0800
ρ_π	0.3000	0.6079	0.2934	0.5886	0.5515	0.5471
ρ_g	0.3000	0.4022	0.3201	0.3722	0.3593	0.3767
ρ_x	0.2500	0.1825	0.2742	0.1979	0.1760	0.2034
ρ_a	0.8500	0.9583	0.5179	0.8694	0.9334	0.9340
ρ_e	0.8500	0.8843	0.8858	0.8838	0.8874	0.8629
$10000\sigma_a^2$	30.0000	29.4853	0.3627	4.8860	13.6777	16.2229
$10000\sigma_e^2$	0.0800	0.0077	0.0037	0.0066	0.0060	0.0069
$10000\sigma_z^2$	5.0000	3.6314	0.4287	0.7947	0.6977	0.7657
$10000\sigma_r^2$	0.5000	0.1041	0.1088	0.0967	0.0857	0.0982



(a) ρ_π (b) $10^4 \times \sigma_a^2$ (c) ρ_π (d) $10^4 \times \sigma_a^2$

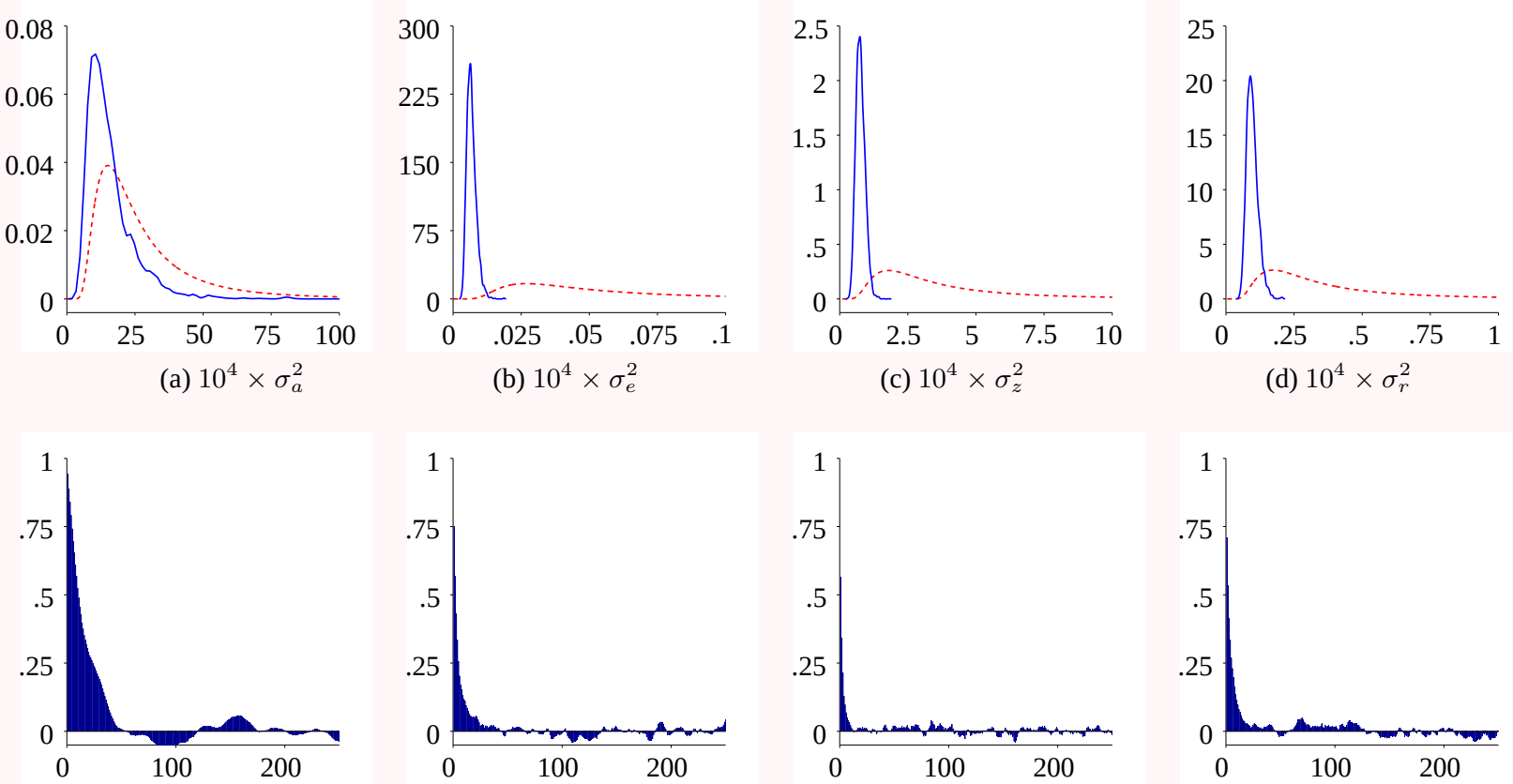
(e) ρ_π (f) $10^4 \times \sigma_a^2$ (g) ρ_π (h) $10^4 \times \sigma_a^2$

Results from Tab-MH

- Chain initialized at prior mean
- Degrees of freedom for t-proposal: $\nu = 15$
- Simulation length 11,000; first 1000 draws discarded as burn-ins

Table 2–Posterior sampling results using the TAB-MH algorithm for the Ireland (2004) model

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



(a) $10^4 \times \sigma_a^2$ (b) $10^4 \times \sigma_e^2$ (c) $10^4 \times \sigma_z^2$ (d) $10^4 \times \sigma_r^2$

(e) $10^4 \times \sigma_a^2$ (f) $10^4 \times \sigma_e^2$ (g) $10^4 \times \sigma_z^2$ (h) $10^4 \times \sigma_r^2$

Marginal likelihood

- We consider the effect of varying the number of blocks (stages), as well as the sample size (n_1) in the reduced MCMC runs, on the marginal likelihood estimate (and the resulting numerical standard error)
- Report results from both two stage and 3 stage decompositions of the posterior ordinate
  1. Based on 5,000, 10,000 and 15,000 draws in the reduced runs
- Also compare the results to the estimate of the marginal likelihood under the RW-MH algorithm (1 stage)
  1. Based on 75,000, 150,000 and 250,000 draws

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

	TaB-MH			RW-MH
n_1	2 stage	3 stage	n_1	1 stage
5,000	1170.08	1170.26	75,000	1169.89
	(0.0324)	(0.0400)		(0.6121)
10,000	1170.18	1170.29	150,000	1170.55
	(0.0268)	(0.0302)		(0.5884)
15,000	1170.15	1170.33	250,000	1170.84
	(0.0216)	(0.0250)		(0.4839)

Example: SW07

- Large scale model: 53 dimensional state vector and 36 parameters
- Locating the posterior mode is challenging to say the least
  - Modal ordinate found using SA is around -877.72 in the log scale
  - In contrast, the modal ordinate reported in SW07 is around -906.29
  - TaB-MH explores even higher regions

Posterior sampling

- Similar values chosen for SA parameters as in Ireland model
- Degrees of freedom in t proposal density set to 10
- Sampler initialized at prior mean and run for 10,000 iterations following a burn-in of 1,000 iterations

SW07: Results

- Posterior ordinate at mean of TaB-MH sample substantially higher than that at the mean of the SW07 RW-MH sample (-871.66 compared to -888.84)
- In effect, $\bar{\pi}$ and $\bar{l}$ significantly different
- 90 percent intervals of the TaB-MH sample wider than that of the RW-MH sample
- Autocorrelations among the sample draws orders of magnitude higher in the RW-MH sample

Table – Summary of posterior ordinates at the mode and mean in the Smets and Wouters (2007) model

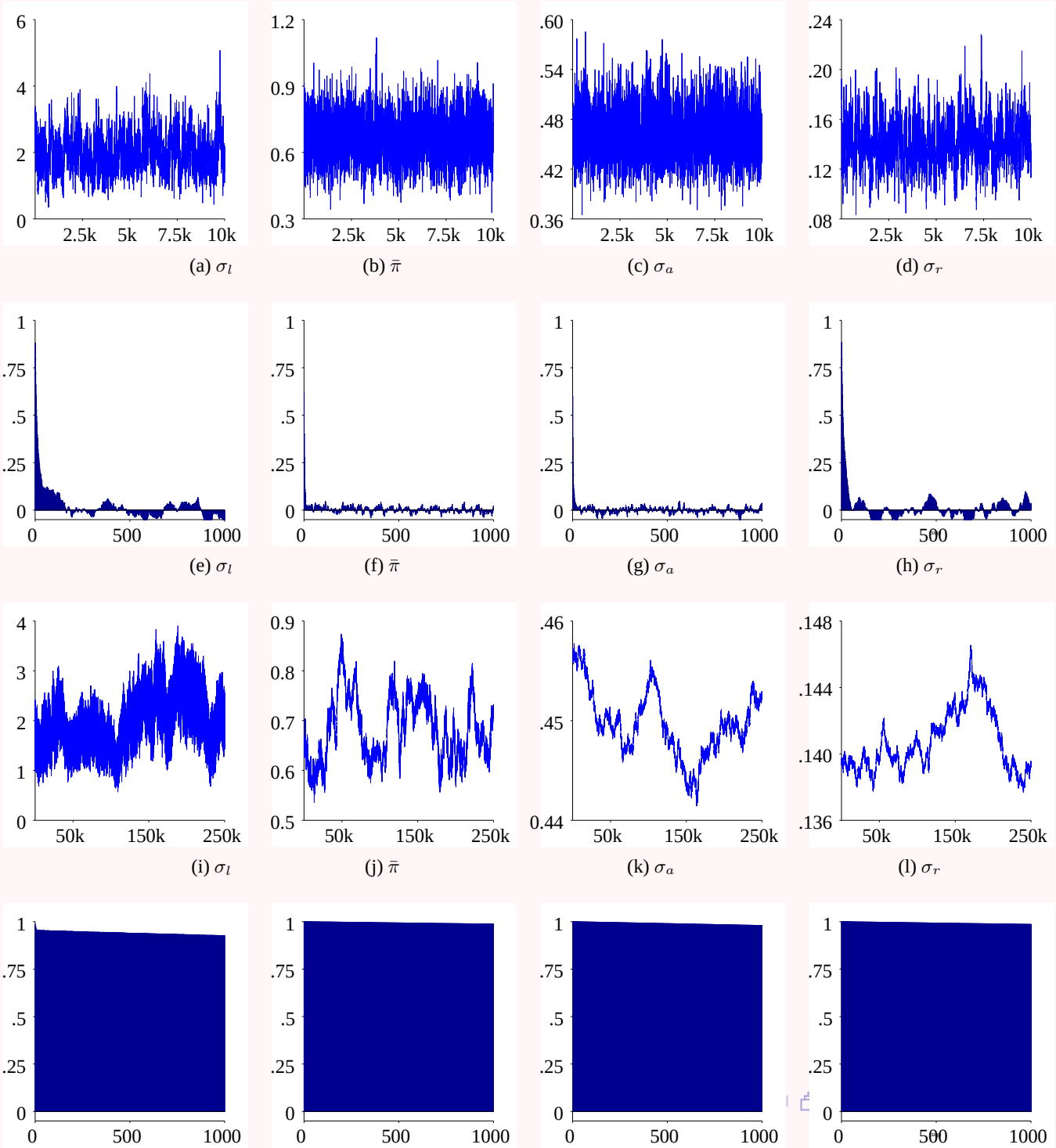
	SW07	SA/TaB-MH
Mode	-906.29	-877.72
Mean	-888.84	-871.66

Table – Posterior summary: structural parameters in SW07 model

	SW07 Posterior		TaB-MH Posterior		
Parameter	Mean	90 percent interval	Mean	90 percent interval	Inefficiency factors
φ	5.74	[3.97,7.42]	5.77	[3.69 ,8.06]	9.73
σ_c	1.38	[1.16,1.59]	1.36	[1.03 ,1.71]	12.14
h	0.71	[0.64,0.78]	0.75	[0.66 ,0.82]	15.64
ξ_w	0.70	[0.60,0.81]	0.65	[0.52 ,0.79]	54.28
σ_l	1.83	[0.91,2.78]	1.98	[0.96 ,3.21]	28.79
ξ_p	0.66	[0.56,0.74]	0.62	[0.49 ,0.75]	45.22
ι_w	0.58	[0.38,0.78]	0.59	[0.31 ,0.83]	8.65
ι_p	0.24	[0.10,0.38]	0.23	[0.08 ,0.41]	18.13
ψ	0.54	[0.36,0.72]	0.59	[0.36 ,0.81]	5.93
Φ	1.60	[1.48,1.73]	1.57	[1.42 ,1.74]	6.44
r_π	2.04	[1.74,2.33]	2.00	[1.64 ,2.37]	11.21
ρ	0.81	[0.77,0.85]	0.80	[0.75 ,0.85]	11.42
r_y	0.08	[0.05,0.12]	0.08	[0.03 ,0.13]	26.12
$r_{\Delta y}$	0.22	[0.18,0.27]	0.23	[0.17 ,0.29]	5.68
$\bar{\pi}$	0.78	[0.61,0.96]	0.66	[0.48 ,0.85]	3.81
$100(\beta^{-1}-1)$	0.16	[0.07,0.26]	0.16	[0.06 ,0.29]	6.09
$\bar{l}$	0.53	[-1.3,2.32]	0.95	[-0.07,2.56]	9.39
$\bar{\gamma}$	0.43	[0.40,0.45]	0.41	[0.37 ,0.46]	9.51
α	0.19	[0.16,0.21]	0.19	[0.15 ,0.23]	4.68

Table – Posterior summary: shock parameters in SW07 model

	SW07 Posterior		TaB-MH Posterior		
Parameter	Mean	90 percent interval	Mean	90 percent interval	Inefficiency factors
σ_a	0.45	[0.41,0.50]	0.46	[0.41,0.53]	4.17
σ_b	0.23	[0.19,0.27]	0.25	[0.18,0.30]	15.23
σ_g	0.53	[0.48,0.58]	0.53	[0.47,0.59]	2.57
σ_I	0.45	[0.37,0.53]	0.43	[0.34,0.55]	33.14
σ_r	0.24	[0.22,0.27]	0.25	[0.22,0.28]	4.30
σ_p	0.14	[0.11,0.16]	0.14	[0.10,0.18]	14.89
σ_w	0.24	[0.20,0.28]	0.26	[0.21,0.32]	11.62
ρ_a	0.95	[0.94,0.97]	0.96	[0.93,0.98]	6.15
ρ_b	0.22	[0.07,0.36]	0.21	[0.04,0.49]	24.42
ρ_g	0.97	[0.96,0.99]	0.98	[0.96,0.99]	7.53
ρ_I	0.71	[0.61,0.80]	0.74	[0.61,0.86]	37.54
ρ_r	0.15	[0.04,0.24]	0.15	[0.04,0.30]	6.64
ρ_p	0.89	[0.80,0.96]	0.89	[0.75,0.98]	48.92
ρ_w	0.96	[0.94,0.99]	0.98	[0.96,1.00]	21.80
μ_p	0.69	[0.54,0.85]	0.66	[0.38,0.84]	38.23
μ_w	0.84	[0.75,0.93]	0.83	[0.63,0.94]	43.72
ρ_{ga}	0.52	[0.37,0.66]	0.50	[0.32,0.69]	2.61



(a) σ_l (b) $\bar{\pi}$ (c) σ_a (d) σ_r

(e) σ_l (f) $\bar{\pi}$ (g) σ_a (h) σ_r

(i) σ_l (j) $\bar{\pi}$ (k) σ_a (l) σ_r

Extension to multi-modal problems

- For simplicity, consider sampling a bimodal distribution
- Assume that the modal values have been found by initial optimization
- Let the location of the two modes be $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$
- Also, let $\mathbf{V}_1$ and $\mathbf{V}_2$ denote the inverse of the negative Hessian at the two modes

- Now TaB-MH algorithm is used as above
- But every few (say a 100) iterations we generate a proposal $\boldsymbol{\theta}^\dagger$ from the mixture density

$$q(\boldsymbol{\theta}|\mathbf{y}) = p\, t(\boldsymbol{\theta}|\boldsymbol{\mu}_1, \mathbf{V}_1, \nu_1) + (1-p)\, t(\boldsymbol{\theta}|\boldsymbol{\mu}_2, \mathbf{V}_2, \nu_2),$$

which we accept with probability

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^\dagger|\mathbf{y}) = \min\left\{ \frac{f(\mathbf{y}|\boldsymbol{\theta}^\dagger)\pi(\boldsymbol{\theta}^\dagger)}{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \frac{p\, t(\boldsymbol{\theta}|\boldsymbol{\mu}_1, \mathbf{V}_1, \nu_1) + (1-p)\, t(\boldsymbol{\theta}|\boldsymbol{\mu}_2, \mathbf{V}_2, \nu_2)}{p\, t(\boldsymbol{\theta}^\dagger|\boldsymbol{\mu}_1, \mathbf{V}_1, \nu_1) + (1-p)\, t(\boldsymbol{\theta}^\dagger|\boldsymbol{\mu}_2, \mathbf{V}_2, \nu_2)}, 1 \right\}$$

Example: Two component mixture of six dimensional normals

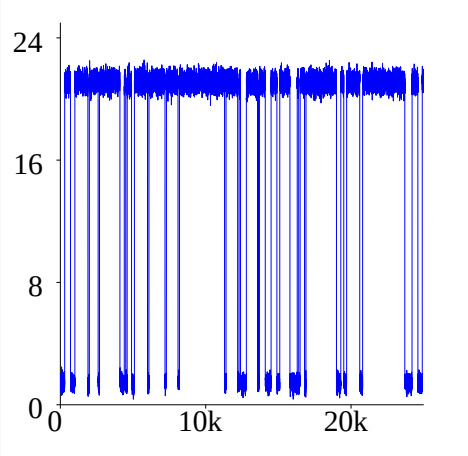
- Pick a modal value ($\boldsymbol{\mu}_1$) from AS07 model and set the other mode ($\boldsymbol{\mu}_2$) to $15 \times \boldsymbol{\mu}_1$

$$\boldsymbol{\mu}_1 : (1.41, 0.81, 0.49, 0.80, 1.07, 0.30)$$
$$\boldsymbol{\mu}_2 : (21.15, 12.15, 7.35, 12.00, 16.05, 4.50)$$

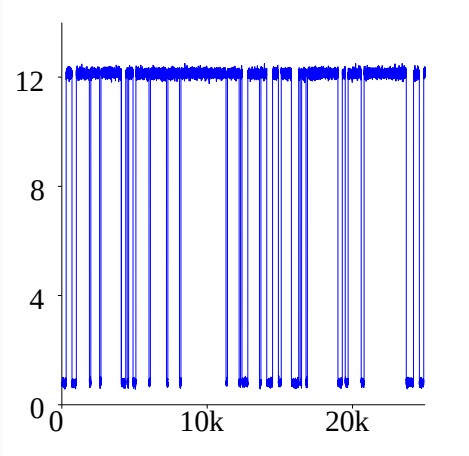
- Also variance equated to the reduced variance at the two modes in AS07 model
- Target density

$$f_X(x) = 0.2\,\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + 0.8\,\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$$

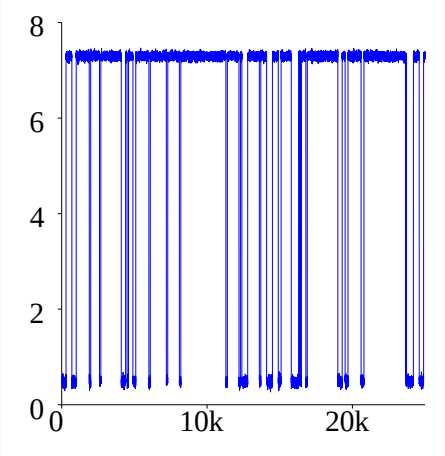
- TaBMJ-MH algorithm sampler initialized at μ_2
- $p = 0.5$, $\nu = 5$ in usual TaB-MH step, $\nu = 5$ in mode jumping step
- Sampler run for 25,000 iterations without any burn-ins (for illustration purposes)



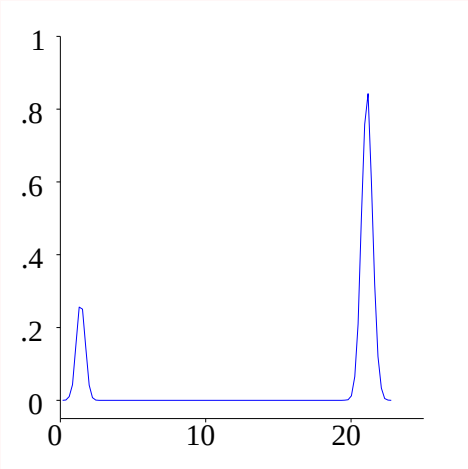
(a) x_1



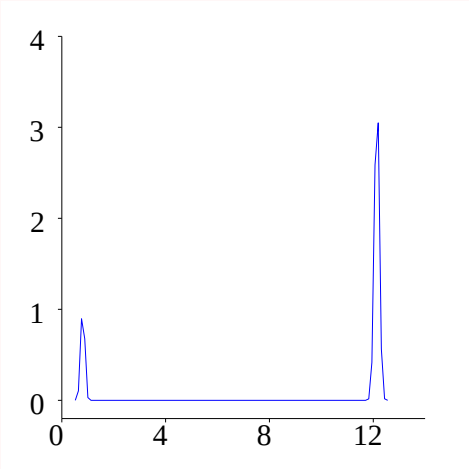
(b) x_2



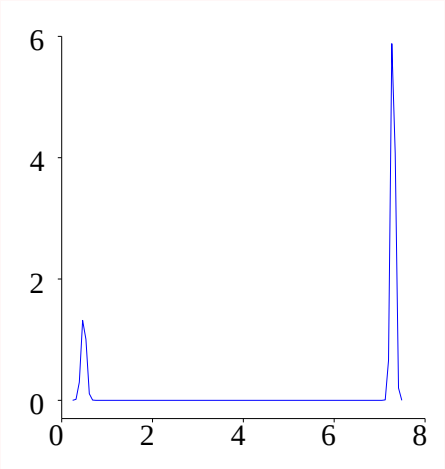
(c) x_3



(d) x_1



(e) x_2



(f) x_3

Example: AS07

- 13 parameter model
- Possibly multi-modal posterior - two distinct separated modal regions
- Difference in (unnormalized) posterior ordinate 8 in log scale
- RW-MH sampler only explores the posterior locally in individual modal regions even when run for a million iterations

- The output growth version of the DSGE model in AS07 is given by

$$
\begin{aligned}
\hat{y}_t &= \mathbb{E}_t \hat{y}_{t+1} + \hat{g}_t - \mathbb{E}_t \hat{g}_{t+1} - \frac{1}{\tau}(\hat{r}_t - \mathbb{E}_t \hat{\pi}_{t+1} - \mathbb{E}_t \hat{z}_{t+1}) \\
\hat{\pi}_t &= \beta \mathbb{E}_t \hat{\pi}_{t+1} + \kappa(\hat{y}_t - \hat{g}_t) \\
\hat{c}_t &= \hat{y}_t - \hat{g}_t \\
\hat{r}_t &= \rho_r \hat{r}_{t-1} + (1-\rho_r)\psi_1 \hat{\pi}_t + (1-\rho_r)\psi_2(\triangle \hat{y}_t + \hat{z}_t) + \varepsilon_{r,t} \\
\hat{g}_t &= \rho_g \hat{a}_{t-1} + \varepsilon_{g,t} \\
\hat{z}_t &= \rho_z \hat{z}_{t-1} + \varepsilon_{z,t}
\end{aligned}
$$

- $n = 80$ observations simulated from the model
- Outcomes assumed to be quarterly observations on
  1. per capita GDP growth rates $\hat{Y}_t = \gamma^Q + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t)$
  2. annualized inflation rates $\pi_t = \pi^A + 400\hat{\pi}_t$
  3. annualized nominal interest rate $r_t = \pi^A + r^A + 4\gamma Q + 400\hat{r}_t$

  where γ^Q, r^A, and π^A are related to the steady states of the relevant variables

SSM

$$\underbrace{\begin{bmatrix} \hat{Y}_t \\ \pi_t \\ r_t \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{bmatrix} \gamma^{(Q)} \\ \pi^A \\ \pi^A + r^A \end{bmatrix}}_{\mathbf{a}} + \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{B}} \mathbf{s}_t$$

$$\mathbf{s}_t = \mathbf{D}(\boldsymbol{\theta})\mathbf{s}_{t-1} + \mathbf{F}(\boldsymbol{\theta})\varepsilon_t$$

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

		Prior		
Parameter	DGP	Density	Mean	Standard deviation
τ	2.00	Gamma	2.00	0.50
κ	0.15	Gamma	0.20	0.10
ψ_1	1.50	Gamma	1.50	0.25
ψ_2	1.00	Gamma	0.50	0.25
ρ_r	0.60	Beta	0.50	0.20
ρ_g	0.95	Beta	0.80	0.10
ρ_z	0.65	Beta	0.66	0.15
r^A	0.40	Gamma	0.50	0.50
π^A	4.00	Gamma	7.00	2.00
γ^Q	0.50	Normal	0.40	0.20
σ_r	0.20	Inverse Gamma	0.50	0.26
σ_g	0.80	Inverse Gamma	1.25	0.65
σ_z	0.45	Inverse Gamma	0.63	0.33

Posterior sampling

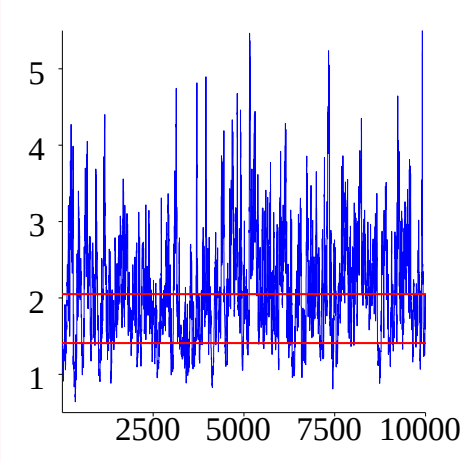
- Sampler initialized at low mode
- TaB-MH step: $\nu = 2$
- TaBMJ-MH step
  - $\nu = 5$
  - Equal probability assigned to both modes
  - Called every $100th$ iteration
- Sampler run for 10,000 iterations without any burn-in

Summary of results

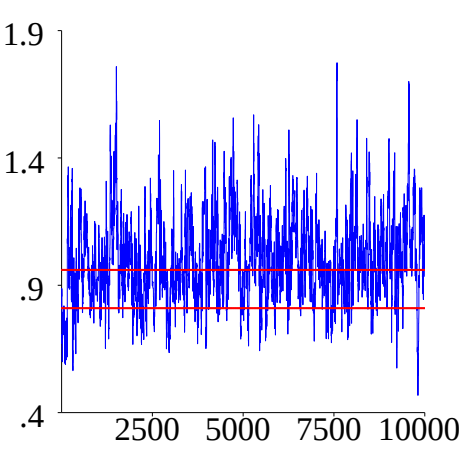
- Jump from the low to the high mode through the TaBMJ-MH step at the 200th iteration
- Reverse jump from the high to the low mode through TaBMJ-MH only once in the 300th iteration
- Occasional visits to the low mode in the TaB-MH steps
- Global exploration of posterior

Table – Posterior sampling results using the TaBMJ-MH algorithm for the An-Schorfheide (2007) model

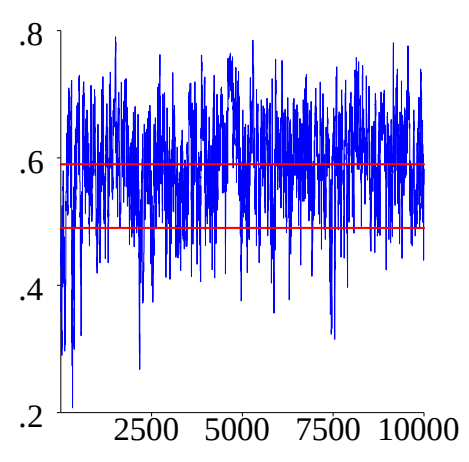
	Posterior modes		Posterior			
Parameter	Mode 1	Mode 2	Mean	Numerical S.E.	90 percent interval	Inefficiency factors
τ	2.05	1.41	2.12	0.0392	[1.04,3.74]	29.91
κ	0.16	0.18	0.17	0.0059	[0.03,0.42]	34.54
ψ_1	1.55	1.57	1.66	0.0149	[1.12,2.42]	19.88
ψ_2	0.96	0.81	1.00	0.0101	[0.70,1.35]	36.38
ρ_r	0.59	0.49	0.59	0.0054	[0.41,0.72]	49.38
ρ_g	0.94	0.97	0.92	0.0033	[0.79,0.98]	41.57
ρ_z	0.58	0.80	0.54	0.0094	[0.21,0.83]	31.57
r^A	0.64	0.62	0.68	0.0083	[0.07,1.43]	5.54
π^A	4.06	4.00	4.16	0.0212	[3.28,5.52]	14.48
γ^Q	0.50	0.54	0.48	0.0050	[0.11,0.80]	8.655
σ_r	0.22	0.24	0.23	0.0012	[0.18,0.32]	11.61
σ_g	0.76	1.07	0.76	0.0120	[0.45,1.33]	28.57
σ_z	0.54	0.30	0.61	0.0111	[0.30,1.01]	38.86



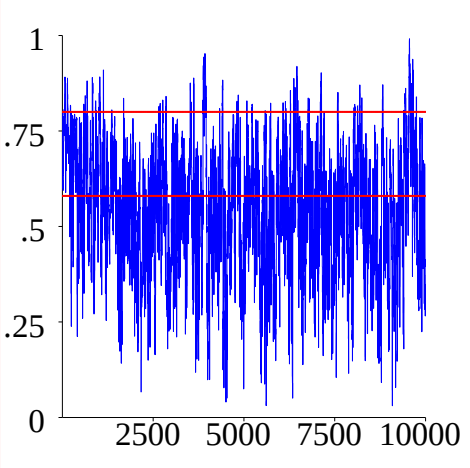
(a) τ



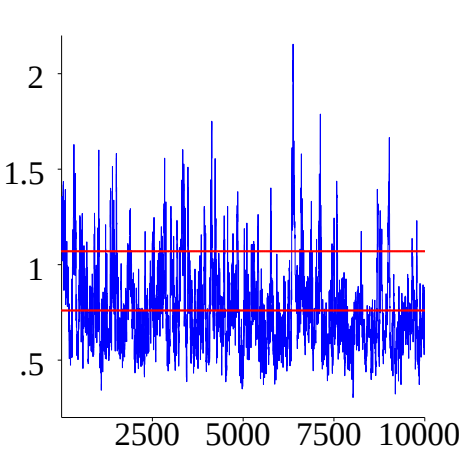
(b) ψ_2



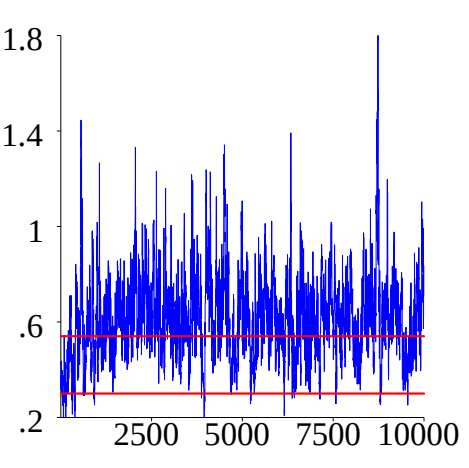
(c) ρ_r



(d) ρ_z



(e) σ_g



(f) σ_z

Conclusion

- Has opened up the possibility of fitting even larger DSGE models than those currently being fit
- Approach can be applied to Bayesian problems in general
- For example, we have applied it successfully to a 168 dimensional theory-driven yield curve model with multiple change points
- Other applications are ongoing