Bayesian Estimation of DSGE Models[1] Chapter 4: Metropolis-Hastings Algorithms for DSGE Models

Ed Herbst[1] Frank Schorfheide[2]

[1]Federal Reserve Board

[2]University of Pennsylvania

February 25, 2016

[1]The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Board of Governors or the Federal Reserve System.

Posterior Inference

- We discussed how to solve a DSGE model;
- and how to compute the likelihood function $p(Y|\theta)$ for a DSGE model.
- According to Bayes Theorem

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int p(Y|\theta)p(\theta)d\theta}$$

- We want to generate draws from posterior...

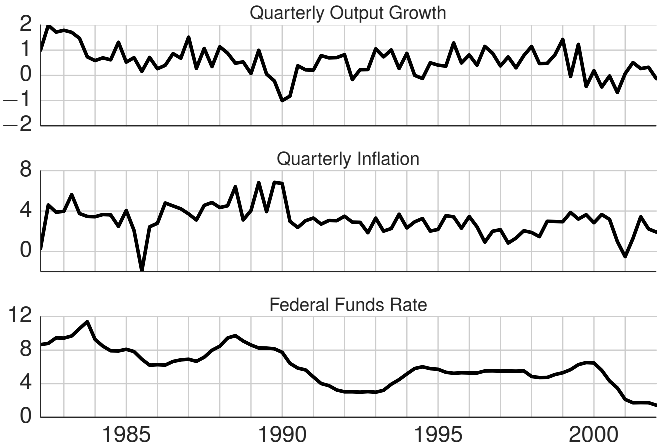
Benchmark Random-Walk Metropolis (RWMH) Algorithm for DSGE Model

1. Use a numerical optimization routine to maximize the log posterior, which up to a constant is given by $\ln p(Y|\theta) + \ln p(\theta)$. Denote the posterior mode by $\hat{\theta}$.
2. Let $\hat{\Sigma}$ be the inverse of the (negative) Hessian computed at the posterior mode $\hat{\theta}$, which can be computed numerically.
3. Draw θ^0 from $N(\hat{\theta}, c_0^2\hat{\Sigma})$ or directly specify a starting value.
4. For $i = 1, \dots, N$:
   - Draw ϑ from the proposal distribution $N(\theta^{i-1}, c^2\hat{\Sigma})$.
   - Let
   $$r(\theta^{i-1}, \vartheta|Y) = \frac{p(Y|\vartheta)p(\vartheta)}{p(Y|\theta^{i-1})p(\theta^{i-1})}.$$
   - Let
   $$\theta^i = \begin{cases} \vartheta & \text{with probability } \min\{1, r(\theta^{i-1}, \vartheta|Y)\} \\ \theta^{i-1} & \text{otherwise} \end{cases}$$

Benchmark Random-Walk Metropolis (RWMH) Algorithm for DSGE Model

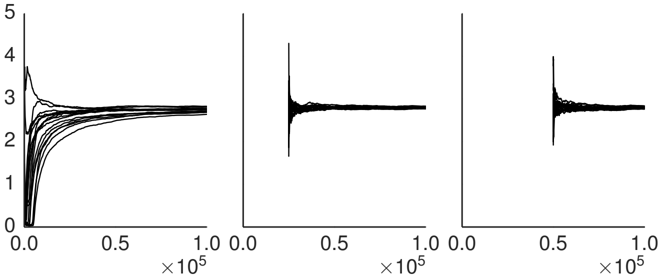
- Steps 1 and 2 are optional.
- If numerical optimization does not work well, one could let $\tilde{\Sigma}$ be a diagonal matrix with prior variances on the diagonal.
- Or, $\tilde{\Sigma}$ could be based on a preliminary run of a posterior sampler.
- It is good practice to run multiple chains based on different starting values.
- For the subsequent illustrations we chose $\tilde{\Sigma} = \mathbb{V}_\pi[h]$, where the posterior variance matrix is obtained from a long MCMC run.

Observables for Small-Scale New Keynesian Model



Notes: Output growth per capita is measured in quarter-on-quarter (Q-o-Q) percentages. Inflation is CPI inflation in annualized Q-o-Q percentages. Federal funds rate is the average annualized effective funds rate for each quarter.

Convergence of Monte Carlo Average $\bar{\tau}_{N|N_0}$



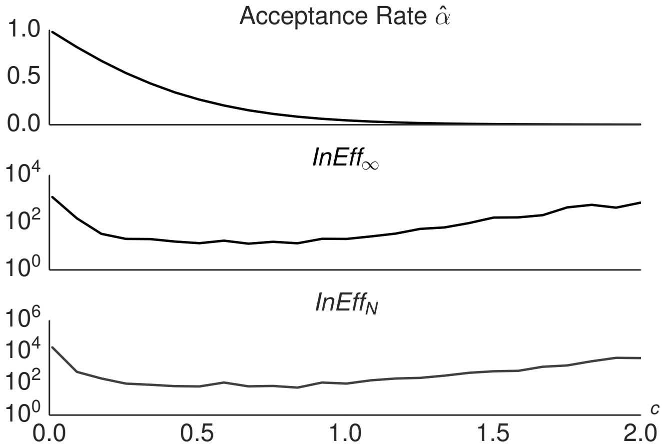
Notes: The $x-$axis indicates the number of draws N.

Posterior Estimates of DSGE Model Parameters

Parameter	Mean	[0.05, 0.95]	Parameter	Mean	[0.05,0.95]
τ	2.83	[1.95, 3.82]	ρ_r	0.77	[0.71, 0.82]
κ	0.78	[0.51, 0.98]	ρ_g	0.98	[0.96, 1.00]
ψ_1	1.80	[1.43, 2.20]	ρ_z	0.88	[0.84, 0.92]
ψ_2	0.63	[0.23, 1.21]	σ_r	0.22	[0.18, 0.26]
$r^{(A)}$	0.42	[0.04, 0.95]	σ_g	0.71	[0.61, 0.84]
$\pi^{(A)}$	3.30	[2.78, 3.80]	σ_z	0.31	[0.26, 0.36]
$\gamma^{(Q)}$	0.52	[0.28, 0.74]			

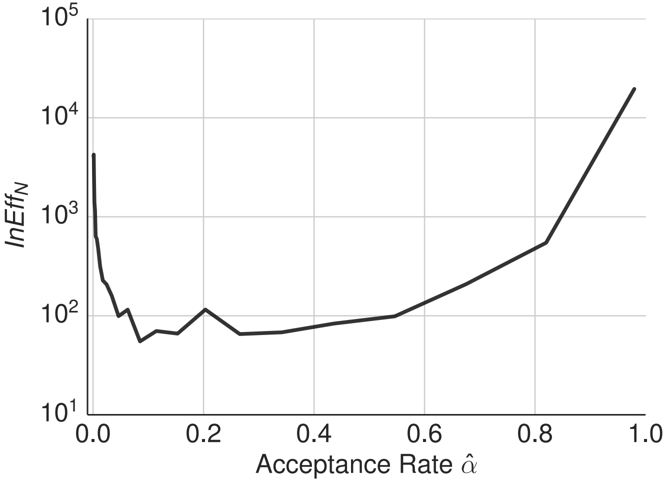
Notes: We generated $N = 100,000$ draws from the posterior and discarded the first 50,000 draws. Based on the remaining draws we approximated the posterior mean and the 5th and 95th percentiles.

Effect of Scaling Constant c



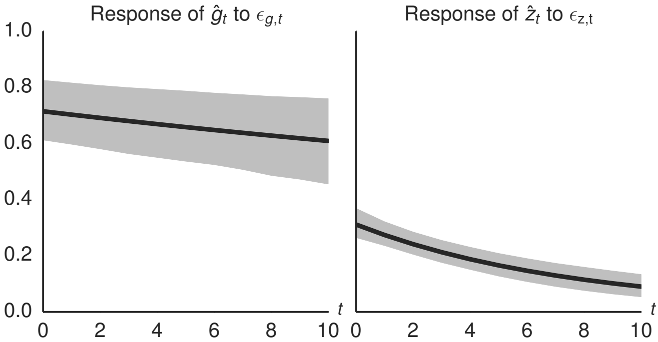
Notes: Results are based on $N_{run} = 50$ independent Markov chains. The acceptance rate (average across multiple chains), HAC-based estimate of $\text{InEff}_\infty[\bar{\tau}]$ (average across multiple chains), and $\text{InEff}_N[\bar{\tau}]$ are shown as a function of the scaling constant c.

Acceptance Rate $\hat{\alpha}$ versus Inaccuracy InEff_N



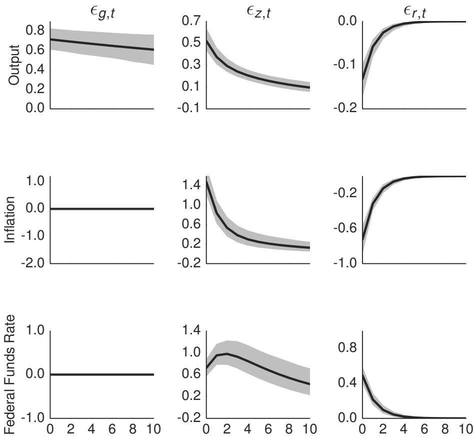
Notes: $\text{InEff}_N[\bar{\tau}]$ versus the acceptance rate $\hat{\alpha}$.

Impulse Responses of Exogenous Processes



Notes: The figure depicts pointwise posterior means and 90% credible bands. The responses are in percent relative to the initial level.

Parameter Transformations: Impulse Responses



Notes: The figure depicts pointwise posterior means and 90% credible bands. The responses of output are in percent relative to the initial level, whereas the responses of inflation and interest rates are in annualized percentages.

Challenges Due to Irregular Posteriors

- A stylized state-space model:

$$y_t = [1\ 1]s_t, \quad s_t = \left[\begin{array}{cc} \phi_1 & 0 \\ \phi_3 & \phi_2 \end{array} \right] s_{t-1} + \left[\begin{array}{c} 1 \\ 0 \end{array} \right] \epsilon_t, \quad \epsilon_t \sim iidN(0,1).$$

  where

- Structural parameters $\theta = [\theta_1, \theta_2]'$, domain is unit square.
- Reduced-form parameters $\phi = [\phi_1, \phi_2, \phi_3]'$

$$\phi_1 = \theta_1^2, \quad \phi_2 = (1 - \theta_1^2), \quad \phi_3 - \phi_2 = -\theta_1\theta_2.$$

Challenges Due to Irregular Posteriors

- $s_{1,t}$ looks like an exogenous technology process.
- $s_{2,t}$ evolves like an endogenous state variable, e.g., the capital stock.

- θ_2 is not identifiable if $\theta_1 = 0$ because θ_2 enters the model only multiplicatively.

- Law of motion of y_t is restricted ARMA(2,1) process:

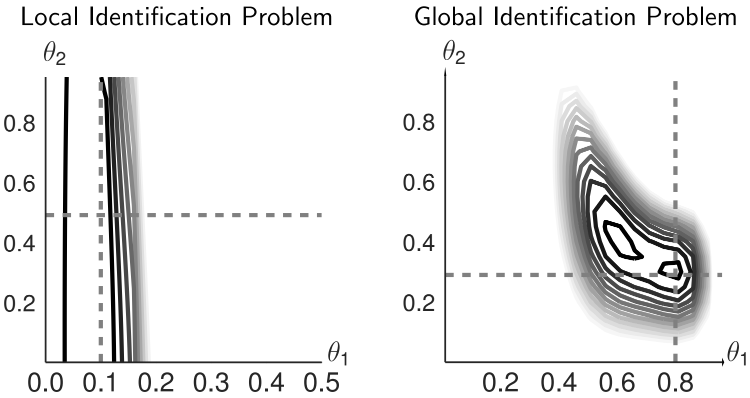
$$\big(1 - \theta_1^2 L\big)\big(1 - (1 - \theta_1^2)L\big) y_t = \big(1 - \theta_1 \theta_2 L\big)\epsilon_t.$$

- Given θ_1 and θ_2, we obtain an observationally equivalent process by switching the values of the two roots of the autoregressive lag polynomial.

- Choose $\tilde{\theta}_1$ and $\tilde{\theta}_2$ such that

$$\tilde{\theta}_1 = \sqrt{1 - \theta_1^2}, \quad \tilde{\theta}_2 = \theta_1 \theta_2 / \tilde{\theta}_1.$$

Posteriors for Stylized State-Space Model



Notes: Intersections of the solid lines indicate parameter values that were used to generate the data from which the posteriors are constructed. Left panel: $\theta_1 = 0.1$ and $\theta_2 = 0.5$. Right panel: $\theta_1 = 0.8$, $\theta_2 = 0.3$.

Improvements to MCMC: Blocking

- In high-dimensional parameter spaces the RWM algorithm generates highly persistent Markov chains.

- What's bad about persistence?

$$\sqrt{n}(\bar{X} - \mathbb{E}[\bar{X}]) \Longrightarrow N\left(0, \frac{1}{n}\sum_{i=1}^{n}\mathbb{V}[X_i] + \frac{1}{n}\sum_{i=1}^{n}\sum_{j \neq i} COV(X_i, X_j)\right).$$

- Potential Remedy:
  - Partition $\theta = [\theta_1, \ldots, \theta_K]$.
  - Iterate over conditional posteriors $p(\theta_k | Y, \theta_{<-k>})$.

- To reduce persistence of the chain, try to find partitions such that parameters are strongly correlated within blocks and weakly correlated across blocks or use random blocking.

Block MH Algorithm

Draw $\theta^0 \in \Theta$ and then for $i = 1$ to N:

1. Create a partition B^i of the parameter vector into N_{blocks} blocks $\theta_1, \dots, \theta_{N_{blocks}}$ via some rule (perhaps probabilistic), unrelated to the current state of the Markov chain.
2. For $b = 1, \dots, N_{blocks}$:
   1. Draw $\vartheta_b \sim q(\cdot| \left[\theta^i_{<b}, \theta^{i-1}_b, \theta^{i-1}_{\geq b}\right])$.
   2. With probability,

$$\alpha = \max\left\{ \frac{p(\left[\theta^i_{<b}, \vartheta_b, \theta^{i-1}_{>b}\right]|Y)q(\theta^{i-1}_b, |\theta^i_{<b}, \vartheta_b, \theta^{i-1}_{>b})}{p(\theta^i_{<b}, \theta^{i-1}_b, \theta^{i-1}_{>b}|Y)q(\vartheta_b|\theta^i_{<b}, \theta^{i-1}_b, \theta^{i-1}_{>b})}, 1 \right\},$$

set $\theta^i_b = \vartheta_b$, otherwise set $\theta^i_b = \theta^{i-1}_b$.

Random-Block MH Algorithm

1. Generate a sequence of random partitions $\{B^i\}_{i=1}^N$ of the parameter vector θ into N_{blocks} equally sized blocks, denoted by θ_b, $b = 1, \ldots, N_{blocks}$ as follows:
    1. assign an $iidU[0,1]$ draw to each element of θ;
    2. sort the parameters according to the assigned random number;
    3. let the b'th block consist of parameters $(b-1)N_{blocks}, \ldots, bN_{blocks}$.[2]
2. Execute Algorithm Block MH Algorithm.

[2] If the number of parameters is not divisible by N_{blocks}, then the size of a subset of the blocks has to be adjusted.

Metropolis-Adjusted Langevin Algorithm

- The proposal distribution of Metropolis-Adjusted Langevin (MAL) algorithm is given by

$$\mu(\theta^{i-1}) = \theta^{i-1} + \frac{c_1}{2} M_1 \frac{\partial}{\partial \theta} \ln p(\theta^{i-1}|Y)\Big|_{\theta=\theta^{i-1}},$$
$$\Sigma(\theta^{i-1}) = c_2^2 M_2.$$

  that is θ^{i-1} is adjusted by a step in the direction of the gradient of the log posterior density function.

- One standard practice is to set $M_1 = M_2 = M$, with

$$M = -\left[\frac{\partial}{\partial \theta \partial \theta'} \ln p(\theta|Y)\Big|_{\theta=\hat{\theta}} \right]^{-1},$$

  where $\hat{\theta}$ is the mode of the posterior distribution obtained using a numerical optimization routine.

Newton MH Algorithm

- Newton MH Algorithm replaces the Hessian evaluated at the posterior mode $\hat{\theta}$ by the Hessian evaluated at θ^{i-1}.

- The proposal distribution is given by

$$
\begin{aligned}
\mu(\theta^{i-1}) &= \theta^{i-1} - s \left[\frac{\partial}{\partial\theta\partial\theta'} \ln p(\theta|Y)\bigg|_{\theta=\theta^{i-1}} \right]^{-1} \\
&\quad \times \frac{\partial}{\partial\theta} \ln p(\theta^{i-1}|Y)\bigg|_{\theta=\theta^{i-1}} \\
\hat{\Sigma}(\theta^{i-1}) &= -c_2^2 \left[\frac{\partial}{\partial\theta\partial\theta'} \ln p(\theta|Y)\bigg|_{\theta=\theta^{i-1}} \right]^{-1}.
\end{aligned}
$$

- It is useful to let s be independently of θ^{i-1}:

$$
c_1 = 2s, \quad s \sim iidU[0, \bar{s}],
$$

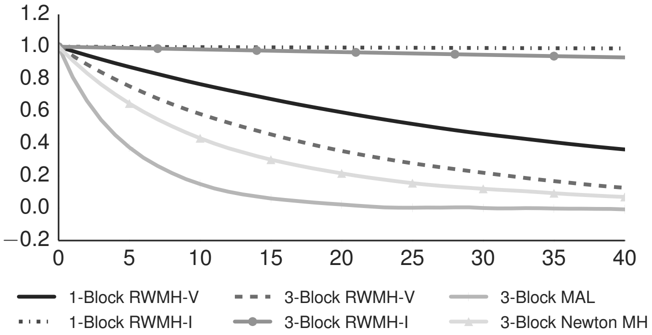
where $\bar{s}$ is a tuning parameter.

Run Times and Tuning Constants for MH Algorithms

Algorithm	Run Time [hh:mm:ss]	Acceptance Rate	Tuning Constants
1-Block RWMH-I	00:01:13	0.28	$c = 0.015$
1-Block RWMH-V	00:01:13	0.37	$c = 0.400$
3-Block RWMH-I	00:03:38	0.40	$c = 0.070$
3-Block RWMH-V	00:03:36	0.43	$c = 1.200$
3-Block MAL	00:54:12	0.43	$c_1 = 0.400, c_2 = 0.750$
3-Block Newton MH	03:01:40	0.53	$\bar{s} = 0.700, c_2 = 0.600$

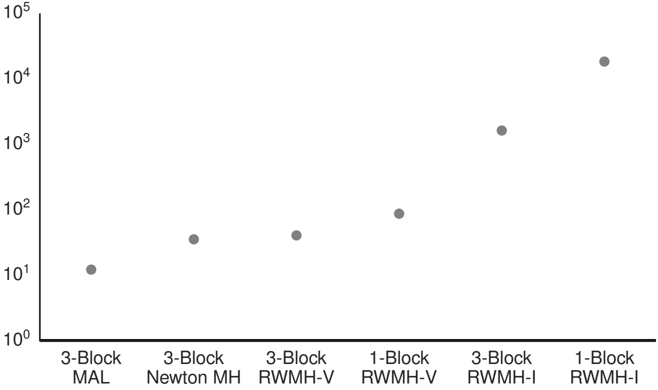
Notes: In each run we generate $N = 100,000$ draws. We report the fastest run time and the average acceptance rate across $N_{run} = 50$ independent Markov chains.

Autocorrelation Function of τ^i



Notes: The autocorrelation functions are computed based on a single run of each algorithm.

Inefficiency Factor $\text{InEff}_N[\bar{\tau}]$



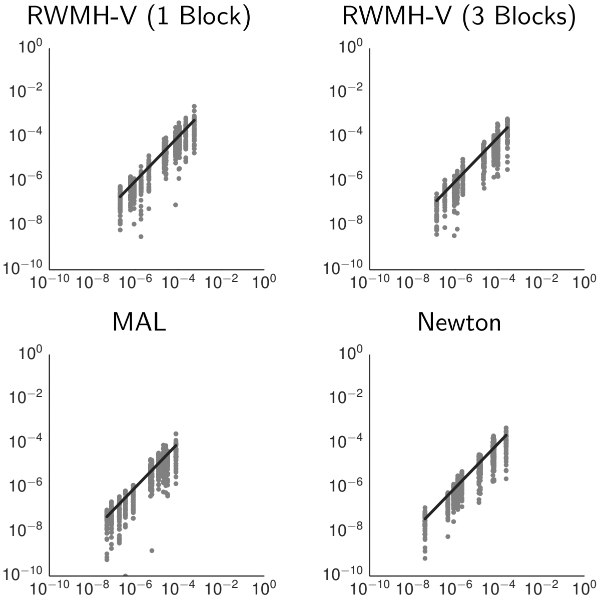
Notes: The small-sample inefficiency factors are computed based on $N_{run} = 50$ independent runs of each algorithm.

IID Equivalent Draws Per Second

$$\textit{iid}\text{-equivalent draws per second} = \frac{N}{\text{Run Time [seconds]}} \cdot \frac{1}{\text{InEff}_N}.$$

- 3-Block MAL: 1.24
- 3-Block Newton MH: 0.13
- 3-Block RWMH-V: 5.65
- 1-Block RWMH-V: 7.76
- 3-Block RWMH-I: 0.14
- 1-Block RWMH-I: 0.04

Performance of Different MH Algorithms



Recall: Posterior Odds and Marginal Data Densities

- Posterior model probabilities can be computed as follows:

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y|\mathcal{M}_i)}{\sum_j \pi_{j,0} p(Y|\mathcal{M}_j)}, \quad j = 1, \dots, 2, \tag{1}$$

- where

$$p(Y|\mathcal{M}) = \int p(Y|\theta, \mathcal{M}) p(\theta|\mathcal{M}) d\theta \tag{2}$$

- Note:

$$\ln p(Y_{1:T}|\mathcal{M}) = \sum_{t=1}^{T} \ln \int p(y_t|\theta, Y_{1:t-1}, \mathcal{M}) p(\theta|Y_{1:t-1}, \mathcal{M}) d\theta$$

- Posterior odds and Bayes Factor

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \underbrace{\frac{\pi_{1,0}}{\pi_{2,0}}}_{\textit{Prior Odds}} \times \underbrace{\frac{p(Y|\mathcal{M}_1)}{p(Y|\mathcal{M}_2)}}_{\textit{Bayes Factor}} \tag{3}$$

Computation of Marginal Data Densities

- Reciprocal importance sampling:
  - Geweke's modified harmonic mean estimator
  - Sims, Waggoner, and Zha's estimator
- Chib and Jeliazkov's estimator
- For a survey, see Ardia, Hoogerheide, and van Dijk (2009).

Modified Harmonic Mean

- Reciprocal importance samplers are based on the following identity:

$$\frac{1}{p(Y)} = \int \frac{f(\theta)}{p(Y|\theta)p(\theta)} p(\theta|Y)d\theta, \tag{4}$$

where $\int f(\theta)d\theta = 1$.

- Conditional on the choice of $f(\theta)$ an obvious estimator is

$$\hat{p}_G(Y) = \left[\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \frac{f(\theta^{(s)})}{p(Y|\theta^{(s)})p(\theta^{(s)})} \right]^{-1}, \tag{5}$$

where $\theta^{(s)}$ is drawn from the posterior $p(\theta|Y)$.

- Geweke (1999):

$$\begin{aligned} f(\theta) &= \tau^{-1}(2\pi)^{-d/2}|V_\theta|^{-1/2} \exp\left[-0.5(\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta})\right] \\ &\times \left\{ (\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta}) \leq F_{\chi^2_d}^{-1}(\tau) \right\}. \end{aligned} \tag{6}$$

Chib and Jeliazkov

- Rewrite Bayes Theorem:

$$p(Y) = \frac{p(Y|\theta)p(\theta)}{p(\theta|Y)}. \tag{7}$$

- Thus,

$$\hat{p}_{CS}(Y) = \frac{p(Y|\tilde{\theta})p(\tilde{\theta})}{\hat{p}(\tilde{\theta}|Y)}, \tag{8}$$

where we replaced the generic θ in (7) by the posterior mode $\tilde{\theta}$.

Chib and Jeliazkov

- Use output of Metropolis-Hastings Algorithm.
- Proposal density for transition $\theta \mapsto \tilde{\theta}$: $q(\theta, \tilde{\theta}|Y)$.
- Probability of accepting proposed draw:

$$\alpha(\theta, \tilde{\theta}|Y) = \min \left\{ 1, \ \frac{p(\tilde{\theta}|Y)/q(\theta, \tilde{\theta}|Y)}{p(\theta|Y)/q(\tilde{\theta}, \theta|Y)} \right\}.$$

- Note that

$$\begin{aligned} & \int \alpha(\theta, \tilde{\theta}|Y) q(\theta, \tilde{\theta}|Y) p(\theta|Y) d\theta \\ &= \int \min \left\{ 1, \ \frac{p(\tilde{\theta}|Y)/q(\theta, \tilde{\theta}|Y)}{p(\theta|Y)/q(\tilde{\theta}, \theta|Y)} \right\} q(\theta, \tilde{\theta}|Y) p(\theta|Y) d\theta \\ &= p(\tilde{\theta}|Y) \int \min \left\{ \frac{p(\theta|Y)/q(\tilde{\theta}, \theta|Y)}{p(\tilde{\theta}|Y)/q(\theta, \tilde{\theta}|Y)}, \ 1 \right\} q(\tilde{\theta}, \theta|Y) d\theta \\ &= p(\tilde{\theta}|Y) \int \alpha(\tilde{\theta}, \theta|Y) q(\tilde{\theta}, \theta|Y) d\theta \end{aligned}$$

Chib and Jeliazkov

- Posterior density at the mode can be approximated as follows

$$\hat{p}(\tilde{\theta}|Y) = \frac{\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \alpha(\theta^{(s)}, \tilde{\theta}|Y) q(\theta^{(s)}, \tilde{\theta}|Y)}{J^{-1} \sum_{j=1}^{J} \alpha(\tilde{\theta}, \theta^{(j)}|Y)}, \tag{9}$$

- $\{\theta^{(s)}\}$ are posterior draws obtained with the the M-H Algorithm;
- $\{\theta^{(j)}\}$ are additional draws from $q(\tilde{\theta}, \theta|Y)$ given the fixed value $\tilde{\theta}$.

MH-Based Marginal Data Density Estimates

Model	Mean(ln $\hat{p}(Y)$)	Std. Dev.(ln $\hat{p}(Y)$)
Geweke ($\tau = 0.5$)	-346.17	0.03
Geweke ($\tau = 0.9$)	-346.10	0.04
SWZ ($q = 0.5$)	-346.29	0.03
SWZ ($q = 0.9$)	-346.31	0.02
Chib and Jeliazkov	-346.20	0.40

Notes: Table shows mean and standard deviation of log marginal data density estimators, computed over $N_{run} = 50$ runs of the RWMH-V sampler using $N = 100,000$ draws, discarding a burn-in sample of $N_0 = 50,000$ draws. The SWZ estimator uses $J = 100,000$ draws to compute $\hat{\tau}$, while the CJ estimators uses $J = 100,000$ to compute the denominator of $\hat{p}(\tilde{\theta}|Y)$.