Introduction to Bayesian Estimation of DSGE Models

Frank Schorfheide
University of Pennsylvania

Bradley Visitor Lectures

October 23, 2017

Small-Scale DSGE Model

- Textbook treatments: Woodford (2003), Gali (2008)
- Intermediate and final goods producers
- Households
- Monetary and fiscal policy
- Exogenous processes
- Equilibrium Relationships

Equilibrium Conditions

- Market clearing:

$$Y_t = C_t + G_t + AC_t \quad \text{and} \quad H_t = N_t.$$

- Aggregate Resource constraint:

$$Y_t = A_t N_t.$$

- Consumption Euler equation and New Keynesian Phillips curve:

$$\begin{aligned}
1 &= \beta \mathbb{E}_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{A_t}{A_{t+1}} \frac{R_t}{\pi_{t+1}} \right] \\
1 &= \phi(\pi_t - \pi) \left[\left(1 - \frac{1}{2\nu} \right) \pi_t + \frac{\pi}{2\nu} \right] \\
&\quad - \phi \beta \mathbb{E}_t \left[\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t} \right)^{-\tau} \frac{Y_{t+1}/A_{t+1}}{Y_t/A_t} (\pi_{t+1} - \pi) \pi_{t+1} \right] \\
&\quad + \frac{1}{\nu} \left[1 - \left(\frac{C_t}{A_t} \right)^{\tau} \right].
\end{aligned}$$

Equilibrium Conditions – Continued

- Central bank adjusts money supply to attain desired interest rate.
- Monetary policy rule:

$$R_t \quad = \quad {R_t^*}^{1-\rho_R} R_{t-1}^{\rho_R} e^{\epsilon_{R,t}}, \quad R_t^* = r\pi^* \left(\frac{\pi_t}{\pi^*}\right)^{\psi_1} \left(\frac{Y_t}{Y_t^*}\right)^{\psi_2}$$

- Fiscal authority consumes fraction of aggregate output: $G_t = \zeta_t Y_t$.
- In the absence of nominal rigidities ($\phi = 0$) aggregate output is given by

$$Y_t^* = (1-\nu)^{1/\tau} A_t g_t,$$

  which is the target level of output that appears in the monetary policy rule.
- Technology:

$$\ln A_t = \ln \gamma + \ln A_{t-1} + \ln z_t, \quad \ln z_t = \rho_z \ln z_{t-1} + \epsilon_{z,t}.$$

- Government spending / aggregate demand: define $g_t = 1/(1-\zeta_t)$; assume

$$\ln g_t = (1-\rho_g) \ln g + \rho_g \ln g_{t-1} + \epsilon_{g,t}.$$

- Monetary policy shock $\epsilon_{R,t}$ is assumed to be serially uncorrelated.

Solving DSGE Models

- Derive nonlinear equilibrium conditions:
  - System of nonlinear expectational difference equations;
  - transversality conditions.
- Find solution(s) of system of expectational difference methods:
  - Global (nonlinear) approximation
  - Local approximation near steady state
- More detail in: Fernandez-Villaverde, Rubio-Ramirez, and Schorfheide (2016) "Solution and Estimation Methods for DSGE Models," prepared for forthcoming *Handbook of Macroeconomics*.

Loglinearization of New Keynesian Model

- Consumption Euler equation:

$$\hat{y}_t = \mathbb{E}_t[\hat{y}_{t+1}] - \frac{1}{\tau}\left(\hat{R}_t - \mathbb{E}_t[\hat{\pi}_{t+1}] - \mathbb{E}_t[\hat{z}_{t+1}]\right) + \hat{g}_t - \mathbb{E}_t[\hat{g}_{t+1}]$$

- New Keynesian Phillips curve:

$$\hat{\pi}_t = \beta\mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa(\hat{y}_t - \hat{g}_t),$$

where

$$\kappa = \tau\frac{1-\nu}{\nu\pi^2\phi}$$

- Monetary policy rule:

$$\hat{R}_t = \rho_R\hat{R}_{t-1} + (1-\rho_R)\psi_1\hat{\pi}_t + (1-\rho_R)\psi_2\left(\hat{y}_t - \hat{g}_t\right) + \epsilon_{R,t}$$

Solving the Linearized DSGE Model

- Linearized DSGE leads to linear rational expectations (LRE) system:

$$\Gamma_0(\theta)s_t = \Gamma_1(\theta)s_{t-1} + \Psi\epsilon_t + \Pi\eta_t \tag{1}$$

  where
  - s_t is a vector of model variables, ϵ_t is a vector of exogenous shocks,
  - η_t is a vector of RE errors with elements $\eta_t^x = \hat{x}_t - \mathbb{E}_{t-1}[\hat{x}_t]$, and
  - s_t contains (among others) the conditional expectation terms $\mathbb{E}_t[\widetilde{x}_{t+1}]$.
- General solution methods for LREs: Blanchard and Kahn (1980), King and Watson (1998), Uhlig (1999), Anderson (2000), Klein (2000), Christiano (2002), Sims (2002).
- Overall the solution in terms of s_t is of the form

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{2}$$

Measurement Equation

- Relate model variables s_t to observables y_t.
- In NK model:

$$\begin{aligned} YGR_t &= \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t) \\ INFL_t &= \pi^{(A)} + 400\hat{\pi}_t \\ INT_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t. \end{aligned}$$

where

$$\gamma = 1 + \frac{\gamma^{(Q)}}{100}, \quad \beta = \frac{1}{1 + r^{(A)}/400}, \quad \pi = 1 + \frac{\pi^{(A)}}{400}.$$

- More generically:

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t \underbrace{+u_t}_{\text{optional}} .$$

Where Are We Headed? ... Bayesian Inference

- DSGE model parameters θ.
- Likelihood function $p(Y|\theta)$; requires that the DSGE model is solved numerically given θ.
- Bayes Theorem

$$p(\theta|Y) \propto p(Y|\theta)p(\theta).$$

State-Space Representation and Likelihood

- Measurement:

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t \underbrace{+u_t}_{\text{optional}}$$

- State transition:

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t$$

- Joint density for the observations and latent states:

$$\begin{aligned} p(Y_{1:T}, S_{1:T}|\theta) &= \prod_{t=1}^{T} p(y_t, s_t | Y_{1:t-1}, S_{1:t-1}, \theta) \\ &= \prod_{t=1}^{T} p(y_t|s_t, \theta)p(s_t|s_{t-1}, \theta). \end{aligned}$$

- Problem: we need the marginal $p(Y_{1:T}|\theta)$.

Filtering - General Idea

- State-space representation of DSGE model

$$\begin{aligned} y_t &= \Psi(s_t, t; \theta) \underbrace{+ u_t, \quad u_t \sim F_u(\cdot;\theta)}_{\text{optional}} \quad \text{measurement} \\ s_t &= \Phi(s_{t-1}, \epsilon_t; \theta), \quad \epsilon_t \sim F_\epsilon(\cdot;\theta) \quad \text{state transition} \end{aligned}$$

- Likelihood function:

$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1}, \theta)$$

- A filter generates a sequence of conditional distributions $s_t|Y_{1:t}$.
- Iterations:
  - Initialization at time $t-1$: $p(s_{t-1}|Y_{1:t-1}, \theta)$
  - Forecasting t given $t-1$:
    1. Transition equation: $p(s_t|Y_{1:t-1}, \theta) = \int p(s_t|s_{t-1}, Y_{1:t-1}, \theta) p(s_{t-1}|Y_{1:t-1}, \theta) ds_{t-1}$
    2. Measurement equation: $p(y_t|Y_{1:t-1}, \theta) = \int p(y_t|s_t, Y_{1:t-1}, \theta) p(s_t|Y_{1:t-1}, \theta) ds_t$
  - Updating with Bayes theorem. Once y_t becomes available:

$$p(s_t|Y_{1:t}, \theta) = p(s_t|y_t, Y_{1:t-1}, \theta) = \frac{p(y_t|s_t, Y_{1:t-1}, \theta) p(s_t|Y_{1:t-1}, \theta)}{p(y_t|Y_{1:t-1}, \theta)}$$

Conditional Distributions for Kalman Filter (Linear Gaussian State-Space Model)

	Distribution	Mean and Variance
$s_{t-1} \| (Y_{1:t-1}, \theta)$	$N\left(\bar{s}_{t-1\|t-1}, P_{t-1\|t-1}\right)$	Given from Iteration $t-1$
$s_t \| (Y_{1:t-1}, \theta)$	$N\left(\bar{s}_{t\|t-1}, P_{t\|t-1}\right)$	$\bar{s}_{t\|t-1} = \Phi_1 \bar{s}_{t-1\|t-1}$ $P_{t\|t-1} = \Phi_1 P_{t-1\|t-1} \Phi_1' + \Phi_\epsilon \Sigma_\epsilon \Phi_\epsilon'$
$y_t \| (Y_{1:t-1}, \theta)$	$N\left(\bar{y}_{t\|t-1}, F_{t\|t-1}\right)$	$\bar{y}_{t\|t-1} = \Psi_0 + \Psi_1 t + \Psi_2 \bar{s}_{t\|t-1}$ $F_{t\|t-1} = \Psi_2 P_{t\|t-1} \Psi_2' + \Sigma_u$
$s_t \| (Y_{1:t}, \theta)$	$N\left(\bar{s}_{t\|t}, P_{t\|t}\right)$	$\bar{s}_{t\|t} = \bar{s}_{t\|t-1} + P_{t\|t-1} \Psi_2' F_{t\|t-1}^{-1} (y_t - \bar{y}_{t\|t-1})$ $P_{t\|t} = P_{t\|t-1} - P_{t\|t-1} \Psi_2' F_{t\|t-1}^{-1} \Psi_2 P_{t\|t-1}$

Prior Elicitation

- Group parameters:
  - steady-state related parameters
  - parameters assoc with exogenous shocks
  - parameters assoc with internal propagation
- Non-sample information $p(\theta|\mathcal{X}^0)$:
  - pre-sample information
  - micro-level information
- To guide the prior for θ, you can ask: what are its implications for observables Y?

Bayesian Estimation – Prior

Name	Domain	Prior		
		Density	Para (1)	Para (2)
Steady State Related Parameters $\theta_{(ss)}$				
$r^{(A)}$	$\mathbb{R}^+$	Gamma	0.50	0.50
$\pi^{(A)}$	$\mathbb{R}^+$	Gamma	7.00	2.00
$\gamma^{(Q)}$	$\mathbb{R}$	Normal	0.40	0.20
Endogenous Propagation Parameters $\theta_{(endo)}$				
τ	$\mathbb{R}^+$	Gamma	2.00	0.50
κ	$[0,1]$	Uniform	0.00	1.00
ψ_1	$\mathbb{R}^+$	Gamma	1.50	0.25
ψ_2	$\mathbb{R}^+$	Gamma	0.50	0.25
ρ_R	$[0,1)$	Uniform	0.00	1.00
Exogenous Shock Parameters $\theta_{(exo)}$				
ρ_G	$[0,1)$	Uniform	0.00	1.00
ρ_Z	$[0,1)$	Uniform	0.00	1.00
$100\sigma_R$	$\mathbb{R}^+$	InvGamma	0.40	4.00
$100\sigma_G$	$\mathbb{R}^+$	InvGamma	1.00	4.00
$100\sigma_Z$	$\mathbb{R}^+$	InvGamma	0.50	4.00

Bayesian Inference

- Ingredients of Bayesian Analysis:
  - Likelihood function $p(Y|\theta)$
  - Prior density $p(\theta)$
  - Marginal data density $p(Y) = \int p(Y|\theta)p(\theta)d\phi$
- Bayes Theorem:

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

- Implementation: usually by generating a sequence of draws (not necessarily iid) from posterior

$$\theta^i \sim p(\theta|Y), \quad i = 1, \ldots, N$$

- Algorithms: direct sampling, accept/reject sampling, importance sampling, Markov chain Monte Carlo sampling, sequential Monte Carlo sampling...

Draws from Posterior

- We will focus on MCMC and SMC algorithms that generate draws $\{\theta^i\}_{i=1}^N$ from posterior distributions of parameters.
- Draws can then be transformed into objects of interest, $h(\theta^i)$, and under suitable conditions a Monte Carlo average of the form

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^{N} h(\theta^i) \approx \mathbb{E}_\pi[h] \int h(\theta)p(\theta|Y)d\theta.$$

- Strong law of large numbers (SLLN), central limit theorem (CLT)...

Bayesian Inference – Decision Making

- The posterior expected loss of decision $\delta(\cdot)$:

$$\rho\big(\delta(\cdot)|Y\big) = \int_\Theta L\big(\theta, \delta(Y)\big)p(\theta|Y)d\theta.$$

- Bayes decision minimizes the posterior expected loss:

$$\delta^*(Y) = \text{argmin}_d \; \rho\big(\delta(\cdot)|Y\big).$$

- Approximate $\rho\big(\delta(\cdot)|Y\big)$ by a Monte Carlo average

$$\bar{\rho}_N\big(\delta(\cdot)|Y\big) = \frac{1}{N}\sum_{i=1}^{N} L\big(\theta^i, \delta(\cdot)\big).$$

- Then compute

$$\delta^*_N(Y) = \text{argmin}_d \; \bar{\rho}_N\big(\delta(\cdot)|Y\big).$$

Bayesian Inference

- Point estimation:
  - Quadratic loss: posterior mean
  - Absolute error loss: posterior median
- Interval/Set estimation $\mathbb{P}_\pi\{\theta \in C(Y)\} = 1 - \alpha$:
  - highest posterior density sets
  - equal-tail-probability intervals

Model Uncertainty

- Assign prior probabilities $\gamma_{j,0}$ to models M_j, $j = 1, \ldots, J$.
- Posterior model probabilities are given by

$$\gamma_{j,T} = \frac{\gamma_{j,0} p(Y|M_j)}{\sum_{j=1}^{J} \gamma_{j,0} p(Y|M_j)},$$

where

$$p(Y|M_j) = \int p(Y|\theta_{(j)}, M_j) p(\theta_{(j)}|M_j) d\theta_{(j)}$$

- Log marginal data densities are one-step-ahead predictive scores:

$$\begin{aligned} &\ln p(Y|M_j) \\ &= \sum_{t=1}^{T} \ln \int p(y_t|\theta_{(j)}, Y_{1:t-1}, M_j) p(\theta_{(j)}|Y_{1:t-1}, M_j) d\theta_{(j)}. \end{aligned}$$

- Model averaging:

$$p(h|Y) = \sum_{j=1}^{J} \gamma_{j,T} p(h_j(\theta_{(j)})|Y, M_j).$$

Markov Chain Monte Carlo (MCMC)

- Main idea: create a sequence of serially correlated draws such that the distribution of θ^i converges to the posterior distribution $p(\theta|Y)$.

Generic Metropolis-Hastings Algorithm

For $i = 1$ to N:

1. Draw ϑ from a density $q(\vartheta|\theta^{i-1})$.
2. Set $\theta^i = \vartheta$ with probability

$$\alpha(\vartheta|\theta^{i-1}) = \min\left\{1,\ \frac{p(Y|\vartheta)p(\vartheta)/q(\vartheta|\theta^{i-1})}{p(Y|\theta^{i-1})p(\theta^{i-1})/q(\theta^{i-1}|\vartheta)}\right\}$$

and $\theta^i = \theta^{i-1}$ otherwise.

Recall $p(\theta|Y) \propto p(Y|\theta)p(\theta)$.

We draw θ^i conditional on a parameter draw θ^{i-1}: leads to Markov transition kernel $K(\theta|\tilde{\theta})$.

Importance Invariance Property

- It can be shown that

$$p(\theta|Y) = \int K(\theta|\tilde{\theta})p(\tilde{\theta}|Y)d\tilde{\theta}.$$

- Write

$$K(\theta|\tilde{\theta}) = u(\theta|\tilde{\theta}) + r(\tilde{\theta})\delta_{\tilde{\theta}}(\theta).$$

- $u(\theta|\tilde{\theta})$ is the density kernel (note that $u(\theta|\cdot)$ does not integrated to one) for accepted draws:

$$u(\theta|\tilde{\theta}) = \alpha(\theta|\tilde{\theta})q(\theta|\tilde{\theta}).$$

- Rejection probability:

$$r(\tilde{\theta}) = \int \left[1 - \alpha(\theta|\tilde{\theta})\right] q(\theta|\tilde{\theta})d\theta = 1 - \int u(\theta|\tilde{\theta})d\theta.$$

Importance Invariance Property

- Reversibility: Conditional on the sampler not rejecting the proposed draw, the density associated with a transition from $\tilde{\theta}$ to θ is identical to the density associated with a transition from θ to $\tilde{\theta}$:

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

- Using the reversibility result, we can now verify the invariance property:

$$\begin{aligned}
\int K(\theta|\tilde{\theta})p(\tilde{\theta}|Y)d\tilde{\theta} &= \int u(\theta|\tilde{\theta})p(\tilde{\theta}|Y)d\tilde{\theta} + \int r(\tilde{\theta})\delta_{\tilde{\theta}}(\theta)p(\tilde{\theta}|Y)d\tilde{\theta} \\
&= \int u(\tilde{\theta}|\theta)p(\theta|Y)d\tilde{\theta} + r(\theta)p(\theta|Y) \\
&= p(\theta|Y)
\end{aligned}$$

A Discrete Example

- Suppose parameter vector θ is scalar and takes only two values:

$$\Theta = \{\tau_1, \tau_2\}$$

- The posterior distribution $p(\theta|Y)$ can be represented by a set of probabilities collected in the vector π, say $\pi = [\pi_1, \pi_2]$ with $\pi_2 > \pi_1$.
- Suppose we obtain ϑ based on transition matrix Q:

$$Q = \begin{bmatrix} q & (1-q) \\ (1-q) & q \end{bmatrix}.$$

Discrete MH Algorithm

- Iteration i: suppose that $\theta^{i-1} = \tau_j$. Based on transition matrix

$$Q = \left[\begin{array}{cc} q & (1-q) \\ (1-q) & q \end{array} \right],$$

determine a proposed state $\vartheta = \tau_s$.

  - With probability $\alpha(\tau_s|\tau_j)$ the proposed state is accepted. Set $\theta^i = \vartheta = \tau_s$.
  - With probability $1 - \alpha(\tau_s|\tau_j)$ stay in old state and set $\theta^i = \theta^{i-1} = \tau_j$.

- Choose (Q terms cancel because of symmetry)

$$\alpha(\tau_s|\tau_j) = \min \left\{ 1, \frac{\pi_s}{\pi_j} \right\}.$$

Discrete MH Algorithm: Transition Matrix

- The resulting chain's transition matrix is:

$$K = \begin{bmatrix} q & (1-q) \\ (1-q)\frac{\pi_1}{\pi_2} & q + (1-q)\left(1 - \frac{\pi_1}{\pi_2}\right) \end{bmatrix}.$$

- Straightforward calculations reveal that the transition matrix K has eigenvalues:

$$\lambda_1(K) = 1, \quad \lambda_2(K) = q - (1-q)\frac{\pi_1}{1-\pi_1}.$$

- Equilibrium distribution is eigenvector associated with unit eigenvalue.
- For $q \in [0, 1)$ the equilibrium distribution is unique.

Convergence

- The persistence of the Markov chain depends on second eigenvalue, which depends on the proposal distribution Q.

- Define the transformed parameter

$$\xi^i = \frac{\theta^i - \tau_1}{\tau_2 - \tau_1}.$$

- We can represent the Markov chain associated with ξ^i as first-order autoregressive process

$$\xi^i = (1 - k_{22}) + \lambda_2(K)\xi^{i-1} + \nu^i.$$

- Conditional on $\xi^i = j$, $j = 0, 1$, the innovation ν^i has support on k_{jj} and $(1 - k_{jj})$, its conditional mean is equal to zero, and its conditional variance is equal to $k_{jj}(1 - k_{jj})$.

Convergence

- Autocovariance function of $h(\theta^i)$:

$$\begin{aligned} & COV(h(\theta^i), h(\theta^{(i-l)})) \\ &= \big(h(\tau_2) - h(\tau_1)\big)^2 \pi_1(1-\pi_1)\left(q - (1-q)\frac{\pi_1}{1-\pi_1}\right)^l \\ &= \mathbb{V}_\pi[h]\left(q - (1-q)\frac{\pi_1}{1-\pi_1}\right)^l \end{aligned}$$

- If $q = \pi_1$ then the autocovariances are equal to zero and the draws $h(\theta^i)$ are serially uncorrelated (in fact, in our simple discrete setting they are also independent).

Convergence

- Define the Monte Carlo estimate

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^{N} h(\theta^i).$$

- Deduce from CLT

$$\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \Longrightarrow N\big(0, \Omega(h)\big),$$

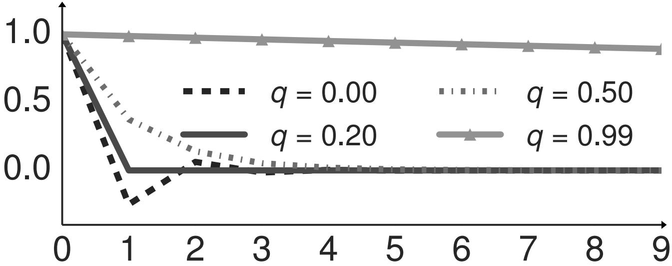
where $\Omega(h)$ is the long-run covariance matrix

$$\Omega(h) = \lim_{L \longrightarrow \infty} \mathbb{V}_\pi[h]\left(1 + 2\sum_{l=1}^{L}\frac{L-l}{L}\left(q - (1-q)\frac{\pi_1}{1-\pi_1}\right)^l\right).$$

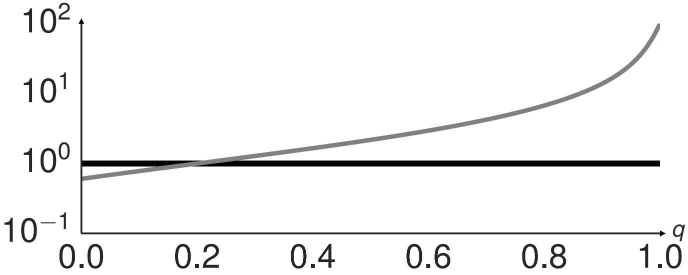
- In turn, the asymptotic inefficiency factor is given by

$$\text{InEff}_\infty = \frac{\Omega(h)}{\mathbb{V}_\pi[h]} = 1 + 2\lim_{L \longrightarrow \infty}\sum_{l=1}^{L}\frac{L-l}{L}\left(q - (1-q)\frac{\pi_1}{1-\pi_1}\right)^l.$$

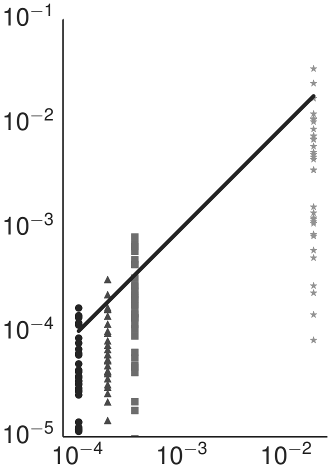
Autocorrelation Function of θ^i, $\pi_1 = 0.2$



Asymptotic Inefficiency InEff_∞, $\pi_1 = 0.2$



Small Sample Variance $\mathbb{V}[\bar{h}_N]$ across Chains versus HAC Estimates of $\Omega(h)$



Solid line is 45-degree line.

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-Hastings (RWMH) Algorithm for DSGE Models

- Initialization:
  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.

- Main Algorithm – For $i = 1, \ldots, N$:
  1. Draw ϑ from the proposal distribution $N(\theta^{i-1}, c^2\hat{\Sigma})$.
  2. Set $\theta^i = \vartheta$ with probability

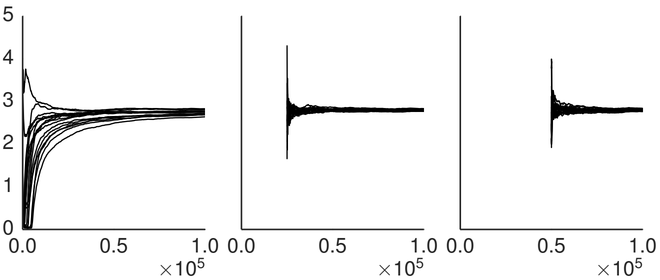
$$\alpha(\vartheta|\theta^{i-1}) = \min \left\{1, \frac{p(Y|\vartheta)p(\vartheta)}{p(Y|\theta^{i-1})p(\theta^{i-1})}\right\}$$

and $\theta^i = \theta^{i-1}$ otherwise.

Benchmark RWMH Algorithm for DSGE Models

- Initialization steps can be modified as needed for particular application.
- If numerical optimization does not work well, one could let $\hat{\Sigma}$ be a diagonal matrix with prior variances on the diagonal.
- Or, $\hat{\Sigma}$ 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 $\hat{\Sigma} = \mathbb{V}_\pi[h]$, where the posterior variance matrix is obtained from a long MCMC run.

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



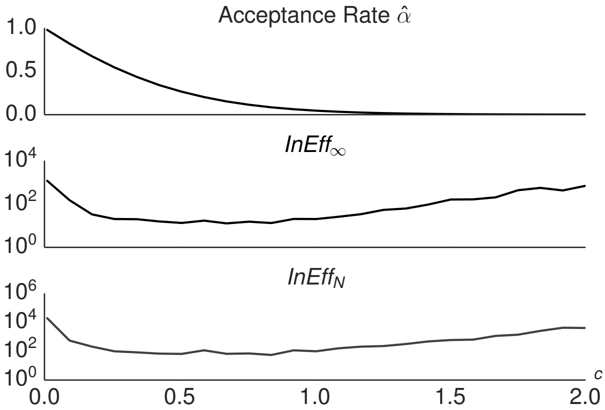
Notes: The $x-$axis indicates the number of draws N. N_0 is set to 0, $25,000$ and $50,000$, respectively.

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]			

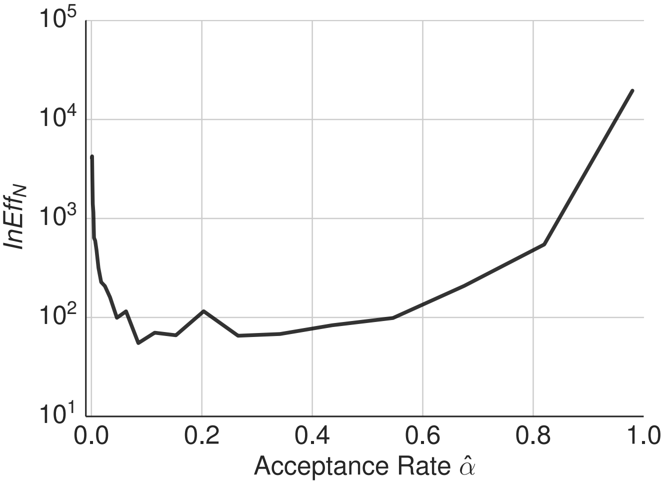
Notes: We generated $N = 100,000$ draws from the posterior and discarded the first 50,000 draws. Based on the remaining draws we approximated the posterior mean and the 5th and 95th percentiles.

Effect of Scaling Constant c



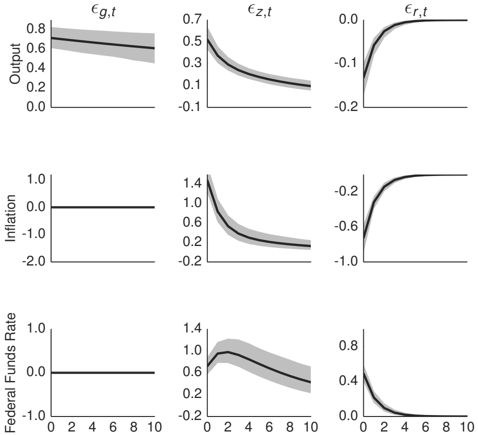
Notes: Results are based on $N_{run} = 50$ independent Markov chains. The acceptance rate (average across multiple chains), HAC-based estimate of $\text{InEff}_\infty[\bar{\tau}]$ (average across multiple chains), and $\text{InEff}_N[\bar{\tau}]$ are shown as a function of the scaling constant c.

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



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

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 $\phi = [\phi_1, \phi_2, \phi_3]'$

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

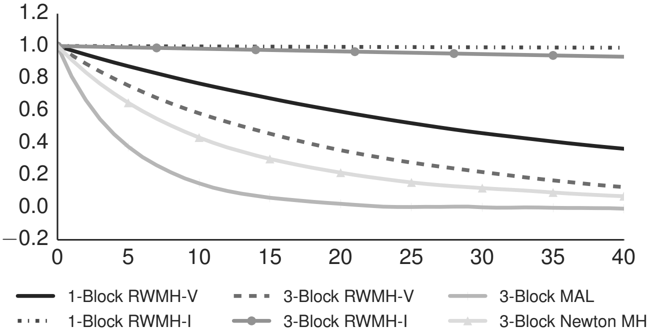
Improvements to MCMC: Blocking

- In high-dimensional parameter spaces the RWMH algorithm generates highly persistent Markov chains.
- What's bad about persistence?

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

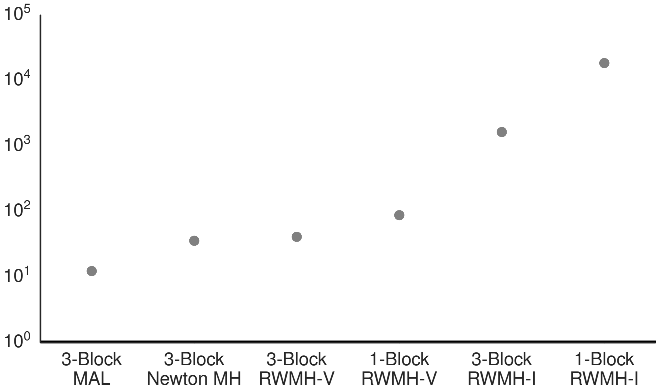
- Potential Remedy:
  - Partition $\theta = [\theta_1, \dots, \theta_K]$.
  - Iterate over conditional posteriors $p(\theta_k | Y, \theta_{<-k>})$.
- To reduce persistence of the chain, try to find partitions such that parameters are strongly correlated within blocks and weakly correlated across blocks or use random blocking.

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

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

Algorithm	Draws Per Second
1-Block RWMH-V	7.76
3-Block RWMH-V	5.65
3-Block MAL	1.24
3-Block RWMH-I	0.14
3-Block Newton MH	0.13
1-Block RWMH-I	0.04

Recall: Posterior Odds and Marginal Data Densities

- Posterior model probabilities can be computed as follows:

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y|\mathcal{M}_i)}{\sum_j \pi_{j,0} p(Y|\mathcal{M}_j)}, \quad j = 1, \ldots, 2, \tag{3}$$

- where

$$p(Y|\mathcal{M}) = \int p(Y|\theta, \mathcal{M}) p(\theta|\mathcal{M}) d\theta \tag{4}$$

- 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

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \underbrace{\frac{\pi_{1,0}}{\pi_{2,0}}}_{\text{Prior Odds}} \times \underbrace{\frac{p(Y|\mathcal{M}_1)}{p(Y|\mathcal{M}_2)}}_{\text{Bayes Factor}} \tag{5}$$

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:

$$\frac{1}{p(Y)} = \int \frac{f(\theta)}{p(Y|\theta)p(\theta)} p(\theta|Y)d\theta, \tag{6}$$

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

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

$$\hat{p}_G(Y) = \left[\frac{1}{N} \sum_{i=1}^{N} \frac{f(\theta^i)}{p(Y|\theta^i)p(\theta^i)} \right]^{-1}, \tag{7}$$

  where θ^i is drawn from the posterior $p(\theta|Y)$.

- Geweke (1999):

$$\begin{aligned} f(\theta) &= \tau^{-1}(2\pi)^{-d/2}|V_\theta|^{-1/2} \exp\left[-0.5(\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta})\right] \\ &\quad \times \left\{ (\theta-\bar{\theta})' V_\theta^{-1}(\theta-\bar{\theta}) \leq F^{-1}_{\chi^2_d}(\tau) \right\}. \end{aligned} \tag{8}$$

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 $\hat{p}(\tilde{\theta}|Y)$.