Bayesian Estimation of DSGE Models

Ed Herbst

Last Updated: October 23, 2018

Small New Keynesian Model

$$
\begin{aligned}
\hat{y}_t &= \mathbb{E}_t[\hat{y}_{t+1}] - \frac{1}{\tau}\left(\hat{R}_t - \mathbb{E}_t[\hat{\pi}_{t+1}] - \mathbb{E}_t[\hat{z}_{t+1}]\right) \\
&\quad + \hat{g}_t - \mathbb{E}_t[\hat{g}_{t+1}] \\
\hat{\pi}_t &= \beta\mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa(\hat{y}_t - \hat{g}_t) \\
\hat{R}_t &= \rho_R \hat{R}_{t-1} + (1-\rho_R)\psi_1\hat{\pi}_t + (1-\rho_R)\psi_2\left(\hat{y}_t - \hat{g}_t\right) + \epsilon_{R,t}
\end{aligned} \tag{1}
$$

States: $s_t = \left[\hat{y}_t, \hat{\pi}_t, \hat{R}_t, \hat{g}_t, \hat{z}_t, \mathbb{E}_t[\hat{y}_{t+1}], \mathbb{E}_t[\hat{\pi}_{t+1}]\right]'$.
Shocks: $\epsilon_t = \left[\epsilon_{z,t}, \epsilon_{g,t}, \epsilon_{R,t}\right]'$. Observables: $y_t = \left[YGR_t, INFL_t, INT_t\right]'$.

$$
\begin{aligned}
YGR_t &= \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t) \\
INFL_t &= \pi^{(A)} + 400\hat{\pi}_t \\
INT_t &= \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t.
\end{aligned} \tag{2}
$$

Parameters $\theta = [\tau, \kappa, \psi_1, \psi_2, \rho_R, \rho_g, \rho_z, r^{(A)}, \pi^{(A)}, \gamma^{(Q)}, \sigma_R, \sigma_g, \sigma_z]'$

Model Solution

- The model in (1) can be cast as in *Linear Rational Expectations* form.

- Solving this system–**ask Gary**–yields (in most cases) the VAR:

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{3}$$

  $\Phi_1(\theta)$ and $\Phi_\epsilon(\theta)$ are functions of the parameters of the DSGE model.

- Can write the observations equation as :

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t + u_t, \tag{4}$$

  allow for a vector of measurement errors, u_t.

- Equations (3) and (4) provide state space representation.

Likelihood

- Let $X_{t_1:t_2} = \{x_{t_1}, x_{t_1+1}, \ldots, x_{t_2}\}$.

- The state-space representation provides a joint density for the observations and latent states given the parameters:

$$p(Y_{1:T}, S_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t, s_t|Y_{1:t-1}, S_{1:t-1}, \theta) \quad (5)$$
$$= \prod_{t=1}^{T} p(y_t|s_t, \theta)p(s_t|s_{t-1}, \theta),$$

- where $p(y_t|s_t, \theta)$ and $p(s_t|s_{t-1}, \theta)$ represent the measurement and state-transition equations, respectively.

- **Problem**: Bayesian Inference has to be based on the likelihood function that is constructed only from the observables, $p(Y_{1:T}|\theta)$

Generic Filter

A filter generates the sequence of conditional distributions $s_t|Y_{1:t}$ and densities $p(y_t|Y_{1:t-1},\theta)$. In turn, the desired likelihood function can be obtained as: $p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1},\theta)$.

Let $p(s_0|Y_{1:0},\theta) = p(s_0|\theta)$. For $t = 1$ to T:

1. From iteration $t-1$ we have $p(s_{t-1}|Y_{1:t-1},\theta)$.
2. Forecasting t given $t-1$:
   2.1 Transition equation:
   $$p(s_t|Y_{1:t-1},\theta) = \int p(s_t|s_{t-1},Y_{1:t-1},\theta)p(s_{t-1}|Y_{1:t-1},\theta)ds_{t-1}$$
   2.2 Measurement equation:
   $$p(y_t|Y_{1:t-1},\theta) = \int p(y_t|s_t,Y_{1:t-1},\theta)p(s_t|Y_{1:t-1},\theta)ds_t$$
3. Updating with Bayes Theorem. Once y_t becomes available:
$$\begin{aligned} p(s_t|Y_{1:t},\theta) &= p(s_t|y_t,Y_{1:t-1},\theta) \\ &= \frac{p(y_t|s_t,Y_{1:t-1},\theta)p(s_t|Y_{1:t-1},\theta)}{p(y_t|Y_{1:t-1},\theta)}. \end{aligned}$$

Kalman Filter

- If the DSGE model is log-linearized and the errors are Gaussian, then the distributions that appear in Generic Filter are Gaussian.

- Maintained Assumption:

$$\begin{aligned} \epsilon_t &\sim iidN(0, \Sigma_\epsilon), \quad u_t \sim iidN(0, \Sigma_u), \\ s_0 &\sim N\big(\bar{s}_{0|0}, P_{0|0}\big). \end{aligned} \tag{6}$$

  common to assume that $\bar{s}_{0|0}$ and $P_{0|0}$ correspond to the invariant distribution associated with s_t.

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

Bayesian Estimation of DSGE Models

The Bayesian Choice

- Bayesian approach: joint distribution over data and parameters.

  Bayesian Model: $p(Y, \theta)$

- Can be factorized into

  Likelihood $\times$ Prior $= p(Y|\theta) \times p(\theta)$

- **Inference**: posterior distribution $p(\theta|Y)$ via Bayes rule

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

- The Bayesian approach prescribes consistency among the beliefs held by an individual and their reasonable relation to any kind of objective data. Learning about θ takes place by updating the prior distribution in light of the data Y.

From Prior to Posterior

- Prior distributions are used to describe the state of knowledge about the parameter vector θ before observing the sample Y.
- In our example, we have to specify a joint probability distribution in 13-dimensional parameter space.

Eliciting prior distributions [Del Negro-Schorfheide (2008)]:

- Group parameters by categories: $\theta_{(ss)}$ (related to steady state), $\theta_{(exo)}$ (related to exogenous processes), $\theta_{(endo)}$ (affects mechanisms but not steady state).

$$
\begin{aligned}
\theta_{(ss)} &= [r^{(A)}, \pi^{(A)}, \gamma^{(Q)}]' \\
\theta_{(exo)} &= [\rho_g, \rho_z, \sigma_g, \sigma_z, \sigma_R]' \\
\theta_{(endo)} &= [\tau, \kappa, \psi_1, \psi_2, \rho_R]'
\end{aligned}
$$

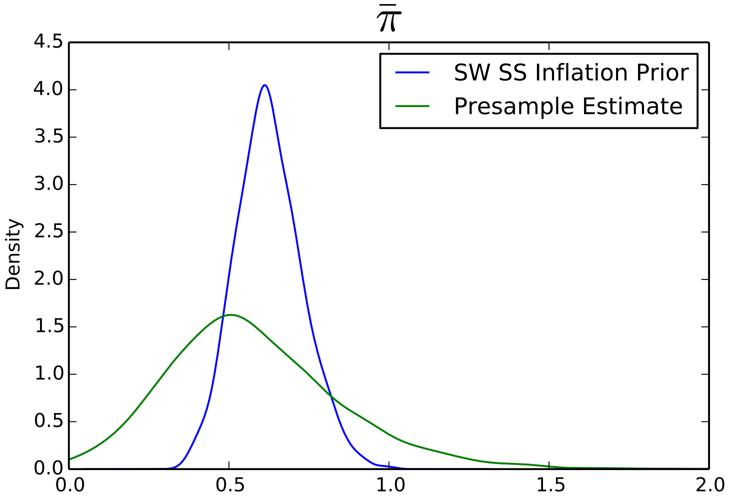
Priors, Continued

- Priors for $\theta_{(ss)}$ are often based on pre-sample averages. If sample starts in 1983:I, the prior distribution for $r^{(A)}$, $\pi^{(A)}$, and $\gamma^{(Q)}$ may be informed by data from the 1970s.
- Priors for $\theta_{(endo)}$ may be partly based on microeconometric evidence.
- Priors for $\theta_{(exo)}$ are the most difficult to specify. You could specific indirectly, by looking at the volatility/autocorrelation of observables implied by $\theta_{(exo)}$ given other parameters.

Above all: Generate draws from the prior distribution of θ; compute important transformations of θ such as steady-state ratios and possibly impulse-response functions or variance decompositions.

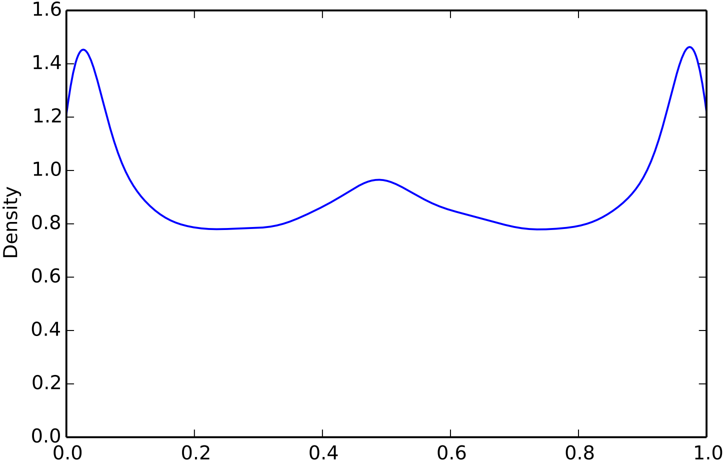
- Marginals may be plausible, while joint is not.
- Nonlinear transformations of uniform variables are not uniform!

Try not to set priors based Y



$\rho = \frac{x^2}{x^2+y^2}, x \sim U[0,1], y \sim U[0,1]$

Density of ρ



Bayesian Estimation of DSGE Models

The main event

- Inference: Need to characterize posterior $p(\theta|Y)$.
- Unfortunately, for many interesting models it is not possible to evaluate the moments and quantiles of the posterior $p(\theta|Y)$ analytically.
- **Rules of game**: we can only numerically evaluate prior $p(\theta)$ and likelihood $p(Y|\theta)$.
- To evaluate posterior moments of function $h(\theta)$, we need numerical techniques.

Look **posterior samplers** that generate sequences of draws θ^i, $i = 1, \ldots, N$ from $p(\theta|Y)$.

- (Monte Carlo) averages of these draws typically follow Strong Law of Large Numbers (SLLN) and (sometimes) Central Limit Theorem (CLT).
- SLLN justifies using averages to approx. moments, CLT characterizes accuracy of approx.

Sampler 1: Importance Sampler

Importance Sampling

$$\pi(\theta) = \frac{f(\theta)}{Z} = \frac{p(Y|\theta)p(\theta)}{p(\theta)} \tag{7}$$

$f(\cdot)$ is the function we can evaluate numerically.

References: Hammersley and Handscomb (1964), Kloek and van Dijk (1978), and Geweke (1989).

Let g be an arbitrary, easy-to-sample pdf over θ (think normal distribution).

Importance sampling (IS) is based on the following identity:

$$\mathbb{E}_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta = \frac{1}{Z}\int_\Theta h(\theta)\frac{f(\theta)}{g(\theta)}g(\theta)d\theta. \tag{8}$$

Since $\mathbb{E}_\pi[1] = 1$,

$$Z = \int_\Theta \frac{f(\theta)}{g(\theta)}g(\theta)d\theta.$$

(Unnormalized) Importance weight:

$$w(\theta) = \frac{f(\theta)}{g(\theta)}$$

Normalized Importance Weight:

$$v(\theta) = \frac{w(\theta)}{\int w(\theta)g(\theta)d\theta} = \frac{w(\theta)}{\int Z\pi(\theta)d\theta} = \frac{w(\theta)}{Z}. \tag{9}$$

Can show:

$$\mathbb{E}_\pi[h(\theta)] = \int v(\theta)h(\theta)g(\theta)d\theta. \tag{10}$$

Algorithm (Importance Sampling)

1. *For* $i = 1$ *to* N, *draw* $\theta^i \overset{iid}{\sim} g(\theta)$ *and compute the unnormalized importance weights*

$$w^i = w(\theta^i) = \frac{f(\theta^i)}{g(\theta^i)}. \tag{11}$$

2. *Compute the normalized importance weights*

$$W^i = \frac{w^i}{\frac{1}{N}\sum_{i=1}^N w^i}. \tag{12}$$

   An approximation of $\mathbb{E}_\pi[h(\theta)]$ *is given by*

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^N W^i h(\theta^i). \tag{13}$$

Note W^i is (slightly) different from v in previous slide.

- Refer to the collection of pairs $\{(\theta^i, W^i)\}_{i=1}^N$ as a particle approximation of $\pi(\theta)$.

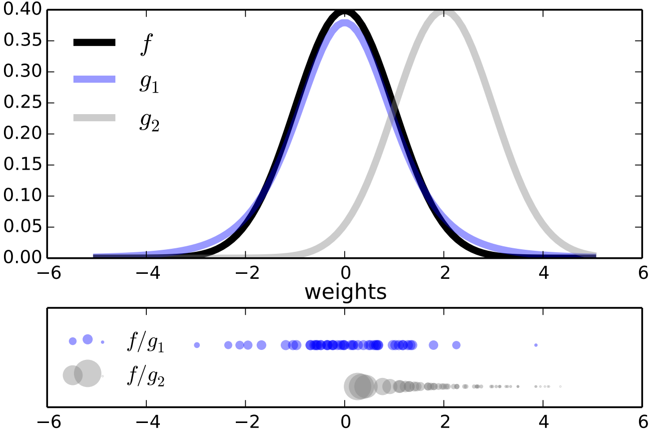
- The accuracy of the approximation is driven by the "closeness" of $g(\cdot)$ to $f(\cdot)$ and is reflected in the distribution of the weights.

- If the distribution of weights is very uneven, the Monte Carlo approximation $\bar{h}$ is inaccurate.

- Uniform weights arise if $g(\cdot) \propto f(\cdot)$, which means that we are sampling directly from $\pi(\theta)$.

Effectiveness of IS depends on similarity of f and g

$$f = \mathcal{N}(0,1), \quad g_1 = t(0,1,5), \quad g_2 = \mathcal{N}(2,1)$$



Only a few draws from $N(2,1)$ have meaningful weight.

$\Longrightarrow$ estimate is based on small sample.

$\Longrightarrow$ estimate will be noisy.

Convergence

- **SLLN:** If $\mathbb{E}_g[|hf/g|] < \infty$ and $\mathbb{E}_g[|f/g|] < \infty$, see Geweke (1989), the Monte Carlo estimate $\bar{h}_N$ defined in (13) converges almost surely (a.s.) to $E_\pi[h(\theta)]$ as $N \longrightarrow \infty$.

- **CLT:** Provided that $\sup_\theta \ \pi(\theta)/g(\theta) < \infty$ and $\mathbb{E}_g[h^2] < \infty$, we can apply a multivariate extension of the Lindeberg-Levy CLT.

  Argument: first order taylor expansion of $\bar{h}_N$ around $\mathbb{E}_{\pi[h]}$, (extremely) tedious algebra.

$$\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \Longrightarrow N\big(0, \Omega(h)\big), \tag{14}$$

where

$$\Omega(h) = \mathbb{V}_g[(\pi/g)(h - \mathbb{E}_\pi[h])].$$

Accuracy

- Assess the accuracy by computing a Monte Carlo approximation $\bar{h}_N$ multiple times and examine its variability across repeated runs of the posterior sampler.

- If $\bar{h}_N$ satisfies a CLT and the number of draws N is sufficiently large, then the variance across repeated runs of the algorithm (provided this variance is finite for the given N) will approximately coincide with the asymptotic variance implied by the CLT.

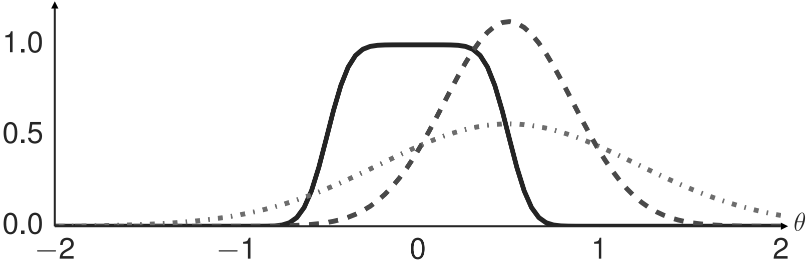
- Define **inefficiency factor** relative to IID sampling,

$$\text{InEff}_{\infty} = \frac{\Omega(h)}{\mathbb{V}_{\pi}[h]}.$$

If Ineff_{∞} ¿ 1 we are worse than iid sampling.

Numerical Illustration

- Let's take a harder $\pi(\theta)$, the set-identified posterior from Moon-Schorfheide (2013).
- Consider **diffuse** and **concentrated** importance sample densities g.



Experiment

- Using various N_draw, generate *IS* approximations for $h(\theta) = \theta$ and $h(\theta) = \theta^2$.

- Calculate estimate of InEff_∞ using $N_{run} = 1000$ Monte Carlo simulations, as well as the exact value [by sampling from $\pi(\theta)$.] Estimates come from:

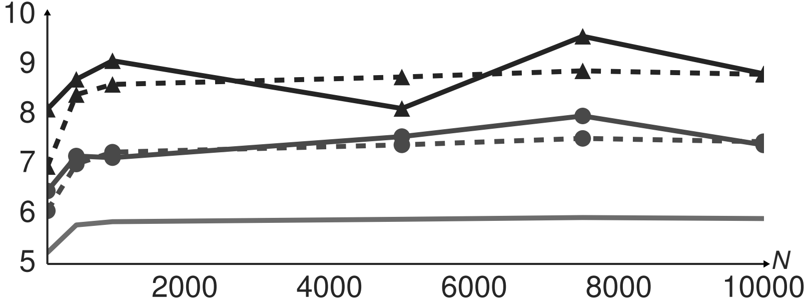
$$\text{InEff}_N = \frac{\mathbb{V}[\bar{h}_N]}{\mathbb{V}_\pi[h]/N}. \tag{15}$$

- Also calculate poor man's version of Inefficiency Factor, because everyone uses it.

$$\text{InEff}_\infty \approx 1 + \mathbb{V}_g[\pi/g]. \tag{16}$$

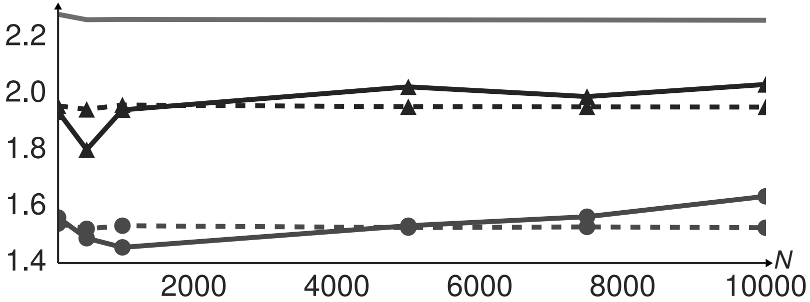
Concetrated IS Density

- solid line = estimates of $\text{lnEff}_\infty[h]$, dashed = truth
- triangles = $h(\theta) = \theta$, circles = $h(\theta) = \theta^2$
- grey line = poor man's inefficiency



Concetrated IS Density

- Solid line = estimates of $\mathrm{InEff}_{\infty}[h]$, dashed = truth
- triangles = $h(\theta) = \theta$, circles = $h(\theta) = \theta^2$
- grey line = poor man's inefficiency



Take aways

- It is important that the importance density g is well-tailored toward the target distribution π!

- Everything is h specific!

- with approximately elliptical posterior, a good importance density can be obtained by centering a fat-tailed t distribution at the mode of π and using a scaled version of the inverse Hessian of $\ln \pi$ at the mode to align the contours of the importance density with the contours of the posterior π.

- Very bad for highly irregular and non-elliptical posteriors...

Sampler 2: Metropolis-Hastings Sampler

The Metropolis-Hastings Algorithm

- Metropolis-Hastings (MH) algorithm belongs to the class of Markov chain Monte Carlo (MCMC) algorithms.
- Algorithm constructs a Markov chain such that the stationary distribution associated with this Markov chain is unique and equals the posterior distribution of interest.
- First version constructed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953). Later generalized by Hastings (1970). Tierney (1994) proved important convergence results for MCMC algorithms.
- Introduction: Chib and Greenberg (1995). Textbook Robert and Casella (2004) or Geweke (2005).

- Importance sampler generates a sequence of independent draws from the posterior distribution $\pi(\theta)$, the MH algorithm generates a sequence of serially correlated draws.
- As long as the correlation in the Markov chain is not too strong, Monte Carlo averages of these draws can accurately approximate posterior means of $h(\theta)$.
- We are going to care a lot about this correlation. Why?

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

A key ingredient is the proposal distribution $q(\vartheta|\theta^{i-1})$, which potentially depends on the draw θ^{i-1} in iteration $i-1$ of the algorithm.

Algorithm (Generic MH 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.

Because $p(\theta|Y) \propto p(Y|\theta)p(\theta)$ we can replace the posterior densities in the calculation of the acceptance probabilities $\alpha(\vartheta|\theta^{i-1})$

This yields a Markov transition kernel $K(\theta|\tilde{\theta})$, where the conditioning value $\tilde{\theta}$ corresponds to the parameter draw from iteration $i-1$.

Convergence

Probability theory for MH is much harder than for IS.

1. Suppose that $\theta^0 \sim g(\cdot)$ and θ^N is obtained by iterating the Markov transition kernel forward N times, then is it true that θ^N is approximately distributed according to $p(\theta|Y)$ and the approximation error vanishes as $N \longrightarrow \infty$?
2. Suppose that (i) is true, is it also true that sample averages of θ^i, $i = 1, \ldots, N$ satisfy a SLLN and a CLT?

Key property: **invariance** of Markov Chain.

$$p(\theta|Y) = \int K(\theta|\tilde{\theta})p(\tilde{\theta}|Y)d\tilde{\theta}. \tag{17}$$

Show this property using **reversibility of the Markov Chain**

Not sufficient for SLLN or CLT, these things depend on q and π.

Look at specific example.

A Specific Example

- Suppose the parameter space is discrete and θ can only take two values: τ_1 and τ_2.

- The posterior distribution then simplifies to two probabilities which we denote as $\pi_l = \mathbb{P}\{\theta = \tau_l | Y\}$, $l = 1, 2$.

- The proposal distribution in Algorithm 2 can be represented as a two-stage Markov process with transition matrix

$$Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}, \tag{18}$$

  where q_{lk} is the probability of drawing $\vartheta = \tau_k$ conditional on $\theta^{i-1} = \tau_l$.
- Assume that

$$q_{11} = q_{22} = q, \quad q_{12} = q_{21} = 1 - q$$

  and that the posterior distribution has the property

$$\pi_2 > \pi_1.$$

Deriving the Transition Kernel

- Suppose that $\theta^{i-1} = \tau_1$. Then with probability q, $\vartheta = \tau_1$. The probability that this draw will be accepted is

$$\alpha(\tau_1|\tau_1) = \min\left\{1, \frac{\pi_1/q}{\pi_1/q}\right\} = 1.$$

- With probability $1 - q$ the proposed draw is $\vartheta = \tau_2$. The probability that this draw will be rejected is

$$1 - \alpha(\tau_2|\tau_1) = 1 - \min\left\{1, \frac{\pi_2/(1-q)}{\pi_1/(1-q)}\right\} = 0$$

  because we previously assumed that $\pi_2 > \pi_1$.
- The probability of a transition from $\theta^{i-1} = \tau_1$ to $\theta^i = \tau_1$ is

$$k_{11} = q \cdot 1 + (1-q) \cdot 0 = q.$$

Transition Kernel, Continued

- Similar reasoning as before

$$K = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} = \begin{bmatrix} q & (1-q) \\ (1-q)\frac{\pi_1}{\pi_2} & q + (1-q)\left(1 - \frac{\pi_1}{\pi_2}\right) \end{bmatrix}.$$

- K has two eigenvalues λ_1 and λ_2:

$$\lambda_1(K) = 1, \quad \lambda_2(K) = q - (1-q)\frac{\pi_1}{1-\pi_1}. \tag{19}$$

  Eigenvector associated with with $\lambda_1(K)$ determines the invariant distribution of the Markov chain (=posterior). If $\lambda_2(K) \neq 1$, this distribution is unique.

  The persistence of the Markov chain is characterized by the eigenvalue $\lambda_2(K)$.

Markov Chain

- We can represent the Markov Chain generated by MH as an AR(1). Define:

$$\xi^i = \frac{\theta^i - \tau_1}{\tau_2 - \tau_1}, \quad \xi^i \in \{0, 1\}.$$

ξ^i follows the first-order autoregressive process

$$\xi^i = (1 - k_{11}) + \lambda_2(K)\xi^{i-1} + \nu^i. \tag{20}$$

Conditional on $\xi^{i-1} = j - 1$, $j = 1, 2$, 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 $k_{jj}(1 - k_{jj})$.

More on Markov Chain

- Persistence of the Markov chain depends on the proposal distribution, which in our discrete example is characterized by the probability q.

- You could get an *iid* sample from the posterior by setting $q = \pi_1$, so $\lambda_2(K) = 0$.)

- OTOH, if $q = 1$, then $\theta^i = \theta^1$ for all i and the equilibrium distribution of the chain is no longer unique.

- General goal of MCMC: keep the persistence of the chain as low as possible.

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

we deduce from a central limit theorem for dependent random variables that

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

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

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

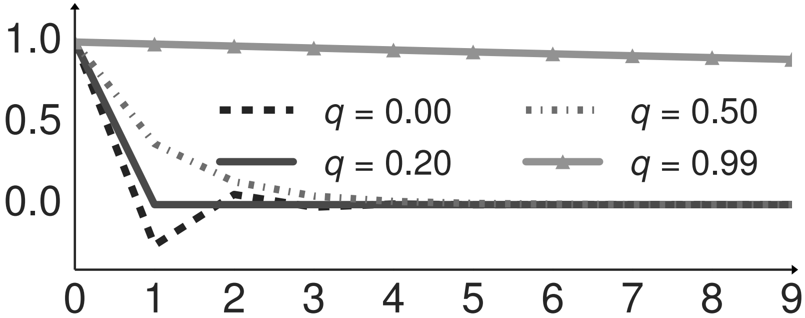
In turn, the asymptotic inefficiency factor is given by

$$\begin{aligned} \mathrm{InEff}_\infty &= \frac{\Omega(h)}{\mathbb{V}_\pi[h]} \qquad (21) \\ &= 1 + 2\lim_{L \longrightarrow \infty} \sum_{l=1}^{L}\frac{L-l}{L}\left(q - (1-q)\frac{\pi_1}{1-\pi_1}\right)^l. \end{aligned}$$

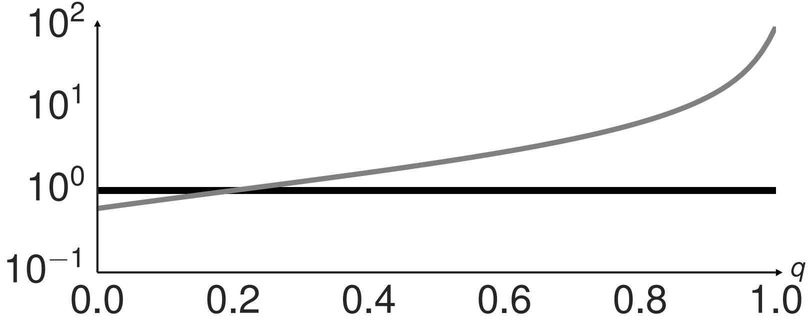
Numerical Example

- Bernoulli distribution ($\tau_1 = 0, \tau_2 = 1$) with $\pi_1 = 0.2$.
- Assess the effectiveness of different MH settings, we vary $q \in [0, 1)$.
- Look at autocorrelation for $q = \{0, 0.2, 0.5, 0.99\}$.
- Ineff_∞ for $q \in [0, 1)$.
- Relationship between across chain variance and within chain (HAC) estimates. This the heart of many convergence statistics.

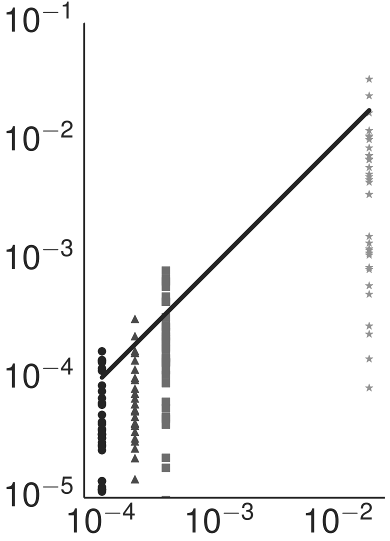
Autocorrelation Functions



Log Inefficiency Factor as function of q



Convergence: within vs across chain variance estimates



Take Aways

- high autocorrelation reflects the fact that it will take a high number of draws to accurately reflect the target distribution

- for large values of q, the variance of Monte Carlo estimates of h drawn from the MH chain are much larger than the variance of estimates derived from *iid* draws

- HAC estimates bracket small-sample estimates, indicating convergence, but they tend to underestimate variance for all q.

How to pick q for a DSGE model?

Random Walk Metropolis-Hastings

Random Walk Metropolis-Hastings

- Most popular q for DSGE Models.
- $q(\vartheta|\theta^{i-1})$ can be expressed as the random walk $\vartheta = \theta^{i-1} + \eta$
- η is normally distributed with mean zero and variance $c^2\hat{\Sigma}$.
- Given the symmetric nature of the proposal distribution, the acceptance probability becomes

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

- Still need to specify c and $\hat{\Sigma}$.

On $\hat{\Sigma}$

- Want $\hat{\Sigma}$ to incorporate information about the posterior.

- One approach: Schorfheide (2000), is to set $\hat{\Sigma}$ to be the negative of the inverse Hessian at the mode of the log posterior, $\hat{\theta}$, obtained by running a numerical optimization .

  This has appealing large sample properties, but can be tedious and innacurate.

- Another (adaptive) approach: use prior variance for a first sequence of posterior draws, the compute the sample covariance matrix and use that as $\hat{\Sigma}$. *Must be fixed eventually*.

- Here we cheat:

$$\text{RWMH-V} : \hat{\Sigma} = \mathbb{V}_{\pi}[\theta].$$

Picking Scaling c

- Goldilocks principal: choose c so that you don't reject too much or too little.
- Roberts, Gelman, and Gilks (1997) have derived a limit (in the size of parameter vector) optimal acceptance rate of 0.234 for a special case (normal posterior).
- Most practitioners target an acceptance rate between 0.20 and 0.40.
- Requites pre-estimation tuning.

Baseline Estimation

Table: Posterior Estimates of DSGE Model Parameters

	Mean	[0.05, 0.95]		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.

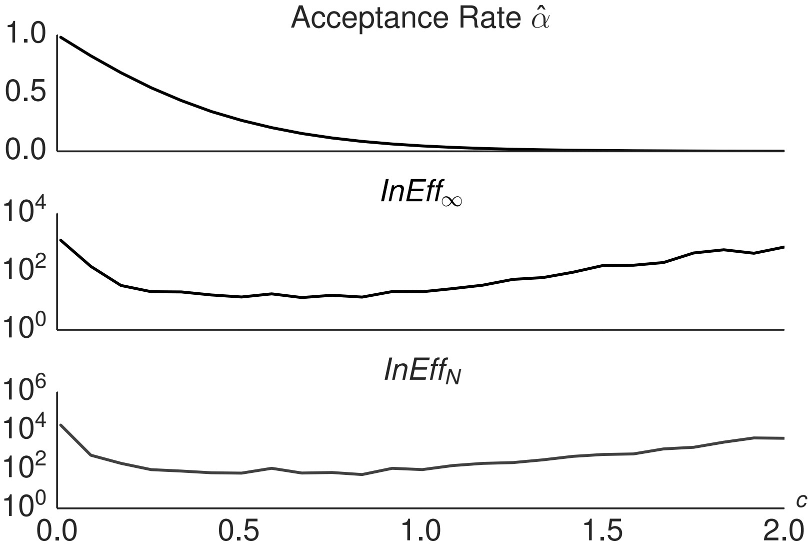
More on c

Vary $c \in (0, 2]$. Look at effect on

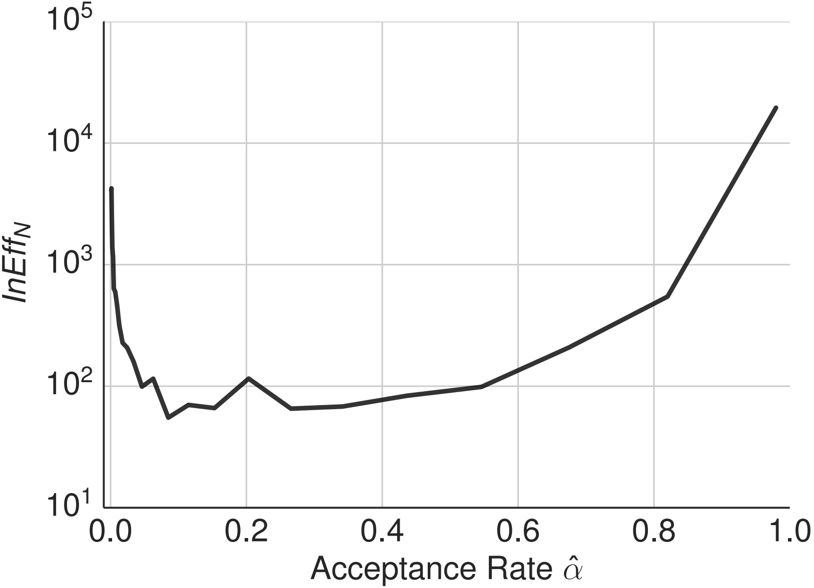
- Acceptance Rate
- $Ineff_\infty$
- $Ineff_N$

What is the relationship between acceptance rate and accuracy?

Effects of Scaling



Acceptance Rate vs. Accuracy



Next Time

- More elaborate MCMC; blocking.
- Where this really breaks down: 3 DSGE Examples.
- An alternative approach: Sequential Monte Carlo

CHIB, S., AND E. GREENBERG (1995): “Understanding the Metropolis-Hastings Algorithm,” *American Statistician*, 49, 327–335.

GEWEKE, J. (1989): “Bayesian Inference in Econometric Models Using Monte Carlo Integration,” *Econometrica*, 57(6), 1317–1399.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons.

HAMMERSLEY, J., AND D. HANDSCOMB (1964): *Monte Carlo Methods*. Methuen and Company.

HASTINGS, W. (1970): “Monte Carlo Sampling Methods Using Markov Chains and Their Applications,” *Biometrika*, 57, 97–109.

KLOEK, T., AND H. K. VAN DIJK (1978): “Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo,” *Econometrica*, 46, 1–20.

METROPOLIS, N., A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER, AND E. TELLER (1953): “Equations of State Calculations by Fast Computing Machines,” *Journal of Chemical Physics*, 21, 1087–1091.

Robert, C. P., and G. Casella (2004): *Monte Carlo Statistical Methods*. Springer.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997): "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *Annals of Applied Probability*, 7(1), 110–120.

Schorfheide, F. (2000): "Loss Function-based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15, 645–670.

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