Bayesian Estimation of DSGE Models

Frank Schorfheide

University of Pennsylvania

Gerzensee Ph.D. Course on Bayesian Macroeconometrics

May 28, 2019

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

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

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

Bayesian Inference

- We previously discussed the evaluation of the likelihood function: given a parameter θ
  - solve the DSGE model to obtain the state-space representation;
  - use the Kalman filter to evaluate the likelihood function.
- Let's talk a bit about prior distributions.

Prior Distributions

- Ideally: probabilistic representation of our knowledge/beliefs before observing sample Y.
- More realistically: choice of prior as well as model are influenced by some observations. Try to keep influence small or adjust measures of uncertainty.
- Views about role of priors:
  1. keep them "uninformative" (???) so that posterior inherits shape of likelihood function;
  2. use them to regularize the likelihood function;
  3. incorporate information from sources other than Y;

Prior Elicitation for DSGE Models

- 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?

Prior Distribution

Name	Domain	Prior Density	Para (1)	Para (2)
Steady-State-Related Parameters $\theta_{(ss)}$				
$100(1/\beta - 1)$	$\mathbb{R}^+$	Gamma	0.50	0.50
$100 \log \pi^*$	$\mathbb{R}^+$	Gamma	1.00	0.50
$100 \log \gamma$	$\mathbb{R}$	Normal	0.75	0.50
λ	$\mathbb{R}^+$	Gamma	0.20	0.20
Endogenous Propagation Parameters $\theta_{(endo)}$				
ζ_p	$[0, 1]$	Beta	0.70	0.15
$1/(1+\nu)$	$\mathbb{R}^+$	Gamma	1.50	0.75

Notes: Marginal prior distributions for each DSGE model parameter. Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$. The joint prior distribution of θ is truncated at the boundary of the determinacy region.

Prior Distribution

Name	Domain	Prior		
		Density	Para (1)	Para (2)
		Exogenous Shock Parameters $\theta_{(exo)}$		
ρ_ϕ	$[0,1)$	Uniform	0.00	1.00
ρ_λ	$[0,1)$	Uniform	0.00	1.00
ρ_z	$[0,1)$	Uniform	0.00	1.00
$100\sigma_\phi$	$\mathbb{R}^+$	InvGamma	2.00	4.00
$100\sigma_\lambda$	$\mathbb{R}^+$	InvGamma	0.50	4.00
$100\sigma_z$	$\mathbb{R}^+$	InvGamma	2.00	4.00
$100\sigma_r$	$\mathbb{R}^+$	InvGamma	0.50	4.00

Notes: Marginal prior distributions for each DSGE model parameter. Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The joint prior distribution of θ is truncated at the boundary of the determinacy region.

Draws from Posterior

- We will focus on Markov chain Monte Carlo (MCMC) 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)...

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 $K(\theta|\tilde{\theta})$.

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

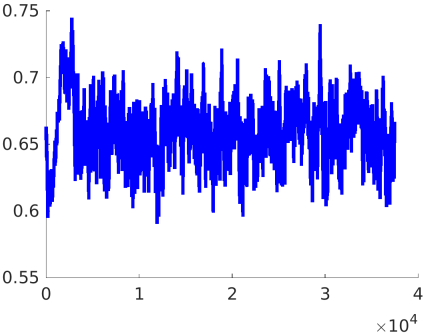
- Initialization steps can be modified as needed for particular application.
- If numerical optimization does not work well, one could let $\hat{\Sigma}$ be a diagonal matrix with prior variances on the diagonal.
- Or, $\hat{\Sigma}$ could be based on a preliminary run of a posterior sampler.
- It is good practice to run multiple chains based on different starting values.

Numerical Illustration

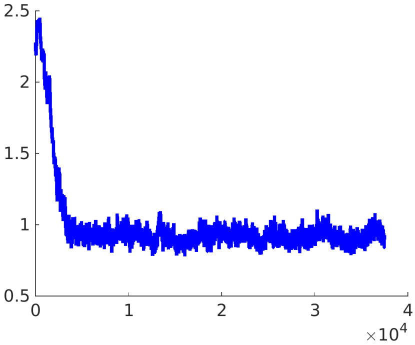
- Generate a single sample of size $T = 80$ from the stylized DSGE model.
- Combine likelihood and prior to form posterior.
- Draws from this posterior distribution are generated using the RWMH algorithm.
- Chain is initialized with a draw from the prior distribution.
- The covariance matrix $\hat{\Sigma}$ is based on the negative inverse Hessian at the mode. The scaling constant c is set equal to 0.075, which leads to an acceptance rate for proposed draws of 0.55.

Parameter Draws from MH Algorithm

ζ_p^i Draws

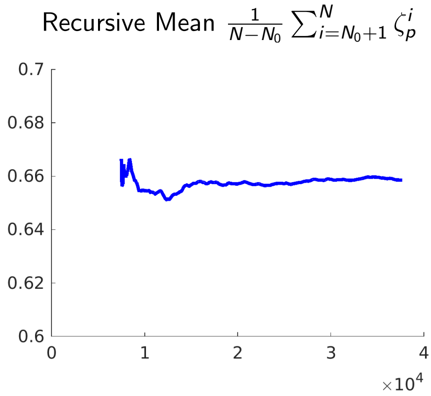


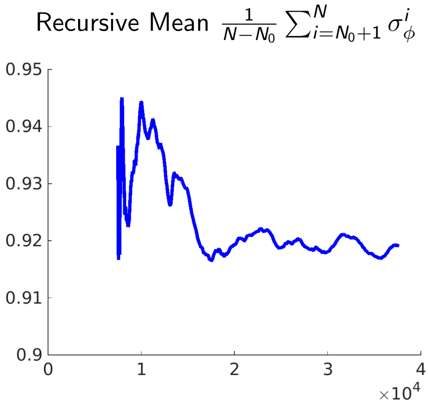
σ_ϕ^i Draws



Notes: The posterior is based on a simulated sample of observations of size $T = 80$. The top panel shows the sequence of parameter draws and the bottom panel shows recursive means.

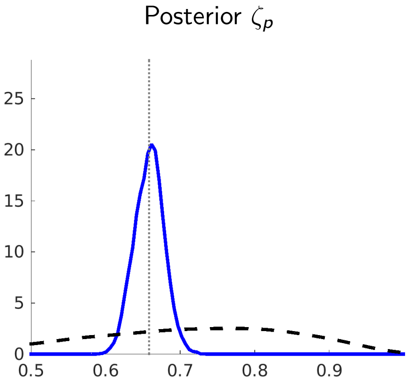
Parameter Draws from MH Algorithm

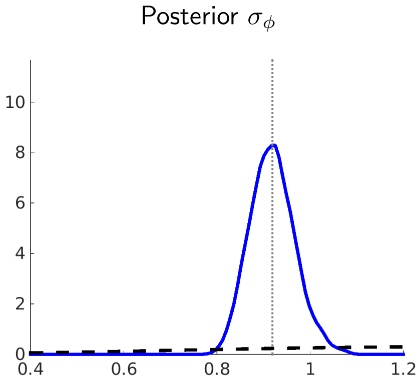




Notes: The posterior is based on a simulated sample of observations of size $T = 80$. The top panel shows the sequence of parameter draws and the bottom panel shows recursive means.

Prior and Posterior Densities





Notes: The dashed lines represent the prior densities, whereas the solid lines correspond to the posterior densities of ζ_p and σ_ϕ. The posterior is based on a simulated sample of observations of size $T = 80$. We generate $N = 37,500$ draws from the posterior and drop the first $N_0 = 7,500$ draws.