Bayesian Inference for DSGE Models

Lawrence J. Christiano

Outline

- State space-observer form.
  - convenient for model estimation and many other things.
- Bayesian inference
  - Bayes' rule.
  - Monte Carlo integation.
  - MCMC algorithm.
  - Laplace approximation

State Space/Observer Form

- Compact summary of the model, and of the mapping between the model and data used in the analysis.
- Typically, data are available in log form. So, the following is useful:
  - If x is steady state of x_t :

$$\begin{aligned} \hat{x}_t &\equiv \frac{x_t - x}{x}, \\ &\implies \frac{x_t}{x} = 1 + \hat{x}_t \\ &\implies \log\left(\frac{x_t}{x}\right) = \log(1 + \hat{x}_t) \approx \hat{x}_t \end{aligned}$$

- Suppose we have a model solution in hand:[1]

$$\begin{aligned} z_t &= Az_{t-1} + Bs_t \\ s_t &= Ps_{t-1} + \epsilon_t,\ E\epsilon_t\epsilon_t' = D. \end{aligned}$$

---

[1]Notation taken from solution lecture notes, http://faculty.wcas.northwestern.edu/~lchrist/course/Korea 2012/lecture on solving rev.pdf

State Space/Observer Form

- Suppose we have a model in which the date t endogenous variables are capital, K_{t+1}, and labor, N_t:

$$z_t = \begin{pmatrix} \hat{K}_{t+1} \\ \hat{N}_t \end{pmatrix},\ s_t = \hat{\varepsilon}_t,\ \epsilon_t = e_t.$$

- Data may include variables in z_t and/or other variables.
  - for example, suppose available data are N_t and GDP, y_t and production function in model is:

$$y_t = \varepsilon_t K_t^{\alpha} N_t^{1-\alpha},$$

    so that

$$\begin{aligned} \hat{y}_t &= \hat{\varepsilon}_t + \alpha \hat{K}_t + (1-\alpha)\hat{N}_t \\ &= \begin{pmatrix} 0 & 1-\alpha \end{pmatrix} z_t + \begin{pmatrix} \alpha & 0 \end{pmatrix} z_{t-1} + s_t \end{aligned}$$

- From the properties of $\hat{y}_t$ and $\hat{N}_t$:

$$Y_t^{data} = \begin{pmatrix} \log y_t \\ \log N_t \end{pmatrix} = \begin{pmatrix} \log y \\ \log N \end{pmatrix} + \begin{pmatrix} \hat{y}_t \\ \hat{N}_t \end{pmatrix}$$

State Space/Observer Form

- Model prediction for data:

$$
\begin{aligned}
Y_t^{data} &= \begin{pmatrix} \log y \\ \log N \end{pmatrix} + \begin{pmatrix} \hat{y}_t \\ \hat{N}_t \end{pmatrix} \\
&= \begin{pmatrix} \log y \\ \log N \end{pmatrix} + \begin{bmatrix} 0 & 1-\alpha \\ 0 & 1 \end{bmatrix} z_t + \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix} z_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} s_t \\
&= a + H\xi_t
\end{aligned}
$$

$$
\xi_t = \begin{pmatrix} z_t \\ z_{t-1} \\ \hat{\varepsilon}_t \end{pmatrix},\ a = \begin{bmatrix} \log y \\ \log N \end{bmatrix},\ H = \begin{bmatrix} 0 & 1-\alpha & \alpha & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}
$$

- The *Observer Equation* may include measurement error, w_t :

$$
Y_t^{data} = a + H\xi_t + w_t,\ Ew_t w_t' = R.
$$

- Semantics: ξ_t is the *state* of the system (do not confuse with the economic state (K_t, ε_t)!).

State Space/Observer Form

- Law of motion of the state, ξ_t (state-space equation):

$$\xi_t = F\xi_{t-1} + u_t,\ Eu_tu_t' = Q$$

$$\begin{pmatrix} z_{t+1} \\ z_t \\ s_{t+1} \end{pmatrix} = \begin{bmatrix} A & 0 & BP \\ I & 0 & 0 \\ 0 & 0 & P \end{bmatrix} \begin{pmatrix} z_t \\ z_{t-1} \\ s_t \end{pmatrix} + \begin{pmatrix} B \\ 0 \\ I \end{pmatrix} \epsilon_{t+1},$$

$$u_t = \begin{pmatrix} B \\ 0 \\ I \end{pmatrix} \epsilon_t,\ Q = \begin{bmatrix} BDB' & 0 & BD \\ 0 & 0 & 0 \\ DB' & & D \end{bmatrix},\ F = \begin{bmatrix} A & 0 & BP \\ I & 0 & 0 \\ 0 & 0 & P \end{bmatrix}.$$

Uses of State Space/Observer Form

- Estimation of θ and forecasting ξ_t and Y_t^{data}
- Can take into account situations in which data represent a mixture of quarterly, monthly, daily observations.
- 'Data Rich' estimation. Could include several data measures (e.g., employment based on surveys of establishments and surveys of households) on a single model concept.
- Useful for solving the following forecasting problems:
  - Filtering (mainly of technical interest in computing likelihood function):

$$P\left[\xi_t | Y_{t-1}^{data}, Y_{t-2}^{data}, ..., Y_1^{data}\right],\ t = 1, 2, ..., T.$$

  - Smoothing:

$$P\left[\xi_t | Y_T^{data}, ..., Y_1^{data}\right],\ t = 1, 2, ..., T.$$

  - Example: 'real rate of interest' and 'output gap' can be recovered from ξ_t using simple New Keynesian model.
- Useful for deriving a model's implications vector autoregressions

Quick Review of Probability Theory

- Two random variables, $x \in (x_1, x_2)$ and $y \in (y_1, y_2)$.
- *Joint distribution*: $p(x, y)$

$$\begin{array}{c|c|c|} & x_1 & x_2 \\ \hline y_1 & p_{11} & p_{12} \\ \hline y_2 & p_{21} & p_{22} \\ \hline \end{array} = \begin{array}{c|c|c|} & x_1 & x_2 \\ \hline y_1 & 0.05 & 0.40 \\ \hline y_2 & 0.35 & 0.20 \\ \hline \end{array}$$

where

$$p_{ij} = probability\left(x = x_i, y = y_j\right).$$

- Restriction:

$$\int_{x,y} p(x, y)\, dxdy = 1.$$

Quick Review of Probability Theory

- *Joint distribution*: $p(x,y)$

$$\begin{array}{cc} & \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{array}{c} y_1 \\ y_2 \end{array} & \begin{array}{|c|c|} \hline p_{11} & p_{12} \\ \hline p_{21} & p_{22} \\ \hline \end{array} \end{array} = \begin{array}{cc} & \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{array}{c} y_1 \\ y_2 \end{array} & \begin{array}{|c|c|} \hline 0.05 & 0.40 \\ \hline 0.35 & 0.20 \\ \hline \end{array} \end{array}$$

- *Marginal distribution* of x : $p(x)$

Probabilities of various values of x without reference to the value of y:

$$p(x) = \begin{cases} p_{11} + p_{21} = 0.40 & x = x_1 \\ p_{12} + p_{22} = 0.60 & x = x_2 \end{cases}.$$

or,

$$p(x) = \int_y p(x,y)\, dy$$

Quick Review of Probability Theory

- *Joint distribution*: $p(x, y)$

$$
\begin{array}{ccc} & x_1 & x_2 \\ y_1 & \multicolumn{1}{|c|}{p_{11}} & \multicolumn{1}{c|}{p_{12}} \\ \hline y_2 & \multicolumn{1}{|c|}{p_{21}} & \multicolumn{1}{c|}{p_{22}} \\ \hline \end{array}
=
\begin{array}{ccc} & x_1 & x_2 \\ y_1 & \multicolumn{1}{|c|}{0.05} & \multicolumn{1}{c|}{0.40} \\ \hline y_2 & \multicolumn{1}{|c|}{0.35} & \multicolumn{1}{c|}{0.20} \\ \hline \end{array}
$$

- *Conditional distribution* of x given y : $p(x|y)$
  - Probability of x given that the value of y is known

$$
p(x|y_1) = \begin{cases} p(x_1|y_1) & \frac{p_{11}}{p_{11}+p_{12}} = \frac{p_{11}}{p(y_1)} = \frac{0.05}{0.45} = 0.11 \\ p(x_2|y_1) & \frac{p_{12}}{p_{11}+p_{12}} = \frac{p_{12}}{p(y_1)} = \frac{0.40}{0.45} = 0.89 \end{cases}
$$

or,

$$
p(x|y) = \frac{p(x, y)}{p(y)}.
$$

Quick Review of Probability Theory

- *Joint distribution*: $p(x, y)$

	x_1	x_2	
y_1	0.05	0.40	$p(y_1) = 0.45$
y_2	0.35	0.20	$p(y_2) = 0.55$
	$p(x_1) = 0.40$	$p(x_2) = 0.60$	

- Mode
  - *Mode of joint distribution* (in the example):

$$\text{argmax}_{x,y} p(x, y) = (x_2, y_1)$$

  - *Mode of the marginal distribution*:

$$\text{argmax}_x p(x) = x_2, \ \text{argmax}_y p(y) = y_2$$

  - Note: mode of the marginal and of joint distribution conceptually different.

Maximum Likelihood Estimation

- State space-observer system:

$$
\begin{aligned}
\xi_{t+1} &= F\xi_t + u_{t+1},\ Eu_t u_t' = Q, \\
Y_t^{data} &= a_0 + H\xi_t + w_t,\ Ew_t w_t' = R
\end{aligned}
$$

- Reduced form parameters, (F, Q, a_0, H, R), functions of θ.
- Choose θ to maximize likelihood, $p\left(Y^{data}|\theta\right)$:

$$
\begin{aligned}
p\left(Y^{data}|\theta\right) &= p\left(Y_1^{data}, ..., Y_T^{data}|\theta\right) \\
&= p\left(Y_1^{data}|\theta\right) \times p\left(Y_2^{data}|Y_1^{data}, \theta\right) \\
&\quad \times \cdots \times \overbrace{p\left(Y_t^{data}|Y_{t-1}^{data} \cdots Y_1^{data}, \theta\right)}^{\text{computed using Kalman Filter}} \\
&\quad \times \cdots \times p\left(Y_T^{data}|Y_{T-1}^{data}, \cdots, Y_1^{data}, \theta\right)
\end{aligned}
$$

- Kalman filter straightforward (see, e.g., Hamilton's textbook).

Bayesian Inference

- Bayesian inference is about describing the mapping from prior beliefs about θ, summarized in $p(\theta)$, to new posterior beliefs in the light of observing the data, Y^{data}.
- General property of probabilities:

$$p\left(Y^{data},\theta\right) = \begin{cases} p\left(Y^{data}|\theta\right) \times p\left(\theta\right) \\ p\left(\theta|Y^{data}\right) \times p\left(Y^{data}\right) \end{cases},$$

which implies Bayes' rule:

$$p\left(\theta|Y^{data}\right) = \frac{p\left(Y^{data}|\theta\right) p\left(\theta\right)}{p\left(Y^{data}\right)},$$

mapping from prior to posterior induced by Y^{data}.

Bayesian Inference

- Report features of the posterior distribution, $p\left(\theta|Y^{data}\right)$.
  - The value of θ that maximizes $p\left(\theta|Y^{data}\right)$, 'mode' of posterior distribution.
  - Compare marginal prior, $p\left(\theta_i\right)$, with marginal posterior of individual elements of θ, $g\left(\theta_i|Y^{data}\right)$:

$$g\left(\theta_i|Y^{data}\right) = \int_{\theta_{j\neq i}} p\left(\theta|Y^{data}\right) d\theta_{j\neq i} \text{ (multiple integration!!)}$$

  - Probability intervals about the mode of θ ('Bayesian confidence intervals'), need $g\left(\theta_i|Y^{data}\right)$.

- Marginal likelihood for assessing model 'fit':

$$p\left(Y^{data}\right) = \int_{\theta} p\left(Y^{data}|\theta\right) p\left(\theta\right) d\theta \text{ (multiple integration)}$$

Markov Chain, Monte Carlo (MCMC) Algorithms

- Among the top 10 algorithms "with the greatest influence on the development and practice of science and engineering in the 20th century".
  - Reference: January/February 2000 issue of Computing in Science & Engineering, a joint publication of the American Institute of Physics and the IEEE Computer Society.'

- Developed in 1946 by John von Neumann, Stan Ulam, and Nick Metropolis (see http://www.siam.org/pdf/news/637.pdf)

MCMC Algorithm: Overview

- compute a sequence, $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}$, of values of the $N \times 1$ vector of model parameters in such a way that

$$\lim_{M \to \infty} Frequency \left[\theta^{(i)} \text{ close to } \theta \right] = p \left(\theta | Y^{data} \right).$$

- Use $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}$ to obtain an approximation for
  - $E\theta$, $Var(\theta)$ under posterior distribution, $p\left(\theta | Y^{data}\right)$
  - $g\left(\theta^{i} | Y^{data}\right) = \int_{\theta_{i \neq j}} p\left(\theta | Y^{data}\right) d\theta d\theta$
  - $p\left(Y^{data}\right) = \int_{\theta} p\left(Y^{data} | \theta\right) p(\theta) d\theta$
  - posterior distribution of any function of θ, $f(\theta)$ (e.g., impulse responses functions, second moments).
- MCMC also useful for computing posterior mode, $\arg\max_{\theta} p\left(\theta | Y^{data}\right)$.

MCMC Algorithm: setting up

- Let $G(\theta)$ denote the log of the posterior distribution (excluding an additive constant):

$$G(\theta) = \log p\left(Y^{data}|\theta\right) + \log p(\theta);$$

- Compute posterior mode:

$$\theta^* = \arg\max_{\theta} G(\theta).$$

- Compute the positive definite matrix, V :

$$V \equiv \left[-\frac{\partial^2 G(\theta)}{\partial\theta\partial\theta'}\right]^{-1}_{\theta=\theta^*}$$

- Later, we will see that V is a rough estimate of the variance-covariance matrix of θ under the posterior distribution.

MCMC Algorithm: Metropolis-Hastings

- $\theta^{(1)} = \theta^*$
- to compute $\theta^{(r)}$, for $r > 1$
  - step 1: select candidate $\theta^{(r)}$, x,

$$\text{draw } \underbrace{x}_{N\times 1} \text{ from } \theta^{(r-1)} + \overbrace{k \times N\left(\underbrace{0}_{N\times 1}, V\right)}^{\text{'jump' distribution'}}, \; k \text{ is a scalar}$$

  - step 2: compute scalar, λ :

$$\lambda = \frac{p\left(Y^{data}|x\right) p\left(x\right)}{p\left(Y^{data}|\theta^{(r-1)}\right) p\left(\theta^{(r-1)}\right)}$$

  - step 3: compute $\theta^{(r)}$:

$$\theta^{(r)} = \begin{cases} \theta^{(r-1)} & \text{if } u > \lambda \\ x & \text{if } u < \lambda \end{cases}, \; u \text{ is a realization from uniform } [0,1]$$

Practical issues

- What is a sensible value for k?
  - set k so that you accept (i.e., $\theta^{(r)} = x$) in step 3 of MCMC algorithm are roughly 23 percent of time
- What value of M should you set?
  - want 'convergence', in the sense that if M is increased further, the econometric results do not change substantially
  - in practice, $M = 10,000$ (a small value) up to $M = 1,000,000$.
  - large M is time-consuming.
    - could use Laplace approximation (after checking its accuracy) in initial phases of research project.
    - more on Laplace below.
- Burn-in: in practice, some initial $\theta^{(i)}$'s are discarded to minimize the impact of initial conditions on the results.
- Multiple chains: may promote efficiency.
  - increase independence among $\theta^{(i)}$'s.
  - can do MCMC utilizing parallel computing (Dynare can do this).

MCMC Algorithm: Why Does it Work?

- Proposition that MCMC works may be surprising.
  - Whether or not it works does *not* depend on the details, i.e., precisely how you choose the jump distribution (of course, you had better use $k > 0$ and V positive definite).
    - Proof: see, e.g., Robert, C. P. (2001), *The Bayesian Choice*, Second Edition, New York: Springer-Verlag.
  - The details may matter by improving the efficiency of the MCMC algorithm, i.e., by influencing what value of M you need.
- Some Intuition
  - the sequence, $\theta^{(1)}, \theta^{(2)}, ..., \theta^{(M)}$, is relatively heavily populated by θ's that have high probability and relatively lightly populated by low probability θ's.
  - Additional intuition can be obtained by positing a simple scalar distribution and using MATLAB to verify that MCMC approximates it well (see, e.g., question 2 in assignment 9).

MCMC Algorithm: using the Results

- To approximate marginal posterior distribution, $g\left(\theta_i | Y^{data}\right)$, of θ_i,
  - compute and display the histogram of $\theta_i^{(1)}, \theta_i^{(2)}, ..., \theta_i^{(M)}$, $i = 1, ..., M$.
- Other objects of interest:
  - mean and variance of posterior distribution θ :

$$E\theta \simeq \bar{\theta} \equiv \frac{1}{M} \sum_{j=1}^{M} \theta^{(j)}, \; Var\left(\theta\right) \simeq \frac{1}{M} \sum_{j=1}^{M} \left[\theta^{(j)} - \bar{\theta}\right] \left[\theta^{(j)} - \bar{\theta}\right]'.$$

MCMC Algorithm: using the Results

- More complicated objects of interest:
  - impulse response functions,
  - model second moments,
  - forecasts,
  - Kalman smoothed estimates of real rate, natural rate, etc.
- All these things can be represented as non-linear functions of the model parameters, i.e., $f(\theta)$.
  - can approximate the distribution of $f(\theta)$ using

$$f\left(\theta^{(1)}\right), \ldots, f\left(\theta^{(M)}\right)$$

$$\begin{aligned} &\rightarrow \quad Ef(\theta) \simeq \bar{f} \equiv \frac{1}{M}\sum_{i=1}^{M} f\left(\theta^{(i)}\right), \\ Var\left(f(\theta)\right) &\simeq \quad \frac{1}{M}\sum_{i=1}^{M}\left[f\left(\theta^{(i)}\right)-\bar{f}\right]\left[f\left(\theta^{(i)}\right)-\bar{f}\right]' \end{aligned}$$

MCMC: Remaining Issues

- In addition to the first and second moments already discused, would also like to have the marginal likelihood of the data.
- Marginal likelihood is a Bayesian measure of model fit.

MCMC Algorithm: the Marginal Likelihood

- Consider the following sample average:

$$\frac{1}{M}\sum_{j=1}^{M}\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)},$$

  where $h(\theta)$ is an arbitrary density function over the $N-$ dimensional variable, θ.

By the law of large numbers,

$$\frac{1}{M}\sum_{j=1}^{M}\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)}\underset{M\rightarrow\infty}{\rightarrow}E\left(\frac{h(\theta)}{p\left(Y^{data}|\theta\right)p(\theta)}\right)$$

MCMC Algorithm: the Marginal Likelihood

$$\frac{1}{M}\sum_{j=1}^{M}\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)}\rightarrow_{M\rightarrow\infty}E\left(\frac{h\left(\theta\right)}{p\left(Y^{data}|\theta\right)p\left(\theta\right)}\right)$$

$$=\int_{\theta}\left(\frac{h\left(\theta\right)}{p\left(Y^{data}|\theta\right)p\left(\theta\right)}\right)\frac{p\left(Y^{data}|\theta\right)p\left(\theta\right)}{p\left(Y^{data}\right)}d\theta \quad =\frac{1}{p\left(Y^{data}\right)}.$$

- When $h(\theta) = p(\theta)$, *harmonic mean estimator of the marginal likelihood*.
- Ideally, want an h such that the variance of

$$\frac{h\left(\theta^{(j)}\right)}{p\left(Y^{data}|\theta^{(j)}\right)p\left(\theta^{(j)}\right)}$$

  is small (recall the earlier discussion of Monte Carlo integration). More on this below.