

Macroeconometrics

Lecture 17 Bayesian estimation using simulation smoother

Tomasz Woźniak

Department of Economics
University of Melbourne

A simple Unobserved Component model

Prior distributions

Derivation of Gibbs sampler

Simulation smoother

Compulsory readings:

Woźniak (2021) Bayesian estimation of simple Unobserved Component models using simulation smoother

Useful readings:

Woźniak (2021) Simulation Smoother using RcppArmadillo, Rcpp Gallery

Objectives.

- ▶ To present Bayesian estimation of state-space models
- ▶ To set the model specification through a prior distribution
- ▶ To derive a Gibbs sampler for Unobserved Component models

Learning outcomes.

- ▶ Deriving full conditional posterior distributions
- ▶ Implementing a simulation smoother
- ▶ Programming a sampler from multivariate normal distribution with special type of the mean and covariance

A simple Unobserved Component model

A simple Unobserved Component model

$$y_t = \tau_t + \epsilon_t \quad (1)$$

$$\tau_t = \mu + \tau_{t-1} + \eta_t \quad (2)$$

$$\epsilon_t = \alpha_1 \epsilon_{t-1} + \cdots + \alpha_p \epsilon_{t-p} + e_t \quad (3)$$

$$\begin{bmatrix} \eta_t \\ e_t \end{bmatrix} \Big| Y_{t-1} \sim \text{iid} \mathcal{N} \left(\mathbf{0}_2, \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix} \right)$$

$$\alpha_p(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$$

$$\alpha_p(z) = 0 : \quad |z| > 1 \quad \forall z \in \mathbb{C}$$

$$\sigma_{\eta e} = \text{Cov}[\eta_t, e_t] = 0$$

τ_0 – an estimated parameter

$$\mathbf{0}_p = (\epsilon_0, \epsilon_{-1}, \dots, \epsilon_{-p+1})'$$

A simple Unobserved Component model

Define $T \times 1$ matrices.

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \quad \tau = \begin{bmatrix} \tau_1 \\ \vdots \\ \tau_T \end{bmatrix} \quad \epsilon = \epsilon_{[1:T]} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_T \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix} \quad I_T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad e_{1:T} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Define matrices.

$$\underset{(2 \times 1)}{\beta} = [\mu \quad \tau_0]' \quad \underset{(T \times 2)}{X_\tau} = [I_T \quad e_{1:T}]$$

$$\underset{(p \times 1)}{\alpha} = [\alpha_1 \quad \dots \quad \alpha_p]' \quad \underset{(T \times p)}{X_\epsilon} = [\epsilon_{[0:(T-1)]} \quad \epsilon_{[-1:(T-2)]} \quad \dots \quad \epsilon_{[(-p+1):(T-p)]}]$$

Define $T \times T$ matrices.

$$H = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \quad H_\alpha = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -\alpha_1 & 1 & 0 & \dots & 0 & 0 \\ -\alpha_2 & -\alpha_1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\alpha_p & -\alpha_{p-1} & -\alpha_{p-2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\alpha_1 & 1 \end{bmatrix}$$

A simple Unobserved Component model

$$y = \tau + \epsilon \quad (1)$$

$$H\tau = \mu I_T + \tau_0 e_{1..T} + \eta \quad (2)$$

$$H\tau = X_\tau \beta + \eta \quad (2)$$

$$\tau = H^{-1}X_\tau \beta + H^{-1}\eta \quad (2)$$

$$H_\alpha \epsilon = e \quad (3)$$

$$\epsilon = H_\alpha^{-1}e \quad (3)$$

$$= X_\epsilon \alpha + e \quad (3)$$

$$\eta \sim \mathcal{N}(\mathbf{0}_T, \sigma_\eta^2 I_T)$$

$$e \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 I_T)$$

To be estimated: $\tau, \epsilon, \beta = (\mu, \tau_0), \alpha, \sigma = (\sigma_\eta^2, \sigma_e^2)$

Prior distributions

Prior distributions

Prior distribution of τ is specified by equation (2)

$$\tau = H^{-1}X_{\tau}\beta + H^{-1}\eta$$

$$\eta \sim \mathcal{N}(\mathbf{0}_T, \sigma_{\eta}^2 I_T)$$

$$H^{-1}\eta \sim \mathcal{N}(\mathbf{0}_T, \sigma_{\eta}^2 (H'H)^{-1})$$

\downarrow

$$\tau|\beta, \sigma_{\eta}^2 \sim \mathcal{N}_T(H^{-1}X_{\tau}\beta, \sigma_{\eta}^2 (H'H)^{-1})$$

$$\propto \exp\left\{-\frac{1}{2}\frac{1}{\sigma_{\eta}^2}(\tau - H^{-1}X_{\tau}\beta)' H'H(\tau - H^{-1}X_{\tau}\beta)\right\}$$

Prior distributions

Prior distribution of ϵ is specified by equation (3)

$$\epsilon = H_{\alpha}^{-1}e$$

$$e \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 I_T)$$

$$H_{\alpha}^{-1}e \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 (H'_{\alpha} H_{\alpha})^{-1})$$

$$\downarrow$$

$$\epsilon | \alpha, \sigma_e^2 \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 (H'_{\alpha} H_{\alpha})^{-1})$$

$$\propto \exp \left\{ -\frac{1}{2} \frac{1}{\sigma_e^2} \epsilon' H'_{\alpha} H_{\alpha} \epsilon \right\}$$

Prior distributions

Prior distributions of $\alpha, \beta, \sigma_\eta^2, \sigma_e^2$ are assumed to be

$$\alpha \sim \mathcal{N}_p(\underline{\alpha}, \underline{V}_\alpha) \mathcal{I}(\alpha \in A) \propto \exp \left\{ -\frac{1}{2} (\alpha - \underline{\alpha})' \underline{V}_\alpha^{-1} (\alpha - \underline{\alpha}) \right\} \mathcal{I}(\alpha \in A)$$

$$\beta \sim \mathcal{N}_2(\underline{\beta}, \underline{V}_\beta) \propto \exp \left\{ -\frac{1}{2} (\beta - \underline{\beta})' \underline{V}_\beta^{-1} (\beta - \underline{\beta}) \right\}$$

$$\sigma_\eta^2 \sim \text{IG2}(\underline{s}, \underline{\nu}) \propto (\sigma_\eta^2)^{-\frac{\underline{\nu}+2}{2}} \exp \left\{ -\frac{1}{2} \frac{\underline{s}}{\sigma_\eta^2} \right\}$$

$$\sigma_e^2 \sim \text{IG2}(\underline{s}, \underline{\nu}) \propto (\sigma_e^2)^{-\frac{\underline{\nu}+2}{2}} \exp \left\{ -\frac{1}{2} \frac{\underline{s}}{\sigma_e^2} \right\}$$

$\alpha \in A$ – set of parameters α for which stationarity holds and

$$\mathcal{I}(\alpha \in A) = \begin{cases} 1 & \text{if } \alpha \in A \\ 0 & \text{otherwise} \end{cases}$$

Prior distributions

$$p(\tau, \epsilon, \alpha, \beta, \sigma) = p(\tau | \beta, \sigma_\eta^2) p(\beta) p(\sigma_\eta^2) p(\epsilon | \alpha, \sigma_e^2) p(\alpha) p(\sigma_e^2)$$

$$\tau | \beta, \sigma_\eta^2 \sim \mathcal{N}(H^{-1} X_\tau \beta, \sigma_\eta^2 (H' H)^{-1})$$

$$\beta \sim \mathcal{N}_2(\underline{\beta}, \underline{V}_\beta)$$

$$\sigma_\eta^2 \sim \text{IG2}(\underline{s}, \underline{\nu})$$

$$\epsilon | \alpha, \sigma_e^2 \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 (H'_\alpha H_\alpha)^{-1})$$

$$\alpha \sim \mathcal{N}_p(\underline{\alpha}, \underline{V}_\alpha) \mathcal{I}(\alpha \in A)$$

$$\sigma_e^2 \sim \text{IG2}(\underline{s}, \underline{\nu})$$

Derivation of Gibbs sampler

Gibbs sampler

is an iterative algorithm at each iteration of which draws are sampled from the following **full-conditional posterior distributions**

$$\tau \sim p(\tau|y, \alpha, \beta, \sigma)$$

$$\epsilon \sim p(\epsilon|y, \alpha, \beta, \sigma)$$

$$\beta \sim p(\beta|y, \tau, \sigma_\eta^2)$$

$$\alpha \sim p(\alpha|y, \epsilon, \sigma_e^2)$$

$$\sigma_\eta^2 \sim p(\sigma_\eta^2|y, \tau, \beta)$$

$$\sigma_e^2 \sim p(\sigma_e^2|y, \epsilon, \alpha)$$

Full-conditional posterior distribution of $\tau|y, \alpha, \beta, \sigma$

Conditional likelihood is based on equation (1)

$$\tau = y - \epsilon$$

$$\epsilon \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 (H'_\alpha H_\alpha)^{-1})$$

$$L(\tau|y, \alpha, \sigma_e^2) \propto \exp\left\{-\frac{1}{2}\sigma_e^{-2}(\tau - y)' H'_\alpha H_\alpha (\tau - y)\right\}$$

Prior distribution is given by

$$p(\tau|\beta, \sigma_\eta^2) \propto \exp\left\{-\frac{1}{2}\frac{1}{\sigma_\eta^2}(\tau - H^{-1}X_\tau\beta)' H' H (\tau - H^{-1}X_\tau\beta)\right\}$$

Full-conditional posterior distribution.

$$\begin{aligned} p(\tau|y, \alpha, \beta, \sigma) &\propto L(\tau|y, \alpha, \sigma_e^2) p(\tau|\beta, \sigma_\eta^2) \\ &= \mathcal{N}_T(\bar{\tau}, \bar{V}_\tau) \end{aligned}$$

$$\begin{aligned} \bar{V}_\tau &= [\sigma_e^{-2} H'_\alpha H_\alpha + \sigma_\eta^{-2} H' H]^{-1} \\ \bar{\tau} &= \bar{V}_\tau [\sigma_e^{-2} H'_\alpha H_\alpha y + \sigma_\eta^{-2} H' X_\tau \beta] \end{aligned}$$

Full-conditional posterior distribution of $\epsilon|y, \alpha, \beta, \sigma$

Conditional likelihood is based on equation (1)

$$\epsilon = y - \tau$$

$$y - \tau \sim \mathcal{N}(y - H^{-1}X_{\tau}\beta, \sigma_{\eta}^2(H'H)^{-1})$$

$$L(\epsilon|y, \beta, \sigma_{\eta}^2) \propto \exp\left\{-\frac{1}{2}\sigma_{\eta}^{-2}\left(\epsilon - (y - H^{-1}X_{\tau}\beta)\right)' H'H\left(\epsilon - (y - H^{-1}X_{\tau}\beta)\right)\right\}$$

Prior distribution is given by

$$p(\epsilon|\alpha, \sigma_e^2) \propto \exp\left\{-\frac{1}{2}\frac{1}{\sigma_e^2}\epsilon'H'_{\alpha}H_{\alpha}\epsilon\right\}$$

Full-conditional posterior distribution.

$$\begin{aligned} p(\epsilon|y, \alpha, \beta, \sigma) &\propto L(\epsilon|y, \beta, \sigma_{\eta}^2)p(\epsilon|\alpha, \sigma_e^2) \\ &= \mathcal{N}_T(\bar{\epsilon}, \bar{V}_{\epsilon}) \end{aligned}$$

$$\begin{aligned} \bar{V}_{\epsilon} &= [\sigma_e^{-2}H'_{\alpha}H_{\alpha} + \sigma_{\eta}^{-2}H'H]^{-1} \\ \bar{\epsilon} &= \bar{V}_{\epsilon}\sigma_{\eta}^{-2}H'H(y - H^{-1}X_{\tau}\beta) \end{aligned}$$

Full-conditional posterior distribution of $\beta|y, \tau, \sigma_\eta^2$

Conditional likelihood is based on equation (2)

$$H\tau - X_\tau\beta = \eta$$

$$\eta \sim \mathcal{N}(\mathbf{0}_T, \sigma_\eta^2 I_T)$$

$$L(\beta|y, \tau, \sigma_\eta^2) \propto \exp\left\{-\frac{1}{2}\sigma_\eta^{-2}(X_\tau\beta - H\tau)'(X_\tau\beta - H\tau)\right\}$$

Prior distribution is given by

$$p(\beta) \propto \exp\left\{-\frac{1}{2}(\beta - \underline{\beta})'\underline{V}_\beta^{-1}(\beta - \underline{\beta})\right\}$$

Full-conditional posterior distribution.

$$\begin{aligned} p(\beta|y, \tau, \sigma_\eta^2) &\propto L(\beta|y, \tau, \sigma_\eta^2) p(\beta) \\ &= \mathcal{N}_2(\bar{\beta}, \bar{V}_\beta) \end{aligned}$$

$$\bar{V}_\beta = [\sigma_\eta^{-2} X_\tau' X_\tau + \underline{V}_\beta^{-1}]^{-1}$$

$$\bar{\beta} = \bar{V}_\beta [\sigma_\eta^{-2} X_\tau' H\tau + \underline{V}_\beta^{-1} \underline{\beta}]$$

Full-conditional posterior distribution of $\alpha|y, \epsilon, \sigma_e^2$

Conditional likelihood is based on equation (3)

$$\epsilon - X_\epsilon \alpha = e$$

$$e \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 I_T)$$

$$L(\alpha|y, \epsilon, \sigma_e^2) \propto \exp\left\{-\frac{1}{2}\sigma_e^{-2}(X_\epsilon \alpha - \epsilon)'(X_\epsilon \alpha - \epsilon)\right\}$$

Prior distribution is given by

$$p(\alpha) \propto \exp\left\{-\frac{1}{2}(\alpha - \underline{\alpha})' \underline{V}_\alpha^{-1}(\alpha - \underline{\alpha})\right\} \mathcal{I}(\alpha \in A)$$

Full-conditional posterior distribution.

$$\begin{aligned} p(\alpha|y, \epsilon, \sigma_e^2) &\propto L(\alpha|y, \epsilon, \sigma_e^2) p(\alpha) \mathcal{I}(\alpha \in A) \\ &= \mathcal{N}_p(\bar{\alpha}, \bar{V}_\alpha) \mathcal{I}(\alpha \in A) \end{aligned}$$

$$\begin{aligned} \bar{V}_\alpha &= [\sigma_e^{-2} X_\epsilon' X_\epsilon + \underline{V}_\alpha^{-1}]^{-1} \\ \bar{\alpha} &= \bar{V}_\alpha [\sigma_e^{-2} X_\epsilon' \epsilon + \underline{V}_\alpha^{-1} \underline{\alpha}] \end{aligned}$$

Full-conditional posterior distribution of $\sigma_\eta^2|y, \tau, \beta$

Conditional likelihood is based on equation (2)

$$H\tau - X_\tau\beta = \eta$$

$$\eta \sim \mathcal{N}(\mathbf{0}_T, \sigma_\eta^2 I_T)$$

$$L(\sigma_\eta^2|y, \tau, \beta) \propto (\sigma_\eta^2)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sigma_\eta^{-2} (X_\tau\beta - H\tau)' (X_\tau\beta - H\tau)\right\}$$

Prior distribution is given by

$$p(\sigma_\eta^2) \propto (\sigma_\eta^2)^{-\frac{\underline{\nu}+2}{2}} \exp\left\{-\frac{1}{2} \frac{\underline{s}}{\sigma_\eta^2}\right\}$$

Full-conditional posterior distribution.

$$p(\sigma_\eta^2|y, \tau, \beta) \propto L(\sigma_\eta^2|y, \tau, \beta) p(\sigma_\eta^2)$$

$$= \mathcal{IG2}(\bar{s}_\eta, \bar{\nu}_\eta)$$

$$\bar{s}_\eta = \underline{s} + (H\tau - X_\tau\beta)'(H\tau - X_\tau\beta)$$

$$\bar{\nu}_\eta = \underline{\nu} + T$$

Full-conditional posterior distribution of $\sigma_e^2|y, \epsilon, \alpha$

Conditional likelihood is based on equation (3)

$$\epsilon - X_\epsilon \alpha = e$$

$$e \sim \mathcal{N}(\mathbf{0}_T, \sigma_e^2 I_T)$$

$$L(\sigma_e^2|y, \epsilon, \alpha) \propto (\sigma_e^2)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sigma_e^{-2} (X_\epsilon \alpha - \epsilon)' (X_\epsilon \alpha - \epsilon)\right\}$$

Prior distribution is given by

$$p(\sigma_e^2) \propto (\sigma_e^2)^{-\frac{\nu+2}{2}} \exp\left\{-\frac{1}{2} \frac{\underline{s}}{\sigma_\eta^2}\right\}$$

Full-conditional posterior distribution.

$$\begin{aligned} p(\sigma_e^2|y, \epsilon, \alpha) &\propto L(\sigma_e^2|y, \epsilon, \alpha) p(\sigma_e^2) \\ &= \mathcal{IG2}(\bar{s}_e, \bar{\nu}_e) \\ \bar{s}_e &= \underline{s} + (\epsilon - X_\epsilon \alpha)' (\epsilon - X_\epsilon \alpha) \\ \bar{\nu}_e &= \underline{\nu} + T \end{aligned}$$

Simulation smoother

Simulation smoother

Sampling from a multivariate normal distribution using a precision matrix

$$\mathcal{N}(D^{-1}b, D^{-1})$$

D^{-1} – an $N \times N$ covariance matrix

D – an $N \times N$ precision matrix that is a tridiagonal

$O(N^3)$ operations needed to invert D with usual computing routines

$O(N)$ operations are required to invert D using dedicated routines for special types of matrices

Simulation smoother

$$\mathcal{N}(D^{-1}b, D^{-1})$$

Let $L = \text{chol}(D)$ be a lower-triangular matrix such that $D = LL'$

Suppose that L^{-1} can be computed efficiently

Compute the mean of the distribution by:

$$L^{-1'}L^{-1}b = (LL')^{-1}b = D^{-1}b$$

Let x denote an $N \times 1$ vector with elements drawn independently from a standard normal distribution

Sample a draw from the target normal distribution

$$L^{-1'}(L^{-1}b + x)$$

The method in the next slide further simplifies the algorithm and bypasses inverting L

Simulation smoother

Let $L \setminus b$ denote the unique solution to the triangular system $Lx = b$ obtained by forward (backward) substitution, that is, $L \setminus b = L^{-1}b$.

Simulation smoother.

Compute $L = \text{chol}(D)$ such that $D = LL'$

Sample $x \sim \mathcal{N}(0_{N \times 1}, I_N)$

Compute a draw from the distribution via the affine transformation:

$$L' \setminus (L \setminus b + x)$$

Simulation smoother in R for tridiagonal D matrix

```
library(mgcv)

N          = dim(D)[1]
lead.diag  = diag(D)
sub.diag   = sdiag(D, -1)

D.chol     = trichol(ld = lead.diag, sd=sub.diag)
D.L        = diag(D.chol$ld)
sdiag(D.L, -1) = D.chol$sd

x          = matrix(rnorm(n*N), ncol=n)
a          = forwardsolve(D.L, b)
draw       = backsolve(t(D.L), a + x)
```

Simulation smoother in R comparison

```
rmvnorm.tridiag.precision = function(n, D, b){  
  N          = dim(D)[1]  
  lead.diag  = diag(D)  
  sub.diag   = sdiag(D, -1)  
  
  D.chol     = trichol(ld = lead.diag, sd=sub.diag)  
  D.L        = diag(D.chol$ld)  
  sdiag(D.L,-1) = D.chol$sd  
  
  x          = matrix(rnorm(n*N), ncol=n)  
  a          = forwardsolve(D.L, b)  
  draw       = backsolve(t(D.L),  
                        matrix(rep(a,n), ncol=n) + x)  
  
  return(draw)  
}
```

Simulation smoother in R comparison

```
rmvnorm.usual = function(n, D, b){  
  N          = dim(D)[1]  
  D.chol     = t(chol(D))  
  variance.chol = solve(D.chol)  
  
  x          = matrix(rnorm(n*N), ncol=n)  
  draw       = t(variance.chol) %*%  
              (matrix(rep(variance.chol%*%b,n), ncol=n) + x)  
  
  return(draw)  
}
```

Simulation smoother in R comparison

```
library(mgcv); library(microbenchmark)
```

```
set.seed(12345)
```

```
T      = 240
```

```
md      = rgamma(T, shape=10, scale=10)
```

```
od      = rgamma(T-1, shape=10, scale=1)
```

```
D       = 2*diag(md)
```

```
sdiag(D, 1) = -od
```

```
sdiag(D, -1) = -od
```

```
b       = as.matrix(rnorm(T))
```

```
microbenchmark(
```

```
  trid  = rmvnorm.tridiag.precision(n=100, D=D, b=b),
```

```
  usual = rmvnorm.nothing.special(n=100, D=D, b=b),
```

```
  check = "equal", setup=set.seed(123456)
```

```
)
```

```
Unit: milliseconds
```

expr	min	lq	mean	median	uq	max	neval
trid	3.489267	4.539225	6.655018	4.666931	4.963046	154.70131	100
usual	13.769023	15.355094	16.746697	15.694483	17.626655	28.79406	100

Simulation smoother in R comparison

```
set.seed(12345)
T      = 720
md     = rgamma(T, shape=10, scale=10)
od     = rgamma(T-1, shape=10, scale=1)
D      = 2*diag(md)
sdiag(D, 1) = -od
sdiag(D, -1) = -od
b      = as.matrix(rnorm(T))
```

```
microbenchmark(
  trid  = rmvnorm.tridiag.precision(n=100, D=D, b=b),
  usual = rmvnorm.nothing.special(n=100, D=D, b=b),
  check = "equal", setup=set.seed(123456)
)
```

Unit: milliseconds

expr	min	lq	mean	median	uq	max	neval
trid	22.46832	26.55893	45.00241	29.93325	35.21184	164.7311	100
usual	249.23538	263.19570	286.56741	270.97134	289.85116	583.5321	100

Bayesian estimation of Unobserved Component models

proceeds via Gibbs sampling with well-specified full conditional posterior distributions

bypasses the application of Kalman filter that requires sequential computer calculations

uses the simulation smoother instead in order to sample draws of latent processes

applies dedicated algorithms for special types of matrices for fast computations