Solving and Estimating a MS-DSGE Model

Francesco Bianchi

Duke University
Bayesian Econometrics is based on the Bayes Theorem:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

Given a set of parameters, \( \theta \in \Theta \subseteq R^k \), a model \( M \), and the data, \( Y^T = \{ Y_t \}_{t=1}^T \), define:

- The **likelihood function**: \( \ell(\theta|Y^T, M) \propto p(Y^T|\theta, M) \)
- The **prior distribution**: \( p(\theta|M) \)
- The **marginal data density**: \( p(Y^T|M) = \int p(Y^T|\theta, M) p(\theta|M) d\theta \)

Then the posterior, \( p(\theta|Y^T, M) \), is given by:

\[ p(\theta|Y^T, M) = \frac{p(Y^T|\theta, M)p(\theta|M)}{p(Y^T|M)} \propto \ell(\theta|Y^T, M)p(\theta|M) \]
The prior $p(\theta|M)$ contains the a-priori researchers’ beliefs about the value of the parameters. It can contain information that comes...

- ...from previous estimates
- ...economic theory ("A model is a probability distribution over outcomes"...)
- ..."arbitrary" researchers’ beliefs
- ...other sources (for example, in the case of macro-models, from micro studies)
The Likelihood

- $p\left(\tilde{Y}^T | \theta, M\right)$ is the observable density, i.e. the probability of the data for a given model $M$ and a vector of parameters $\theta$
- $Y^T$ is the observable realization of the random vector $\tilde{Y}^T$
- The likelihood function is any function $\ell\left(\theta | Y^T, M\right) \propto p\left(Y^T | \theta, M\right)$
- In other words, the likelihood function is a kernel of $p\left(Y^T | \theta, M\right)$, i.e. it is proportional to the probability of the data $Y^T$ for a particular set of parameters $\theta$ and a model $M$
- Note that $p\left(Y^T | \theta, M\right)$ is a specific number only after we observe a realization of $\tilde{Y}^T$
Combining the data density with the prior density we obtain a joint distribution for parameters and data:

\[ p \left( \tilde{Y}^T, \theta | M \right) = p \left( \tilde{Y}^T | \theta, M \right) p(\theta | M) \]  

(1)

The marginal density in \( \tilde{Y}^T \) is obtained integrating w.r.t. \( \theta \):

\[ p \left( \tilde{Y}^T | M \right) = \int_{\Theta} p \left( \tilde{Y}^T, \theta | M \right) d\theta \]  

(2)

This represents the a priori density of the data implied by model \( M \).
The Marginal data density

- If we replace $\tilde{Y}^T$ with $Y^T$ in (2), we obtain the marginal data density or **marginal likelihood**:

$$p \left( Y^T | M \right) = \int_\Theta p \left( Y^T, \theta | M \right) d\theta = \int_\Theta p \left( Y^T | \theta, M \right) p(\theta | M) d\theta$$  \hspace{1cm} (3)

- The term marginal likelihood comes from the fact that if we define the likelihood as $\ell \left( \theta | Y^T, M \right) = p \left( Y^T | \theta, M \right)$, (3) becomes:

$$p \left( Y^T | M \right) = \int_\Theta \ell \left( \theta | Y^T, M \right) p(\theta | M) d\theta$$  \hspace{1cm} (4)

- Note that this is not true in general because the likelihood is defined as any function $\ell \left( \theta | Y^T, M \right)$ such that $\ell \left( \theta | Y^T, M \right) \propto p \left( Y^T | \theta, M \right)$. Therefore (4) requires that all constants of integration are carried forward in $\ell \left( \theta | Y^T, M \right)$.
The posterior $p\left(\theta | Y^T, M\right)$ combines the a-priori information with the information arising from the data:

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

As the name suggests, the posterior contains the *a posteriori* information about $\theta$

Note that:

$$p\left(\theta | Y^T, M\right) \propto \ell \left(\theta | Y^T, M\right) p\left(\theta | M\right)$$
Experiments and Evidence

- An **experiment** is characterized by the triplet $E = \{ \tilde{Y}^T, \theta, p(\tilde{Y}^T|\theta, M) \}$, where a realization of the random variable $\tilde{Y}^T$, whose distribution is given by $p(\tilde{Y}^T|\theta, M)$, is observed.

- The **evidence** about $\theta$ that arises from the experiment $E$ and its realization $\tilde{Y}^T = Y^T$, denoted with $Ev(E, Y^T)$, is any inference, conclusion or report concerning the parameter $\theta$ based on the the experiment $E$ and the realization $\tilde{Y}^T = Y^T$. 
The likelihood principle

- The likelihood principle asserts that all the information about $\theta$ that can be obtained from an experiment is summarized by the likelihood function.

Formally:

**Definition**

**Likelihood principle.**

Given two experiments $E_j = \{\tilde{Y}_j^T, \theta, p(\tilde{Y}_j^T | \theta, M_j)\}$, $j = 1, 2$, and their realizations, $Y_1^T$ and $Y_2^T$, if

$$\ell(\theta; Y_1^T, M_1) \propto \ell(\theta; Y_2^T, M_2)$$

then:

$$Ev_1(E_1, Y_1^T) = Ev_2(E_2, Y_2^T)$$
Consider the following two experiments (Geweke (2005), page 50):

\[
\begin{array}{|c|c|c|c|}
\hline
E_1 & P(y = 1|\theta) & P(y = 2|\theta) & P(y = 3|\theta) \\
\hline
\theta = 0 & .9 & .05 & .05 \\
\theta = 1 & .09 & .055 & .855 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|}
\hline
E_2 & P(y = 1|\theta) & P(y = 2|\theta) & P(y = 3|\theta) \\
\hline
\theta = 0 & 0.260 & 0.730 & 0.010 \\
\theta = 1 & 0.026 & 0.803 & 0.171 \\
\hline
\end{array}
\]

If you were able to choose which experiment to carry out, which one would you choose?
Suppose that we observe $y_1 = 1$ and $y_2 = 1$ 
But $(\ell(\theta = 0 | y_1 = 1, E_1), \ell(\theta = 1 | y_1 = 1, E_1)) = (.9, .09)$ and 
$(\ell(\theta = 0 | y_2 = 1, E_2), \ell(\theta = 1 | y_2 = 1, E_2)) = (.260, .026)$
Then, according to the likelihood principle $Ev_1(E_1, y_1) = Ev_2(E_2, y_2)$
This is true for all values of $y_1$ and $y_2$ (please make sure you can verify this)
According to the likelihood principle whenever $\ell(\theta; Y_1^T, M_1) \propto \ell(\theta; Y_2^T, M_2)$
we have $Ev_1(E_1, y_1) = Ev_2(E_2, y_2)$
Therefore, the two experiments convey the same information.
Consider the following classical test for the null $H_0 : \theta = 0$

\[
\begin{cases}
    \text{accept if } y_i = 1 \\
    \text{reject otherwise}
\end{cases}
\]

It can be verified that this a most powerful test.

However, type I error is 0.1 under $E_1$ against 0.74 under $E_2$ while type II error is 0.09 under $E_1$ against 0.026 under $E_2$.

- Type I error: something that should have been accepted was rejected
- Type II error: something that should have been rejected was accepted

Overall, very different answers from the two experiments.
- Experiment 1 is preferred from an ex-ante prospective
- This is highlighted by the better behavior of the errors
- However, once an observation is available, ex-ante precision is irrelevant
- Bayesian econometrics is based on the posterior probabilities. The likelihood principle always holds
A conjugate prior is a prior that delivers a posterior distribution that belongs to the same parametric family.

It has the same form as a likelihood function computed on a set of dummy observations generated using the same model.

It is extremely convenient because it delivers a known distribution as posterior.

A typical example: The standard linear regression model.

However, in many interesting cases it is not possible to obtain a known distribution for the posterior distribution. An important example:

**DSGE models**
Once linearized, the model can be expressed as a system of equations:

\begin{align}
\tilde{R}_t & = \psi_\pi \tilde{\pi}_t + \psi_y (\tilde{y}_t - \tilde{y}_{t-1}) + \sigma_R \epsilon_{R,t} \\
\tilde{\pi}_t & = \beta E_t (\tilde{\pi}_{t+1}) + \kappa (\tilde{y}_t - \tilde{g}_t - \mu^{-1} \tilde{a}_t) \\
\tilde{y}_t & = E_t (\tilde{y}_{t+1}) - \tau^{-1} (\tilde{R}_t - E_t (\tilde{\pi}_{t+1})) + \left(1 - \rho_g\right) \tilde{g}_t \\
\tilde{a}_t & = \rho_a \tilde{a}_{t-1} + \sigma_a \epsilon_{a,t} \\
\tilde{g}_t & = \rho_g \tilde{g}_{t-1} + \sigma_g \epsilon_{g,t}
\end{align}

where $\beta = 1/r$ and $\kappa = \tau \frac{1-v}{v \varphi \Pi^2}$.
Define the parameter vector $\theta$ as

$$\theta = [\tau, \kappa, \psi_\pi, \psi_y, \rho_g, \rho_z, r, \pi]'$$

and the DSGE state vector $S_t$ and the vector of shocks $\epsilon_t$ as

$$S_t = [\tilde{y}_t, \tilde{\pi}_t, \tilde{R}_t, g_t, z_t, E_t(\tilde{y}_{t+1}), E_t(\tilde{\pi}_{t+1})]'$$
$$\epsilon_t = [\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{a,t}]', \quad \epsilon_t \sim N(0, I), \quad Q = diag(\sigma_R, \sigma_g, \sigma_a)$$
We can rewrite the system of equations (5)-(9) as

\[ \Gamma_0 (\theta) S_t = \Gamma_1 (\theta) S_{t-1} + C + \Psi Q\epsilon_t + \Pi\eta_t \]

where the vector \( \eta_t \) contains the expectation errors (i.e. \( \eta_{t}^x = x_t - E_{t-1} (x_t) \)). The model can be solved using gensys (http://sims.princeton.edu/yftp/gensys/). This returns a first order VAR in the DSGE state vector:

\[ S_t = T(\theta) S_{t-1} + R(\theta) Q\epsilon_t \]  \hspace{1cm} (10)

Here we assume that a solution exists and is unique.
The law of motion of the DSGE state vector can be combined with an observation equation:

\[ X_t = D(\theta) + ZS_t + Uv_t \]  \hspace{1cm} (11)

\[ v_t \sim N(0, I), \quad U = \text{diag} \left( \sigma_y^2, \sigma_\pi^2, \sigma_r^2 \right) \]  \hspace{1cm} (12)

\[ X_t = \begin{bmatrix} GDP_t \\ \text{INFL}_t \\ \text{FFR}_t \end{bmatrix}, \quad D(\theta) = \begin{bmatrix} 0 \\ 4\pi \\ 4(\pi + r) \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \]

where \( v_t \) is a vector of observation errors and \( X_t \) includes HP detrended output, annualized quarterly inflation, and the Federal Funds Rate. Then the \textbf{Kalman filter} is used to evaluate the likelihood \( \ell \left( \theta, M, \sigma_\zeta | Y^T \right) \).
We’ll focus on Bayesian methods (other methods: ML, Calibration, GMM, …)

- The parameters of the model are treated as random variables and the likelihood is combined with a prior distribution
- The prior can incorporate information arising from...
  - previous estimations
  - researcher’s beliefs
  - economic theory
- Inference is based on the posterior distribution
- Different models compared according to the marginal data density
• Metropolis-Hasting provides a way to deal with the fact that we cannot draw directly from the posterior distribution.

• To capture the intuition behind the algorithm, we will consider a simple example.

• Suppose we have an m-state Markov process $x_t$, with states denoted by $S = \{s_1, ..., s_m\}$, and transition probabilities given by $P = [p_{ij}]$, where $p_{ij}$ represents the probability of moving from state $i$ to state $j$.

• Let $\pi_t = [\pi_t(1), ..., \pi_t(m)]$ be the vector containing the probabilities of being in state $i$ at time $t$.

• Then $\pi_{t+1} = \pi_t P$. We say that $\pi$ is the equilibrium distribution if $\pi = \pi P$. 

Francesco Bianchi (Duke University)
A Markov chain is reversible if probability of observing $i \rightarrow j$ is the same as $j \rightarrow i$:

$$\pi_i p_{ij} = \pi_j p_{ji}$$

A chain that is reversible has an equilibrium distribution $\pi$ because

$$(\pi P)_j = \sum_{i=1}^{m} \pi_i p_{ij} = \sum_{i=1}^{m} \pi_j p_{ji} = \pi_j$$

This means that one can start from any $\pi_0$ and eventually the chain will converge to $\pi$
The Metropolis-Hastings algorithm constructs the transition matrix $P$ from a transition matrix $Q$ such that $P$ has the desired equilibrium distribution $\pi$.

$\pi$ is the target distribution (the posterior $p(\theta | Y^T)$). Suppose it is not possible to sample from $\pi$, whereas it is easy to draw from the distribution $Q$.

Suppose that at time $t$ we are in state $s_i$. Based on a draw from $Q$, we should move to state $s_j$ and we do that probability $\alpha_{ij} = \min[1; (\pi_j q_{ji}) / (\pi_i q_{ij})]$.

The chain is reversible and has an equilibrium distribution:

$$\pi_i p_{ij} = \pi_i \min[1; (\pi_j q_{ji}) / (\pi_i q_{ij})]q_{ij}$$

$$= \min[\pi_i q_{ij}; \pi_j q_{ji}] = \pi_j \min[(\pi_i q_{ij}) / (\pi_j q_{ji}); 1]q_{ji} = \pi_j p_{ji}$$

Weak regularity conditions on $Q$ can ensure that the equilibrium distribution $\pi$ is unique and that the chain is convergent.
Suppose the distribution $Q$ is "symmetric", i.e. $q_{ji} = q_{ij}$

One common case: $Q$ is a normal

Suppose that at time $t$ we are in state $s_i$. Based on a draw from $Q$, we should move to state $s_j$. Now the acceptance probability is simply $\alpha_{ij} = \min[1; \pi_j / \pi_i]$.

This was the original algorithm used by Metropolis et al. (1953), whereas the generalization presented before is due to Hastings (1970)
Formally:

- The basic ingredients of the Metropolis-Hastings algorithm are:
  - An arbitrary transition probability density function $q(\vartheta; \theta)$ indexed by $\theta \in \Theta$ and with density argument $\vartheta$
  - An arbitrary starting value $\theta_0 \in \Theta$

- A candidate value $\vartheta$ for $\theta^m$ is drawn from $q(\vartheta; \theta^{m-1})$ and...

- ...it is accepted $(\theta^m = \vartheta)$ with probability

$$\alpha(\theta^m; \vartheta) = \min \left\{ \frac{p(\vartheta)}{p(\theta^{m-1})} \frac{q(\theta^{m-1}; \vartheta)}{q(\vartheta; \theta^{m-1})}, 1 \right\}$$

- If the candidate draw is rejected, set $\theta^m = \theta^{m-1}$
Metropolis Random Walk Algorithm

- When $q(\theta^{m-1}; \vartheta) = q(\vartheta; \theta^{m-1})$ we get the Metropolis Random Walk Algorithm (for example, when the proposal distribution is a normal)
- The candidate value $\vartheta$ is accepted ($\theta^m = \vartheta$) with probability

$$\alpha (\theta^{m-1}; \vartheta) = \min \left\{ \frac{p(\vartheta)}{p(\theta^{m-1})}, 1 \right\}$$

- This is the algorithm that is often used when estimating DSGE models
Implementation of the Metropolis algorithm

The practical implementation involves the following steps:

1. Use a numerical optimization routine to maximize the (log) posterior and find the posterior mode $\bar{\theta}$

2. Draw the starting value $\theta_0$ from $N(\bar{\theta}, c_0 \Sigma)$, where $\Sigma$ is the inverse of the Hessian computed at the posterior mode $\bar{\theta}$ and $c_0$ is a scale factor

3. For $m = 1, ..., nrep$, draw $\vartheta$ from the proposal distribution $N(\theta^{m-1}, c\Sigma)$ where and $c$ is a scale factor used to obtain the desired acceptance rate. The candidate draw is accepted with probability

$$\alpha \left( \theta^{m-1}; \vartheta \right) = \min \left\{ \frac{p(\vartheta | Y^T)}{p(\theta^{m-1} | Y^T)}, 1 \right\}$$
Suppose we are interested in the expected value of a function $h(\theta)$:

$$E \left( h(\theta) \mid Y^T \right) = \int h(\theta) p(\theta \mid Y^T) \, d\theta$$

The draws obtained with the Metropolis algorithm can be used to approximate the integral:

$$E \left( h(\theta) \mid Y^T \right) \approx \frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} h(\theta^i)$$
Evaluate the likelihood: Kalman filter

Two basic ingredients:

1. A transition equation:
   \[ s_t = A s_{t-1} + \epsilon_t \]

2. An observation equation
   \[ y_t = H s_t + \nu_t \]

- Assumptions
  \[
  \begin{bmatrix}
  \epsilon_t \\
  \nu_t 
  \end{bmatrix}
  \sim N
  \begin{bmatrix}
  0 \\
  0 
  \end{bmatrix},
  \begin{bmatrix}
  \Omega & 0 \\
  0 & \Xi 
  \end{bmatrix}
  \]

- \( \epsilon_t \) and \( \nu_t \) independent of past \( y \) and \( s \).
Joint Normals and conditional distributions

Suppose we have a jointly normal random vector split into two pieces, $X_1$ and $X_2$

$$
\begin{bmatrix}
X_1 \\
X_2
\end{bmatrix}
\sim \mathcal{N}
\left(
\begin{bmatrix}
\mu_1 \\
\mu_2
\end{bmatrix},
\begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix}
\right)
$$

Then the conditional of $X_1$ given $X_2$ is

$$
X_1|X_2 \sim \mathcal{N}
\left(
\mu_1 + \beta \left( X_2 - \mu_2 \right), \Omega
\right)
$$

$$
\beta = \Sigma_{12} \Sigma_{22}^{-1}, \quad \Omega = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}
$$
In our case...

- Suppose that given the information at time $t$,

$$ s_t \sim N(\mu_t, \Sigma_t) $$

- Then the joint distribution of $s_t$ and $y_t$ is given by

$$ \begin{bmatrix} s_{t+1} \\ y_{t+1} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} A\Sigma_t A' + \Omega \\ H\Sigma_t A' + H\Omega \end{bmatrix}, \begin{bmatrix} A\mu_t \\ HA\mu_t \end{bmatrix}', \begin{bmatrix} A\Sigma_t A' H' + \Omega H' \\ H\Sigma_t A' H' + H\Omega H' + \Xi \end{bmatrix} \right) $$
Applying the formulas for the conditional distribution of a normal

\[ s_{t+1} \mid y_{t+1} \sim N \left( \mu_{t+1}, \Sigma_{t+1} \right) \]

\[
\mu_{t+1} = A\mu_t + (A\Sigma_t A' H' + \Omega H') \left( HA\Sigma_t A' H' + H\Omega H' + \Xi \right)^{-1} (y_{t+1} - HA\mu_t)
\]

\[
\Sigma_{t+1} = A\Sigma_t A' + \Omega - (A\Sigma_t A' H' + \Omega H') \left( HA\Sigma_t A' H' + H\Omega H' + \Xi \right)^{-1} (HA\Sigma_{11} A' + H\Omega)
\]
At each date, the Kalman filter provides an algorithm to form a normal distribution for $y_{t+1}|\ell_t$, $p(y_{t+1}|\ell_t)$.

Then the pdf of the entire observed sample is

$$p \left( y^T | \ell_0 \right) = \prod_{t=1}^{T} p \left( y_t | \ell_{t-1} \right)$$

The initial distribution $p \left( y_1 | \ell_0 \right)$ is determined by an initial Gaussian prior on the initial state $s_0$. 


Initial conditions

- We can use unconditional mean and covariance matrix when $s_t$ is stationary:

$$s_{0|0} = c + As_{0|0}$$

$$s_{0|0} = (I - A)^{-1} c = 0$$

$$\text{Cov} (s_t) = ACov (s_{t-1}) A' + \Omega$$

$$\Sigma_{0|0} = A\Sigma_{0|0} A' + \Omega$$

$$\text{vec} \left( \Sigma_{0|0} \right) = (A \otimes A) \text{vec} \left( \Sigma_{0|0} \right) + \text{vec} \left( \Omega \right)$$

$$\text{vec} \left( \Sigma_{0|0} \right) = [I - (A \otimes A)]^{-1} \text{vec} \left( \Omega \right)$$

- Otherwise, make a wild guess for $s_{0|0}$ and choose a very large $\Sigma_{0|0}$
The log likelihood density is then just the sum of the log \( p(y_t|\omega_{t-1}) \) terms:

\[
-.5 (y_t - \hat{y}_t)' \Phi_t^{-1} (y_t - \hat{y}_t) - .5 \log |\Phi_t|
\]

where \( \hat{y}_t = HA\mu_t \) and \( \Phi_t = HA\Sigma_t A' H' + H\Omega H' + \Xi \)

- In DSGE models \( A, \Omega, H, \) and \( \Xi \) are transformations of the underlying parameters entering the model.
- Then, the Kalman Filter is executed to evaluate the likelihood at many values of the underlying parameters.
- In Bayesian econometrics, the likelihood is then combined with the priors on the structural parameters.
The Kalman filter returns an estimate for $s_t|\ell_{t-1}$

However, we might be interested in $s_t|\ell_T$, i.e. an estimate for the underlying state that takes into account all the information available to the econometrician

The Kalman filter returns an estimate for $s_T|\ell_T$

Then a backward procedure can be used to obtain smoothed estimates for the remaining states or to make draws from the posterior
The joint distribution of $s_t$ and $s_{t+1}$

$$\begin{bmatrix} s_{t+1} \\ s_t \end{bmatrix} | \ell_t \sim \mathcal{N} \left( \begin{bmatrix} A\mu_{t|t} \\ \mu_{t|t} \end{bmatrix}, \begin{bmatrix} A\Sigma_{t|t} A' + \Omega & A\Sigma_{t|t} \\ A\Sigma_{t|t} & \Sigma_{t|t} \end{bmatrix} \right)$$

That means that given a draw for $s_{t+1}$, we can derive $s_t | s_{t+1}$
We start drawing $s_T$ from a Normal with mean $\mu_{T|T}$ and variance $\Sigma_{T|T}$. These are returned from the last step of the Kalman filter.

We then proceed backward drawing $s_t|s_{t+1}$ from a Normal with mean $\mu_{t|t,s_{t+1}}$ and covariance matrix $\Sigma_{t|t,s_{t+1}}$.

The value $s_{t+1}$ represents another piece of information originating from the observation equation:

$$s_{t+1} = As_t + \epsilon_{t+1}$$

Applying the formulas for the conditional Normal we have:

$$\mu_{t|t,s_{t+1}} = \mu_{t|t} + \Sigma_{t|t} A' (A \Sigma_{t|t} A' + \Omega)^{-1} (s_{t+1} - A \mu_{t|t})$$

$$\Sigma_{t|t,s_{t+1}} = \Sigma_{t|t} - \Sigma_{t|t} A' (A \Sigma_{t|t} A' + \Omega)^{-1} A \Sigma_{t|t}$$
Smotherer (optional)

- We just showed that given a draw $s_{t+1}$

$$\mu_{t|t,s_{t+1}} = \mu_{t|t} + \Sigma_{t|t} A' \left( A\Sigma_{t|t} A' + \Omega \right)^{-1} (s_{t+1} - A\mu_{t|t})$$

- Then it easy to show that:

$$\mu_{t|t,s_{t+1}} = \mu_{t|T,s_{t+1}}$$

i.e. knowing the entire sample realization of $y^T$ does not add any information if we already know $s_{t+1}$

- Then

$$\mu_{t|T} = \mu_{t|t} + \Sigma_{t|t} A' \left( A\Sigma_{t|t} A' + \Omega \right)^{-1} (\mu_{t+1|T} - A\mu_{t|t})$$
The mean squared error associated to the smoothed estimates is:

\[
 s_t - \mu_{t|T} = s_t - \mu_{t|t} - K_t \left( \mu_{t+1|T} - A\mu_{t|t} \right)
\]

where \( K_t = \Sigma_{t|t} A' \left( A\Sigma_{t|t} A' + \Omega \right)^{-1} \)

Then:

\[
 s_t - \mu_{t|T} + K_t \mu_{t+1|T} = s_t - \mu_{t|t} + K_t A\mu_{t|t}
\]

Taking cross products and expectations:

\[
 E \left[ \left( s_t - \mu_{t|T} \right) \left( s_t - \mu_{t|T} \right)' \right] + K_t E \left[ \mu_{t+1|T} \mu_{t+1|T}' \right] K_t' =
\]

\[
 = E \left[ \left( s_t - \mu_{t|t} \right) \left( s_t - \mu_{t|t} \right)' \right] + K_t E \left[ \mu_{t+1|t} \mu_{t+1|t}' \right] K_t'
\]

From which:

\[
 \Sigma_{t|T} = \Sigma_{t|t} + K_t \left( \Sigma_{t+1|T} - \Sigma_{t+1|t} \right) K_t'
\]
In deriving the last relation we have used the fact that:

\[ - \left[ E \left[ \mu_{t+1|T} \mu_{t+1|T} \right] + E \left[ \mu_{t+1|t} \mu_{t+1|t} \right] \right] \]

\[ = -E \left[ \mu_{t+1|T} \mu_{t+1|T} \right] + E \left[ \mu_{t+1|t} \mu_{t+1|t} \right] + E \left[ s_{t+1} s_{t+1}' \right] - E \left[ s_{t+1} s_{t+1}' \right] \]

\[ = E \left[ (s_{t+1} - \mu_{t+1|T}) (s_{t+1} - \mu_{t+1|T})' \right] - E \left[ (s_{t+1} - \mu_{t+1|t}) (s_{t+1} - \mu_{t+1|t})' \right] \]

\[ = \Sigma_{t+1|T} - \Sigma_{t+1|t} \]
Applications

There are at least three popular applications:

1. Factor models
2. DSGE models
3. Time varying VAR
DSGE model

Observation equation:

\[ y_t = ZS_t + \nu_t \]

Law of motion:

\[ S_t = T(\theta) S_{t-1} + R(\theta) \epsilon_t \]

\[ \epsilon_t \sim N(0, Q) \]

A prototypical new-Keynesian model with regime changes

Once linearized, the model can be expressed as a system of equations:

\[
\begin{align*}
\tilde{R}_t &= \psi_{\pi,\xi_t} \tilde{\pi}_t + \psi_{y,\xi_t} (\tilde{y}_t - \tilde{y}_{t-1}) + \sigma_{R,\xi_t} \epsilon_{R,t} \\
\tilde{\pi}_t &= \beta E_t(\tilde{\pi}_{t+1}) + \kappa(\tilde{y}_t - \tilde{g}_t - \mu^{-1}\tilde{a}_t) \\
\tilde{y}_t &= E_t(\tilde{y}_{t+1}) - \tau^{-1}(\tilde{R}_t - E_t(\tilde{\pi}_{t+1})) + \left(1 - \rho_g\right)\tilde{g}_t \\
\tilde{a}_t &= \rho_a \tilde{a}_{t-1} + \sigma_{a,\xi_t} \epsilon_{a,t} \\
\tilde{g}_t &= \rho_g \tilde{g}_{t-1} + \sigma_{g,\xi_t} \epsilon_{g,t}
\end{align*}
\]

(13, 14, 15, 16, 17)

where \( \beta = 1/r \) and \( \kappa = \frac{\tau(1-v)}{\nu \phi \Pi^2} \). See Davig and Leeper (2007), Davig and Doh (2012), Bianchi (2012)
Allowing for Markov-switching regimes

The unobserved states \( \xi_{sp,t} \) and \( \xi_{vo,t} \) follow two independent Markov chains. The probability of moving from one state to another is given by:

\[
P[\xi_{sp,t} = i | \xi_{sp,t-1} = j] = h_{ij}^{sp} \quad (18)
\]

\[
P[\xi_{vo,t} = i | \xi_{vo,t-1} = j] = h_{ij}^{vo} \quad (19)
\]

The model is now described by (13)-(19).

The parameter set extended to include different regimes for volatilities and monetary policy:

\( \psi_{\pi,1} \cdots \psi_{\pi,m} \)
If we collect all of the structural parameters and the stochastic volatilities in two vectors $\theta^{sp}$ and $\theta^{vo}$ and define the DSGE state vector $S_t$, we can rewrite the system of equations as:

$$
\Gamma_0 \left( \zeta_t^{sp}, \theta^{sp} \right) S_t = \Gamma_1 \left( \zeta_t^{sp}, \theta^{sp} \right) S_{t-1} + \Psi Q \left( \zeta_t^{vo}, \theta^{vo} \right) \epsilon_t + \Pi \eta_t
$$

$Q \left( \zeta_t^{vo}, \theta^{vo} \right) = \text{diag} \left( \theta^{vo} \right)$ is a diagonal matrix collecting the volatilities of the shocks conditional on the regime $\zeta_t^{vo}$.
The model can be solved following Davig and Leeper (2007) or Farmer, Waggoner, and Zha (2010):

\[ S_t = T (\xi^s_t, \theta^s, H^m) S_{t-1} + R (\xi^s_t, \theta^s, H^m) Q (\xi^v_t, \theta^v) \epsilon_t \]

The benchmark model assumes \( H^m = H^s \)

The law of motion can be combined with a system of observation equations to obtain a model cast in state space form:

\[
\begin{align*}
Y_t &= D (\theta^s) + Z S_t + U v_t \\
S_t &= T (\xi^s_t) S_{t-1} + R (\xi^s_t) Q (\xi^v_t) \epsilon_t \\
H_{i,j}^s &= p (\xi^s_t = i | \xi^s_{t-1} = j) \\
H_{i,j}^v &= p (\xi^v_t = i | \xi^v_{t-1} = j)
\end{align*}
\]
Kim’s approximation of the likelihood

- Using the standard Kalman filter is computationally challenging: We need to infer the paths of the Markov chains \( (\zeta^{vo}, T \text{ and } \zeta^{sp}, T) \) and of the DSGE state vector \( (S^T) \)
- A possible solution consists of keeping track of a limited number of paths (Schorfheide (2005))
- Alternatively, Kim’s approximation of the likelihood can be used to keep the number of possible paths low
Gibbs sampling in general

- We are interested in characterizing the distribution $p(\theta)$
- Partition the parameter vector $\theta$ into $B$ blocks: $\theta = [\theta_1, \ldots, \theta_B]$
- The Gibbs sampling algorithm turns out to be useful whenever it is hard to draw directly from $p(\theta)$, whereas it is relatively easy to draw a subset of parameters $\theta_i$ conditioning on the subvector $\theta_{-i} = [\ldots, \theta_{i-1}, \ldots, \theta_{i+1}, \ldots]$
- In the models will consider $p(\theta)$ will always be the posterior $p(\theta|Y^T, M)$, however the methodology can be applied to any density
Gibbs sampling algorithm for MS-DSGE

While \( n < n_{\text{sim}} \):

1. Given \( S_{n-1}^T, \theta_{n-1}^{vo}, \) and \( H_{n-1}^{vo} \), use Bayesian updating to get a filtered estimate of \( \tilde{\zeta}_n^{vo}, T \) and then draw a sequence for \( \tilde{\zeta}_n^{vo}, T \).

2. Given \( \tilde{\zeta}_n^{vo}, T \), \( H_n^{vo} \) can be drawn according to a Dirichlet distribution.

3. Conditional on \( \tilde{\zeta}_n^{vo}, T \), evaluate the likelihood of the state space form model using a modified Kalman filter. Draw \( H_n^{sp} \) and \( \theta_n^{sp} \) using a Metropolis-Hastings algorithm. This step also returns filtered estimates for the joint distribution of \( \tilde{\zeta}_n^{sp}, T \) and \( \tilde{S}_n^T \).

4. Use a backward procedure to draw \( \tilde{\zeta}_n^{sp}, T \) and \( S_n^T \).

5. Conditional on a draw for the DSGE states \( S_n^T \), observation errors and model innovations are observable. Draw each of the elements of \( \theta_n^{vo} \) using an inverse gamma.
Counterfactual simulations

For each draw we can compute a sequence of shocks: $\varepsilon^T$

- Therefore, for each draw we can reconstruct a counterfactual path for the macroeconomic variables changing
  1. Regime sequence (always hawkish)
  2. Policy makers’ behavior (more hawkish when hawkish)
  3. Agents’ beliefs (agents can be more optimistic/pessimistic, other regimes can be introduced)

$$S_t = T \left( \hat{\zeta}^{sp}_t, \hat{\theta}^{sp}_t, \hat{H}^m_t \right) S_{t-1} + R \left( \hat{\zeta}^{sp}_t, \hat{\theta}^{sp}_t, \hat{H}^m_t \right) Q \left( \zeta^{vo}_t, \theta^{vo}_t \right) \varepsilon_t$$

- Counterfactual simulations are robust to the Lucas’ critique because model is re-solved taking into account the change in agents’ information set.