Bayesian inference based only on simulated likelihood

Thomas Flury and Neil Shephard

Oxford-Man Institute, Department of Economics, University of Oxford

24.09.2009
Overview

- Likelihood based inference, simulated maximum likelihood.
- “New” Bayesian approach: static models
- Particle filters
- Dynamic models
Likelihood based inference

• Interested in parameter inference based on likelihoods.
• Problem: An analytical expression is often not available in economic models.
• Solution: Simulated maximum likelihood (SML) estimation.
• This goes back at least to Lerman and Manski (1981) and Diggle and Gratton (1984)
• Want to unbiasedly estimate the likelihood using $M$ independent and identically distributed (i.i.d.) draws.
• Draws are kept common as $\theta$ varies.
• Maximize log of this estimate

$$\hat{\theta}_{M,T} = \arg \max_\theta \sum_{t=1}^{T} \frac{1}{M} \sum_{i=1}^{M} \tilde{f} \left( y_t, x_t, u_t^{(i)} ; \theta \right)$$

• For $M$ fixed, and $T \to \infty$ this is inconsistent.
Drawbacks of SML

• For a sample of size $T$, for i.i.d. data, we need $M \to \infty$ for this maximum simulation likelihood estimator to be consistent.

• and $\sqrt{T}/M \to 0$ to have the same distribution as the maximum likelihood estimator.

• Could have rather large simulation demands.

• How close is $\sqrt{T}/M$ to zero?

• SML requires continuity in the estimated log-likelihood.

• In practice: Outcome varies in function of random numbers used.
New approach

- For many economic models the issue of needing $M$ to be large can be entirely sidestepped — while still maintaining efficiency.
- We first saw this new approach in the context of dynamic models in a paper in statistical theory by Andrieu, Doucet and Holenstein (2009).
- Much simpler than SML. Most convergence speed problems can be sidestepped.
- Price to pay: Bayesian inference.
Bayesian inference

- Object of interest is the posterior distribution of $\theta$ given observed data.
- Bayes theorem

$$f(\theta | y) = \frac{f(y | \theta) f(\theta)}{f(y)} \propto f(y | \theta) f(\theta)$$

- Analytical expression is often not available in economic models:
- Neither for $f(\theta | y)$, nor for $f(y | \theta)$. 
Monte Carlo Bayesian inference

- Solution: approximate $f(\theta|y)$ by a large number of draws from it.
- Monte Carlo integration

$$E^N [h(\theta) \mid y] = \frac{1}{N} \sum_{i=1}^{N} h(\theta^{(i)})$$

- So we can carry out inference by sampling from

$$f(\theta|y) \propto f(y|\theta)f(\theta),$$

where $f(\theta)$ is a prior.
New approach

- We need to calculate the likelihood $f(y|\theta)$, but here we assume that all we have is a simulation based estimator
  \[
  \hat{f}_u(y|\theta),
  \]
- ...which is unbiased
  \[
  E_u \left[ \hat{f}_u(y|\theta) \right] = f(y|\theta),
  \]
  where we average over the simulation denoted by the multivariate $u$.
- This estimator is itself a density function.
New approach

- Think of the simulation estimator $\hat{f}$ as being based on an auxiliary variable:

$$\hat{f}_u(y|\theta) = g(y, u|\theta),$$

- $g$ is a joint density which, when marginalised over $u$, delivers $f(y|\theta)$.
- Massive implications econometrically, because now we can carry out inference by sampling from

$$g(u, \theta|y) \propto g(y, u|\theta)f(\theta),$$

- This simulation based Bayesian method delivers draws

$$\left(u^{(1)}, \theta^{(1)}\right), \left(u^{(2)}, \theta^{(2)}\right), \ldots, \left(u^{(N)}, \theta^{(N)}\right)$$
New approach

- Throwing away the $u$ samples leaves us with

$$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)}$$

which are from $f(\theta|y)$.

- We can use these samples to approximate the posterior median, which is an efficient estimator (in the classical sense) of $\theta$ by the Bernstein-von Mises Theorem.

- The sampling can be carried out using generic Markov chain Monte Carlo (MCMC) algorithms.
Generic MCMC

- Suppose we have $\theta^{(i-1)}$.
- Use a proposal density $q(\theta^{(i)} | \theta^{(i-1)})$ to sample $\theta^{(i)}$.
- Decide if or not to accept the proposed $\theta^{(i)}$.
- The acceptance probability is given by

$$q = \min \left[ \frac{f(\theta^{(i)} | y) q(\theta^{(i-1)} | \theta^{(i)})}{f(\theta^{(i-1)} | y) q(\theta^{(i)} | \theta^{(i-1)})}, 1 \right]$$
• Or using Bayes:

\[
q = \min \left[ \frac{f(y | \theta^{(i)})}{f(y | \theta^{(i-1)})} \frac{f(\theta^{(i)})}{f(\theta^{(i-1)})} \frac{q(\theta^{(i-1)} | \theta^{(i)})}{q(\theta(i) | \theta^{(i-1)})}, 1 \right]
\]

• Draw \( V \sim U(0, 1) \) and if \( V > q \) set

\[
\theta^{(i)} = \theta^{(i-1)}.
\]

• Under very weak conditions (e.g. Chib (2001)) the sequence \( \{ \theta^{(i)} \} \)
  for \( i = 1, \ldots, N \) converges to samples from \( f(\theta | y) \) as \( N \to \infty \).
MCMC with estimated likelihood

- The only difference here:
- After drawing \( \theta^{(i)} \), draw the uniformly distributed \( u^{(i)} \) and compute

\[
\hat{L}^{(i)} = \hat{f}_u (y | \theta^{(i)})
\]

- The acceptance probability is given by

\[
q = \min \left[ \frac{\hat{L}^{(i)}}{\hat{L}^{(i-1)}} \frac{f(\theta^{(i)})}{f(\theta^{(i-1)})} \frac{q(\theta^{(i-1)} | \theta^{(i)})}{q(\theta^{(i)} | \theta^{(i-1)})}, 1 \right], \quad V \sim U(0, 1).
\]

- If \( V > q \) set

\[
\left( \hat{L}^{(i)}, \theta^{(i)} \right) = \left( \hat{L}^{(i-1)}, \theta^{(i-1)} \right).
\]
Static model: Binary choice

- Labour force participation of \( T = 753 \) women.
- Data from Mroz (1987).
- Assume that \( y_t = 0 \) if \( y_t^* \leq 0 \) and \( y_t = 1 \) if \( y_t^* > 0 \), where

\[
y_t^* = \beta_0 + \beta_1 \text{nwifeinc}_t + \beta_2 \text{educ}_t + \beta_3 \text{exper}_t + \beta_4 \text{exper}^2_t + \beta_5 \text{age}_t + \beta_6 \text{kidslt6}_t + \beta_7 \text{kidsge6}_t + \varepsilon_t
\]

- We model

\[
\Pr(y_t = 1| x_t, \beta, \psi) = p_t = \Pr(x_t'\beta + \varepsilon_t \geq 0), \quad \varepsilon_t | x_t \sim F_t(\psi),
\]

- and write

\[
p_t = \Pr (x_t'\beta + \varepsilon_t \geq 0) = \Pr (-\varepsilon_t \leq x_t'\beta) = F (x_t'\beta | \psi)
\]
Static model: Binary choice

• For the simulation based estimator of $p_t$ we draw

$$
\varepsilon_t^{(j)} \sim i.i.d. N(0, \sigma^2_\varepsilon) \quad j = 1, \ldots, M
$$

• and compute

$$
\hat{p}_t = \frac{1}{M} \sum_{j=1}^{M} 1_{x_t' \beta + \varepsilon_t^{(j)} \geq 0},
$$

• The estimate of the likelihood is given by

$$
f(y|\beta, \psi) = \prod_{t=1}^{T} \hat{p}_t^{y_t} (1 - \hat{p}_t)^{1-y_t}.
$$

• Use this estimator inside a MCMC algorithm to make inference on $\beta$. 
Static model: Binary choice, $N = 100000$
Static model: Binary choice, $N = 100000$
## Static model: Binary choice

<table>
<thead>
<tr>
<th>Parameter</th>
<th>exact likelihood</th>
<th>MCMC</th>
<th>$M = 1,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>mcse</td>
<td>Pr</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.295</td>
<td>0.033</td>
<td>0.42</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.012</td>
<td>0.000</td>
<td>0.41</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.130</td>
<td>0.001</td>
<td>0.41</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.124</td>
<td>0.001</td>
<td>0.41</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-0.002</td>
<td>0.000</td>
<td>0.41</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-0.053</td>
<td>0.001</td>
<td>0.41</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.868</td>
<td>0.004</td>
<td>0.43</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>0.035</td>
<td>0.001</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Static model: Binary choice

<table>
<thead>
<tr>
<th></th>
<th>$M = 2,000$</th>
<th></th>
<th>$M = 4,000$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean mcse Pr inef</td>
<td></td>
<td>mean mcse Pr inef</td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.349 .035 .33 462</td>
<td></td>
<td>0.376 .030 .36 420</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.012 .000 .33 35</td>
<td></td>
<td>-0.012 .000 .36 42</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.130 .001 .33 209</td>
<td></td>
<td>0.126 .001 .36 168</td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.123 .001 .32 160</td>
<td></td>
<td>0.124 .001 .36 197</td>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-0.002 .000 .33 162</td>
<td></td>
<td>-0.002 .000 .36 189</td>
<td></td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-0.054 .001 .33 458</td>
<td></td>
<td>-0.054 .001 .37 321</td>
<td></td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-0.876 .004 .34 128</td>
<td></td>
<td>-0.871 .004 .37 67</td>
<td></td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>0.032 .002 .33 94</td>
<td></td>
<td>0.032 .001 .36 77</td>
<td></td>
</tr>
</tbody>
</table>
• We assume we have some observations

\[ y = (y_1, y_2, \ldots, y_T) \]

• We want to make Bayesian inference on some unknown parameters \( \theta \).

• We consider an underlying non-linear and non-Gaussian state-space model of the following type:
Dynaimc models: Assumptions

• We can compute the measurement density

\[ f(y_t|\alpha_t, \mathcal{F}_{t-1}, \theta), \quad t = 1, 2, ..., T, \]

where \( \alpha_t \) is the unobserved state and \( \mathcal{F}_{t-1} = y_1, y_2, \ldots, y_{t-1} \) is the natural filtration.

• We can simulate from the random variable

\[ \alpha_t|\alpha_{t-1}, \mathcal{F}_{t-1}, \theta, \quad t = 1, 2, ..., T, \]

where we assumed that we can also draw from the initial condition \( \alpha_0|\mathcal{F}_0, \theta \).

• We can compute the prior \( f(\theta) \).
Dynaimc models: Assumptions

- We do not assume we can compute $f(\alpha_{t+1} | \alpha_t, F_t, \theta)$.
- In economics and finance models we can often only simulate from it.
- We do not assume that such simulations are continuous with respect to $\theta$.
- A large number of intractable econometric models are of this form.
- For example: DSGE models, and some (continuous time) stochastic volatility models.
Dynamic models: Assumptions

- The predictive decomposition

\[ f(y|\mathcal{F}_0, \theta) = \prod_{t=1}^{T} f(y_t|\mathcal{F}_{t-1}, \theta). \]

is key to the success of the Kalman filter and the use of hidden Markov models, where the predictive distributions \( f(y_t|\mathcal{F}_{t-1}, \theta) \) can be computed exactly using recursive formulae.

- In the models considered here we need to use simulation to unbiasedly estimate \( f(y_t|\mathcal{F}_{t-1}, \theta) \).

- This will be carried out using a particle filter, whose recursive structure will allow us to calculate an \textit{unbiased} estimator of \( f(y|\mathcal{F}_0, \theta) \).

- Use this as the basis for inference using an MCMC algorithm; Andrieu, Doucet Holenstein (2009).
Intro to SMC: Classical Monte Carlo

- Want the expectation of a function $h$ of $X \sim f(x \mid y)$.

$$E_f[h(X)] = \int h(x)f(x \mid y)dx$$

- If we can sample $x \sim f(\cdot \mid y)$ then we can use $M$ such (i.i.d.) samples $\{x^{(i)}\}_{i=1}^{M}$ to estimate $E_f[h(X)]$:

$$\hat{I}_M = \frac{1}{M} \sum_{i=1}^{M} h(x^{(i)})$$

- $\hat{I}_M \to E_f[h(X)]$ by the S.L.L.N.

- We typically want to compute the posterior expectation of the hidden state given $y_{0:T}$.

- But we cannot sample from $f(x_{0:T} \mid y_{0:T})$. 
Intro to SMC: Importance Sampling

- It is not necessary for us to sample according to $f$.

$$E_f[h(X)] = \int h(x)f(x|y)dx = \int h(x)\frac{f(x|y)}{g(x)}g(x)dx$$

as long as $f(x|y) > 0 \Rightarrow g(x) > 0$.

- If we sample $x \sim g(\cdot)$ then we can use $M$ such (i.i.d.) samples $\{x^{(i)}\}_{i=1}^M$ to estimate $E_f[h(X)]$:

$$\hat{I}_M = \frac{1}{M} \sum_{i=1}^M h(x^{(i)})\frac{f(x^{(i)}|y)}{g(x^{(i)})}$$

- $\hat{I}_M \to E_f[h(X)]$ by the S.L.L.N.

- Problem solved? Not really: Need to recompute after arrival of new observation. Very difficult and inefficient as high-dimensional (usually large $T$).
Intro to SMC: Sequential importance sampling

- Hard to come up with a good proposal distribution.
- Recall the sequential structure of our problem at hand.
- Many real-life applications require fast on-line estimation of hidden states.
- Note that we can evaluate the density $f(x_0:T|y_0:T)$ pointwise up to a normalizing constant:

$$f(x_0:T|y_0:T) \propto f(x_0)f(y_0|x_0) \prod_{t=1}^{T} f(y_t|x_t)f(x_t|x_{t-1})$$

- Hence, by choosing proposal density

$$q(x_0:T|y_0:T) = q(x_0|y_0) \prod_{t=1}^{T} q(x_t|x_{t-1}, y_t)$$

with sequential structure, we can calculate weights sequentially.
Sequential Importance Sampling (generic)

1. At time $t = 0$:
   - For $i = 1, \ldots, M$, sample $x_0^{(i)} \sim q(x_0 | y_0)$
   - For $i = 1, \ldots, M$, evaluate the importance weights:
     \[
     w_0(x_0^{(i)}) = \frac{f(x_0^{(i)}, y_0)}{q(x_0^{(i)} | y_0)} = \frac{f(y_0 | x_0^{(i)}) f(x_0^{(i)})}{q(x_0^{(i)} | y_0)}
     \]

2. For $t = 1, \ldots, T$:
   - For $i = 1, \ldots, M$, sample $x_t^{(i)} \sim q(x_t | y_t, x_t^{(i)})$ and set $x_{0:t}^{(i)} = (x_{0:t-1}^{(i)}, x_t^{(i)})$
   - For $i = 1, \ldots, M$, evaluate the importance weights:
     \[
     w_t(x_{0:t}^{(i)}) = w_{t-1}(x_{0:t-1}^{(i)}) \frac{f(x_{0:t}^{(i)}, y_0:t)}{f(x_{0:t-1}^{(i)}, y_0:t-1) q(x_t^{(i)} | y_t, x_t^{(i)})} = w_{t-1}(x_{0:t-1}^{(i)}) \frac{f(y_t | x_t^{(i)}) f(x_t^{(i)} | x_t^{(i)})}{q(x_t^{(i)} | y_t, x_t^{(i)})}
     \]
Degeneracy

- SIS has serious flaw – degeneracy even for small $T$.
- All the weights eventually become negligible except for one - and this particle need not even be good.
- Solution: resampling - interaction between particles.
Resampling

- Kill “useless” particles and duplicate useful particles by resampling.
- Draw $N$ times with replacement from the empirical distribution

$$
\hat{f}_M(x_{0:t}|y_{0:t}) = \sum_{i=1}^{M} W_t^{(i)} \delta_{x_{0:t}^{(i)}}(x_{0:t})
$$

where $W_t^{(i)} = \frac{w_t(x_{0:t}^{(i)})}{\sum_{j=1}^{M} w_t(x_{0:t}^{(j)})}$.

- After resampling some particles have disappeared and some appear multiple times

$$
\tilde{f}_M(x_{0:t}|y_{0:t}) = \frac{1}{M} \sum_{i=1}^{M} n_t^{(i)} \delta_{x_{0:t}^{(i)}}(x_{0:t})
$$

where $n_t^{(i)} \in \{0, 1, \ldots, M\}$ and $\sum_{i=1}^{M} n_t^{(i)} = M$.

- For convergence, it is sufficient for $E[n_t^{(i)}] = MW_t^{(i)}$. 

Sequential Importance Sampling / Resampling (generic)

1. At time $t = 0$:
   - For $i = 1, \ldots, M$, sample $\tilde{x}_0^{(i)} \sim q(x_0 | y_0)$
   - For $i = 1, \ldots, M$, evaluate the importance weights:
     \[
     w_0(\tilde{x}_0^{(i)}) = \frac{f(\tilde{x}_0^{(i)}, y_0)}{q(\tilde{x}_0^{(i)} | y_0)} = \frac{f(y_0 | \tilde{x}_0^{(i)}) f(\tilde{x}_0^{(i)})}{q(\tilde{x}_0^{(i)} | y_0)}
     \]

2. For times $t = 1, \ldots, T$:
   - For $i = 1, \ldots, M$, sample $\tilde{x}_t^{(i)} \sim q(x_t | y_t, x_{t-1}^{(i)})$ and set $\tilde{x}_{0:t}^{(i)} = (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)})$
   - For $i = 1, \ldots, M$, evaluate the importance weights:
     \[
     w_t(\tilde{x}_{0:t}^{(i)}) = w_{t-1}^{(i)} \frac{f(y_t | \tilde{x}_t^{(i)}) f(\tilde{x}_t^{(i)} | x_{t-1}^{(i)})}{q(\tilde{x}_t^{(i)} | y_t, x_{t-1}^{(i)})}
     \]
   - Normalize the importance weights.
   - Depending on some criteria, resample the particles. Set $w_t^{(i)} = \frac{1}{N}$ for $i = 1, \ldots, M$. 
Why it works

- The weights no longer degenerate.
- Intuitively, we replicate promising particles and discard poorer ones. This allows us to explore the space more effectively.
- There is some theory that shows how for well-behaved distributions, the variance is many orders of magnitude lower than SIS.
- At each time $t$ we have an empirical distribution $\hat{f}_M(x_{0:t}|y_{0:t})$ approximating $f(x_{0:t}|y_{0:t})$.
- Byproduct:

\[
\hat{f}(y|\mathcal{F}_0, \theta) = \prod_{t=1}^{T} \left( \frac{1}{M} \sum_{i=1}^{M} w_t^{(i)} \right)
\]

if we resample at every time step, and

\[
E \left[ \hat{f}(y|\mathcal{F}_0, \theta) \right] = f(y|\mathcal{F}_0, \theta)
\]
Real implementations

- $q(x_t|y_t, x_{t-1})$ can be anything.
- Easiest choice: $f(x_t | x_{t-1})$.
- Optimal choice: Minimises variance of weights. (Impossible.)
- Decision when to resample is usually based on effective sample size.
- Resampling criteria are usually variants of multinomial sampling
  - Stratified resampling
  - Residual resampling
  - Systematic resampling
Linear Gaussian state-space model

• Consider the Gaussian linear model (e.g. Harvey (1989) and Durbin and Koopman 2001)

\[
\begin{align*}
y_t &= \mu + \alpha_t + \sigma_\varepsilon \varepsilon_t, \\
\alpha_{t+1} &= \phi \alpha_t + \sigma_\eta \eta_t,
\end{align*}
\]

\[
\begin{pmatrix}
\varepsilon_t \\
\eta_t
\end{pmatrix}
\sim_{\text{i.i.d.}} N(0, I_2),
\]

where \( \alpha_0 \sim N\left(0, \sigma_\eta^2 / (1 - \phi^2)\right) \).

• \( \theta = (\mu, \log \sigma_\varepsilon^2, \phi, \log \sigma_\eta^2)' \)

• \( \mu \) controls the unconditional mean of \( y_t \), \( \phi \) the autocorrelation and \( \sigma_\eta^2 \) the variance of the latent process.

• The likelihood can be computed using the Kalman filter and this will serve us as a benchmark.
Linear Gaussian state-space model

- Impact of $M$ and $N$. 

![Graphs showing the impact of M and N on μ|y for different values of M and N.](image)
Linear Gaussian state-space model

- Impact of $M$ and $N$.
Tuning PMCMC

- How to choose $M$ and the variances of the random-walk proposals?
- $M$ is “large enough” when the speed with which the acceptance probabilities increase with $M$ starts to slow down.
- Alternative (referee): Set $M$ to obtain a specified level of the variance of the log-likelihood estimate, for a given $\theta$.
- (But the variance can change quite a bit in function of $\theta$)
Tuning PMCMC

• Then tune the proposal variances to get the desired levels for the acceptance probabilities.

• If one ends up having to decrease variances by a lot to get acceptance probabilities of around 40% for long chains this is an indication that $M$ is not sufficiently large.

• It is helpful to always keep an eye on the ACFs: If one needs small proposal variances to get acceptance probabilities of 40% and observes highly autocorrelated chains at the same time this is another strong indicator that $M$ is too small.
Gaussian discrete time stochastic volatility model

- The stock returns are assumed to follow the process

\[ y_t = \mu + \exp \{ \beta_0 + \beta_1 \alpha_t \} \varepsilon_t \]

- and the stochastic volatility factor

\[ \alpha_{t+1} = \phi \alpha_t + \eta_t, \]

where

\[
\begin{pmatrix}
\varepsilon_t \\
\eta_t
\end{pmatrix}
\overset{i.i.d.}{\sim}
\mathcal{N}
\left(
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix}
\right),
\alpha_0 \sim \mathcal{N}
\left(
0,
\frac{1}{1 - \phi^2}
\right)
\]

- Want to make inference on \( \theta = (\mu, \beta_0, \beta_1, \phi, \rho) \).
Dynamic model: stochastic volatility, $M = 2000$, $N = 100000$
Dynamic model: stochastic volatility, $M = 2000$, $N = 100000$
Dynamic model: stochastic volatility, \( M = 2000, \quad N = 100000 \)
### Dynamic model: parameter estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Mcse</th>
<th>Pr</th>
<th>Posterior covariance and correlation</th>
<th>Inferf</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.042</td>
<td>0.000</td>
<td>0.410</td>
<td>0.000</td>
<td>-0.686</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-0.141</td>
<td>0.001</td>
<td>0.395</td>
<td>-0.001</td>
<td>0.006</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.080</td>
<td>0.000</td>
<td>0.398</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.982</td>
<td>0.000</td>
<td>0.424</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.742</td>
<td>0.000</td>
<td>0.427</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
DSGE models

- A particle filter inside an MCMC algorithm can also be used to estimate DSGE models.
- Fernandez-Villaverde and Rudio-Ramirez (2007) were the first to consider using particle filters to perform parameter inference.
- They do SML with a particle filter.
- Particle filter estimate is a discontinuous function of $\theta$.
- PMCMC works.
DSGE model

- In a toy DSGE model, the objective function is

\[
\max_{\{C_t, L_t\}_{t=0}^{\infty}} E_0 \left[ \sum_{t=0}^{\infty} \beta^t \{ \log (C_t) + \psi \log (1 - L_t) \} \right], \quad \beta \in (0, 1), \psi > 0,
\]

- and the constraints are

\[
Y_t = A_t K_t^\alpha L_t^{1-\alpha}
\]
\[
K_{t+1} = (1 - \delta) K_t + U_t I_t
\]
\[
C_t + I_t = Y_t
\]
\[
\log A_t = \rho_a \log A_{t-1} + \sigma_a \eta_{a,t}
\]
\[
\log U_t = \rho_u \log U_{t-1} + \sigma_u \eta_{u,t}
\]
DSGE model

• The (second-order) solution is

\[
\hat{k}_{t+1} = h_{x,1}\hat{x}_t + \frac{1}{2}\hat{x}_t' h_{xx,1}\hat{x}_t + \frac{1}{2} h_{\sigma\sigma,1}\sigma^2
\]

\[
\hat{a}_t = \rho_a\hat{a}_{t-1} + \sigma\eta_{a,t}
\]

\[
\hat{u}_t = \rho_u\hat{u}_{t-1} + \sigma_u\eta_{u,t}
\]

and

\[
\hat{c}_t = g_{x,1}\hat{x}_t + \frac{1}{2}\hat{x}_t' g_{xx,1}\hat{x}_t + \frac{1}{2} g_{\sigma\sigma,1}\sigma^2
\]

\[
\hat{l}_t = g_{x,2}\hat{x}_t + \frac{1}{2}\hat{x}_t' g_{xx,2}\hat{x}_t + \frac{1}{2} g_{\sigma\sigma,2}\sigma^2
\]

• By specifying (an) observation question(s) we obtain a non-linear state space model.
How does PMCMC work here?

Draw $\theta^{(i)} \sim q(\theta^{(i)} | \theta^{(i-1)})$.

Compute $K_{ss}, A_{ss}, U_{ss}, C_{ss}, L_{ss}$ given $\theta^{(i)}$.

Use perturbation methods to find numerical values for $h_x, g_x, h_{xx}, g_{xx}, h_{\sigma\sigma}, g_{\sigma\sigma}$.

Run the particle filter on the state-space system to obtain $\hat{L}(\theta^{(i)})$.

Accept or reject $\theta^{(i)}$. 

Estimating the likelihood of a DSGE model
Difficulties

- Small number of observations.
- Very spiky likelihood, i.e. low acceptance probabilities.
- Almost static states.
- Need large number of particles.
- Very slow!
- Parallelisable...
Conclusion

• Using estimated likelihoods as the basis for approximate maximum likelihood estimation can have flaws, as emphasised in the literature.

• We note that the effect of estimation can be removed by replacing the maximisation of the likelihood by placing the estimated likelihood inside a MCMC algorithm.

• The theory of this is very simple.

• The estimators are pretty general as they just need one to be able to simulate from the dynamics of the model to be able to implement it.