The Econometrics of DSGE Models

Giuseppe Ragusa
Luiss University

EIEF
Lecture 9-10: Twilight zone of DSGE models (Particle Filter)

March 27, 2017

To go beyond the linear approximation

The general form

$$y_t = \phi(x_t, \theta) + u_t, \quad u_t \sim P_\theta(\cdot)$$
$$x_{t+1} = \xi(x_t, \theta) + \varepsilon_t, \quad \varepsilon_t \sim F_\theta(\cdot)$$

- The function ϕ and ξ are generated numerically by solution methods.
- The objective is to estimate θ
  - Extended Kalman Filter
  - Unscented Kalman Filter
  - Particle Filter

Particle Filter

Needed:

- Particle methods assume $\{X_t\}_{t\geq 1}$ and the observations $\{Y_t\}_{t\geq 1}$ satisfy the following setup:
  1. $X_1, X_2, \ldots$ is a first order Markov process such that
  $$X_t | (X_{t-1} = x_{t-1}) \sim f_\theta(\cdot | x_{t-1})$$
  with an initial distribution
  $$X_1 \sim \mu_\theta(x_1).$$
  2. Given $\{X_t\}_{t\geq 1}$, the observations $\{Y_t\}_{t\geq 1}$ are statistically independent, and their marginal densities are given
  $$Y_t | (X_t = x_t) \sim g_\theta(y_t | x_t),$$

Particle Filter (DSGE)

- Fernandez-Villaverde, Jesus, and Juan F. Rubio-Ramirez. "Estimating macroeconomic models: A likelihood approach." The Review of Economic Studies 74.4 (2007): 1059-1087.
- Flury, Thomas, and Neil Shephard. "Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models." Econometric Theory 27.5 (2011): 933.
- Fernandez-Villaverde, Jesus. "The econometrics of DSGE models." SERIEs 1.1-2 (2010): 3-49.

Particle Filter (Other application)

Stochastic Volatility Model

$$X_t = \alpha X_{t-1} + \sigma V_t$$
$$Y_t = \beta \exp\left(X_t/2\right) W_t$$

where

$$V_n \sim N(0,1), \quad W_n \sim N(0,1).$$

In this case, we have:

$$\theta = (\alpha, \sigma, \beta),\ \mu_\theta(x) = N\left(x; 0, \frac{\sigma^2}{1-\alpha^2}\right)$$

$$f_\theta(x_t|x_{t-1}) = N(x_t; \alpha x_{t-1}, \sigma^2),\ g_\theta(y_t|x_t) = N\left(y; 0, \beta^2 \exp(x)\right).$$

The Bayesian model

The equations (and conditional independence)

$$X_t | (X_{t-1} = x_{t-1}) \sim f_\theta(\cdot | x_{t-1}) \tag{1}$$

$$Y_t | (X_t = x_t) \sim g_\theta(y_t | x_t) \tag{2}$$

define a Bayesian model in which
equation (1) defines a prior distribution of the process $\{X_t\}_{t \geq T}$, that is

$$p_\theta(x_{1:T}) = \mu(x_1) \prod_{t=2}^{T} f_\theta(x_t | x_{t-1})$$

equation (2) defines the likelihood function

$$p_\theta(y_{1:T} | x_{1:T}) = \prod_{t=1}^{T} g_\theta(y_t | x_t)$$

Inference about $X_{1:T}$

(given θ)
In such a Bayesian context, inference about $X_{1:T}$ given θ and a realization of the observations $Y_{1:T} = y_{1:T}$ relies on the posterior distribution

$$p_\theta(x_{1:T}|y_{1:T}) = \frac{p_\theta(x_{1:T}, y_{1:T})}{p_\theta(y_{1:T})} = \frac{p_\theta(x_{1:T})p_\theta(y_{1:T}|x_{1:T})}{p_\theta(y_{1:T})}$$

where

$$p_\theta(y_{1:T}) = \int p_\theta(x_{1:T})p_\theta(y_{1:T}|x_{1:T})dx_{1:T}.$$

Filtering and Marginal likelihood computation

- **[Filtering] Sequential** approximation of $\{p(\theta, x_{1:t}|y_{1:t})\}_{t\geq 1}$ for a **given** θ, that is

$$
\begin{aligned}
t = 1 &\mapsto p_\theta(x_1|y_1) \\
t = 2 &\mapsto p_\theta(x_{1:2}|y_{1:2}) \\
\vdots = &\qquad \vdots \\
t = T &\mapsto p_\theta(x_{1:T}|y_{1:T})
\end{aligned}
$$

- **[Marginal likelihood] Sequential** approximation of the marginal likelihood

$$
\begin{aligned}
t = 1 &\mapsto p_\theta(y_1) \\
t = 2 &\mapsto p_\theta(y_{1:2}) \\
\vdots = &\qquad \vdots \\
t = T &\mapsto {\color{red} p_\theta(y_{1:T})}
\end{aligned}
$$

Filtering

Key recursion (I)

Notice that

$$p_\theta(x_{1:T}|y_{1:T}) = \frac{p_\theta(x_{1:T}, y_{1:T})}{p_\theta(y_{1:T})}$$

The **unnormalized posterior**

$$\begin{aligned} p_\theta(x_{1:T}, y_{1:T}) &= p_\theta(x_{1:T-1}, y_{1:T-1}) p_\theta(x_T, y_T | x_{1:T-1}, y_{T-1}) \\ &= p_\theta(x_{1:T-1}, y_{1:T-1}) p_\theta(y_T | x_{1:T-1}, x_T, y_{T-1}) p_\theta(x_T | x_{1:T-1}, y_{T-1}) \\ &= {\color{red} p_\theta(x_{1:T-1}, y_{1:T-1}) g_\theta(y_T | x_T) f_\theta(x_T | x_{T-1})} \end{aligned} \tag{3}$$

where the last equality follows from the Markovian assumption on $\{X_t\}_{t\geq 1}$ and conditional independence of $\{Y_t\}_{t\geq T}$ from $\{X_t\}_{t\geq 1}$.

Filtering

Key recursion (ii)

Thus, at any $t \geq 1$

$$p_\theta(x_{1:t}, y_{1:t}) = p_\theta(x_{1:t-1}, y_{1:t-1}) g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1})$$

implies that the posterior $p(x_{1:T}|y_{1:T})$ satisfies the following **recursion**

$$\begin{aligned} p_\theta(x_{1:t}|y_{1:t}) &= p_\theta(x_{1:t-1}|y_{1:t-1}) \frac{g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1}) p(y_{1:t-1})}{p(y_{1:t})} \\ &= \frac{p_\theta(x_{1:t-1}|y_{1:t-1}) g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1})}{p(y_t|y_{1:t-1})}, \end{aligned}$$

where

$$\begin{aligned} p(y_t|y_{1:t-1}) &= \int \left[\int p_\theta(x_{1:t-1}|y_{1:t-1}) dx_{1:t-2} \right] g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1}) dx_{t-1:t} \\ &= \int p_\theta(x_{t-1}|y_{1:t-1}) g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1}) dx_{t-1:t} \end{aligned}$$

Filtering

Key recursion (iii)

Integrating out $x_{1:t-1}$ from $p_\theta(x_{1:t}|y_{1:t})$ we obtain the so called prediction and updating step. Notice

$$
\begin{aligned}
\int p_\theta(x_{1:t}|y_{1:t})dx_{1:t-1} &= \int \frac{p_\theta(x_{1:t-1}|y_{1:t-1})g_\theta(y_t|x_t)f_\theta(x_t|x_{t-1})}{p_\theta(y_t|y_{1:t-1})}dx_{1:t-1} \\
&= \frac{g_\theta(y_t|x_t)\int\left[\int p_\theta(x_{1:t-1}|y_{1:t-1})dx_{1:t-2}\right]f_\theta(x_t|x_{t-1})dx_{t-1}}{p_\theta(y_t|y_{1:t-1})} \\
&= \frac{g_\theta(y_t|x_t)\int p_\theta(x_{t-1}|y_{1:t-1})f_\theta(x_t|x_{t-1})dx_{t-1}}{p_\theta(y_t|y_{1:t-1})}.
\end{aligned}
$$

Thus,

$$p_\theta(x_t|y_{1:t}) = \frac{g_\theta(y_t|x_t)p_\theta(x_t|y_{1:t-1})}{p_\theta(y_t|y_{1:t-1})} \quad \textbf{(Updating step)}$$

where

$$p_\theta(x_t|y_{1:t-1}) = \int f_\theta(x_t|x_{t-1})p_\theta(x_{t-1}|y_{1:t-1})dx_{t-1}, \quad \textbf{(Prediction Step)}$$

Particle filter

$$p_\theta(x_t|y_{1:t}) = \frac{g_\theta(y_t|x_t)p_\theta(x_t|y_{1:t-1})}{p_\theta(y_t|y_{1:t-1})} \; \textbf{(Updating step)}$$

$$p_\theta(x_t|y_{1:t-1}) = \int f_\theta(x_t|x_{t-1})p_\theta(x_{t-1}|y_{1:t-1})dx_{t-1}, \; \textbf{(Prediction Step)}$$

- The computation in the prediction and update steps cannot be carried out analytically
- Hence the need of approximate methods such as Monte Carlo sampling.
- Sequential importance sampling (SIS) is the most basic Monte Carlo method used for this purpose
- SIS actually approximate $p_\theta(x_{1:t}|y_{1:t})$ rather than just $p_\theta(x_t|y_{1:t})$
- SIS is based on the Importance Sampling (IS).

Importance sampling (IS)

- Recall that in the importance sample one approximates a target distribution

$$\pi_t(x_{1:t}) = \gamma_t(x_{1:t})/Z_t,$$

  where

$$Z_t = \int \gamma_t(x_{1:t}) dx_{1:t}.$$

- Drawn from a proposal distribution $X^i_{1:t} \sim q_t(x_{1:t})$, $i = 1, \ldots, N$, and weight each sample $x^i_{1:t}$ by

$$w_t(x^i_{1:t}) = \frac{\gamma_t(x^i_{1:t})}{q_t(x^i_{1:t})}, \; W^i_t = \frac{w_t(x^i_{1:t})}{\sum_{i=1}^N w_t(x^i_{1:t})}.$$

- We saw that we can approximate

$$\hat{\pi}_t(x_{1:t}) = \sum_{i=1}^N W^i_t \delta_{X^i_{1:t}}(x_{1:t})$$

  and given a function $h(\cdot)$

$$\int h(x_{1:t})\hat{\pi}_t(x_{1:t}) d_{x_{1:t}} = \sum_{i=1}^N W^i_t h(x_{1:t}) \xrightarrow{p} \int h(x_{1:t})\pi_t(x_{1:t}) dx_{1:t}.$$

Sequential importance sampling

- The SIS is based on the idea of using IS sequential
- Select an importance distribution which has the following structure

$$\begin{aligned} q_t(x_{1:t}) &= q_{t-1}(x_{1:t-1})q_t(x_t|x_{1:t-1}) \\ &= q_1(x_1)\prod_{k=2}^{t} q_k(x_k|x_{1:k-1}) \end{aligned}$$

- This means that to obtain **particles** $X^i_{1:t} \sim q_t(x_{1:t})$, we sample $X^i_1 \sim q_1(x_1)$, $X^i_2 \sim q_2(x_2|X^i_1)$, $X^i_3 \sim q_3(x_3|X^i_2, X^i_1)$,...,$X^i_k \sim q_k(x_k|X^i_{1:k-1})$
- The unnormalized weights can be computed recursively

$$w_t(x_{1:t}) = \frac{\gamma_t(x_{1:t})}{q_t(x_{1:t})} = \frac{\gamma_{t-1}(x_{1:t-1})}{q_{t-1}(x_{1:t-1})} \frac{\gamma_t(x_{1:t})}{\gamma_{t-1}(x_{1:t-1})q_t(x_t|x_{1:t-1})}$$

or

$$w_t(x_{1:t}) = w_{t-1}(x_{1:t-1})\alpha_t(x_{1:t})$$

Sequential importance sampling

At time $t = 1$

1. Sample $X_1^i \sim q_1(x_1)$
2. Compute the weights $w_1(X_1^i)$ and $W_1^i \propto w_1(X_1^i)$

At time $t \geq 2$

1. Sample $X_t^i \sim q_t(x_t | X_{1:t-1}^i)$
2. Compute the weights

$$w_t(X_{1:t}^i) = w_{t-1}(X_{1:t-1})\alpha_t(X_{1:t})$$
$$W_t^i \propto w_t(X_{1:t}^i)$$

The choice of $q_t(x_t|x_{1:t-1})$

- The choice of the importance distribution $q_t(x_t|x_{1:t-1})$ at each t is important for the performance of SIS
- A sensible strategy consists of selecting it so as to minimize variance of the weights
- This is achieved by selecting
$$q_t^*(x_t|x_{1:t-1}) = \pi_t(x_t|x_{1:t-1})$$
- It is not always possible to sample from q^*, but it could serve as a guide for selecting

Particle filter is SIS

Now let

$$\pi_t(x_{1:t}) = \frac{\gamma_t(x_{1:t})}{Z_t}$$

where

$$\gamma_t(x_{1:t}) = p_\theta(x_{1:t}, y_{1:t}),\ Z_t = \int p_\theta(x_{1:t}, y_{1:t}) dx_{1:t}.$$

The weights of SIS are

$$w_t(x_{1:t}) = w_{t-1}(x_{1:t}) \frac{p_\theta(x_{1:t}, y_{1:t})}{p_\theta(x_{1:t-1}, y_{1:t-1}) q_t(x_t | x_{1:t-1})}$$

but from (3)

$$p_\theta(x_{1:t}, y_{1:t}) = p_\theta(x_{1:t-1}, y_{1:t-1}) g_\theta(y_t | x_t) f_\theta(x_t | x_{t-1})$$

it follows that

$$w_t(x_{1:t}) = w_{t-1}(x_{1:t}) \frac{g_\theta(y_t | x_t) f_\theta(x_t | x_{t-1})}{q_t(x_t | x_{1:t-1})}$$

Particle filter is SIS

- The “optimal” importance distribution is

$$q_t^*(x_t|x_{1:t-1}) = \pi_t(x_t|x_{1:t-1})$$

which in the case of the particle filter

$$\begin{aligned} q_t^*(x_t|x_{1:t-1}, y_{1:t}) &= \pi_t(x_t|x_{1:t-1}, y_{1:t}) \\ &= p_\theta(x_t|x_{t-1}, y_t) \\ &= \frac{g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1})}{p_\theta(y_t|x_{t-1})} \end{aligned}$$

and the associated incremental weight function α is

$$\alpha_t^*(x_{1:t}) = \frac{g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1})}{q_t^*(x_t|x_{1:t-1}, y_{1:t})} = 1$$

Particle filter (I)

At time $t = 1$

1. Sample $X_1^i \sim q_1(x_1)$
2. Compute the weights $w_1(X_1^i) = \frac{\mu(X_1^i)g(y_1|X_1^i)}{q(X_1^i|y_1)}$ and $W_1^i \propto w_1(X_1^i)$

At time $t \geq 2$

1. Sample $X_t^i \sim q_t(x_t|y_t, X_{1:t-1}^i)$
2. Compute the weights

$$\alpha_t(X_{1:t}^i) = \frac{g_\theta(y_t|X_t^i)f_\theta(X_t^i|X_{t-1}^i)}{q_t(X_t^i|X_{1:t-1}^i, y_{1:t})}$$

$$w_t(X_{1:t}^i) = w_{t-1}(X_{1:t-1}^i)\alpha_t(X_{1:t})$$
$$W_t^i \propto w_t(X_{1:t}^i)$$

Particle filter (II)

- Applying the previous algorithm, we obtain at time T

$$\hat{p}_\theta(x_{1:T}|y_{1:T}) = \sum_{i=1}^{N} W_T^i \delta_{X_{1:n}^i}(x_{1:T})$$

$$\hat{p}_\theta(y_t|y_{1:t-1}) = \sum_{i=1}^{N} w_{t-1}^i \alpha_t(X_{1:t})$$

Since

$$\log p_\theta(y_{1:T}) = \sum_{i=1}^{T} \log p(y_t|y_{1:t-1})$$

the log-likelihood can be approximated

$$\log \widehat{p_\theta(y_{1:T})} = \sum_{i=1}^{T} \log \hat{p}(y_t|y_{1:t-1})$$

Sequential importance resampling

- SIS provides estimates whose variance increases, typically exponentially, with T
- Resampling techniques are a key ingredient which (partially) solve this problem
- The idea of resampling is instead of constructing an SIS estimate based on sampling from $\hat{p}_\theta(x_{1:T}|y_{1:T})$
- In practice this amounts to sample particles at each step (or at certain point) of the algorithm

Particle filter with resampling

At time $t = 1$

1. Sample $X_1^i \sim q_1(x_1)$
2. Compute the weights $w_1(X_1^i) = \frac{\mu(X_1^i)g(y_1|X_1^i)}{q(X_1^i|y_1)}$ and $W_1^i \propto w_1(X_1^i)$
3. Reample $\{W_1^i, X_1^i\}$ to obtain N equally weighted particles $\{N^{-1}, \bar{X}^i\}$

At time $t \geq 2$

1. Sample $X_t^i \sim q_t(x_t|y_t, \bar{X}_{1:t-1}^i)$ and set $X_{1:t}^i \leftarrow (\bar{X}_{1:t-1}^i, X_t^i)$
2. Compute the weights

$$\alpha_t(X_{1:t}) = \frac{g_\theta(y_t|X_t^i)f_\theta(X_t^i|X_{t-1}^i)}{q_t(X_t^i|X_{1:t-1}^i, y_{1:t})}$$

$$w_t(X_{1:t}^i) = \alpha_t(X_{1:t})$$
$$W_t(X_{1:t}^i) \propto w_t(X_{1:t}^i)$$

3. Resample $\{W_t^i, X_{1:t}^i\}$ to obtain N new equally weighted particles $\{N^{-1}, \bar{X}_{1:t}^i\}$

Proposal distribution

- The proposal distribution, generically denoted $q_t(X_t^i | X_{1:t-1}^i, y_{1:t})$ in the algorithm above, needs to be chosen to make the algorithm feasible.
- A common choice is to set
$$q_t(x_t^i | x_{1:t-1}^i, y_{1:t}) = f_\theta(x_t | x_{t-1}).$$
- With choice
$$\alpha_t(x_{1:t}) = g_\theta(y_t | x_t)$$
- Sequential Importance Resampling (SIR) filters with transition prior probability distribution as importance function are commonly known as *bootstrap filter* or *condensation algorithm*.