Bayesian Estimation of DSGE Models[1]
Chapter 10: Particle Filters & SMC Samplers

Ed Herbst[1] Frank Schorfheide[2]

[1]Federal Reserve Board

[2]University of Pennsylvania

February 9, 2016

[1]The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Board of Governors or the Federal Reserve System.

Using SMC to Approximate Likelihood and Posterior – SMC^2

- Start from SMC algorithm...
  - replace actual likelihood $p(Y_{1:T}|\theta)$
  - by particle filter approximation $\hat{p}(Y_{1:T}|\theta)$

  in the correction and mutation steps of SMC algorithm.
- Key Idea from particle MCMC literature (Andrieu, Doucet, and Holenstein, 2010): let
$$\hat{p}(Y_{1:t_n}|\theta_n) = g(Y_{1:t_n}|\theta_n, U_{1:t_n}).$$
  where $U_{1:t_n} \sim p(U_{1:t_n})$ is an array of *iid* uniform random variables generated by the particle filter.
- Important Result: Particle filter delivers an unbiased estimate of the incremental weight $p(Y_{t_{n-1}+1:t_n}|\theta)$:
$$p(Y_{1:t_n}|\theta_n) = \int g(Y_{1:t_n}|\theta_n, U_{1:t_n}) p(U_{1:t_n}) dU_{1:t_n}.$$

Particle System for SMC^2 Sampler After Stage n

- Data tempering instead of likelihood tempering: $\pi_n^D(\theta) = p(\theta|Y_{1:t_n})$ – avoids nonlinear transformation $[p(Y_{1:T}|\theta)]^{\phi_n - \phi_{n-1}}$ of likelihood increment.
- To simplify notation, we add one observation at a time, $n = t$, and write θ_t and $\pi_t(\cdot)$.

Parameter	State			
(θ_n^1, W_n^1)	$(s_{t_n}^{1,1}, \mathcal{W}_{t_n}^{1,1})$	$(s_{t_n}^{1,2}, \mathcal{W}_{t_n}^{1,2})$	$\cdots$	$(s_{t_n}^{1,M}, \mathcal{W}_{t_n}^{1,M})$
(θ_n^2, W_n^2)	$(s_{t_n}^{2,1}, \mathcal{W}_{t_n}^{2,1})$	$(s_{t_n}^{2,2}, \mathcal{W}_{t_n}^{2,2})$	$\cdots$	$(s_{t_n}^{2,M}, \mathcal{W}_{t_n}^{2,M})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
(θ_n^N, W_n^N)	$(s_{t_n}^{N,1}, \mathcal{W}_{t_n}^{N,1})$	$(s_{t_n}^{N,2}, \mathcal{W}_{t_n}^{N,2})$	$\cdots$	$(s_{t_n}^{N,M}, \mathcal{W}_{t_n}^{N,M})$

Why Does SMC^2 Work?

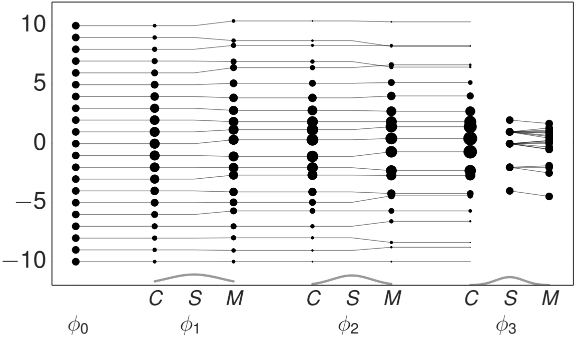
- Work on enlarged probability space that includes sequence of random vectors $U^i_{1:t-1}$ that underlies the simulation approximation of the particle filter.
- At the end of iteration $t-1$:
    - Swarm of state particles $\{s^{i,j}_{t-1}, \mathcal{W}^{i,j}_{t-1}\}_{j=1}^M$ that represents the distribution $p(s_{t-1}|Y_{1:t-1}, \theta^i_{t-1})$.
    - For each parameter value θ^i_{t-1} there is PF approx of the likelihood: $\hat{p}(Y_{1:t-1}|\theta^i_{t-1}) = g(Y_{1:t-1}|\theta^i_{t-1}, U^i_{1:t-1})$.
    - Parameter Particles $\{\theta^i_{t-1}, U^i_{1:t-1}, W^i_{t-1}\}_{i=1}^N$, approximate:

$$\int\int h(\theta, U_{1:t-1}) p(U_{1:t-1}) p(\theta|Y_{1:t-1}) dU_{1:t-1} d\theta \approx \frac{1}{N}\sum_{i=1}^N h(\theta^i_{t-1}, U^i_{1:t-1}) W^i_{t-1}.$$

Recall...

- $\pi_n(\theta)$ is represented by a swarm of particles $\{\theta_n^i, W_n^i\}_{i=1}^N$:

$$\bar{h}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} W_n^i h(\theta^i) \xrightarrow{a.s.} \mathbb{E}_{\pi_n}[h(\theta_n)].$$



- C is Correction; S is Selection; and M is Mutation.

Correction Step of Parameter SMC

- Write particle filter approximation of the likelihood increment as

$$\tilde{w}_t^i = \hat{p}(y_t|Y_{1:t-1}, \theta_{t-1}^i) = g(y_t|Y_{1:t-1}, U_{1:t}^i, \theta_{t-1}^i).$$

- By induction: $\frac{1}{N}\sum_{i=1}^N h(\theta_{t-1}^i)\tilde{w}_t^i W_{t-1}^i$ approximates

$$\begin{aligned}
&\int\int h(\theta) g(y_t|Y_{1:t-1}, U_{1:t}, \theta) p(U_{1:t}) p(\theta|Y_{1:t-1}) dU_{1:t} d\theta \\
&\quad = \int h(\theta)\left[\int g(y_t|Y_{1:t-1}, U_{1:t}, \theta) p(U_{1:t}) dU_{1:t}\right] p(\theta|Y_{1:t-1}) d\theta \\
&\quad = \int h(\theta) p(y_t|\theta, Y_{1:t-1}) p(\theta|Y_{1:t-1}) d\theta.
\end{aligned}$$

- Last equality follows unbiasedness of PF approximation.

Mutation Step of Parameter SMC

- Recall: mutation is based on Metropolis-Hastings algorithm...
- For each particle i we have:
  - a proposed parameter value ϑ_t^i;
  - an associated particle filter approximation of the likelihood:

$$\hat{p}(Y_{1:t}|\vartheta_t^i) = g(Y_{1:t}|\vartheta_t^i, U_{1:t}^*).$$

  - based on sequence of random vectors $U_{1:t}^*$ drawn from the distribution $p(U_{1:t})$;
- Key Insight: the densities $p(U_{1:t}^i)$ and $p(U_{1:t}^*)$ cancel from the formula for the acceptance probability $\alpha(\vartheta_t^i|\hat{\theta}_t^i)$:

$$\begin{aligned} \alpha(\vartheta|\theta^{i-1}) &= \min \left\{ 1, \frac{\frac{g(Y|\vartheta,U^*)p(U^*)p(\vartheta)}{q(\vartheta|\theta^{(i-1)})p(U^*)}}{\frac{g(Y|\theta^{(i-1)},U^{(i-1)})p(U^{(i-1)})p(\theta^{(i-1)})}{q(\theta^{(i-1)}|\theta^*)p(U^{(i-1)})}} \right\} \\ &= \min \left\{ 1, \frac{\hat{p}(Y|\vartheta)p(\vartheta)\big/q(\vartheta|\theta^{(i-1)})}{\hat{p}(Y|\theta^{(i-1)})p(\theta^{(i-1)})\big/q(\theta^{(i-1)}|\vartheta)} \right\}. \end{aligned}$$

Application to Small-Scale DSGE Model

- Results are based on $N_{run} = 20$ runs of the SMC^2 algorithm with $N = 4,000$ particles.
- D is data tempering and L is likelihood tempering.
- KF is Kalman filter, CO-PF is conditionally-optimal PF with $M = 400$, BS-PF is bootstrap PF with $M = 40,000$. CO-PF and BS-PF use data tempering.
- Accuracy:
  - Note that under *iid* sampling under suitable regularity conditions

$$\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \Longrightarrow N\big(0, \mathbb{V}_\pi[h]\big)$$

  - Inefficiency factor:

$$\text{InEff}_N = \frac{\mathbb{V}[\bar{h}_N]}{\mathbb{V}_\pi[h]/N}$$

Accuracy of SMC^2 Approximations

	Posterior Mean (Pooled)				Inefficiency Factors				Std Dev of Means			
	KF(L)	KF(D)	CO-PF	BS-PF	KF(L)	KF(D)	CO-PF	BS-PF	KF(L)	KF(D)	CO-PF	BS-PF
τ	2.65	2.67	2.68	2.53	1.51	10.41	47.60	6570	0.01	0.03	0.07	0.76
κ	0.81	0.81	0.81	0.70	1.40	8.36	40.60	7223	0.00	0.01	0.01	0.18
ψ_1	1.87	1.88	1.87	1.89	3.29	18.27	22.56	4785	0.01	0.02	0.02	0.27
ψ_2	0.66	0.66	0.67	0.65	2.72	10.02	43.30	4197	0.01	0.02	0.03	0.34
ρ_r	0.75	0.75	0.75	0.72	1.31	11.39	60.18	14979	0.00	0.00	0.01	0.08
ρ_g	0.98	0.98	0.98	0.95	1.32	4.28	250.34	21736	0.00	0.00	0.00	0.04
ρ_z	0.88	0.88	0.88	0.84	3.16	15.06	35.35	10802	0.00	0.00	0.00	0.05
$r^{(A)}$	0.45	0.46	0.44	0.46	1.09	26.58	73.78	7971	0.00	0.02	0.04	0.42
$\pi^{(A)}$	3.32	3.31	3.31	3.56	2.15	40.45	158.64	6529	0.01	0.03	0.06	0.40
$\gamma^{(Q)}$	0.59	0.59	0.59	0.64	2.35	32.35	133.25	5296	0.00	0.01	0.03	0.16
σ_r	0.24	0.24	0.24	0.26	0.75	7.29	43.96	16084	0.00	0.00	0.00	0.06
σ_g	0.68	0.68	0.68	0.73	1.30	1.48	20.20	5098	0.00	0.00	0.00	0.08
σ_z	0.32	0.32	0.32	0.42	2.32	3.63	26.98	41284	0.00	0.00	0.00	0.11
$\ln p(Y)$	-358.75	-357.34	-356.33	-340.47					0.120	1.191	4.374	14.49

Computational Considerations

- The SMC^2 results are obtained by utilizing 40 processors.
- We parallelized the likelihood evaluations $\hat{p}(Y_{1:t}|\theta_t^i)$ for the θ_t^i particles rather than the particle filter computations for the swarms $\{s_t^{i,j}, \mathcal{W}_t^{i,j}\}_{j=1}^M$.
- The run time for the SMC^2 with conditionally-optimal PF ($N = 4,000$, $M = 400$) is 23:24 [mm:ss] minutes, where as the algorithm with bootstrap PF ($N = 4,000$ and $M = 40,000$) runs for 08:05:35 [hh:mm:ss] hours.
- Due to memory constraints we re-computed the entire likelihood for $Y_{1:t}$ in each iteration.
- Our sequential (data-tempering) implementation of the SMC^2 algorithm suffers from particle degeneracy in the intial stages, i.e., for small sample sizes.