Sequential Monte Carlo Methods for DSGE Models [1]

Ed Herbst* Frank Schorfheide+

*Federal Reserve Board

+University of Pennsylvania, PIER, CEPR, and NBER

February 6, 2018

[1] Material at http://edherbst.net/teaching/indiana-minicourse.The views expressed in this presentation are those of the presenters and do not necessarily reflect the views of the Federal Reserve Board of Governors or the Federal Reserve System.

Some References

These lectures use material from our joint work:

- "Tempered Particle Filtering," 2016, *PIER Working Paper*, 16-017
- *Bayesian Estimation of DSGE Models*, 2015, Princeton University Press
- "Sequential Monte Carlo Sampling for DSGE Models," 2014, *Journal of Econometrics*

Some Background

- DSGE model: dynamic model of the macroeconomy, indexed by θ – vector of preference and technology parameters. Used for forecasting, policy experiments, interpreting past events.

- Bayesian analysis of DSGE models:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \propto p(Y|\theta)p(\theta).$$

- Computational hurdles: numerical solution of model leads to state-space representation $\Longrightarrow$ likelihood approximation $\Longrightarrow$ posterior sampler.

- "Standard" approach for *(linearized) models* (Schorfheide, 2000; Otrok, 2001):
  - Model solution: log-linearize and use linear rational expectations system solver.
  - Evaluation of $p(Y|\theta)$: Kalman filter
  - Posterior draws θ^i: MCMC

Sequential Monte Carlo (SMC) Methods

SMC can help to

Lecture 1

- approximate the posterior of θ: Chopin (2002) ... Durham and Geweke (2013) ... Creal (2007), Herbst and Schorfheide (2014)

Lecture 2

- approximate the likelihood function (particle filtering): Gordon, Salmond, and Smith (1993) ... Fernandez-Villaverde and Rubio-Ramirez (2007)
- or both: SMC^2: Chopin, Jacob, and Papaspiliopoulos (2012) ... Herbst and Schorfheide (2015)

Lecture 1

Sampling from Posterior

- DSGE model posteriors are often non-elliptical, e.g., multimodal posteriors may arise because it is difficult to
  - disentangle internal and external propagation mechanisms;
  - disentangle the relative importance of shocks.

- Economic Example: is wage growth persistent because
  1. wage setters find it very costly to adjust wages?
  2. exogenous shocks affect the substitutability of labor inputs and hence markups?

Sampling from Posterior

- If posterior distributions are irregular, standard MCMC methods can be inaccurate (examples will follow).
- SMC samplers often generate more precise approximations of posteriors in the same amount of time.
- SMC can be parallelized.
- SMC = importance sampling on steroids $\Longrightarrow$ We will first review importance sampling.

Introduction

- Unfortunately, "standard" MCMC can be inaccurate, especially in medium and large-scale DSGE models:
  - disentangling importance of internal versus external propagation mechanism;
  - determining the relative importance of shocks.

- Previously: Modify MCMC algorithms to overcome weaknesses: blocking of parameters; tailoring of (mixture) proposal densities

- Now, we use sequential Monte Carlo (SMC) (more precisely, sequential importance sampling) instead:
  - Better suited to handle irregular and multimodal posteriors associated with large DSGE models.
  - Algorithms can be easily parallelized.

- SMC = Importance Sampling on Steriods. We build on
  - Theoretical work: Chopin (2004); Del Moral, Doucet, Jasra (2006)
  - Applied work: Creal (2007); Durham and Geweke (2011, 2012)

Importance Sampling

- Approximate $\pi(\cdot)$ by using a different, tractable density $g(\theta)$ that is easy to sample from.
- For more general problems, posterior density may be unnormalized. So we write

$$\pi(\theta) = \frac{p(Y|\theta)p(\theta)}{p(Y)} = \frac{f(\theta)}{\int f(\theta)d\theta}.$$

- Importance sampling is based on the identity

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta = \frac{\int_\Theta h(\theta)\frac{f(\theta)}{g(\theta)}g(\theta)d\theta}{\int_\Theta \frac{f(\theta)}{g(\theta)}g(\theta)d\theta}.$$

- (Unnormalized) importance weight:

$$w(\theta) = \frac{f(\theta)}{g(\theta)}.$$

Importance Sampling

1. For $i = 1$ to N, draw $\theta^i \overset{iid}{\sim} g(\theta)$ and compute the unnormalized importance weights

$$w^i = w(\theta^i) = \frac{f(\theta^i)}{g(\theta^i)}.$$

2. Compute the normalized importance weights

$$W^i = \frac{w^i}{\frac{1}{N}\sum_{i=1}^N w^i}.$$

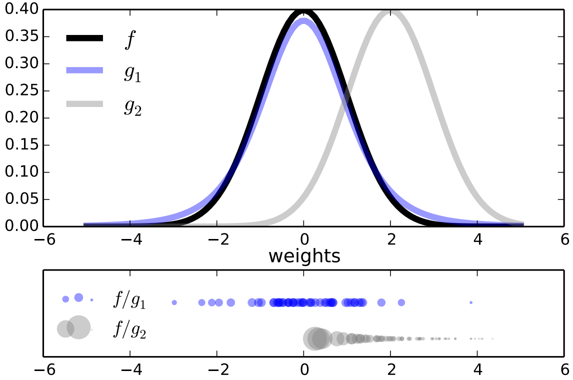
An approximation of $\mathbb{E}_\pi[h(\theta)]$ is given by

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^N W^i h(\theta^i).$$

Illustration

If θ^i's are draws from $g(\cdot)$ then

$$\mathbb{E}_\pi[h] \approx \frac{\frac{1}{N}\sum_{i=1}^N h(\theta^i)w(\theta^i)}{\frac{1}{N}\sum_{i=1}^N w(\theta^i)}, \quad w(\theta) = \frac{f(\theta)}{g(\theta)}.$$



Accuracy

- Since we are generating *iid* draws from $g(\theta)$, it's fairly straightforward to derive a CLT:

$$\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \Longrightarrow N\big(0, \Omega(h)\big), \quad \text{where} \quad \Omega(h) = \mathbb{V}_g[(\pi/g)(h - \mathbb{E}_\pi[h])].$$

- Using a crude approximation (see, e.g., Liu (2008)), we can factorize $\Omega(h)$ as follows:

$$\Omega(h) \approx \mathbb{V}_\pi[h]\big(\mathbb{V}_g[\pi/g] + 1\big).$$

  The approximation highlights that the larger the variance of the importance weights, the less accurate the Monte Carlo approximation relative to the accuracy that could be achieved with an *iid* sample from the posterior.

- Users often monitor

$$ESS = N\frac{\mathbb{V}_\pi[h]}{\Omega(h)} \approx \frac{N}{1 + \mathbb{V}_g[\pi/g]}.$$

From Importance Sampling to Sequential Importance Sampling

- In general, it's hard to construct a good proposal density $g(\theta)$,
- especially if the posterior has several peaks and valleys.
- Idea - Part 1: it might be easier to find a proposal density for

$$\pi_n(\theta) = \frac{[p(Y|\theta)]^{\phi_n} p(\theta)}{\int [p(Y|\theta)]^{\phi_n} p(\theta) d\theta} = \frac{f_n(\theta)}{Z_n}.$$

  at least if ϕ_n is close to zero.

- Idea - Part 2: We can try to turn a proposal density for π_n into a proposal density for π_{n+1} and iterate, letting $\phi_n \longrightarrow \phi_N = 1$.

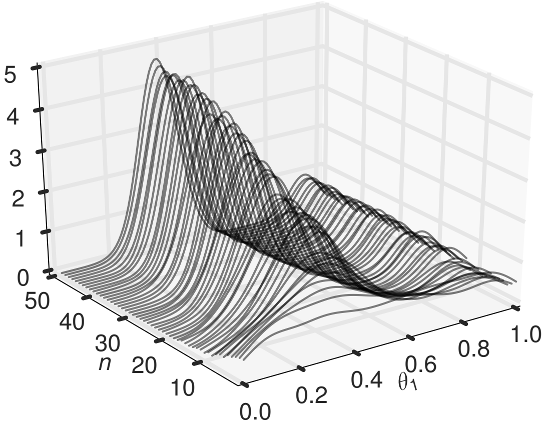
Illustration:

- Our state-space model:

$$y_t = [1\ 1]s_t, \quad s_t = \begin{bmatrix} \theta_1^2 & 0 \\ (1-\theta_1^2) - \theta_1\theta_2 & (1-\theta_1^2) \end{bmatrix} s_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t.$$

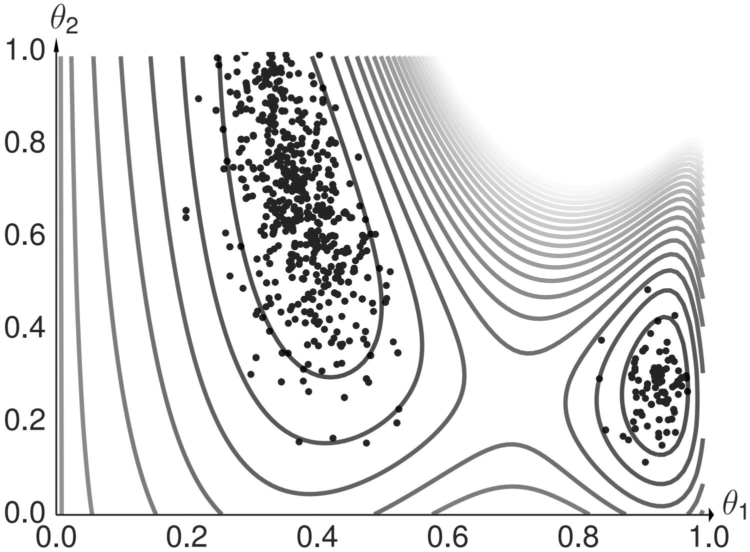
- Innovation: $\epsilon_t \sim iidN(0,1)$.
- Prior: uniform on the square $0 \le \theta_1 \le 1$ and $0 \le \theta_2 \le 1$.
- Simulate $T = 200$ observations given $\theta = [0.45, 0.45]'$, which is observationally equivalent to $\theta = [0.89, 0.22]'$

Illustration: Tempered Posteriors of θ_1

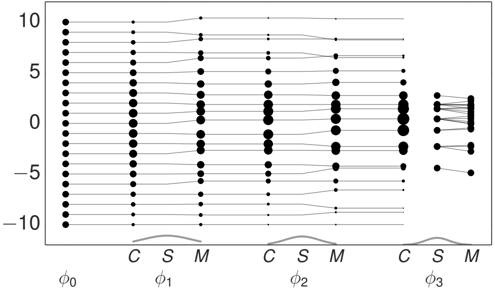


$$\pi_n(\theta) = \frac{[p(Y|\theta)]^{\phi_n} p(\theta)}{\int [p(Y|\theta)]^{\phi_n} p(\theta) d\theta} = \frac{f_n(\theta)}{Z_n}, \quad \phi_n = \left(\frac{n}{N_\phi}\right)^\lambda$$

Illustration: Posterior Draws



SMC Algorithm: A Graphical Illustration



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

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

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

SMC Algorithm

1. **Initialization.** ($\phi_0 = 0$). Draw the initial particles from the prior: $\theta_1^i \stackrel{iid}{\sim} p(\theta)$ and $W_1^i = 1$, $i = 1, \ldots, N$.
2. **Recursion.** For $n = 1, \ldots, N_\phi$,
   1. **Correction.** Reweight the particles from stage $n-1$ by defining the incremental weights
   $$\tilde{w}_n^i = [p(Y|\theta_{n-1}^i)]^{\phi_n - \phi_{n-1}} \tag{1}$$
   and the normalized weights
   $$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}, \quad i = 1, \ldots, N. \tag{2}$$
   An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by
   $$\tilde{h}_{n,N} = \frac{1}{N}\sum_{i=1}^N \tilde{W}_n^i h(\theta_{n-1}^i). \tag{3}$$
   2. **Selection.**

SMC Algorithm

1. **Initialization.**
2. **Recursion.** For $n = 1, \dots, N_\phi$,
   1. **Correction.**
   2. **Selection. (Optional Resampling)** Let $\{\hat{\theta}\}_{i=1}^N$ denote N *iid* draws from a multinomial distribution characterized by support points and weights $\{\theta_{n-1}^i, \tilde{W}_n^i\}_{i=1}^N$ and set $W_n^i = 1$. An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by

$$\hat{h}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} W_n^i h(\hat{\theta}_n^i). \tag{4}$$

   3. **Mutation.** Propagate the particles $\{\hat{\theta}_i, W_n^i\}$ via N_{MH} steps of a MH algorithm with transition density $\theta_n^i \sim K_n(\theta_n|\hat{\theta}_n^i;\zeta_n)$ and stationary distribution $\pi_n(\theta)$. An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by

$$\bar{h}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} h(\theta_n^i) W_n^i. \tag{5}$$

Remarks

- Correction Step:
  - reweight particles from iteration $n-1$ to create importance sampling approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$
- Selection Step: the resampling of the particles
  - (good) equalizes the particle weights and thereby increases accuracy of subsequent importance sampling approximations;
  - (not good) adds a bit of noise to the MC approximation.
- Mutation Step: changes particle values
  - adapts particles to posterior $\pi_n(\theta)$;
  - imagine we don't do it: then we would be using draws from prior $p(\theta)$ to approximate posterior $\pi(\theta)$, which can't be good!

Theoretical Properties

- Goal: strong law of large numbers (SLLN) and central limit theorem (CLT) as $N \longrightarrow \infty$ for every iteration $n = 1, \ldots, N_\phi$.
- Regularity conditions:
  - proper prior;
  - bounded likelihood function;
  - $2 + \delta$ posterior moments of $h(\theta)$.
- Idea of proof (Chopin, 2004): proceed recursively
  - Initialization: SLLN and CLT for *iid* random variables because we sample from prior.
  - Assume that $n-1$ approximation (with normalized weights) yields

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} h(\theta^i_{n-1})W^i_{n-1} - \mathbb{E}_{\pi_{n-1}}[h(\theta)]\right) \Longrightarrow N\big(0, \Omega_{n-1}(h)\big)$$

  - Show that

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} h(\theta^i_{n})W^i_{n} - \mathbb{E}_{\pi_{n}}[h(\theta)]\right) \Longrightarrow N\big(0, \Omega_{n}(h)\big)$$

Theoretical Properties: Correction Step

- Suppose that the $n-1$ approximation (with normalized weights) yields

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} h(\theta^i_{n-1})W^i_{n-1} - \mathbb{E}_{\pi_{n-1}}[h(\theta)]\right) \Longrightarrow N\big(0,\Omega_{n-1}(h)\big)$$

- Then

$$\sqrt{N}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} h(\theta^i_{n-1})[p(Y|\theta^i_{n-1})]^{\phi_n-\phi_{n-1}}W^i_{n-1}}{\frac{1}{N}\sum_{i=1}^{N}[p(Y|\theta^i_{n-1})]^{\phi_n-\phi_{n-1}}W^i_{n-1}} - \mathbb{E}_{\pi_n}[h(\theta)]\right)$$
$$\Longrightarrow \quad N\big(0,\tilde{\Omega}_n(h)\big)$$

where

$$\tilde{\Omega}_n(h) = \Omega_{n-1}\big(v_{n-1}(\theta)(h-\mathbb{E}_{\pi_n}[h])\big) \quad v_{n-1}(\theta) = [p(Y|\theta)]^{\phi_n-\phi_{n-1}}\frac{Z_{n-1}}{Z_n}$$

- This step relies on likelihood evaluations from iteration $n-1$ that are already stored in memory.

Theoretical Properties: Selection / Resampling

- After resampling by drawing from iid multinomial distribution we obtain

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} h(\hat{\theta}_i)W_n^i - \mathbb{E}_{\pi_n}[h]\right) \Longrightarrow N\big(0, \hat{\Omega}(h)\big),$$

where

$$\hat{\Omega}_n(h) = \tilde{\Omega}(h) + \mathbb{V}_{\pi_n}[h]$$

- Disadvantage of resampling: it adds noise.
- Advantage of resampling: it equalizes the particle weights, reducing the variance of $v_n(\theta)$ in $\tilde{\Omega}_{n+1}(h) = \Omega_n\big(v_n(\theta)(h - \mathbb{E}_{\pi_{n+1}}[h]\big)$.

Theoretical Properties: Mutation

- We are using the Markov transition kernel $K_n(\theta|\hat{\theta})$ to transform draws $\hat{\theta}_n^i$ into draws θ_n^i.
- To preserve the distribution of the $\hat{\theta}_n^i$'s it has to be the case that

$$\pi_n(\theta) = \int K_n(\theta|\hat{\theta})\pi_n(\hat{\theta})d\hat{\theta}.$$

- It can be shown that the overall asymptotic variance after the mutation is the sum of
  - the variance of the approximation of the conditional mean $\mathbb{E}_{K_n(\cdot|\theta_{n-1})}[h(\theta)]$ which is given by

$$\hat{\Omega}\big(\mathbb{E}_{K_n(\cdot|\theta_{n-1})}[h(\theta)]\big);$$

  - a weighted average of the conditional variance $\mathbb{V}_{K_n(\cdot|\theta_{n-1})}[h(\theta)]$:

$$\int W_{n-1}(\theta_{n-1})v_{n-1}(\theta_{n-1})\mathbb{V}_{K_n(\cdot|\theta_{n-1})}[h(\theta)]\pi_{n-1}(\theta_{n-1}).$$

- This step is *embarassingly parallelizable*, well designed for single instruction, multiple data (SIMD) processing.

More on Transition Kernel in Mutation Step

- Transition kernel $K_n(\theta|\hat{\theta}_{n-1};\zeta_n)$: generated by running M steps of a Metropolis-Hastings algorithm.

- Lessons from DSGE model MCMC:
  - blocking of parameters can reduces persistence of Markov chain;
  - mixture proposal density avoids "getting stuck."

- Blocking: Partition the parameter vector θ_n into N_{blocks} equally sized blocks, denoted by $\theta_{n,b}$, $b = 1,\ldots,N_{blocks}$. (We generate the blocks for $n = 1,\ldots,N_\phi$ randomly prior to running the SMC algorithm.)

- Example: random walk proposal density:

$$\vartheta_b|(\theta^i_{n,b,m-1},\theta^i_{n,-b,m},\Sigma^*_{n,b}) \quad \sim \quad N\left(\theta^i_{n,b,m-1}, c^2_n\Sigma^*_{n,b}\right).$$

Adaptive Choice of $\zeta_n = (\Sigma_n^*, c_n)$

- Infeasible adaption:
  - Let $\Sigma_n^* = \mathbb{V}_{\pi_n}[\theta]$.
  - Adjust scaling factor according to

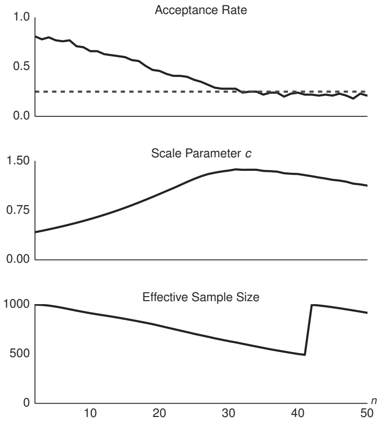
    $$c_n = c_{n-1} f\big(1 - R_{n-1}(\zeta_{n-1})\big),$$

    where $R_{n-1}(\cdot)$ is population rejection rate from iteration $n-1$ and

    $$f(x) = 0.95 + 0.10 \frac{e^{16(x-0.25)}}{1 + e^{16(x-0.25)}}.$$

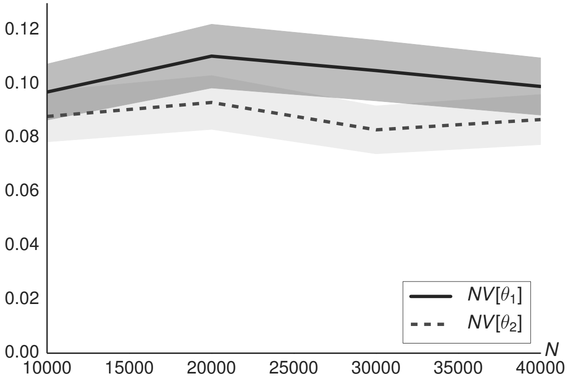
- Feasible adaption – use output from stage $n-1$ to replace ζ_n by $\hat{\zeta}_n$:
  - Use particle approximations of $\mathbb{E}_{\pi_n}[\theta]$ and $\mathbb{V}_{\pi_n}[\theta]$ based on $\{\theta_{n-1}^i, \tilde{W}_n^i\}_{i=1}^N$.
  - Use actual rejection rate from stage $n-1$ to calculate $\hat{c}_n = \hat{c}_{n-1} f\big(\hat{R}_{n-1}(\hat{\zeta}_{n-1})\big)$.

Adaption of SMC Algorithm for Stylized State-Space Model



Notes: The dashed line in the top panel indicates the target acceptance rate of 0.25.

Convergence of SMC Approximation for Stylized State-Space Model



Notes: The figure shows $N\mathbb{V}[\bar{\theta}_j]$ for each parameter as a function of the number of particles N. $\mathbb{V}[\bar{\theta}_j]$ is computed based on $N_{run} = 1,000$ runs of the SMC algorithm with $N_\phi = 100$. The width of the bands is $(2 \cdot 1.96)\sqrt{3/N_{run}}(N\mathbb{V}[\bar{\theta}_j])$.

More on Resampling

- So far, we have used *multinomial resampling*. It's fairly intuitive and it is straightforward to obtain a CLT.
- But: *multinominal resampling is not particularly efficient*.
- The Herbst-Schorfheide book contains a section on alternative resampling schemes (*stratified resampling*, *residual resampling*...)
- These alternative techniques are designed to achieve a variance reduction.
- Most resampling algorithms are not parallelizable because they rely on the normalized particle weights.

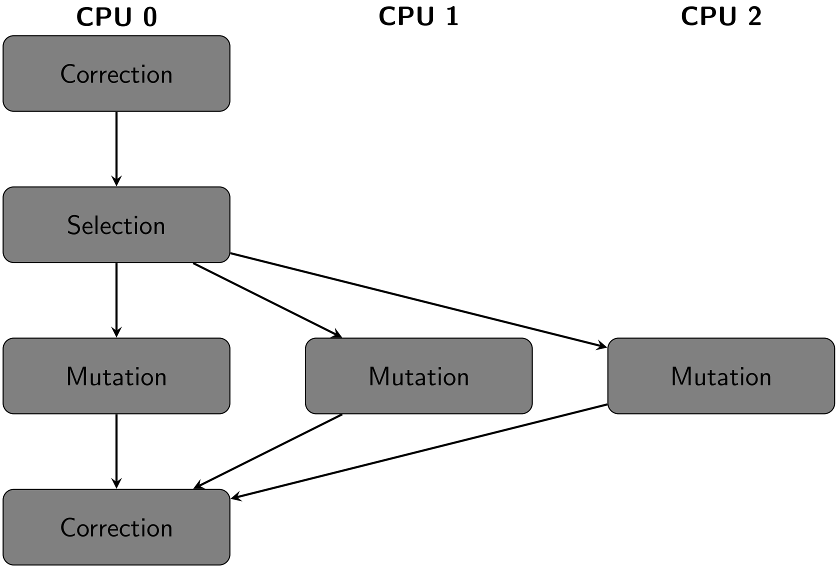
Running Time – It's all about Mutation

- The most time consuming part of (any of) these algorithms, is **evaluating the likelihood function,** which occurs in the mutation step.
- But each particle is *mutated independently* of the other particles.
- This is extremely easy to parallelize.

How I do it – distributed memory parallelization in `Fortran`

- Use Message Passing Interface (MPI) to scatter particles across many processors (CPUs).
- Execute mutuation across processors.
- Use MPI to gather the newly mutated particles.

Could be better with more programming.


CPU 0
CPU 1
CPU 2
Correction
Selection
Mutation
Mutation
Mutation
Correction

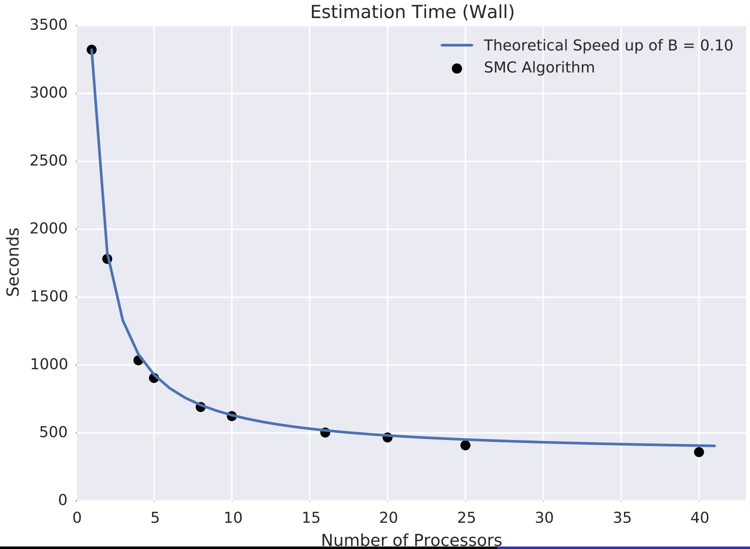
How well does this work?

- The extent to which HPC can help us is determined by the amount of algorithm that can be executed in parallel vs. serial.
- Suppose a fraction $B \in [0,1]$ must executed in serial fashion for a particular algorithm.
- **Amdahls Law**: Theoretical gain from using N processors in an algorithm is given by:

$$R(N) = B + \frac{1}{N}(1 - B)$$

- Question: What is B for our SMC algorithm?
  Answer: about 0.1!

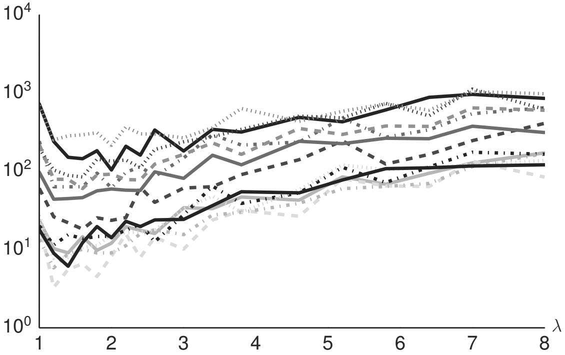
Gains from Parallelization



Application 1: Small Scale New Keynesian Model

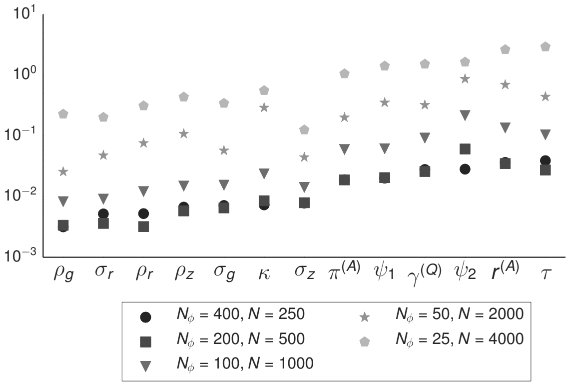
- We will take a look at the effect of various tuning choices on accuracy:
  - Tempering schedule λ: $\lambda = 1$ is linear, $\lambda > 1$ is convex.
  - Number of stages N_ϕ versus number of particles N.

Effect of λ on Inefficiency Factors $\text{InEff}_N[\bar{\theta}]$



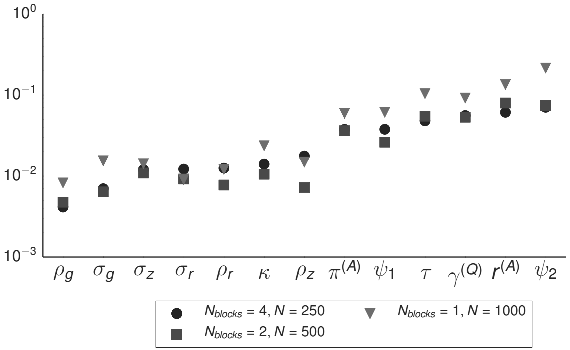
Notes: The figure depicts hairs of $\text{InEff}_N[\bar{\theta}]$ as function of λ. The inefficiency factors are computed based on $N_{run} = 50$ runs of the SMC algorithm. Each hair corresponds to a DSGE model parameter.

Number of Stages N_ϕ vs Number of Particles N



Notes: Plot of $\mathbb{V}[\bar{\theta}]/\mathbb{V}_\pi[\theta]$ for a specific configuration of the SMC algorithm. The inefficiency factors are computed based on $N_{run} = 50$ runs of the SMC algorithm. $N_{blocks} = 1$, $\lambda = 2$, $N_{MH} = 1$.

Number of blocks N_{blocks} in Mutation Step vs Number of Particles N



Notes: Plot of $\mathbb{V}[\bar{\theta}]/\mathbb{V}_\pi[\theta]$ for a specific configuration of the SMC algorithm. The inefficiency factors are computed based on $N_{run} = 50$ runs of the SMC algorithm. $N_\phi = 100$, $\lambda = 2$, $N_{MH} = 1$.

A Few Words on Posterior Model Probabilities

- Posterior model probabilities

$$\pi_{i,T} = \frac{\pi_{i,0}\, p(Y_{1:T}|\mathcal{M}_i)}{\sum_{j=1}^{M} \pi_{j,0}\, p(Y_{1:T}|\mathcal{M}_j)}$$

  where

$$p(Y_{1:T}|\mathcal{M}_i) = \int p(Y_{1:T}|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)}|\mathcal{M}_i) d\theta_{(i)}$$

- For any model:

$$\ln p(Y_{1:T}|\mathcal{M}_i) = \sum_{t=1}^{T} \ln \int p(y_t|\theta_{(i)}, Y_{1:t-1}, \mathcal{M}_i) p(\theta_{(i)}|Y_{1:t-1}, \mathcal{M}_i) d\theta_{(i)}$$

- Marginal data density $p(Y_{1:T}|\mathcal{M}_i)$ arises as a by-product of SMC.

Marginal Likelihood Approximation

- Recall $\tilde{w}_n^i = [p(Y|\theta_{n-1}^i)]^{\phi_n - \phi_{n-1}}$.
- Then

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} \tilde{w}_n^i W_{n-1}^i &\approx \int [p(Y|\theta)]^{\phi_n - \phi_{n-1}} \frac{p^{\phi_{n-1}}(Y|\theta)p(\theta)}{\int p^{\phi_{n-1}}(Y|\theta)p(\theta)d\theta} d\theta \\
&= \frac{\int p(Y|\theta)^{\phi_n} p(\theta)d\theta}{\int p(Y|\theta)^{\phi_{n-1}} p(\theta)d\theta}
\end{aligned}
$$

- Thus,

$$
\prod_{n=1}^{N_\phi}\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{w}_n^i W_{n-1}^i\right) \approx \int p(Y|\theta)p(\theta)d\theta.
$$

SMC Marginal Data Density Estimates

	$N_\phi = 100$		$N_\phi = 400$	
N	Mean($\ln \hat{p}(Y)$)	SD($\ln \hat{p}(Y)$)	Mean($\ln \hat{p}(Y)$)	SD($\ln \hat{p}(Y)$)
500	-352.19	(3.18)	-346.12	(0.20)
1,000	-349.19	(1.98)	-346.17	(0.14)
2,000	-348.57	(1.65)	-346.16	(0.12)
4,000	-347.74	(0.92)	-346.16	(0.07)

Notes: Table shows mean and standard deviation of log marginal data density estimates as a function of the number of particles N computed over $N_{run} = 50$ runs of the SMC sampler with $N_{blocks} = 4$, $\lambda = 2$, and $N_{MH} = 1$.

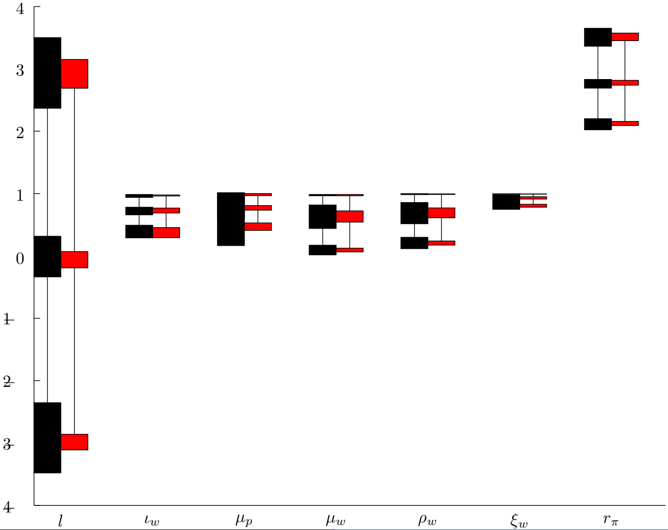
Application 2: Estimation of Smets and Wouters (2007) Model

- Benchmark macro model, has been estimated many (many) times.
- "Core" of many larger-scale models.
- 36 estimated parameters.
- RWMH: 10 million draws (5 million discarded); SMC: 500 stages with 12,000 particles.
- We run the RWM (using a particular version of a parallelized MCMC) and the SMC algorithm on 24 processors for the same amount of time.
- We estimate the SW model twenty times using RWM and SMC and get essentially identical results.

Application 2: Estimation of Smets and Wouters (2007) Model

- More interesting question: how does quality of posterior simulators change as one makes the priors more diffuse?
- Replace Beta by Uniform distributions; increase variances of parameters with Gamma and Normal prior by factor of 3.
- Motivation:
  - SW priors might be considered implausible because they seem to be informed by in-sample information.
  - Del Negro and Schorfheide (2008): inference about wage and price stickiness is very sensitive to priors.
  - Müller (2011) finds that posterior is sensitive to small shifts in prior mean.
  - Del Negro and Schorfheide (2013) report a strong effect of priors for steady state parameters on forecast performance.
  - Posterior odds in favor of specification with diffuse prior exp(28).

SW Model with DIFFUSE Prior: Estimation stability RWH (black) versus SMC (red)



A Measure of Effective Number of Draws

- Suppose we could generate *iid* N_{eff} draws from posterior, then

$$\hat{\mathbb{E}}_\pi[\theta] \stackrel{approx}{\sim} N\left(\mathbb{E}_\pi[\theta], \frac{1}{N_{eff}}\mathbb{V}_\pi[\theta]\right).$$

- We can measure the variance of $\hat{\mathbb{E}}_\pi[\theta]$ by running SMC and RWM algorithm repeatedly.

- Then,

$$N_{eff} \approx \frac{\mathbb{V}_\pi[\theta]}{\mathbb{V}\left[\hat{\mathbb{E}}_\pi[\theta]\right]}$$

Effective Number of Draws

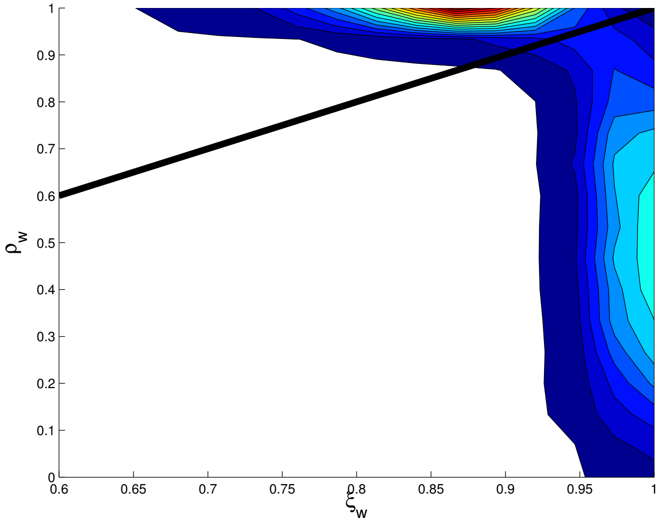
	SMC			RWMH		
Parameter	Mean	STD(Mean)	N_{eff}	Mean	STD(Mean)	N_{eff}
σ_l	3.06	0.04	1058	3.04	0.15	60
l	-0.06	0.07	732	-0.01	0.16	177
ι_p	0.11	0.00	637	0.12	0.02	19
h	0.70	0.00	522	0.69	0.03	5
Φ	1.71	0.01	514	1.69	0.04	10
r_π	2.78	0.02	507	2.76	0.03	159
ρ_b	0.19	0.01	440	0.21	0.08	3
φ	8.12	0.16	266	7.98	1.03	6
σ_p	0.14	0.00	126	0.15	0.04	1
ξ_p	0.72	0.01	91	0.73	0.03	5
ι_w	0.73	0.02	87	0.72	0.03	36
μ_p	0.77	0.02	77	0.80	0.10	3
ρ_w	0.69	0.04	49	0.69	0.09	11
μ_w	0.63	0.05	49	0.63	0.09	11
ξ_w	0.93	0.01	43	0.93	0.02	8

A Closer Look at the Posterior: Two Modes

Parameter	Mode 1	Mode 2
ξ_w	0.844	0.962
ι_w	0.812	0.918
ρ_w	0.997	0.394
μ_w	0.978	0.267
Log Posterior	-804.14	-803.51

- Mode 1 implies that wage persistence is driven by extremely exogenous persistent wage markup shocks.
- Mode 2 implies that wage persistence is driven by endogenous amplification of shocks through the wage Calvo and indexation parameter.
- SMC is able to capture the two modes.

A Closer Look at the Posterior: Internal ξ_w versus External ρ_w Propagation



Stability of Posterior Computations: RWH (black) versus SMC (red)