Particle Filtering with Dynare

S. Adjemian[1,2] F. Karamé[1,2]

[1]Université du Mans (GAINS-TEPP, FR CNRS 2042 and IRA)

[2]Dynare Team, Cepremap

Dynare Summer School

Motivations

- DSGE models are now largely widespread among academic and institutional macro-economists.
- These models are often approximated by a first order Taylor approximation around the steady state.
- The reduced form solution has a linear state-space representation, whose likelihood can be computed with the Kalman filter
- The computational gain comes at the cost that not all questions can be addressed with a linear(ized) model which by nature displays the certainty equivalence property.
- By approximating the model, we also loose some information about the deep parameters to be estimated.
- This paper reviews the econometric approach and their implementation in Dynare for handling DSGE models beyond the (log-)linear approximation.

The reduced form model with a first-order approximation is:

$$z_t = \bar{z}(\theta) + g_z(\theta)\hat{z}_{t-1} + g_u(\theta)u_t$$

which is a (stacked) linear state-space model. Inference on the latent variables and/or parameters is possible with the Kalman filter and MLE.

The reduced form model with a second-order approximation is:

$$\begin{aligned} z_t = \bar{z}(\theta) &+ g_z(\theta)\hat{z}_{t-1} + g_u(\theta)u_t \\ &+ 0.5\, g_{\sigma\sigma}(\theta) \\ &+ 0.5\, g_{yy}(\theta)\ (\hat{z}_{t-1} \otimes \hat{z}_{t-1}) \\ &+ 0.5\, g_{uu}(\theta)\ (u_t \otimes u_t) \\ &+ 0.5\, g_{uz}(\theta)\ (\hat{z}_{t-1} \otimes u_t) \end{aligned}$$

Outline of the talk

The state/space representation

Assume the reduced form of the DSGE model can be cast into the following representation:

$$s_t = f(s_{t-1}, \varepsilon_t; \theta)$$
$$y_t = g(s_t; \theta) + e_t$$

In the sequel the sample is denoted $y_{1:T} = \{y_t\}_{t=1}^T$. More generally $x_{1:t} = \{x_i\}_{i=1}^t$ denotes the set of variables x up to time t.

- What can we say about the latent variables? $\Rightarrow$ Characterize the distribution of s_t conditional on the sample up to time $t-1$ and t (namely $p(s_t|y_{1:t-1})$ and $p(s_t|y_{1:t})$).
- What can we say about the likelihood of an observation? $\Rightarrow$ Characterize the distribution of y_t conditional on the states at time t (s_t) or the sample up to time $t-1$ (resp. $p(y_t|s_t)$ and $p(y_t|y_{1:t-1})$) .

Assumptions

The reduced form DSGE model displays the following properties:

- the first-order Markov relationship on states variables:

  $$p(s_t \,|s_{0:t-1}, y_{1:t-1}) = p(s_t \,|s_{t-1})$$

  unknown but can be easily drawn in by simulating state equations $s_t = f(s_{t-1}, \varepsilon_t; \theta)$.

- the conditional independence of observations:

  $$p(y_t \,|y_{1:t-1}, s_{0:t}) = p(y_t \,|s_t)$$

  can be (easily) evaluated since measurement errors $e_t (= y_t - g(s_t; \theta))$ are assumed gaussian.

The optimal Bayesian approach

Suppose that the parameters, θ, and the distribution of the states at time $t-1$ conditional on the sample up to time $t-1$, $p(s_{t-1}|y_{1:t-1})$, are known.

Given these informations, we want to characterize how we update our beliefs about the latent variables ($p(s_t|y_{1:t-1})$) when a new observation y_t becomes available, *i.e.* characterize $p(s_t|y_{1:t})$.

Traditionally, two steps are implemented to obtain this relation:

- The prediction step:

$$p(s_t|y_{1:t-1}) = \int p(s_t|s_{t-1})\, p(s_{t-1}|y_{1:t-1}) ds_{t-1}$$

- The updating step:

$$p(s_t\,|y_{1:t}) = \frac{p(y_t\,|s_t)\,p(s_t\,|y_{1:t-1})}{p(y_t\,|y_{1:t-1})}$$

  where $p(y_t\,|y_{1:t-1}) = \int p(y_t\,|s_t)\,p(s_t\,|y_{1:t-1})\,ds_t$ acts as a constant of integration.

Skip proof

The prediction step

$$\begin{aligned} p(s_t\,|y_{1:t-1}) &= \int p(s_t, s_{t-1}\,|y_{1:t-1})\, ds_{t-1} \\ &= \int p(s_t\,|s_{t-1}, y_{1:t-1})\, p(s_{t-1}\,|y_{1:t-1})\, ds_{t-1} \\ &= \int p(s_t\,|s_{t-1})\, p(s_{t-1}\,|y_{1:t-1})\, ds_{t-1} \end{aligned}$$

$p(s_t\,|s_{t-1})$ is unknown but can be easily drawn in by simulating state equations $s_t = f(s_{t-1}, \varepsilon_t; \theta)$.

The updating step

$$
\begin{aligned}
p(s_t | y_{1:t}) &= \frac{p(y_{1:t} | s_t)\, p(s_t)}{p(y_{1:t})} \\
&= \frac{p(y_t, y_{1:t-1} | s_t)\, p(s_t)}{p(y_t, y_{1:t-1})} \\
&= \frac{p(y_t | y_{1:t-1}, s_t)\, p(y_{1:t-1} | s_t)\, p(s_t)}{p(y_t | y_{1:t-1})\, p(y_{1:t-1})} \\
&= \frac{p(y_t | y_{1:t-1}, s_t)\, p(y_{1:t-1} | s_t)\, p(s_t)}{p(y_t | y_{1:t-1})\, p(y_{1:t-1})} \\
&= \frac{p(y_t | y_{1:t-1}, s_t) p(s_t | y_{1:t-1})}{p(y_t | y_{1:t-1})} \\
&= \frac{p(y_t | s_t) p(s_t | y_{1:t-1})}{p(y_t | y_{1:t-1})}
\end{aligned}
$$

$p(y_t | s_t)$ can be (easily) evaluated since measurement errors are assumed additive and gaussian.

- Suppose a linear/gaussian state/space representation:

$$s_t = f(s_{t-1}, \varepsilon_t; \theta) = Ds_{t-1} + E + \varepsilon_t$$
$$y_t = g(s_t; \theta) + e_t = As_t + B$$

  - $p(s_t | y_{1:t-1})$, $p(s_t | y_{1:t})$, $p(y_t | s_t)$ and $p(y_t | y_{1:t-1})$ are all Gaussian.
  - Only their conditional expectancy and variance have to be tracked;
  - This is what the Kalman filter do analytically:

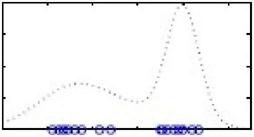
$$\begin{aligned}
\mathbb{E}(s_t|y_{1:t-1}) &= D\mathbb{E}(s_{t-1}|y_{1:t-1}) + E \\
\mathbb{V}(s_t|y_{1:t-1}) &= D\mathbb{V}(s_{t-1}|y_{1:t-1})D^{'} + Q \\
\mathbb{E}(y_t|y_{1:t-1}) &= A\mathbb{E}(s_t|y_{1:t-1}) + B \\
\mathbb{V}(y_t|y_{1:t-1}) &= A\mathbb{V}(s_t|y_{1:t-1})A^{'} \\
\mathbb{E}(s_t|y_{1:t}) &= \mathbb{E}(s_t|y_{1:t-1}) + K_t[y_t - \mathbb{E}(y_t|y_{1:t-1})] \\
\mathbb{V}(s_t|y_{1:t}) &= (I_m - K_t A)\mathbb{V}(s_t|y_{1:t-1})
\end{aligned}$$

  with $K_t = \mathbb{V}(s_t|y_{1:t-1})A^{'}[\mathbb{V}(y_t|y_{1:t-1})]^{-1}$ the Kalman filter gain.

- Otherwise, approximations are required.

Numerical integration (or perfect sampling)

We assume that the continuous distribution of s_t conditional on $y_{1:t}$ can be approximated by a set of particles $\left\{s_t^{(i)}\right\}_{i=1:N}$ with weights $\left\{w_t^{(i)}\right\}_{i=1:N}$ (summing to one).



Any moments of this distribution can be approximated by a weighted average as follows:

$$\mathbb{E}_{p(s_t|y_{1:t})}\left[h(s_t)\right] = \int h(s_t)p(s_t|y_{1:t})\mathrm{d}s_t$$
$$\approx \sum_{i=1}^{N} w_t^{(i)} h\left(s_t^{(i)}\right) = \frac{1}{N}\sum_{i=1}^{N} h\left(s_t^{(i)}\right)$$

The idea

Problem: $p(s_t|y_{1:t})$ is unknown.

Solution: **importance sampling**, *i.e.* choose an easy-to-sample *proposal* distribution $q(s_t|y_{1:t})$ (informative on the target distribution)

$$\begin{aligned}\mathbb{E}_{p(s_t|y_{1:t})}\left[h(s_t)\right] &= \int h(s_t)\frac{p(s_t|y_{1:t})}{q(s_t|y_{1:t})}q(s_t|y_{1:t})\mathrm{d}s_t \\ &= \mathbb{E}_{q(s_t|y_{1:t})}\left[\tilde{w}_t(s_t)h(s_t)\right]\end{aligned}$$

with *normalized* weights defined as:

$$\tilde{w}_t(s_t) \equiv \frac{p(s_t|y_{1:t})}{q(s_t|y_{1:t})}$$

These weights can be viewed as importance ratios. Skip details

Importance sampling

We can equivalently rewrite these weights, by reversing the conditioning in the numerator with the Bayes theorem, as:

$$\tilde{w}_t(s_t) \equiv \frac{p(y_{1:t}|s_t)p(s_t)}{p(y_{1:t})q(s_t|y_{1:t})}$$

Removing the constant of integration (with respect to s_t), we define the *unormalized* weights:

$$\hat{w}_t(s_t) \equiv \frac{p(y_{1:t}|s_t)p(s_t)}{q(s_t|y_{1:t})}$$

A modification that avoids the calculation of $p(y_{1:t})$:

$$\mathbb{E}_{p(s_t|y_{1:t})}\left[h(s_t)\right] = \frac{\mathbb{E}_{q(s_t|y_{1:t})}\left[\hat{w}_t(s_t)h(s_t)\right]}{\mathbb{E}_{q(s_t|y_{1:t})}\left[\hat{w}_t(s_t)\right]}$$

with *unormalized* weights defined as:

$$\hat{w}_t(s_t) \equiv \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_t\,|y_{1:t})} \propto \tilde{w}_t(s_t) \equiv \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_t\,|y_{1:t})\,p(y_{1:t})}$$

Importance sampling

$$\tilde{w}_t(s_t) \equiv \frac{p(s_t\,|y_{1:t})}{q(s_t\,|y_{1:t})} = \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_t\,|y_{1:t})\,p(y_{1:t})} \propto \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_t\,|y_{1:t})} \equiv \hat{w}_t(s_t)$$

A modification that avoids the calculation of $p(y_{1:t})$:

$$\begin{aligned} p(y_{1:t}) &= \int \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_t\,|y_{1:t})} q(s_t\,|y_{1:t}\,)ds_t \\ &= \int \hat{w}_t(s_t) q(s_t\,|y_{1:t})\,ds_t \\ &= \mathbb{E}_{q(s_t|y_{1:t})}\left[\hat{w}_t(s_t)\right] \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{p(s_t|y_{1:t})}\left[h(s_t)\right] &= \frac{1}{p(y_{1:t})} \int h(s_t) \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_t\,|y_{1:t})} q(s_t\,|y_{1:t}) ds_t \\ &= \frac{\mathbb{E}_{q(s_t|y_{1:t}\,)}\left[\hat{w}_t(s_t) h(s_t)\right]}{\mathbb{E}_{q(s_t|y_{1:t})}\left[\hat{w}_t(s_t)\right]} \end{aligned}$$

In practice

Draw $\left\{\tilde{s}_t^{(i)}\right\}_{i=1:N}$ from the proposal $q(s_t|y_{1:t})$ and calculate their respective unormalized weights $\left\{\hat{w}_t^{(i)}\right\}_{i=1:N}$.

$$\begin{aligned}\mathbb{E}_{p(s_t|y_{1:t})}\left[h(s_t)\right] &= \mathbb{E}_{q(s_t|y_{1:t})}\left[\tilde{w}_t(s_t)h(s_t)\right] \\ &= \frac{\mathbb{E}_{q(s_t|y_{1:t})}\left[\hat{w}_t(s_t)h(s_t)\right]}{\mathbb{E}_{q(s_t|y_{1:t})}\left[\hat{w}_t(s_t)\right]} \\ &\approx \frac{\frac{1}{N}\sum_{i=1}^{N}\hat{w}_t^{(i)}h(\tilde{s}_t^{(i)})}{\frac{1}{N}\sum_{i=1}^{N}\hat{w}_t^{(i)}} = \sum_{i=1}^{N}\tilde{w}_t^{(i)}h(\tilde{s}_t^{(i)})\end{aligned}$$

if we define the normalized weights as:

$$\tilde{w}_t^{(i)} = \frac{\hat{w}_t^{(i)}}{\sum_{i=1}^{N}\hat{w}_t^{(i)}}$$

Sequential Importance sampling

A recursive approach can be implemented by choosing a *proposal* verifying:

$$q(s_t\,|y_{1:t}) = q(s_t|s_{t-1}, y_t)\; q(s_{t-1}\,|y_{1:t-1})$$

s_t is simply drawn in $q(s_t|s_{t-1}, y_t)$.

The (unormalized) weights can be recursively calculated as:

$$\hat{w}_t(s_t) \propto \tilde{w}_{t-1}(s_{t-1}) \frac{p(y_t\,|s_t)\, p(s_t\,|s_{t-1})}{q(s_t\,|s_{t-1}, y_t)}$$

▸ Skip proof

Unnormalized weights are defined by:

$$\begin{aligned}
\hat{w}_t(s_t) &\equiv \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_t\,|y_{1:t})}\,\frac{p(y_{1:t-1}\,|s_{t-1}\,)p(s_{t-1})}{p(y_{1:t-1}\,|s_{t-1})\,p(s_{t-1})} \\
&= \frac{p(y_{1:t}\,|s_t)\,p(s_t)}{q(s_{t-1}\,|y_{1:t-1})\,q\,(s_t|s_{t-1},y_t)}\,\frac{p(y_{1:t-1}\,|s_{t-1}\,)p(s_{t-1})}{p(y_{1:t-1}\,|s_{t-1})\,p(s_{t-1})} \\
&= \hat{w}_{t-1}(s_{t-1})\,\frac{p(y_{1:t}\,|s_t)\,p(s_t)}{p(y_{1:t-1}\,|s_{t-1})\,p(s_{t-1})q\,(s_t|s_{t-1},y_t)}
\end{aligned}$$

Since $p(s_t) = p(s_0)\prod_{i=1}^{t} p(s_i\,|s_{i-1})$ and $p(y_{1:t}\,|s_t) = \prod_{i=1}^{t} p(y_i\,|s_i)$:

$$\frac{p(y_{1:t}\,|s_t)\,p(s_t)}{p(y_{1:t-1}\,|s_{t-1})\,p(s_{t-1})} = p(y_t\,|s_t)\,p(s_t\,|s_{t-1})$$

We obtain:

$$\begin{aligned}
\hat{w}_t(s_t) &= \hat{w}_{t-1}(s_{t-1})\,\frac{p(y_t\,|s_t\,)p(s_t\,|s_{t-1}\,)}{q\,(s_t|s_{t-1},y_t)} \\
&\propto \tilde{w}_{t-1}(s_{t-1})\,\frac{p(y_t\,|s_t)\,p(s_t\,|s_{t-1})}{q\,(s_t|s_{t-1},y_t)}
\end{aligned}$$

A generic particle filter

For $t = 1, \ldots, T$ and $i = 1, \ldots, N$,

At time t, knowing $\left\{s_{t-1}^{(i)}, w_{t-1}^{(i)}\right\}_{i=1:N}$ $(\approx p(s_{t-1}|y_{1:t-1}))$:

- Draw $\left\{\tilde{s}_t^{(i)}\right\}_{i=1:N}$ from $q\left(s_t|s_{t-1}^{(i)}, y_t\right)$.
- Evaluate the weights: $\hat{w}_t^{(i)} = w_{t-1}^{(i)} \frac{p(y_t|\tilde{s}_t^{(i)})p(\tilde{s}_t^{(i)}|s_{t-1}^{(i)})}{q(\tilde{s}_t^{(i)}|s_{t-1}^{(i)}, y_t)}$.
- Use $\left\{\tilde{s}_t^{(i)}, \tilde{w}_t^{(i)} = \frac{\hat{w}_t^{(i)}}{\sum_{i=1}^N \hat{w}_t^{(i)}}\right\}_{i=1:N}$ as $\left\{s_t^{(i)}, w_t^{(i)}\right\}_{i=1:N}$ for the next filter iteration.

Degeneracy problem and resampling

- No update here for the moment...
- As t increases, all-but-one particles have negligible weights (essentially in large samples).
- That is the reason why systematic resampling was initially proposed in the literature (Gordon *et al.*, 1993).
- It consists in randomly drawing with replacement particles in their empirical distribution $\left\{\tilde{s}_t^{(i)}, \tilde{w}_t^{(i)}\right\}_{i=1:N}$.
- It amounts to discard particles with low weights and replicate particles with high weights to focus on interesting areas of the distribution using a constant number of particles.
- Negative side-effets
  - Impoverishment of the particles swarm.
  - It can render a filter iteration relatively time-consuming.
  - For estimation (see later).

Illustration

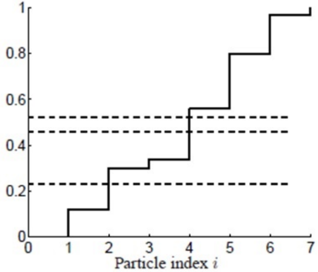


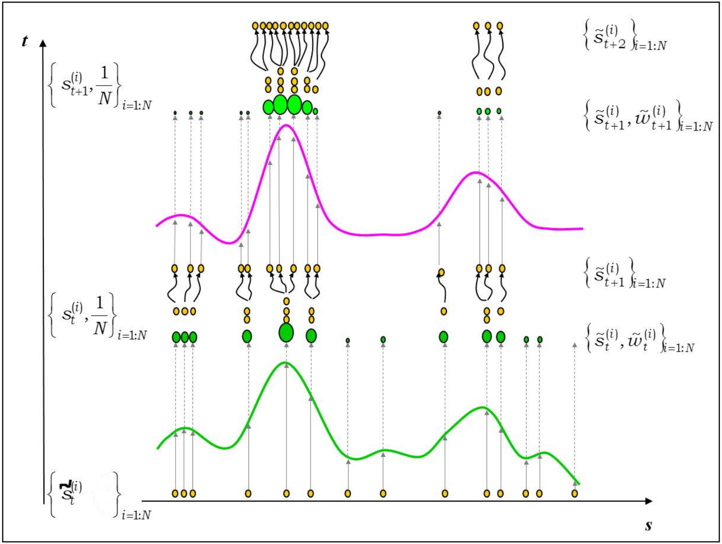
Figure 4.1: *Illustrating the resampling step in the particle filter. The new set of particles is obtained by first generating M sorted uniformly distributed random numbers, three of which are shown by the dashed lines in the figure. These are then associated with a particle guided by the cumulative sum of the normalized importance weights. In the figure particle number 2 is chosen once and particle number 4 is chosen twice.*

Sequential Importance-Sampling with Systematic Resampling

For $t = 1, \dots, T$ and $i = 1, \dots, N$,

At time t, knowing $\left\{s_{t-1}^{(i)}, w_{t-1}^{(i)}\right\}_{i=1:N}$ $(\approx p(s_{t-1}|y_{1:t-1}))$:

- Draw $\left\{\tilde{s}_t^{(i)}\right\}_{i=1:N}$ from $q\left(s_t|s_{t-1}^{(i)}, y_t\right)$.
- Evaluate the weights: $\hat{w}_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(y_t|\tilde{s}_t^{(i)})p(\tilde{s}_t^{(i)}|s_{t-1}^{(i)})}{q(\tilde{s}_t^{(i)}|s_{t-1}^{(i)}, y_t)}$.
- Resample $\left\{\tilde{s}_t^{(i)}, \tilde{w}_t^{(i)} = \frac{\hat{w}_t^{(i)}}{\sum_{i=1}^N \hat{w}_t^{(i)}}\right\}_{i=1:N} \mapsto \left\{s_t^{(i)}, w_t^{(i)} = \frac{1}{N}\right\}_{i=1:N}$ $(\approx p(s_t|y_{1:t}))$.


t
$\left\{ s_{t+1}^{(i)}, \frac{1}{N} \right\}_{i=1:N}$
$\left\{ \tilde{s}_{t+2}^{(i)} \right\}_{i=1:N}$
$\left\{ \tilde{s}_{t+1}^{(i)}, \tilde{w}_{t+1}^{(i)} \right\}_{i=1:N}$
$\left\{ \tilde{s}_{t+1}^{(i)} \right\}_{i=1:N}$
$\left\{ s_{t}^{(i)}, \frac{1}{N} \right\}_{i=1:N}$
$\left\{ \tilde{s}_{t}^{(i)}, \tilde{w}_{t}^{(i)} \right\}_{i=1:N}$
$\left\{ \tilde{s}_{t}^{(i)} \right\}_{i=1:N}$
s

The sample likelihood

$$p(y_{1:T}\,|\theta\,) = p(y_1\,|s_0;\theta\,)p(s_0\,|\theta\,)\prod_{t=2}^{T} p(y_t\,|y_{1:t-1};\theta\,)$$

with

$$p(y_t\,|y_{1:t-1};\theta\,) \approx \sum_{i=1}^{N} \hat{w}_t^{(i)}$$

If $q(s_t|s_{t-1},y_t;\theta) = p(s_t|s_{t-1};\theta)$ and no pre-selection step:

$$p(y_t|y_{1:t-1};\theta) \approx \sum_{i=1}^{N} \tilde{w}_{t-1}^{(i)} p(y_t|\tilde{s}_t^{(i)};\theta)$$

In case of systematic resampling, since $\tilde{w}_{t-1}^{(i)} = 1/N$:

$$p(y_t|y_{1:t-1};\theta) \approx \frac{1}{N}\sum_{i=1}^{N} p(y_t|\tilde{s}_t^{(i)};\theta)$$

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t) = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(c_t^{\theta}(1-l_t)^{1-\theta}\right)^{1-\tau}}{1-\tau}$$

subject to

$$y_t = c_t + i_t$$
$$y_t = a_t k_t^{\alpha} l_t^{1-\alpha}$$
$$k_{t+1} = i_t + (1-\delta) k_t$$
$$\log(a_{t+1}) = \rho \log(a_t) + \epsilon_t$$

▸ Skip details

$$\frac{1-\theta}{\theta}\frac{c_t}{1-l_t} = (1-\alpha)\frac{y_t}{l_t}$$

$$\beta\mathbb{E}_t\left[\frac{\left(c_{t+1}^{\theta}(1-l_{t+1})^{1-\theta}\right)^{1-\tau}}{c_{t+1}}\left(1-\delta+\alpha\frac{y_{t+1}}{k_{t+1}}\right)\right] = \frac{\left(c_t^{\theta}(1-l_t)^{1-\theta}\right)^{1-\tau}}{c_t}$$

$$k_{t+1} = i_t + (1-\delta)k_t$$

$$i_t = y_t - c_t$$

$$y_t = a_t k_t^{\alpha} l_t^{1-\alpha}$$

$$\log(a_t) = \rho\log(a_{t-1}) + \varepsilon_t$$

$$\bar{k} = \frac{-(\alpha-1)\alpha^{\frac{1}{1-\alpha}}\beta^{\frac{1}{1-\alpha}}\left(\left(\beta(\delta-1)+1\right)^{\frac{\alpha}{\alpha-1}}\right)\theta}{-\alpha\delta\beta+\delta\beta+\alpha\theta\beta-\beta-\alpha\theta+1}$$

$$\bar{a} = 1$$

$$\bar{l} = \frac{(\alpha-1)(\beta(\delta-1)+1)\theta}{\alpha\theta+\beta((\alpha-1)\delta-\alpha\theta+1)-1}$$

$$\bar{y} = \bar{k}^{\alpha}\bar{l}^{1-\alpha}$$

$$\bar{i} = \delta\bar{k}$$

$$\bar{c} = \bar{y}-\bar{i}$$

Parameters		Benchmark	Risky	Extreme	*Prior*
Capital elasticity in the production function	α	0.4	0.4	0.4	$U_{[0;1]}$
Consumption-leisure substitution rate	θ	0.357	0.357	0.357	$U_{[0;1]}$
Discount factor	β	0.99	0.99	0.99	$U_{[0.75;1]}$
Depreciation rate of capital	δ	0.02	0.02	0.02	$U_{[0;0.05]}$
Persistence of technological shock	ρ	0.95	0.95	0.95	$U_{[0;1]}$
Intertemporal substitution	τ	2	50	50	$U_{[0;100]}$
Standard deviation of technological shock	σ_ϵ	0.007	0.035	0.035	$U_{[0;.1]}$
Standard deviation	σ_y	0.00158	0.0175	0.000158	$U_{[0;0.1]}$
of	σ_l	0.0011	0.00312	0.0011	$U_{[0;0.1]}$
measurement errors	σ_i	0.000866	0.00465	0.000866	$U_{[0;0.1]}$
Datafile:		benchmark	risky	extreme	

Declaration and assignment

```
var k A c l i y;

varexo e_a;

parameters alp bet tet tau delt rho ;

alp = 0.4;
bet = 0.99;
tet = 0.357 ;
tau = 50 ;
delt = 0.02;
rho = 0.95;
```

Model declaration

```
model;
c = ((1 - alp)*tet/(1-tet))*A*(1-l)*((k(-1)/l)^alp) ;
y = A*(k(-1)^alp)*(l^(1-alp)) ;
i = y-c ;
k = (1-delt)*k(-1) + i ;
log(A) = rho*log(A(-1)) + e_a ;
(((c^(tet))*((1-l)^(1-tet)))^(1-tau))/c -
bet*((((c(+1)^(tet))*((1-l(+1))^(1-tet)))^(1-tau))/c(+1))*
(1 -delt+alp*(A(1)*(k^alp)*(l(1)^(1-alp)))/k)=0 ;
end;

shocks;
var e_a; stderr 0.035;
end;

steady;
```

Measurement errors

Be careful: the measurement errors should not be declared in the `shocks` block (as exogenous variables). The `estimated_params` interface should be used to declare that the observed variables have measurement errors.

Priors

```
estimated_params;
alp, uniform_pdf,,, 0.0001, 1;
bet, uniform_pdf,,, 0.75, 0.999;
tet, uniform_pdf,,, 0.0001, 1;
tau, uniform_pdf,,, 0.0001, 100;
delt, uniform_pdf,,, 0.0001, 0.05;
rho, uniform_pdf,,, 0.0001, 0.999;
stderr e_a, uniform_pdf,,, 0.00001, 0.1;
stderr y, uniform_pdf,,, 0.00001, 0.1;
stderr l, uniform_pdf,,, 0.00001, 0.1;
stderr i, uniform_pdf,,, 0.00001, 0.1;
end;
```

Starting values

```
estimated_params_init;
alp, 0.4;
bet, 0.99;
tet, 0.357 ;
tau, 50;
delt, 0.02;
rho, 0.95;
stderr e_a, .035;
stderr y, .0175;
stderr l, .00312;
stderr i, .00465;
end;

varobs y l i;
```

Sequential Importance-Sampling with Systematic Resampling

For $t = 1, \dots, T$ and $i = 1, \dots, N$,

At time t, knowing $\left\{s_{t-1}^{(i)}, w_{t-1}^{(i)}\right\}_{i=1:N}$ $(\approx p(s_{t-1}|y_{1:t-1}))$:

- Draw $\left\{\tilde{s}_t^{(i)}\right\}_{i=1:N}$ from $q\left(s_t|s_{t-1}^{(i)}, y_t\right)$.
- Evaluate the weights: $\hat{w}_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(y_t|\tilde{s}_t^{(i)})p(\tilde{s}_t^{(i)}|s_{t-1}^{(i)})}{q(\tilde{s}_t^{(i)}|s_{t-1}^{(i)}, y_t)}$.
- Resample $\left\{\tilde{s}_t^{(i)}, \tilde{w}_t^{(i)} = \frac{\hat{w}_t^{(i)}}{\sum_{i=1}^N \hat{w}_t^{(i)}}\right\}_{i=1:N} \mapsto \left\{s_t^{(i)}, w_t^{(i)} = \frac{1}{N}\right\}_{i=1:N}$ $(\approx p(s_t|y_{1:t}))$.

For simplicity sake, a usual choice for the proposal is:

$$q(s_t|s_{t-1}, y_t) = p(s_t|s_{t-1})$$

Draw particles in $p(s_t|s_{t-1})$: $\tilde{s}_t^{(i)} = f(s_{t-1}^{(i)}, \varepsilon_t^{(i)}; \theta)$

The weight expression simplifies:

$$\begin{aligned}\hat{w}_t^{(i)} &\propto \tilde{w}_{t-1}^{(i)} \frac{p(y_t|\tilde{s}_t^{(i)})p(\tilde{s}_t^{(i)}|s_{t-1}^{(i)})}{q(\tilde{s}_t^{(i)}|s_{t-1}^{(i)}, y_t)} \\ &= \tilde{w}_{t-1}^{(i)}\, p(y_t|\tilde{s}_t^{(i)})\end{aligned}$$

Easy to write since measurement errors are assumed additive and gaussian (remember $y_t = g(s_t; \theta) + e_t$):

$$p(y_t|\tilde{s}_t^{(i)}) = (2\pi)^{-\frac{\dim(y_t)}{2}} \, |P_e|^{-\frac{1}{2}} \, \exp\left\{-\frac{1}{2}\left[y_t - g(\tilde{s}_t^{(i)}; \theta)\right]' P_e^{-1} \left[y_t - g(\tilde{s}_t^{(i)}; \theta)\right]\right\}$$

sometimes difficult to calculate due to $\mathbb{V}(e_t) \equiv P_e$.

with Dynare

```
estimation(         order=   2,
        number_of_particles=   [5000] ,
               resampling =   [systematic], none, generic,
        resampling_method =   [kitagawa], residual, smooth,
               mh_replic =   0,
             mode_compute=   7 or 8) ;
```

- Can be viewed as a way to reintroduce y_t in the proposal.
- Add a pre-selection (resampling) step on the past particles based on the predictive likelihood when it is more informative than the transition distribution.
- Only the most promising particles will enter in the calculation of the current particle *proposal* distribution.
- Implement the previous filter on pre-selected past particles (modify slightly current particle weights).
- The resampling step on current particles becomes optional.

- Suppose we know $\left\{s_{t-1}^{(i)}, w_{t-1}^{(i)}\right\}_{i=1:N}$ $(\approx p(s_{t-1}|y_{1:t-1}))$.
- Approximate the state predictive density $\left\{\bar{s}_t^{(i)}\right\}_{i=1:N}$ such as $\bar{s}_t^{(i)} = f(s_{t-1}^{(i)}, 0; \theta)$
- Define $\hat{\tau}_{t-1}^{(i)} \equiv p(y_t|\bar{s}_t^{(i)})\, w_{t-1}^{(i)}$ and resample $\left\{s_{t-1}^{(i)}, \tilde{\tau}_{t-1}^{(i)} = \frac{\hat{\tau}_{t-1}^{(i)}}{\sum_{i=1}^{N} \hat{\tau}_{t-1}^{(i)}}\right\}_{i=1:N}$ $\mapsto \left\{s_{t-1}^{(k^l)}, \frac{1}{N}\right\}_{1:N}$.
- Draw $\left\{\tilde{s}_t^{(l)}\right\}_{l=1:N}$ from $q\left(s_t|s_{t-1}^{(k^l)}, y_t\right)$.
- Evaluate the weights: $$\hat{w}_t^{(l)} \propto \underbrace{\frac{w_{t-1}^{(k^l)}}{\tilde{\tau}_{t-1}^{(k^l)}}}_{\text{stage 1 weights}} \underbrace{\frac{1}{N} \frac{p(y_t|\tilde{s}_t^{(l)})p(\tilde{s}_t^{(l)}|s_{t-1}^{(k^l)})}{q(\tilde{s}_t^{(l)}|s_{t-1}^{(k^l)}, y_t)}}_{\text{stage 2 weights}}.$$
- We get $\left\{\tilde{s}_t^{(l)}, \tilde{w}_t^{(l)} = \frac{\hat{w}_t^{(l)}}{\sum_{l=1}^{N} \hat{w}_t^{(l)}}\right\}_{l=1:N} = \left\{s_t^{(l)}, w_t^{(l)}\right\}_{l=1:N}$ (no resampling).

- If $q(s_t|s_{t-1}, y_t) = p(s_t|s_{t-1})$, the weight becomes

$$\hat{w}_t^{(l)} \propto \frac{1}{N} \frac{p(y_t|\tilde{s}_t^{(l)})}{p(y_t|\tilde{s}_t^{(k^l)})} \sum_{i=1}^{N} \hat{\tau}_{t-1}^{(i)}$$

- Optional resampling in $\left\{ \tilde{s}_t^{(l)}, \tilde{w}_t^{(l)} = \frac{\hat{w}_t^{(l)}}{\sum_{l=1}^{N} \hat{w}_t^{(l)}} \right\}_{l=1:N}$ => the SIR filter with a pre-selection step.

with Dynare

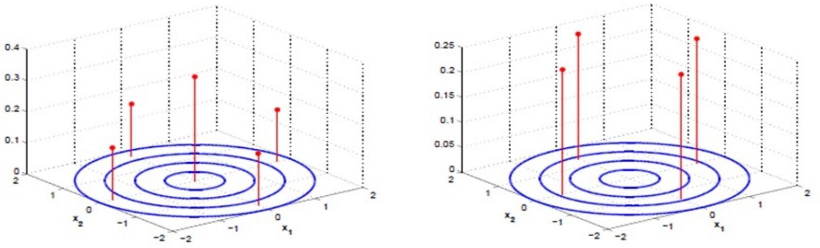
```
estimation(            order=  2,
             filter_algorithm=  apf,
          number_of_particles=  [5000] ,
                  resampling =  [systematic],none,
                   mh_replic =  0,
                mode_compute=  7 or 8) ;
```

	DGP	Kalman	SIR	APF
α	0.400	0.4046 (0.0032)	0.4003	0.4007
θ	0.357	0.3566 (0.0013)	0.3574	0.3574
β	0.990	0.9977 (0.0000)	0.9891	0.9890
δ	0.020	0.0215 (0.0007)	0.0199	0.0199
ρ	0.950	0.9743 (0.0017)	0.9508	0.9491
τ	50.000	54.3064 (11.9222)	49.9102	49.9216
σ_{ϵ}	0.035	0.0438 (0.0026)	0.0355	0.0365
σ_y	0.0175	0.0204 (0.0012)	0.0181	0.0179
σ_l	0.00312	0.0037 (0.0002)	0.0310	0.0310
σ_i	0.00465	0.0000 (0.0000)	0.0047	0.0046
Posterior Kernel		1,155.302	1,187.612	1,187.221
Proposal approximation		-	Particles	Particles
Distribution approximation		-	Particles	Particles
N		-	10,000	10,000
Option		order=1	order=2	order=2
Resampling		-	systematic	none

Gaussian approximations

- The distribution of particles is assumed to be Gaussian.
- Built with Monte-Carlo approximations or Gaussian sparse grids.
  - the Cubature points (Arasaratnam and Haykin, 2009)
  - the Unscented sigma-points (Julier and Ulhmann, 1997)
  - the Smolyak quadrature (Winschel and Kratzig, 2010)
- Pros: fast and easy to implement
  - drastically reduces the number of operations wrt Monte Carlo.
  - focus on the mean and variance of a unique Gaussian distribution
  - no resampling necessary (in theory)
- Cons: extreme approximation, highly inaccurate in case of multimodal distributions.

Illustration in a bivariate case



(a) Sigma point set for the UKF

(b) Third-degree spherical-radial cubature point set for the CKF

- Insert the nonlinear model in the Kalman filter: the moments of all gaussian densities are numerically approximated.
- Suppose the state distribution is Gaussian so that we only track $\bar{s}_{t|t}$ and $P_{s_{t|t}}$.
- Suppose we know $N(s_{t-1}; \bar{s}_{t-1|t-1}, P_{s_{t-1|t-1}})$ at date t.
- Approximate (s_{t-1}, ε_t) with sigma-points $\left\{\chi_j, W_j\right\}_{j=1:L}$ where $\chi_j = \left(\chi_j^{s_{t-1}}, \chi_j^{\varepsilon}\right)$.
- Deduce the densities of s_t and y_t using the (nonlinear) state-space representation:

$$\begin{aligned} \chi_j^{s_t} &= f(\chi_j^{s_{t-1}}, \chi_j^{\varepsilon}; \theta) \\ \chi_j^{y_t} &= g(\chi_j^{s_t}; \theta) \end{aligned}$$

Empirical moments computation

$$\bar{s}_{t|t-1} = \sum_{j=1}^{L} W_j \, \chi_j^{s_t}$$

$$P_{s_{t|t-1}} = \sum_{j=1}^{L} W_j \left(\chi_j^{s_t} - \bar{s}_{t|t-1}\right)\left(\chi_j^{s_t} - \bar{s}_{t|t-1}\right)'$$

$$\bar{y}_{t|t-1} = \sum_{j=1}^{L} W_j \, \chi_j^{y_t}$$

$$P_{y_{t|t-1}} = \sum_{j=1}^{L} W_j \left(\chi_j^{y_t} - \bar{y}_{t|t-1}\right)\left(\chi_j^{y_t} - \bar{y}_{t|t-1}\right)' + P_e$$

$$P_{s_{t|t-1}, y_{t|t-1}} = \sum_{j=1}^{L} W_j \left(\chi_j^{s_t} - \bar{s}_{t|t-1}\right)\left(\chi_j^{y_t} - \bar{y}_{t|t-1}\right)'$$

Kalman update step

- Calculate the Kalman gain $K_t = P_{s_{t|t-1},y_{t|t-1}} \; P^{-1}_{y_{t|t-1}}$ and deduce the *posterior* moments of state variables:

$$
\begin{aligned}
\bar{s}_{t|t} &= \bar{s}_{t|t-1} + K_t \left(y_t - \bar{y}_{t|t-1} \right) \\
P_{s_{t|t}} &= P_{s_{t|t-1}} - K_t P_{y_{t|t-1}} K_t{}'
\end{aligned}
$$

- Of course imperfect since the Kalman updating step derived for a gaussian/linear framework is implemented in a nonlinear framework.

- Approximative likelihood: $p(y_t|y_{1:t-1};\theta) \approx N(y_t;\bar{y}_{t|t-1},P_{y_{t|t-1}})$.

- Better idea: $p(y_t|y_{1:t-1};\theta) \approx \sum_{j=1}^{J} W_j \, N(y_t;\chi_j^{y_t},P_e)$.

with Dynare

```
estimation(                  order=  2,
                  filter_algorithm=  nlkf,
            proposal_approximation=  cubature, [unscented], montecarlo,
                       mh_replic =  0,
                     mode_compute=  4,8) ;
```

	DGP	Likelihood of the mean			Mean of the Likelihood		
α	0.400	0.3411	0.3390	0.3407	0.3997	0.3785	0.3830
θ	0.357	0.3344	0.3334	0.3343	0.3560	0.3477	0.3486
β	0.990	0.9956	0.9957	0.9955	0.9907	0.9934	0.9941
δ	0.020	0.0101	0.0095	0.0101	0.0194	0.0145	0.0147
ρ	0.950	0.9786	0.9793	0.9785	0.9585	0.9690	0.9692
τ	50.000	20.3817	21.6798	20.3379	49.1147	51.8689	48.2707
σ_{ϵ}	0.035	0.0410	0.0413	0.0411	0.0352	0.0320	0.0362
σ_y	0.0175	0.0178	0.0176	0.0177	0.0192	0.0616	0.0606
σ_l	0.00312	0.0031	0.0031	0.0031	0.0030	0.0034	0.0035
σ_i	0.00465	0.0000	0.0000	0.0001	0.0001	0.0367	0.0363
Posterior Kernel		1,200.734	1,200.804	1,200.249	1,205.795	980.1886	984.4534
Approximation		Particles	Unscented	Cubature	Particles	Unscented	Cubature
N		10,000	-	-	10,000	-	-
Option		-	$\begin{cases}\alpha = 1\\ \beta = 2\\ \kappa = 1\end{cases}$	$\begin{cases}\alpha = 1\\ \beta = 0\\ \kappa = 0\end{cases}$	-	$\begin{cases}\alpha = 1\\ \beta = 2\\ \kappa = 1\end{cases}$	$\begin{cases}\alpha = 1\\ \beta = 0\\ \kappa = 0\end{cases}$

α	0.400	0.3702	0.3816	0.2623	0.3733	0.3802	0.3988	0.3785	0.3930
θ	0.357	0.3463	0.3484	0.3026	0.3452	0.3478	0.3563	0.3477	0.3486
β	0.990	0.9932	0.9943	0.9986	0.9955	0.9943	1.0000	0.9934	0.9941
δ	0.020	0.0140	0.0145	0.0006	0.0135	0.0141	0.0178	0.0145	0.0147
ρ	0.950	0.9678	0.9696	0.9909	0.9724	0.9707	0.9625	0.9690	0.9692
τ	50.000	32.2965	46.3625	99.1066	30.7917	47.1475	15.6499	51.8689	48.2707
σ_{ϵ}	0.035	0.0379	0.0367	0.0364	0.0425	0.0367	0.0342	0.0320	0.0362
σ_y	0.0175	0.0528	0.0605	0.0616	0.0589	0.0604	0.0620	0.0616	0.0606
σ_l	0.00312	0.0027	0.0035	0.0038	0.0034	0.0035	0.0035	0.0034	0.0035
σ_i	0.00465	0.0249	0.0362	0.0372	0.0350	0.0362	0.0369	0.0367	0.0363
Posterior Kernel		944.028	985.045	965.945	985.821	985.4500	980.002	980.1886	984.4534
Approximation		unscented	unscented	unscented	unscented	unscented	unscented	unscented	Cubature
α		1	1	1	1	1	1	1	1
β		0	0	0	2	2	2	2	0
κ		-1	-0.1	0.1	-1	-0.1	0	1	0

Intuition

- Suppose a gaussian-based distribution for the states.
- Use the nonlinear Kalman filter *posterior* as *proposal*.
- Adds an extra layer of temporary particles (or sparse grid) with weights to the nonlinear Kalman filter step.
- Two approximations here: gaussian and gaussian-mixture (more general).

Gaussian (particle) filters

- Suppose we know $N(s_{t-1};\tilde{s}_{t-1},P_{s_{t-1}})$ at date t.
- Use the nonlinear Kalman filter step to build $N(s_t;\bar{s}_{t|t-1},P_{s_{t|t-1}})$ and $N(s_t;\bar{s}_{t|t},P_{s_{t|t}})$, resp. the *prior* and *posterior* distributions.
- Draw $\left\{\tilde{s}_t^{(i)}\right\}_{i=1:N}$ in $N(s_t;\bar{s}_{t|t},P_{s_{t|t}})$.
- Calculate their weights: $\hat{w}_t^{(i)} = \frac{1}{N}\frac{p(y_t|\tilde{s}_t^{(i)})\ N(\tilde{s}_t^{(i)};\bar{s}_{t|t-1},P_{s_{t|t-1}})}{N(\tilde{s}_t^{(i)};\bar{s}_{t|t},P_{s_{t|t}})}$
- Compute the moments from particles (eventually resampled):

$$\tilde{s}_t = \sum_{i=1}^{N}\tilde{w}_t^{(i)}\tilde{s}_t^{(i)} \quad \text{and} \quad P_{s_t} = \sum_{i=1}^{N}\tilde{w}_t^{(i)}(\tilde{s}_t^{(i)}-\tilde{s}_t)(\tilde{s}_t^{(i)}-\tilde{s}_t)'$$

with Dynare

estimation(	order=	2,
	filter_algorithm=	gf,
	proposal_approximation=	cubature, [unscented], montecarlo,
	distribution_approximation=	cubature, [unscented], montecarlo,
	number_of_particles=	[5000] ,
	resampling =	[systematic], none, generic,
	resampling_method =	[kitagawa], residual, smooth,
	mh_replic =	0,
	mode_compute=	depends if resampling or not) ;

Skip GMF

	DGP	GF					
α	0.400	0.3918	0.3995	0.3990	0.3615	0.3580	0.3503
θ	0.357	0.3529	0.3559	0.3557	0.3409	0.3394	0.3363
β	0.990	0.9913	0.9901	0.9904	0.9946	0.9952	0.9966
δ	0.020	0.0178	0.0195	0.0194	0.0119	0.0109	0.0095
ρ	0.950	0.9618	0.9581	0.9590	0.9743	0.9764	0.9802
τ	50.000	38.0264	37.8044	37.8781	30.9630	33.9064	28.8161
σ_{ϵ}	0.035	0.0402	0.0422	0.0413	0.0418	0.0396	0.0418
σ_y	0.0175	0.0179	0.0178	0.0179	0.0176	0.0170	0.0171
σ_l	0.00312	0.0031	0.0031	0.0031	0.0031	0.0031	0.0031
σ_i	0.00465	0.0065	0.0067	0.0065	0.0058	0.0083	0.0075
Posterior Kernel		1,185.813	1,180.992	1,181.274	1,189.404	1,174.907	1,176.602
Proposal approximation		Cubature	Unscented	Particles	Unscented	Unscented	Cubature
Distribution approximation		Particles	Particles	Particles	Particles	Unscented	Cubature
N		10,000	10,000	10,000	10,000	-	-
Resampling		systematic	systematic	systematic	none	-	-

Gaussian-mixture (particle) filters

- Kotesha and Djuric (2003), van der Merwe and Wan (2009)
- Limits the size of the Gaussian-mixture for states.
- A G-Gaussian mixture $\left\{\alpha^{(g)}, \mu^{(g)}, P^{(g)}\right\}_{g=1:G}$ is G distributions $N(s; \mu^{(g)}, P^{(g)})$ combined with weights $\alpha^{(g)}$(with $\sum_{g=1}^{G} \alpha^{(g)} = 1$).

Suppose we have

$$p_{GM(G)}(s_{t-1}|y_{1:t-1}) = \sum_{g=1}^{G} \alpha_{t-1}^{(g)} N(s_{t-1}; \mu_{t-1}^{(g)}, P_{t-1}^{(g)})$$

$$p_{GM(H)}(\varepsilon_t) = \sum_{h=1}^{H} \beta_t^{(h)} N(\varepsilon_t; \mu_{\varepsilon_t}^{(h)}, Q_t^{(h)})$$

We track $\{\alpha_t^{(g)}, \mu_t^{(g)}, P_t^{(g)}\}$.

- Define $g'(= 1, \dots, G' = GH)$.
- The nonlinear Kalman filter is implemented on each combinated element of the mixtures to build the *prior* and *posterior* Gaussian-mixture approximations:

$$p_{GM(G')}(s_t \,|y_{1:t-1}\,) = \sum_{g'=1}^{G'} \alpha_t^{(g')} N(s_t; \mu_t^{(g')}, P_t^{(g')})$$

$$p_{GM(G')}(s_t \,|y_{1:t}\,) = \sum_{g''=1}^{G'} \alpha_t^{(g'')} N(s_t; \mu_t^{(g'')}, P_t^{(g'')})$$

- The *posterior* Gaussian-mixture $p_{GM(G')}(s_t|y_{1:t})$ is used as proposal.

- Draw current particles $\left\{\tilde{s}_t^{(i)}\right\}_{i=1:N}$ in $p_{GM(G')}(s_t|y_{1:t})$
- The weight $\hat{w}_t^{(i)}$ is provided by:

$$\hat{w}_t^{(i)} = \frac{1}{N} \frac{p(y_t|\tilde{s}_t^{(i)})\, p_{GM(G')}(\tilde{s}_t^{(i)}|y_{1:t-1})}{p_{GM(G')}(\tilde{s}_t^{(i)}|y_{1:t})}$$

- Enventually resample and fit a new G-Gaussian mixture $\left\{\alpha^{(g)}, \mu^{(g)}, P^{(g)}\right\}_{g=1:G}$ on particles $\left\{\tilde{s}_t^{(i)}, \tilde{w}_t^{(i)} = \frac{\hat{w}_t^{(i)}}{\sum_{i=1}^N \hat{w}_t^{(i)}}\right\}_{i=1:N}$ to avoid explosion.

with Dynare

estimation(	order=	2,
	filter_algorithm=	gmf,
	proposal_approximation=	cubature,[unscented],
	distribution_approximation=	cubature,[unscented], montecarlo,
	number_of_particles=	[5000] ,
	resampling =	[systematic], none, generic,
	resampling_method =	[kitagawa], residual, smooth,
	mh_replic =	0,
	mode_compute=	depends if resampling or not) ;

Intuition

- No particular distributional assumption on states.
- Combine each particle with a Gaussian approximation on structural shocks.
- Still use the nonlinear Kalman filter *posterior* as *proposal*, but for each particle.

- Suppose we know $\left\{s_{t-1}^{(i)}, w_{t-1}^{(i)}\right\}_{i=1:N}$.
- Approximate ε_t with sigma-points $\left\{\chi_j^{\varepsilon}, W_j\right\}_{j=1:L}$:

$$\chi_j^{s_t^{(i)}} = f(s_{t-1}^{(i)}, \chi_j^{\varepsilon}; \theta)$$
$$\chi_j^{y_t^{(i)}} = g(\chi_j^{s_t^{(i)}}; \theta)$$

- Proposal: the Kalman *posterior*. $\forall i \in [1, N]$, draw $\tilde{s}_t^{(i)}$ in $N(s_t; \bar{s}_{t|t}^{(i)}, P_{s_{t|t}}^{(i)})$
- Approximated weights:

$$\hat{w}_t^{(i)} \approx w_{t-1}^{(i)} N(y_t; \bar{y}_{t|t-1}^{(i)}, P_{y_{t|t-1}}^{(i)}) \text{ a la Amisano and Tristani (2010)}$$
$$\hat{w}_t^{(i)} \approx w_{t-1}^{(i)} \frac{p(y_t|\tilde{s}_t^{(i)}) N(\tilde{s}_t^{(i)}; \bar{s}_{t|t-1}^{(i)}, P_{s_{t|t-1}}^{(i)})}{N(\tilde{s}_t^{(i)}; \bar{s}_{t|t}^{(i)}, P_{s_{t|t}}^{(i)})} \text{ a la Murray et al. (2013)}$$

- Resample $\left\{\tilde{s}_t^{(i)}, \tilde{w}_t^{(i)} = \frac{\hat{w}_t^{(i)}}{\sum_{i=1}^N \hat{w}_t^{(i)}}\right\}_{i=1:N} \mapsto \left\{s_t^{(i)}, w_t^{(i)} = \frac{1}{N}\right\}_{i=1:N}$.

Calling estimation

estimation(	order=	2,
	filter_algorithm=	cpf,
	proposal_approximation=	cubature,[unscented],montecarlo,
	number_of_particles=	[5000] ,
	cpf_weights =	[amisanotristani], murrayjonesparslow,
	mh_replic =	0,
	mode_compute=	7 or 8) ;

The sample likelihood

$$p(y_{1:T}\,|\theta\,) = p(y_1\,|s_0;\theta\,)p(s_0\,|\theta\,)\prod_{t=2}^{T} p(y_t\,|y_{1:t-1};\theta\,)$$

with

$$p(y_t\,|y_{1:t-1};\theta\,) \approx \sum_{i=1}^{N} \hat{w}_t^{(i)} = \sum_{i=1}^{N} w_{t-1}^{(i)} \frac{p(y_t|\tilde{s}_t^{(i)};\theta)p(\tilde{s}_t^{(i)}|s_{t-1}^{(i)};\theta)}{q(\tilde{s}_t^{(i)}|s_{t-1}^{(i)},y_t;\theta)}$$

If $q(s_t|s_{t-1},y_t;\theta) = p(s_t|s_{t-1};\theta)$ and no pre-selection step:

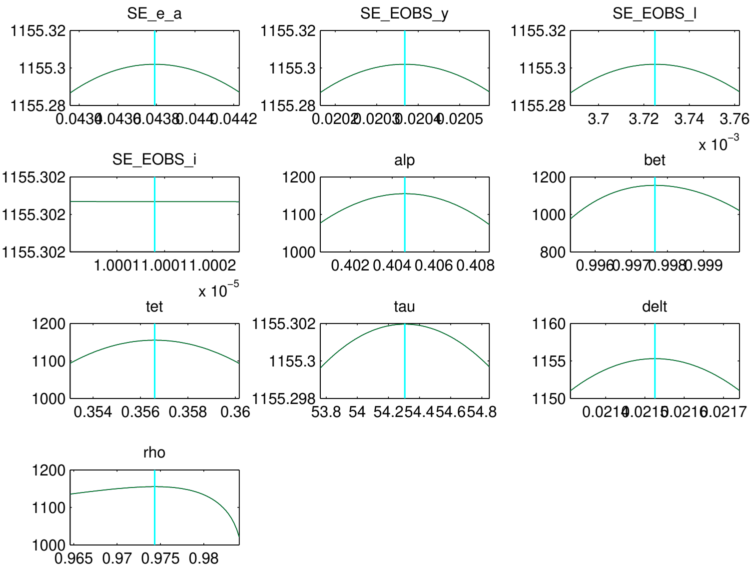
$$p(y_t|y_{1:t-1};\theta) \approx \sum_{i=1}^{N} \tilde{w}_{t-1}^{(i)} p(y_t|\tilde{s}_t^{(i)};\theta)$$

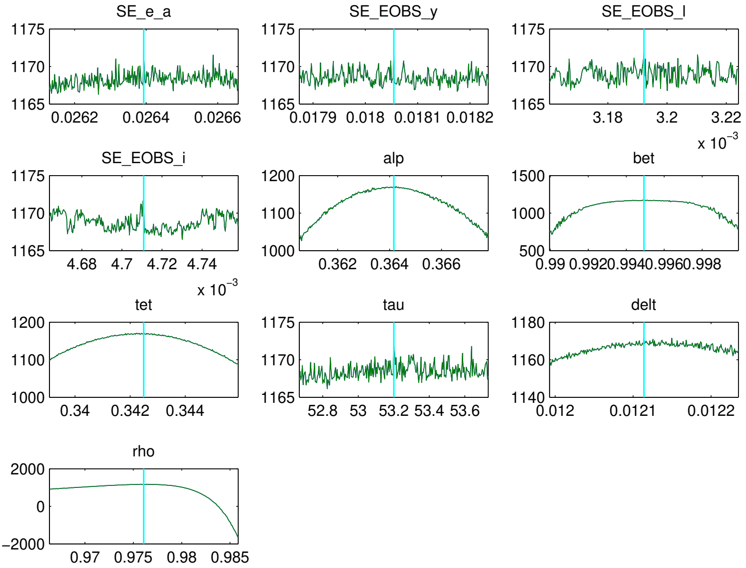
In case of systematic resampling, since $\tilde{w}_{t-1}^{(i)} = 1/N$:

$$p(y_t|y_{1:t-1};\theta) \approx \frac{1}{N}\sum_{i=1}^{N} p(y_t|\tilde{s}_t^{(i)};\theta)$$

The limits of the likelihood approach

- Resampling is necessary to avoid degeneracy...
- ... but complicates inference through maximum likelihood because it renders the likelihood criterion nonsmooth.
- Even when the seed for random draws is fixed across the simulations, the traditional likelihood estimator depends on both resampled particles and the unknown parameters. A small change in the parameters value will cause a small change in the importance weights that will potentially generate a different set of resampled particles. This produces a discontinuity in the likelihood criterion.


SE_e_a
SE_EOBS_y
SE_EOBS_l
x 10^-3
SE_EOBS_i
x 10^-5
alp
bet
tet
tau
delt
rho


SE_e_a
SE_EOBS_y
SE_EOBS_l
SE_EOBS_i
alp
bet
tet
tau
delt
rho

Solutions

- No resample and implement maximum likelihood if possible.
- Resample and implement maximum likelihood with no gradient-based maximization method (downhill simplex, S.A., CMAES, ...).
- Implement a MCMC approach to build the *posterior* distribution of parameters (Andrieu et al. (2010)).
- Implement an online approach, *i.e.* consider parameters as extra states.
- Use a smooth resampling method (but requires extensions to multivariate state variables).
- SMC methods: coming soon

Particle Marginal Metropolis-Hastings (P-MMH)

- Build $p(\theta | y_{1:T})$, the *posterior* distribution of parameters θ.
- As previously, it is a function of the sample likelihood $p(y_{1:T} | \theta)$ and the *priors* on parameters $p(\theta)$:

$$p(\theta | y_{1:T}) \propto p(y_{1:T} | \theta) p(\theta)$$

- In the case of a linear model, the sample likelihood $p(y_{1:T} | \theta)$ is provided by the Kalman filter.
- In the case of a nonlinear model, an unbiased approximation of the sample likelihood $p(y_{1:T} | \theta)$ is provided by particle filtering (Delmoral, 2004).

Particle Random-Walk Metropolis algorithm (Andrieu *et al.*, 2010)

- For $j = 1, \ldots, M$ (set large), define a candidate as:

$$\theta_j^\star = \theta_{j-1} + \epsilon_j$$

  with $\epsilon_j \sim N(0, \gamma_{RW} V(\theta_0))$ and γ_{RW} set in order to obtain an acceptance ratio around 24%.

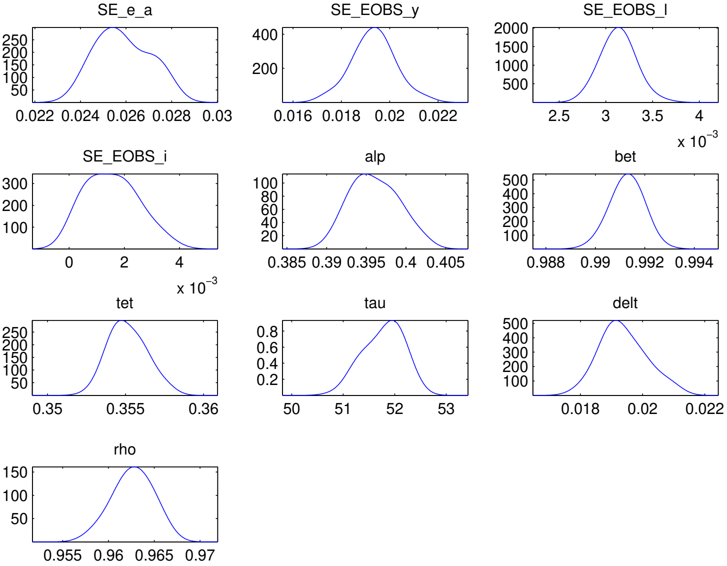
- The *posterior* distribution of deep parameters can be approximated thanks to the acceptation rule:

$$\theta_j = \begin{cases} \theta_j^\star & \text{if } \mathrm{U}_{[0,1]} \leqslant \min\left\{1, \frac{p(\theta_j^\star | y_{1:T})}{p(\theta_{j-1} | y_{1:T})}\right\} \\ \theta_{j-1} & \text{otherwise} \end{cases}$$

with Dynare

estimation(	order=	2,
	filter_algorithm=	[sis], nlkf, gf, gmf, apf, cpf,
	proposal_approximation=	cubature, [unscented],
	distribution_approximation=	cubature, [unscented], montecarlo,
	number_of_particles=	[5000] ,
	resampling =	[systematic], none, generic,
	resampling_method =	[kitagawa], residual, smooth,
	mh_replic =	[20000],
	mode_file=	xx)

	DGP	Kalman (order 1)	NLKF	SIR
α	0.400			0.3960 [0.3908; 0.4003]
θ	0.357			0.3551 [0.3533; 0.3571]
β	0.990			0.9913 [0.9902; 0.9923]
δ	0.020			0.0194 [0.0181; 0.0205]
ρ	0.950			0.9626 [0.9590; 0.9662]
τ	50.000			51.7763 [51.1611; 52.3328]
σ_ϵ	0.035			0.0259 [0.0242; 0.0277]
σ_y	0.0175			0.0193 [0.0175; 0.0206]
σ_l	0.00312			0.0031 [0.0028; 0.0034]
σ_i	0.00465			0.0015 [0.0000; 0.0028]
Distribution approximation		-	-	Particles
N		-	-	10,000
Options		mh_replic=5,000	mh_replic=5,000	mh_replic=5,000
Options		mh_nblck=10	mh_nblck=10	mh_nblck=10
Resampling		-	-	systematic


SE_e_a
SE_EOBS_y
SE_EOBS_l
x 10^-3
SE_EOBS_i
x 10^-3
alp
bet
tet
tau
delt
rho

The online approach (Liu and West, 2001)

- Unknown parameters are considered as extra state variables to allow *online* evaluation. Combines the auxiliary particle filter with an assumed importance density $p(\theta_t|\theta_{t-1})$ for parameters.

- Produces time-varying parameters and thus adds noise to the parameter estimates. To reduce the effect of the artificial variability, the authors adopt a kernel shrinkage technique.

- Pros:
  - Requires only one pass over the sample.
  - Also works if order=1.

- Cons:
  - Two resampling steps (on past and current particles)
  - No strong theoretical foundations but recent developments.

Suppose we have $\left\{s_{t-1}^{(i)}, \theta_{t-1}^{(i)}, w_{t-1}^{(i)}\right\}_{i=1:N}$. Calculate:

$$\bar{\theta}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} \theta_{t-1}^{(i)}$$

$$m_{t-1}^{(i)} = a\, \theta_{t-1}^{(i)} + (1-a)\, \bar{\theta}_{t-1}$$

$$V_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} (\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})(\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})'$$

$$\bar{s}_t^{(i)} = f(s_{t-1}^{(i)}, 0; m_{t-1}^{(i)})$$

1. The index k^l is obtained from sampling in $\hat{\tau}_{t-1}^{(i)} \propto w_{t-1}^{(i)} p(y_t|\bar{s}_t^{(i)}, m_{t-1}^{(i)})$.
2. $\left\{\tilde{\theta}_t^{(l)}\right\}_{l=1:N}$ are drawn from $N(m_{t-1}^{(k^l)}, b^2 V_{t-1})$.
3. $\left\{\tilde{s}_t^{(l)}\right\}_{l=1:N}$ are drawn from $p(s_t|s_{t-1}^{(k^l)}, \tilde{\theta}_t^{(l)})$.
4. The weights are calculated as: $\hat{w}_t^{(l)} \propto p(y_t|\tilde{s}_t^{(l)}, \tilde{\theta}_t^{(l)}) \frac{w_{t-1}^{(k^l)}}{\tilde{\tau}_{t-1}^{(k^l)}}$ and normalized.

- The shrinkage technique (based on parameter a) is used to produce slowly time-varying parameters and also to limit the variability.

- δ is the key parameter that conditions the shrinkage and the smoothness parameters a and b:

$$a = \frac{3\delta - 1}{2\delta}$$
$$b^2 = 1 - a^2$$

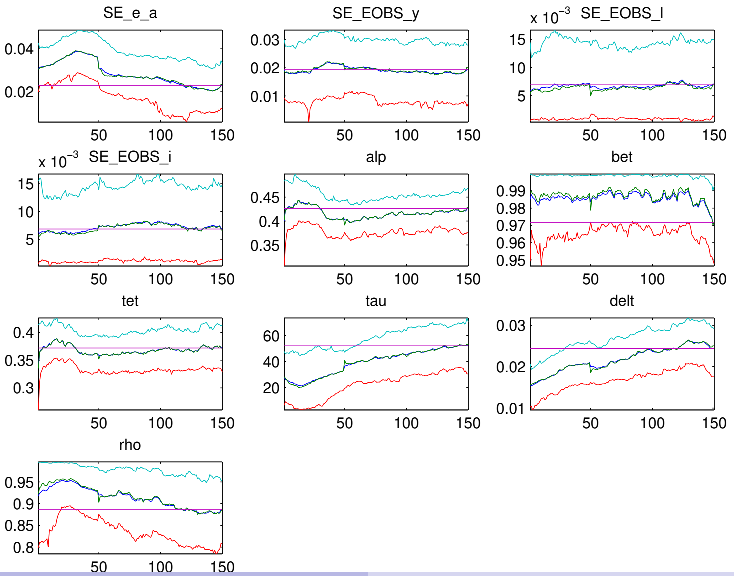
- δ is generally chosen in the range $[0.9; 0.99]$.

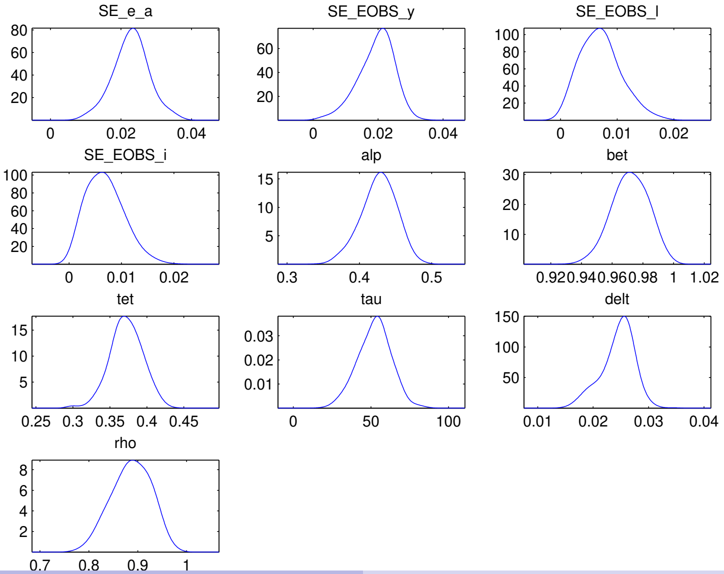
with Dynare

```
estimation(             order=   1 or 2,
           number_of_particles=  [5000] ,
                   resampling =  [systematic], none,
                  mode_compute=  11) ;
```

and fix δ with options_.particle.**liu_west_delta** = 0.9

	DGP	Online Particle filter	
α	0.400	0.42843	0.43621
		[0.39500.4621]	[0.4071; 0.4852]
θ	0.357	0.36648	0.36603
		[0.3004; 0.4193]	[0.3281; 0.4474]
β	0.990	0.96926	0.97778
		[0.9278; 0.9963]	[0.9178; 0.9982]
δ	0.020	0.02473	0.02481
		[0.0184; 0.0323]	[0.0176; 0.0345]
ρ	0.950	0.67283	0.75078
		[0.5210; 0.8282]	[0.5762; 0.8620]
τ	50.000	40.37328	66.82140
		[20.7203; 54.2846]	[30.5316; 94.5316]
σ_{ϵ}	0.035	0.01825	0.02783
		[0.0023; 0.0340]	[0.0095; 0.0403]
σ_y	0.0175	0.07411	0.05547
		[0.0510; 0.0949]	[0.0364; 0.0735]
σ_l	0.00312	0.03323	0.02182
		[0.0124; 0.0482]	[0.0034; 0.0436]
σ_i	0.00465	0.07194	0.07294
		[0.0475; 0.0943]	[0.0560; 0.0918]
Posterior Kernel		-	-
Proposal approximation		-	
Distribution approximation		Particles	Particles
N		10,000	10,000
Mixture		-	-
Option		$\delta = 0.9$	$\delta = 0.9$
Resampling		none	systematic


SE_e_a
SE_EOBS_y
x 10^-3 SE_EOBS_l
x 10^-3 SE_EOBS_i
alp
bet
tet
tau
delt
rho


SE_e_a
SE_EOBS_y
SE_EOBS_l
SE_EOBS_i
alp
bet
tet
tau
delt
rho

in test

- propose a sampler to build $p(\theta|y_{1:T})$, the *posterior* distribution of parameters θ.

- uses the following proposal for a sequence of N_ϕ parameters $\phi_n \in [0,1]$ (with $n \in [0, N_\phi]$):

$$\pi_n(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)^{\phi_n} p(\theta)$$

- $p(y_{1:T}|\theta)^{\phi_n}$ is the tempered likelihood.
- $\phi_0 = 0 => \pi_0(\theta|y_{1:T}) = p(\theta)$: draw in the priors for sampler initialization.
- $\phi_{N_\phi} = 1 => \pi_{N_\phi}(\theta|y_{1:T}) = p(y_{1:T}|\theta)\, p(\theta)$: corresponds to the final step of the sampler.
- $p(y_{1:T}|\theta)$ is provided by the chosen filter.

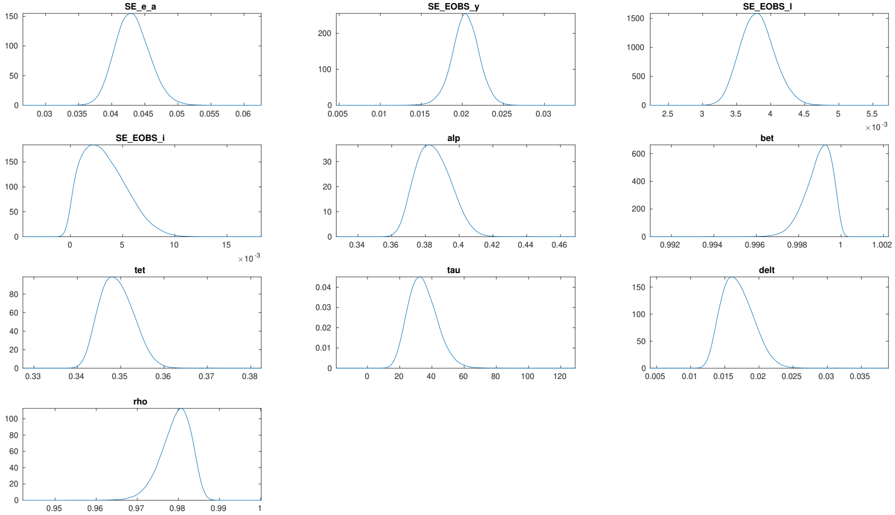
Initialization: $n = 0$: draw $\left\{\theta_0^{(i)}, w_0^{(i)} = 1/N\right\}_{i=1:N}$ in the *prior* distribution.
Choose N_ϕ and λ and define: $\phi_n = (\frac{n}{N_\phi})^\lambda$
At step $n \in [1, N_\phi]$, given $\left\{\theta_{n-1}^{(i)}, w_{n-1}^{(i)}\right\}_{i=1:N}$:

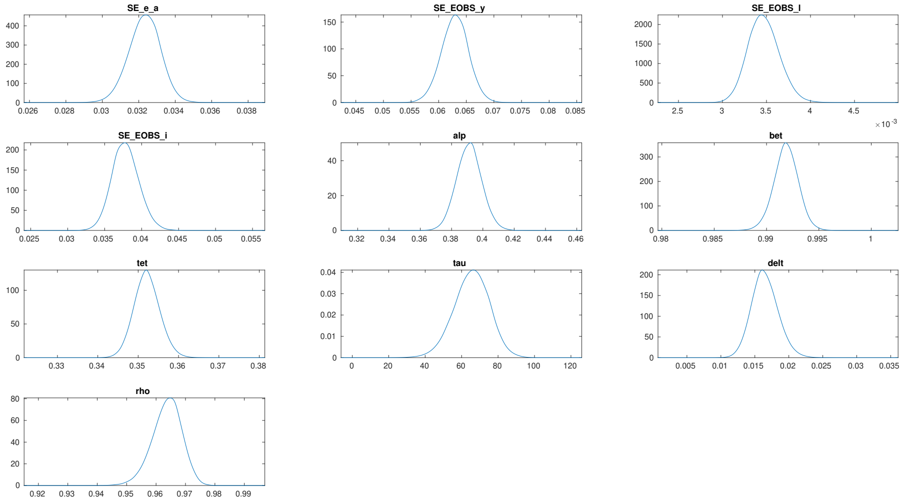
- Correction: $\hat{w}_n^{(i)} \propto w_{n-1}^{(i)} p(y_{1:T} \left| \theta_n^{(i)} \right.)^{\phi_n - \phi_{n-1}}$
- Selection (optional resampling):

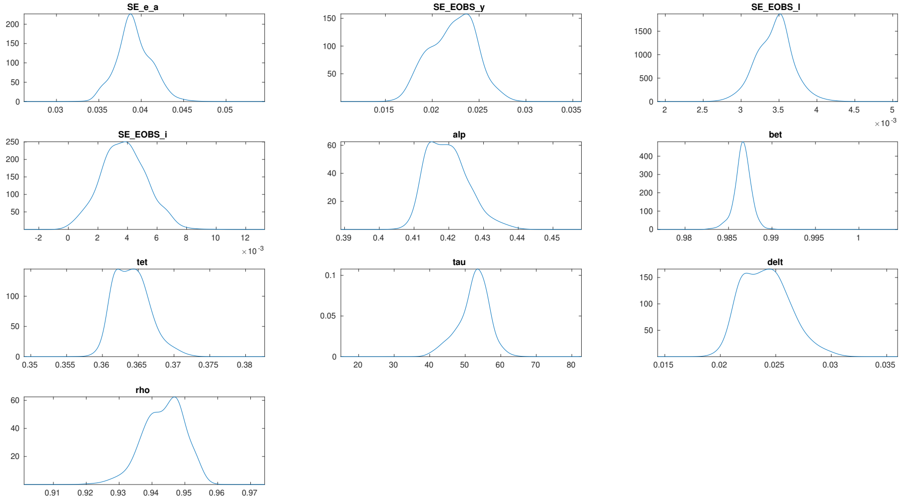
$$\left\{\theta_{n-1}^{(i)}, \tilde{w}_n^{(i)} = \frac{\hat{w}_n^{(i)}}{\sum_{i=1}^N \hat{w}_n^{(i)}}\right\}_{i=1:N} -> \left\{\tilde{\theta}_n^{(i)}, w_n^{(i)} = 1/N\right\}_{i=1:N}$$

- Mutation: perform one MH step: $\left\{\tilde{\theta}_n^{(i)}\right\}_{i=1:N} -> \left\{\theta_n^{(i)}\right\}_{i=1:N}$ (the acceptance ratio is calculated using $\pi_{\phi_n}(\theta \left| y_{1:T} \right.)$).

	DGP	Kalman (order 1)	NLKF	SIR
α	0.400	0.3846	0.39108	0.43621
		[0.3674; 0.4052]	[0.3757; 0.4070]	[0.4071; 0.4852]
θ	0.357	0.3491	0.35211	0.36603
		[0.3427; 0.3569]	[0.3461; 0.3585]	[0.3281; 0.4474]
β	0.990	0.9989	0.99184	0.97778
		[0.9975; 0.9998]	[0.9895; 0.9941]	[0.9178; 0.9982]
δ	0.020	0.0169	0.01646	0.02481
		[0.0132; 0.0218]	[0.0129; 0.0205]	[0.0176; 0.0345]
ρ	0.950	0.9792	0.96355	0.75078
		[0.9711; 0.9950]	[0.9530; 0.9725]	[0.5762; 0.8620]
τ	50.000	34.3492	65.29801	66.82140
		[19.5312; 53.2281]	[45.6439; 83.163]	[30.5316; 94.5316]
σ_ϵ	0.035	0.04311	0.03231	0.02783
		[0.0385; 0.0484]	[0.0306; 0.0339]	[0.0095; 0.0403]
σ_y	0.0175	0.02037	0.06290	0.05547
		[0.0171; 0.0235]	[0.0580; 0.0677]	[0.0364; 0.0735]
σ_l	0.00312	0.00380	0.00347	0.02182
		[0.0034; 0.0043]	[0.0031; 0.0038]	[0.0034; 0.0436]
σ_i	0.00465	0.00327	0.03779	0.07294
		[0.0002; 0.0078]	[0.0344; 0.0415]	[0.0560; 0.0918]
States distribution approximation		-	Particles	Particles
N		-	10,000	10,000
Resampling		-	-	systematic
Param. distribution approximation		100,000	100,000	10,000


SE_e_a
SE_EOBS_y
SE_EOBS_l
SE_EOBS_i
alp
bet
tet
tau
delt
rho


SE_e_a
SE_EOBS_y
SE_EOBS_l
×10^-3
SE_EOBS_i
alp
bet
tet
tau
delt
rho


SE_e_a
0 50 100 150 200
0.03 0.035 0.04 0.045 0.05
SE_EOBS_y
50 100 150
0.015 0.02 0.025 0.03 0.035
SE_EOBS_l
0 500 1000 1500
2 2.5 3 3.5 4 4.5 5
×10^-3
SE_EOBS_i
50 100 150 200 250
-2 0 2 4 6 8 10 12
×10^-3
alp
20 40 60
0.39 0.4 0.41 0.42 0.43 0.44 0.45
bet
0 100 200 300 400
0.98 0.985 0.99 0.995 1
tet
0 50 100
0.35 0.355 0.36 0.365 0.37 0.375 0.38
tau
0 0.05 0.1
20 30 40 50 60 70 80
delt
50 100 150
0.015 0.02 0.025 0.03 0.035
rho
20 40 60
0.91 0.92 0.93 0.94 0.95 0.96 0.97

In a nutshell

Filter	Proposal choice	Proposal approximation	State approximation	Resampling
Bootstrap	Transition distribution	Particles	Particles	Yes
Auxiliary	Pre-selected particles for states transition distribution	Particles	Particles	Yes,No
NL Kalman	-	-	Particles or sparse grids	-
Gaussian	Gaussian on states and errors	Sparse grid / Kalman *posterior*	Particles or sparse grids	Yes,No
Gaussian-Mixture	Gaussian-Mixtures on states and/or errors	Sparse grid / Kalman *posterior*	Particles or sparse grids	Yes,No
Conditional	Particles for states, sparse grids for errors	Kalman *posterior*	Particles	Yes

References

Amisano G. and Tristani O. (2010), Euro Area Inflation Persistence in an Estimated Nonlinear DSGE Model, Journal of Economic Dynamics and Control, 34, 1837-1858.

An S. and Schorfheide F. (2007), Bayesian Analysis of DSGE Models, Econometric Reviews 26(2-4), 113-172.

Andrieu C., Doucet A. and Holenstein R. (2010), Particle Markov Chain Monte Carlo Methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(3), 269-342.

Arasaratnam I. and Haykin S. (2009a), Cubature Kalman Filters, IEEE Transactions on Automatic Control, 54(6), 1254-1269.

Arasaratnam I. and Haykin S. (2009b), Hybrid Cubature Filter: Theory and Tracking Application, McMaster University, Technical Report CSL2009:4, 129.

Arulampalam S., Maskell S., Gordon N. and Clapp T. (2002), A Tutorial on Particle Filters for on-line Non-linear / Non-gaussian Bayesian Tracking, IEEE Trans, Signal Process, 50, 241-254.

Cappé O., Godsill S.J. and Moulines E. (2007), An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo, Proceedings of the IEEE, 95(5), 899-924.

Creal D. (2009), A Survey of Sequential Monte Carlo Methods for Economics and Finance, Econometric Reviews, 31(3), 245-296.

Del Moral P. (2004), Feynman Kac Formulae: Genealogical and Interacting Particle Systems with Applications, New-York Springer.

Doucet A., Freitas J.G. and Gordon J. (2001), Sequential Monte Carlo Methods in Practice, Springer Verlag, New York.

Doucet A., Godsill S. and Andrieu C. (2000), On Sequential Monte Carlo Sampling Methods for Bayesian Filtering, Statistics and Computing, 10, 197-208.

Doucet, A. and Johansen A. (2009), A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later, The Oxford Handbook of Nonlinear filtering, Oxford University Press.

Fernandez-Villaverde J. and Rubio-Ramirez J.F. (2005), Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood, Journal of Applied Econometrics 20, 891-910.

Fernandez-Villaverde, J. and Rubio-Ramirez J.F. (2007), Estimating Macroeconomic Models: a Likelihood Approach, The Review of Economic Studies 74(4), 1059-1087.

References

Fernandez-Villaverde, J., Rubio-Ramirez J.F. and Schorfheide F. (2015), Solution and Estimation Methods for DSGE Models, Handbook of Macroeconomics, Vol.2.

Gordon N., Salmond D. and Smith A.F.M. (1993), Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation, IEE Proceedings-F, 140, 107-113.

Herbst E. and Schorfheide F. (2015), Bayesian Estimation of DSGE Models, online version.

Julier S.J. and Uhlmann J.K. (1997), A New Extension of the Kalman Filter to Nonlinear Systems, Proceedings of AeroSense, the 11th Int. Symp. on Aerospace/Defense Sensing, Simulation and Controls.

Kantas N., Doucet A., Singh S.S., Maciejowski J., and Chopin N. (2015), On Particle Methods for Parameter Estimation in State-Space Models, Statistical Science, 30(3), 328-351.

Kitagawa G. (1987), NonGaussian State Space Modeling of Nonstationary Time Series, Journal of the American Statistical Association 82(400), 1023-1063.

Kotecha J.H. and Djuric P.M. (2003a), Gaussian Particle Filtering, IEEE transactions on signal processing, 51(10),2592-2601.

Kotecha J.H. and Djuric P.M. (2003b), Gaussian Sum Particle Filtering, IEEE transactions on signal processing, 51(10), 2602-2612.

Liu J. and West M. (2001), Combined Parameter and State Estimation in Simulation-Based Filtering, in Sequential Monte Carlo Methods in Practice, eds Doucet, Freitas and Gordon, Springer Verlag, New York.

Malik S. and Pitt M. (2011), Particle Filters for Continuous Likelihood Evaluation and Maximisation, Journal of Econometrics, 165(2), 190-209.

Murray L.M., Jones E.M. and Parslow J. (2013), On Disturbance State-Space Models and the Particle Marginal Metropolis-Hastings Sampler, working paper, arXiv:1202.6159v3.

Pitt M. and Shephard N. (1999), Filtering via Simulation: Auxiliary Particle Filters. Journal of the American Statistical Association, 94(446), 590-599.

van der Merwe R. and Wan E. (2003), Gaussian Mixture Sigma-Point Particle Filters for Sequential Probabilistic Inference in Dynamic State-Space Models, mimeo.

Winschel V. and Kraltzig M. (2010), Solving, Estimating, and Selecting Nonlinear Dynamic Models without the Curse of Dimensionality, Econometrica, 78(2), 803-821.

The third-degree spherical-radial cubature

Proposed by Arasaratnam and Haykin (2009)

$$\{\chi_j\}_{j=1}^{2n} = \{-\sqrt{n}\, I_n \quad \sqrt{n}\, I_n\}$$

$$\{W_j\}_{j=1}^{L} = \left\{\frac{1}{2n}\right\}_{j=1}^{2n}$$

The unscented sigma-points

Proposed by Julier and Ulhmann (2003)

$$\{\chi_j\}_{j=0}^{2n} = \left\{0_{(n,1)} \quad -\sqrt{n+\lambda}\, I_n \quad \sqrt{n+\lambda}\, I_n\right\}$$

with $\lambda = \alpha^2(n+\kappa) - n$ and the weights employed to calculate respectively first- and second-order moments:

$$W_0^m = \frac{\lambda}{n+\lambda} \text{ and } \left\{W_j^m\right\}_{j=1}^{2n} = \frac{1}{2(n+\lambda)}$$

$$W_0^c = \frac{\lambda}{n+\lambda} + 1 - \alpha^2 + \beta \text{ and } \left\{W_j^c\right\}_{j=1}^{2n} = \frac{1}{2(n+\lambda)}$$

α and κ determine the spread of the sigma-points and β characterizes the (non-gaussian) distribution.

By default, we set $\alpha = \kappa = 1$ and $\beta = 2$.