Efficient Likelihood Evaluation of State-Space Representations

David N. DeJong*, Hariharan Dharmarajan**, Roman Liesenfeld***, Guilherme V. Moura****, Jean-Francois Richard*
*Pitt  **Bates and White  ***Kiel  ****Santa Catarina

August 2011
Likelihood evaluation and filtering using state-space representations featuring departures from:

- Linearity
- Normality
Motivation

- In the linear/normal case, exact likelihood evaluations are available analytically via the Kalman filter.
Motivation

- In the linear/normal case, exact likelihood evaluations are available analytically via the Kalman filter.
- However, linear/normal characterizations of economic phenomenon are often inadequate or inappropriate, thus necessitating the implementation of numerical approximation techniques.
Motivation

- In the linear/normal case, exact likelihood evaluations are available analytically via the Kalman filter.
- However, linear/normal characterizations of economic phenomenon are often inadequate or inappropriate, thus necessitating the implementation of numerical approximation techniques.
- Example: In working with DSGE models, linear approximations are problematic for conducting likelihood analysis (Fernandez-Villaverde and Rubio-Ramirez, 2005 *JAE*; 2007 *REStud*)
Numerical approximation techniques do exist for achieving likelihood evaluation and filtering for state-space models.
Numerical approximation techniques do exist for achieving likelihood evaluation and filtering for state-space models.

F-V and R-R used the Bootstrap Particle Filter of Gordon, Salmond, Smith (1993 IEEE Proceedings) in their evaluation of DSGE models.
Motivation, cont.

- Numerical approximation techniques do exist for achieving likelihood evaluation and filtering for state-space models.
- However, the BPF and popular extensions turn out to be numerically inefficient in empirically relevant settings.
State Space Representations

State-transition equation:

\[ s_t = \gamma(s_{t-1}, v_t) \]

Associated density:

\[ f(s_t | s_{t-1}) \]

Measurement equation:

\[ y_t = \delta(s_t, u_t) \]

Associated density:

\[ f(y_t | s_t) \]

Initialization:

\[ f(s_0) \]
State Space Representations, cont.

**Objective:** evaluate the likelihood function

\[
f (Y_T) = \prod_{t=1}^{T} f (y_t | Y_{t-1}),
\]

where \( f (y_1 | Y_0) \equiv f(y_1) \).
State Space Representations, cont.

Time-\( t \) likelihoods are evaluated via marginalization of measurement densities:

\[
f (y_t \mid Y_{t-1}) = \int f (y_t \mid s_t) f (s_t \mid Y_{t-1}) \, ds_t.
\]

Marginalization requires the evaluation of \( f(s_t \mid Y_{t-1}) \):

\[
f (s_t \mid Y_{t-1}) = \int f (s_t \mid s_{t-1}) f (s_{t-1} \mid Y_{t-1}) \, ds_{t-1},
\]

where

\[
f (s_t \mid Y_t) = \frac{f (y_t, s_t \mid Y_{t-1})}{f (y_t \mid Y_{t-1})} = \frac{f (y_t \mid s_t) f (s_t \mid Y_{t-1})}{f (y_t \mid Y_{t-1})}.
\]
Substituting
\[ \int f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}) \, ds_{t-1} \]
for
\[ f(s_t|Y_{t-1}), \]
the likelihood and filtering densities are
\[
\begin{align*}
  f(y_t|Y_{t-1}) &= \int \int f(y_t|s_t) f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}) \, ds_{t-1} \, ds_t, \\
  f(s_t|Y_t) &= \frac{f(y_t|s_t)}{f(y_t|Y_{t-1})} \cdot \int f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}) \, ds_{t-1}. 
\end{align*}
\]
Goal: approximate using efficient numerical techniques.
Importance Sampling

Our goal is to calculate integrals of the form

\[ G(Y) = \int_{\Theta} \phi(\theta; Y) \, d\theta. \]

Special case (e.g., posterior moment):

\[ G(Y) = \int_{\Theta} \phi(\theta; Y) \, p(\theta|Y) \, d\theta. \]

If it is possible to obtain pseudo-random drawings \( \theta_i \) from \( p(\theta|Y) \), by the law of large numbers

\[ \frac{1}{N} \sum_{i=1}^{N} \phi(\theta_i; Y) \]

converges in probability to \( G(Y) \). We refer to \( \frac{1}{N} \sum_{i=1}^{N} \phi(\theta_i; Y) \) as the **Monte Carlo** estimate of \( G(Y) \).
The MC estimate of the standard deviation of $G(Y)$ is given by

$$
\bar{\sigma}_N (G(Y)) = \left[ \left( \frac{1}{N} \sum_i^N \phi(\theta_i; Y)^2 \right) - \overline{G(Y)}_N^2 \right]^{1/2}.
$$

The numerical standard error associated with $G(Y)_N$ is given by

$$
s.e. \left( \overline{G(Y)}_N \right) = \frac{\bar{\sigma}_N (G(Y))}{\sqrt{N}}.
$$

Thus for $N = 10,000$, $s.e. \left( \overline{G(Y)}_N \right)$ is 1% of the size of $\bar{\sigma}_N (G(Y))$. 
IS, cont.

If $p(\theta|Y)$ is unavailable as a sampler, one remedy is to augment the targeted integrand with an **importance sampling** distribution $g(\theta|a)$:

$$
G(Y) = \int_{\Theta} \frac{\phi(\theta; Y)}{g(\theta|a)} g(\theta|a) d\theta
= \int_{\Theta} \frac{\phi(\theta; Y) p(\theta|Y)}{g(\theta|a)} g(\theta|a) d\theta.
$$

**Key requirements:**

- Support of $g(\theta|a)$ must span that of $\phi(\theta|Y)$
If \( p(\theta|Y) \) is unavailable as a sampler, one remedy is to augment the targeted integrand with an **importance sampling** distribution \( g(\theta|a) \):

\[
G(Y) = \int_{\Theta} \frac{\phi(\theta; Y)}{g(\theta|a)} g(\theta|a) \, d\theta = \int_{\Theta} \frac{\phi(\theta; Y) p(\theta|Y)}{g(\theta|a)} g(\theta|a) \, d\theta.
\]

**Key requirements:**

- Support of \( g(\theta|a) \) must span that of \( \phi(\theta|Y) \)
- \( E[G(Y)] \) must exist and be finite.
If \( p(\theta | Y) \) is unavailable as a sampler, one remedy is to augment the targeted integrand with an importance sampling distribution \( g(\theta | a) \):

\[
G(Y) = \int_{\Theta} \frac{\phi(\theta; Y)}{g(\theta | a)} g(\theta | a) \, d\theta
= \int_{\Theta} \frac{\phi(\theta; Y) p(\theta | Y)}{g(\theta | a)} g(\theta | a) \, d\theta.
\]

**Key requirements:**

- Support of \( g(\theta | a) \) must span that of \( \phi(\theta | Y) \)
- \( E[G(Y)] \) must exist and be finite.
- \( g(\theta | a) \) must be implementable as a sampler.
MC estimate of $G(Y)$:

$$
\bar{G}(Y)_N = \frac{1}{N} \sum_{i=1}^{N} \omega_i ; \quad \omega_i = \frac{\varphi(\theta_i|Y)}{g(\theta_i|a)}
$$
MC estimate of $G(Y)$:

$$
\overline{G(Y)}_N = \frac{1}{N} \sum_{i=1}^{N} \omega_i ; \quad \omega_i = \frac{\varphi(\theta_i|Y)}{g(\theta_i|a)}
$$

MC estimate of the standard deviation of $\omega$ w.r.t. $g(\theta|a)$:

$$
\overline{\sigma}_N (\omega(\theta, Y)) = \left[ \left( \frac{1}{N} \sum \omega_i^2 \right) - \overline{G(Y)}_N^2 \right]^{1/2}
$$
IS, cont.

- MC estimate of $G(Y)$:
  \[ \overline{G(Y)}_N = \frac{1}{N} \sum_{i=1}^{N} \omega_i; \quad \omega_i = \frac{\varphi(\theta_i|Y)}{g(\theta_i|a)} \]

- MC estimate of the standard deviation of $\omega$ w.r.t. $g(\theta|a)$:
  \[ \bar{\sigma}_N(\omega(\theta, Y)) = \left[ \left( \frac{1}{N} \sum \omega_i^2 \right) - \overline{G(Y)}_N^2 \right]^{1/2} \]

- Numerical standard error associated with $\overline{G(\theta, Y)}_N$:
  \[ s.e. \left( \overline{G(Y)}_N \right)_I = \frac{\bar{\sigma}_N(\omega(\theta, Y))}{\sqrt{N}}. \]
MC estimate of $G(Y)$:

$$\overline{G(Y)}_N = \frac{1}{N} \sum_{i=1}^{N} \omega_i ; \quad \omega_i = \frac{\varphi(\theta_i|Y)}{g(\theta_i|a)}$$

MC estimate of the standard deviation of $\omega$ w.r.t. $g(\theta|a)$:

$$\overline{\sigma}_N(\omega(\theta, Y)) = \left[ \left( \frac{1}{N} \sum \omega_i^2 \right) - \overline{G(Y)}_N^2 \right]^{1/2}.$$

Numerical standard error associated with $G(\theta, Y)_N$:

$$s.e. \left( \overline{G(Y)}_N \right)_i = \frac{\overline{\sigma}_N(\omega(\theta, Y))}{\sqrt{N}}.$$

**Note:** Variability in $\omega$ translates into increased n.s.e. (numerical inefficiency)
For the special case in which the integrand factorizes as

\[ \phi(\theta|Y) = \phi(\theta; Y) p(\theta|Y), \]

\[ \overline{G(Y)}_N = \frac{1}{N} \sum_{i=1}^{N} \phi(\theta_i; Y) w_i, \]

\[ w_i = \frac{p(\theta_i|Y)}{g(\theta_i|a)} \]

\[ \sigma_N(G(Y)) = \left[ \left( \frac{1}{N} \sum_{i}^{N} \phi(\theta_i; Y)^2 w_i \right) - \overline{G(Y)}_N^2 \right]^{1/2}. \]
Return to the problem of approximating

\[
f(y_t | Y_{t-1}) = \int \int f(y_t | s_t) f(s_t | s_{t-1}) f(s_{t-1} | Y_{t-1}) \, ds_{t-1} \, ds_t,
\]

\[
f(s_t | Y_t) = \frac{f(y_t | s_t)}{f(y_t | Y_{t-1})} \cdot \int f(s_t | s_{t-1}) f(s_{t-1} | Y_{t-1}) \, ds_{t-1},
\]

where

\[
f(s_0) \equiv f(s_0 | Y_0).
\]
SIS/SIR algorithms employ importance samplers to approximate the integrands

\[
f (y_t | s_t) f (s_t | s_{t-1}) f (s_{t-1} | Y_{t-1}),
\]

\[
f (s_t | s_{t-1}) f (s_{t-1} | Y_{t-1}), \quad t = 1, \ldots, T.
\]

In representing

\[
f (s_{t-1} | Y_{t-1}),
\]

particle-based filters use discrete approximations of the form

\[
\hat{f} (s_{t-1} | Y_{t-1}) = \sum_{i=1}^{N} \omega^i_{t-1} \cdot \delta (s_{t-1} - s^i_{t-1}),
\]

where \( \{s^i_{t}\}_{i=1}^{N} \) is a swarm of particles and \( \{\omega^i_{t}\}_{i=1}^{N} \) their associated weights.
Example Algorithm: Bootstrap Particle Filter (Gordon, Salmond, Smith, 1993)

Likelihood evaluation:

- Inherit \( \hat{f}(s_{t-1} | Y_{t-1}) = \{s^i_{t-1}\}_{i=1}^N \) from period-\((t-1)\) step.
Example Algorithm: Bootstrap Particle Filter (Gordon, Salmond, Smith, 1993)

Likelihood evaluation:

- Inherit \( \hat{f}(s_{t-1} \mid Y_{t-1}) = \{s_i^{i-1}\}_{i=1}^N \) from period-\((t - 1)\) step.
- For each \( s_{t-1}^i \), draw \( s_t^i \) from \( f(s_t \mid s_{t-1}^i) \). The resulting swarm represents MC draws from the joint density

\[
f(s_t \mid s_{t-1}) f(s_{t-1} \mid Y_{t-1}).
\]
Example Algorithm: Bootstrap Particle Filter (Gordon, Salmond, Smith, 1993)

Likelihood evaluation:

- Inherit $\hat{f}(s_{t-1}|Y_{t-1}) = \{s^i_{t-1}\}_{i=1}^N$ from period-$(t - 1)$ step.
- For each $s^i_{t-1}$, draw $s^i_t$ from $f(s_t|s^i_{t-1})$. The resulting swarm represents MC draws from the joint density

$$f(s_t|s^i_{t-1}) f(s^i_{t-1}|Y_{t-1}).$$

- Approximate the period-$t$ likelihood value:

$$\hat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N f(y_t|s^i_t).$$
Example Algorithm: Bootstrap Particle Filter (Gordon, Salmond, Smith, 1993)

Likelihood evaluation:

- Inherit \( \hat{f}(s_{t-1}|Y_{t-1}) = \{s_{t-1}^i\}_{i=1}^N \) from period-\((t-1)\) step.
- For each \( s_{t-1}^i \), draw \( s_t^i \) from \( f(s_t|s_{t-1}^i) \). The resulting swarm represents MC draws from the joint density \( f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}) \).
- Approximate the period-\(t\) likelihood value:
  \[
  \hat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N f(y_t|s_t^i).
  \]
- Reference:
  \[
  f(y_t|Y_{t-1}) = \int \int f(y_t|s_t) f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}) ds_{t-1} ds_t.
  \]
BPF: Filtering

Recall the filtering density is given by

\[
f (s_t | Y_t) = \frac{f (y_t | s_t)}{f (y_t | Y_{t-1})} \cdot \int f (s_t | s_{t-1}) f (s_{t-1} | Y_{t-1}) \, ds_{t-1}.
\]

- For each particle in the swarm \( \{ s_t^i \}_{i=1}^N \), assign the weight

\[
\omega_t^i = \frac{f (y_t | s_t^i)}{\hat{f}_N (y_t | Y_{t-1})}.
\]
**BPF: Filtering**

Recall the filtering density is given by

\[
f(s_t|Y_t) = \frac{f(y_t|s_t)}{f(y_t|Y_{t-1})} \cdot \int f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}) \, ds_{t-1}.
\]

- For each particle in the swarm \( \{s^i_t\}_{i=1}^N \), assign the weight

\[
\omega^i_t = \frac{f(y_t|s^i_t)}{\hat{f}_N(y_t|Y_{t-1})}.
\]

- The joint swarm \( \{s^i_t, \omega^i_t\}_{i=1}^N \) yields the approximation

\[
\hat{f}(s_t|Y_t) = \sum_{i=1}^N \omega^i_t \cdot \delta(s_t - s^i_t).
\]
Drawing from \( \{s^i_t\}_{i=1}^N \) with replacement, with the probability of drawing \( s^i_t \) given by \( \omega^i_t \) (i.e., resampling), yields the approximation

\[
\hat{f}(s_t | Y_t) = \sum_{i=1}^{N} \delta(s_t - s^i_t).
\]
SIS/SIR, cont.

- Drawing from \( \{s_t^i\}_{i=1}^N \) with replacement, with the probability of drawing \( s_t^i \) given by \( \omega_t^i \) (i.e., resampling), yields the approximation

  \[
  \hat{f}(s_t|Y_t) = \sum_{i=1}^{N} \delta(s_t - s_t^i). 
  \]

- Passing the resampled \( \{s_t^i\}_{i=1}^N \) to the period-\((t + 1)\) stage and continuing through period \( T \) completes the algorithm.
Proposal densities are “blind” (Pitt & Shephard, 1999 *JASA*): do not incorporate period-\(t\) information. Can lead to “sample impoverishment.”
Shortcomings of the SPF

- Proposal densities are “blind” (Pitt & Shephard, 1999 *JASA*): do not incorporate period-$t$ information. Can lead to “sample impoverishment.”

- In the context of the foregoing, the sampler

\[
f(s_t|s_{t-1})f(s_{t-1}|Y_{t-1})
\]

may be poorly-suited as an approximation of the targeted integrand

\[
f(y_t|s_t)f(s_t|s_{t-1})f(s_{t-1}|Y_{t-1}).
\]
Shortcomings of the SPF

- Proposal densities are “blind” (Pitt & Shephard, 1999 *JASA*): do not incorporate period-\(t\) information. Can lead to “sample impoverishment.”

- In the context of the foregoing, the sampler

\[
f(s_t \mid s_{t-1})f(s_{t-1} \mid Y_{t-1})
\]

may be poorly-suited as an approximation of the targeted integrand

\[
f(y_t \mid s_t)f(s_t \mid s_{t-1})f(s_{t-1} \mid Y_{t-1}).
\]

- Likelihood approximations are discontinuous functions of model parameters (Pitt, 2002); thus likelihood maximization problematic.
Basic Idea Behind Extensions

Extensions seek to incorporate period-$t$ information in constructing proposal densities used to generate candidate particles: adaption.

- Auxiliary/Adapted particle filters (Pitt and Shephard, 1999 *JASA*; Pitt, 2002 Warwick WP; Smith and Santos, 2006, *JBES*) rely on Taylor series approximations of targeted densities to incorporate this information.
Basic Idea Behind Extensions

Extensions seek to incorporate period-$t$ information in constructing proposal densities used to generate candidate particles: adaption.

- Auxiliary/Adapted particle filters (Pitt and Shephard, 1999 *JASA*; Pitt, 2002 Warwick WP; Smith and Santos, 2006, *JBES*) rely on Taylor series approximations of targeted densities to incorporate this information.

- We achieve adaption via efficient importance sampling (EIS, Richard and Zhang, 2007 *JoE*).
Extensions seek to incorporate period-$t$ information in constructing proposal densities used to generate candidate particles: adaption.

- Auxiliary/Adapted particle filters (Pitt and Shephard, 1999 *JASA*; Pitt, 2002 Warwick WP; Smith and Santos, 2006, *JBES*) rely on Taylor series approximations of targeted densities to incorporate this information.

- We achieve adaption via efficient importance sampling (EIS, Richard and Zhang, 2007 *JoE*).
  - EIS yields global approximations of targeted densities.
Basic Idea Behind Extensions

Extensions seek to incorporate period-$t$ information in constructing proposal densities used to generate candidate particles: adaption.

- Auxiliary/Adapted particle filters (Pitt and Shephard, 1999 *JASA*; Pitt, 2002 Warwick WP; Smith and Santos, 2006, *JBES*) rely on Taylor series approximations of targeted densities to incorporate this information.

- We achieve adaption via efficient importance sampling (EIS, Richard and Zhang, 2007 *JoE*).
  - EIS yields global approximations of targeted densities.
  - Approximations are optimal: they minimize the numerical error associated with approximated integrals.
Basic Idea Behind Extensions

Extensions seek to incorporate period-$t$ information in constructing proposal densities used to generate candidate particles: adaption.

- Auxiliary/Adapted particle filters (Pitt and Shephard, 1999 *JASA*; Pitt, 2002 Warwick WP; Smith and Santos, 2006, *JBES*) rely on Taylor series approximations of targeted densities to incorporate this information.

- We achieve adaption via efficient importance sampling (EIS, Richard and Zhang, 2007 *JoE*).
  - EIS yields global approximations of targeted densities.
  - Approximations are optimal: they minimize the numerical error associated with approximated integrals.
  - Approximations are continuous functions of model parameters.
Goal: Minimize the n.s.e. associated with the approximation of

\[ G(Y) = \int_\Theta \varphi(\theta | Y) d\theta. \]

Approach: Introduce importance sampling distribution \( g(\theta | a) \):

\[
G(Y) = \int_\Theta \frac{\varphi(\theta; Y)}{g(\theta | a)} \varphi(\theta | Y) \, d\theta \\
= \int_\Theta \frac{\phi(\theta; Y) \, p(\theta | Y)}{g(\theta | a)} g(\theta | a) \, d\theta.
\]

Optimality: Tailor \( g(\theta | a) \) (via the specification of \( a \)) to minimize n.s.e.
EIS, cont.

Writing

\[ g(\theta|a) = \frac{k(\theta;a)}{\chi(a)}, \]

\[ \chi(a) = \int_{\Theta} k(\theta; Y) \, d\theta, \]

n.s.e. is (approximately) minimized via iterations on

\[ (\hat{a}_{k+1}, \hat{c}_{k+1}) = \arg \min \overline{Q}_N(a, c; Y|\hat{a}_k), \]

\[ \overline{Q}_N(a, c; Y|\hat{a}_k) = \frac{1}{N} \sum_{i=1}^{N} d^2 \left( \theta_i^k, a, c, Y \right), \]

\[ d \left( \theta_i^k, a, c, Y \right) = \ln \varphi \left( \theta_i^k | Y \right) - c - \ln k \left( \theta_i^k; a \right). \]

The term \( c \) is a normalizing constant that controls for factors in \( \varphi \) and \( g \) that do not depend upon \( \theta \).
Specialized to the problem at hand, the **EIS filter** implements an optimized importance sampler

\[ g_t(s_t, s_{t-1}|\hat{a}) \equiv g_t(s_t, s_{t-1}) \]

tailored to the targeted integrand

\[ \varphi_t(s_t, s_{t-1}) = f(y_t|s_t) f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}) , \]

subject to the complication that we lack an analytical expression for

\[ f(s_{t-1}|Y_{t-1}) . \]
Corresponding IS weights and likelihood estimates are

\[
\omega_t(s_t, s_{t-1}) = \frac{\varphi_t(s_t, s_{t-1})}{g_t(s_t, s_{t-1})},
\]

\[
\hat{f}(y_t | Y_{t-1}) = \frac{1}{N} \sum_{i=1}^{N} \omega_t(s_t^i, s_{t-1}^i).
\]
To approximate \( f(s_t|Y_t) \), we partition the EIS sampler as

\[
g_t(s_t, s_{t-1}) = g_t(s_t) g_t(s_{t-1}|s_t).
\]

The period-\( t \) filtering density is then

\[
f(s_t|Y_t) = \frac{g_t(s_t)}{f(y_t|Y_{t-1})} \int \omega_t(s_t, s_{t-1}) g_t(s_{t-1}|s_t) \, ds_{t-1}.
\]

For any given \( s_t \), the integral in \( s_{t-1} \) can be approximated by

\[
\bar{\omega}_t(s_t) = \frac{1}{N} \sum_{i=1}^{N} \omega_t(s_t, s^i_{t-1}(s_t)),
\]

where \( \{s^i_{t-1}(s_t)\}_{i=1}^{N} \) denotes draws from \( g_t(s_{t-1}|s_t) \).
The period-\(t\) filtering density obtains as

\[
\hat{f}(s_t|Y_t) = g_t(s_t)\tilde{\omega}_t(s_t), \quad \text{with}
\]

\[
\tilde{\omega}_t(s_t) = \frac{\sum_{i=1}^{N} \omega_t(s_t, s_{t-1}^i(s_t))}{\sum_{i=1}^{N} \omega_t(s_t^i, s_{t-1}^i)}.
\]
Thus, the absence of an analytical expression for $f(s_{t-1}|Y_{t-1})$ boils down to the need to evaluate $\tilde{\omega}_{t-1}(s_{t-1})$.

Potential approaches:

- Explore the accuracy of a constant-weight approximation for $\tilde{\omega}_{t-1}(s_{t-1})$, in which case $\hat{f}(s_t|Y_t) = g_t(s_t)$.
- Approximate $\tilde{\omega}_{t-1}(s_{t-1})$ numerically (e.g., via interpolation).
In period $t$, inherit

$$\hat{f}(s_{t-1}|Y_{t-1}) = g_{t-1}(s_{t-1}) = f_d^N(s_{t-1}|b_{t-1}, P_{t-1}),$$

and tailor

$$g_t(s_t, s_{t-1}|\hat{a}) \equiv g_t(s_t, s_{t-1}) = g_t(s_t) g_t(s_{t-1}|s_t)$$

to the targeted integrand

$$\varphi_t(s_t, s_{t-1}) = f(y_t|s_t) f(s_t|s_{t-1}) f(s_{t-1}|Y_{t-1}).$$
To initialize $g_t(s_t, s_{t-1} | a_0)$, set

$$
\hat{f}(y_t | s_t) = f^n_N(y_t | \kappa_t + H_t s_t, \Sigma),
\hat{f}(s_t | s_{t-1}) = f^d_N(s_t | c_t + F_t s_{t-1}, \Psi),
$$

and obtain the multivariate Normal density

$$
g_t(s_t, s_{t-1} | a_0) = \hat{f}(y_t | s_t) \hat{f}(s_t | s_{t-1}) \hat{f}(s_{t-1} | Y_{t-1}).
$$

Note: the initialized importance sampler is the extended Kalman filter of Durbin and Koopman (1997 *Biometrika*).

Then optimize $\hat{a}$ via EIS to obtain $g_t(s_t, s_{t-1} | \hat{a})$.

Finally, marginalize to obtain $g_t(s_t)$, which is the approximated filtering density passed to period $(t + 1)$. 
Example: F-V/R-R (2005 JAE) RBC Model

\[
\max_{c_t, l_t} U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_{t}^\varphi l_{t}^{1-\varphi})^{1-\varphi}}{1 - \varphi},
\]

subject to

\[
\begin{align*}
y_t &= z_t k_t^\alpha n_t^{1-\alpha}, \\
1 &= n_t + l_t, \\
y_t &= c_t + i_t, \\
k_{t+1} &= i_t + (1 - \delta) k_t, \\
z_t &= z_0 e^{g_t} e^{\omega_t}, \quad \omega_t = \rho \omega_{t-1} + \varepsilon_t.
\end{align*}
\]
State Transition Equations:

\[
\left(1 + \frac{g}{1 - \alpha}\right) k'(k_t, z_t) = i(k_t, z_t) + (1 - \delta) k_t
\]

\[
\log z_t = (1 - \rho) \log(z_0) + \rho \log z_{t-1} + \varepsilon_t.
\]

Observation Equations:

\[
x_t = x(k_t, z_t) + u_{x,t}, \quad x = y, i, n,
\]

\[
u_{x,t} \sim N(0, \sigma^2_x).
\]
Experiment 1: Generate artificial realizations of \( \{y_t, i_t, n_t\} \), \( T = 100 \) from the RBC model.

Using alternative filters, recover likelihood estimates at actual parameter values.

Note: Std. dev.s of measurement errors for \( \{y_t, i_t, n_t\} \) : [0.000158, 0.0011, 0.000866] ; in turn, \( \sigma_\varepsilon = 0.007 \).
### Table 3.A. MC Means and Standard Deviations of Log-Likelihood Estimates

<table>
<thead>
<tr>
<th></th>
<th>BPF(1M)</th>
<th></th>
<th>EISF(100/200)</th>
<th></th>
<th>EKFIS(100/200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>NSE</td>
<td>Mean</td>
<td>NSE</td>
<td>Mean</td>
</tr>
<tr>
<td>RBC art</td>
<td>1299.9492</td>
<td>0.5287</td>
<td>1300.0674</td>
<td>0.0005</td>
<td>1283.8614</td>
</tr>
<tr>
<td>RBC act</td>
<td>2305.2687</td>
<td>0.9139</td>
<td>2305.6589</td>
<td>0.0151</td>
<td>2305.2988</td>
</tr>
<tr>
<td>SOE art</td>
<td>1293.8399</td>
<td>0.7134</td>
<td>1294.1197</td>
<td>0.0229</td>
<td>1253.0159</td>
</tr>
<tr>
<td>SOE act</td>
<td>1718.1814</td>
<td>0.1699</td>
<td>1718.3298</td>
<td>0.0165</td>
<td>1696.6785</td>
</tr>
</tbody>
</table>
Table 3.B. Bias Corrected Means of Log-Likelihood Estimates and Auxiliary Statistics

<table>
<thead>
<tr>
<th></th>
<th>BPF(1M) asympt.</th>
<th>BFF(1M) fin.-sample</th>
<th>EISF(100/200) asympt.</th>
<th>VARCOF asympt.</th>
<th>t-stat. asympt.</th>
<th>t-stat. fin.-samp</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBC art</td>
<td>1300.0890</td>
<td>1300.0716</td>
<td>1300.0674</td>
<td>0.000051</td>
<td>0.4078</td>
<td>0.0794</td>
</tr>
<tr>
<td>RBC act</td>
<td>2305.6863</td>
<td>2305.5804</td>
<td>2305.6590</td>
<td>0.00063</td>
<td>0.2986</td>
<td>-0.8588</td>
</tr>
<tr>
<td>SOE art</td>
<td>1294.0944</td>
<td>1294.0713</td>
<td>1294.1200</td>
<td>0.0043</td>
<td>-0.3572</td>
<td>-0.6781</td>
</tr>
<tr>
<td>SOE act</td>
<td>1718.1958</td>
<td>1718.1958</td>
<td>1718.3299</td>
<td>0.0029</td>
<td>-7.8961</td>
<td>-7.8961</td>
</tr>
</tbody>
</table>
Figure 1. Log-likelihood approximations.
Figure 2. Log-likelihood sections of the SOE model for the actual data.