Efficient Likelihood Evaluation of State-Space Representations

David N. DeJong*, Roman Liesenfeld**, Guilherme V. Moura***, Jean-Francois Richard*, Hariharan Dharmarajan****
*University of Pittsburgh **Universität Kiel ***VU University ****Bates White LLC

August 2010

Objective

Likelihood evaluation and filtering for state-space representations featuring departures from:

- Linearity
- Normality

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 known as sequential Monte Carlo (SMC) methods.
- Example: In working with DSGE models, linear approximations are problematic for conducting likelihood analysis (Fernandez-Villaverde and Rubio-Ramirez, 2005 *JAE;* 2009 *REStud*)

Sketch of Literature on SMC Methods

- SMC methods employ importance sampling densities to construct numerical approximations of integrals that arise in pursuit of likelihood evaluation and filtering.
- Typically, importance samplers are based on *discrete* approximations of filtering densities. The individual elements of these samplers are known as particles; the approximations they represent collectively are known as *particle swarms*.

SMC Methods, cont.

- Baseline methods construct time-t approximations of filtering densities absent information on the time-t observables y_t. Such methods are termed as being *unadapted*. Leading examples include Handschin and Mayne, 1969 *Intl. J. of Control;* Handschin, 1970 *Automatica;* Gordon, Salmond, and Smith, 1993 *IEEE Proceedings.*
- Baseline methods are relatively *easy to implement*, and yield *unbiased estimates*; however, they can be *numerically inefficient*.
- Refinements seek to achieve improvements in numerical efficiency by taking y_t into account in constructing time-t samplers. The pursuit of such improvements is known as *adaption*. A prominent example of an adapted algorithm is the auxiliary particle filter of Pitt and Shephard, 1999 *JASA.*

SMC Methods, cont.

- To date, adaption has been pursued subject to the constraint that the discrete support of the filtering density constructed in period $t-1$ is taken as given and fixed in period t. We refer to the imposition of this constraint as the pursuit of *conditional adaption*.
- The approach to filtering we propose here is implemented absent this constraint: our objective is to pursue *unconditional adaption*.

SMC Methods, cont.

- Specifically, we use *continuous approximations* of filtering densities as an input to the construction of time-t importance samplers designed to generate optimal (in terms of numerical efficiency) global approximations to targeted integrands.
- The approximations fully account for the information conveyed by y_t, and are constructed using the methodology of efficient importance sampling (EIS) developed by Richard and Zhang, 2007 *J. of Econometrics*.
- Resulting likelihood approximations are continuous functions of model parameters, greatly enhancing the pursuit of parameter estimation.

State Space Representations

State-transition equation:

$$s_t = \gamma(s_{t-1}, Y_{t-1}, \upsilon_t)$$

Associated density:

$$f(s_t | s_{t-1}, Y_{t-1})$$

Measurement equation:

$$y_t = \delta(s_t, Y_{t-1}, u_t)$$

Associated density:

$$f(y_t | s_t, Y_{t-1})$$

Initialization:

$$f(s_0)$$

State Space Representations, cont.

Objective: evaluate the likelihood function

$$f\left(Y_T\right) = \prod_{t=1}^{T} f\left(y_t \middle| Y_{t-1}\right),$$

where $f\left(y_1 \middle| Y_0\right) \equiv f(y_1)$.
Time-t likelihoods are evaluated via marginalization of measurement densities:

$$f\left(y_t \middle| Y_{t-1}\right) = \int f\left(y_t \middle| s_t, Y_{t-1}\right) f\left(s_t \middle| Y_{t-1}\right) ds_t.$$

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

$$f\left(s_t \middle| Y_{t-1}\right) = \int f\left(s_t \middle| s_{t-1}, Y_{t-1}\right) f\left(s_{t-1} \middle| Y_{t-1}\right) ds_{t-1},$$

where

$$f\left(s_t \middle| Y_t\right) = \frac{f\left(y_t, s_t \middle| Y_{t-1}\right)}{f\left(y_t \middle| Y_{t-1}\right)} = \frac{f\left(y_t \middle| s_t, Y_{t-1}\right) f\left(s_t \middle| Y_{t-1}\right)}{f\left(y_t \middle| Y_{t-1}\right)}.$$

Particle Filters: General Principle

- Period-t computation inherently requires the evaluation of

$$f\left(y_t|Y_{t-1}\right) = \int\int f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}(s_{t-1}|Y_{t-1}) ds_{t-1}$$

- Particle filters rely upon approximations in the form of a mixture-of-Dirac measures associated with the period-$(t-1)$ swarm $\{s_{t-1}^i\}_{i=1}^N$ which is fixed in period-t :

$$\widehat{f}(s_{t-1}|Y_{t-1}) = \sum_{i=1}^{N} \omega_{t-1}^i \cdot \delta_{s_{t-1}^i}\left(s_{t-1}\right),$$

  where $\delta_{s_{t-1}^i}\left(s\right)$ denotes the Dirac measure at point s_{t-1}^i, and ω_{t-1}^i the weight associated with particle s_{t-1}^i.

- This approximation effectively solves the (inner) integration in s_{t-1}, yielding

$$f\left(y_t|Y_{t-1}\right) = \sum_{i=1}^{N} \omega_{t-1}^i \int f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}^i, Y_{t-1}\right) ds_t.$$

Unadapted Filters

Period-t algorithm:

- Inherit $\widehat{f}\left(s_{t-1}|Y_{t-1}\right)$, represented using $\{\omega^i_{t-1}, s^i_{t-1}\}_{i=1}^N$, from the period-$(t-1)$ step.
- Approximate $f\left(s_t|Y_{t-1}\right)$: for each s^i_{t-1}, draw s^i_t from $f\left(s_t|s^i_{t-1}, Y_{t-1}\right)$, yielding

$$\widehat{f}\left(y_t|Y_{t-1}\right) = \sum_{i=1}^N \omega^i_{t-1} \cdot f\left(y_t|s^i_t, Y_{t-1}\right).$$

- Approximate $\widehat{f}\left(s_t|Y_t\right)$ as

$$\widehat{f}\left(s_t|Y_t\right) = \sum_{i=1}^N \omega^i_t \delta_{s^i_t}\left(s_t\right),$$

  where the (posterior) weights ω^i_t obtain from the (prior) weights ω^i_{t-1} by application of Bayes' theorem:

$$\omega^i_t = \omega^i_{t-1} \cdot \frac{f\left(y_t|s^i_t, Y_{t-1}\right)}{\widehat{f}\left(y_t|Y_{t-1}\right)}.$$

Conditional Adaptation

- The measurement density incorporates the assumption that y_t is independent of s_{t-1} given (s_t, Y_{t-1}); this implies

$$f\left(y_t | s_t, Y_{t-1}\right) \cdot f\left(s_t | s_{t-1}, Y_{t-1}\right) = f\left(s_t | s_{t-1}, Y_t\right) \cdot f\left(y_t | s_{t-1}, Y_{t-1}\right)$$

- When this factorization is analytically tractable, it is possible to achieve conditionally optimal adaption:

$$\begin{aligned} f\left(y_t | Y_{t-1}\right) &= \int \int f\left(s_t | s_{t-1}, Y_t\right) \cdot f\left(y_t | s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}\left(s_{t-1} | Y_{t-1}\right) d \\ &= \int f\left(y_t | s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}\left(s_{t-1} | Y_{t-1}\right) ds_{t-1} \\ &= \sum_{i=1}^{N} \omega_{t-1}^{i} \cdot f\left(y_t | s_{t-1}^{i}, Y_{t-1}\right). \end{aligned}$$

Conditional Adaptation: Implementation

- To implement, for each particle s_{t-1}^i, draw a particle s_t^i from $f\left(s_t|s_{t-1}^i, Y_t\right)$. The corresponding weights are given by

$$\omega_t^i = \omega_{t-1}^i \cdot \frac{f\left(y_t|s_{t-1}^i, Y_{t-1}\right)}{\widehat{f}\left(y_t|Y_{t-1}\right)}.$$

- Key difference relative to unadapted filters: the draws of s_t are conditional on y_t. Since ω_t^i does not depend on s_t^i, but only on s_{t-1}^i, its *conditional* variance is zero given $\{s_{t-1}^i\}_{i=1}^N$. This is referenced as the optimal sampler following Zaritskii et al., 1975 *Automation and Remote Control;* Akaski and Kumamoto, 1977 *Automatica*.
- Since the factorization

$$f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) = f\left(s_t|s_{t-1}, Y_t\right) \cdot f\left(y_t|s_{t-1}, Y_{t-1}\right)$$

is tractable only in special cases, this sampler represents a theoretical rather than an operational benchmark.

Approximate Conditional Optimality

- Attempts at approximating conditional optimality follow from the interpretation of

$$f\left(y_t | Y_{t-1}\right) = \sum_{i=1}^{N} \omega_{t-1}^{i} \int f\left(y_t | s_t, Y_{t-1}\right) \cdot f\left(s_t | s_{t-1}^{i}, Y_{t-1}\right) ds_t$$

  as a mixed integral in (s_t, k_t), where k_t denotes the index of particles, and follows the multinomial distribution $MN\left(N, \{\omega_{t-1}^{i}\}_{i=1}^{N}\right)$.
- The likelihood integral may then be evaluated via importance sampling, relying upon a mixed density kernel of the form

$$\gamma_t(s,k) = \omega_{t-1}^{k} \cdot p_t(s,k) \cdot f\left(s_t | s_{t-1}^{k}, Y_{t-1}\right).$$

- Pitt and Shephard (1993 *JASA*) pursue conditional optimality by specifying $p_t(s,k)$ as

$$p_t(s,k) = f\left(y_t | \mu_t^{k}, Y_{t-1}\right), \qquad \mu_t^{k} = E(s_t | s_{t-1}^{k}, Y_{t-1}).$$

Unconditional Optimality

- Returning to the period-t likelihood integral

$$f\left(y_t|Y_{t-1}\right) = \int\int f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}(s_{t-1}|Y_{t-1})ds_{t-1}$$

consider the theoretical factorization

$$f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}\left(s_{t-1}|Y_{t-1}\right) = f\left(s_t, s_{t-1}|Y_t\right) \cdot f\left(y_t|\right.$$

- If analytically tractable, $f\left(s_t, s_{t-1}|Y_t\right)$ would be the unconditionally optimal (fully adapted) sampler for the likelihood integral, as a single draw from it would produce an estimate of $f\left(y_t|Y_{t-1}\right)$ with zero MC variance.
- The period-t filtering density would then obtain by marginalization with respect to s_{t-1} :

$$f\left(s_t|Y_t\right) = \int f\left(s_t, s_{t-1}|Y_t\right) ds_{t-1}.$$

Unconditional Optimality, cont.

- Our goal: approximate unconditional optimality by constructing importance samplers in (s_{t-1}, s_t) for the likelihood integral

$$f\left(y_t \middle| Y_{t-1}\right) = \int\int f\left(y_t \middle| s_t, Y_{t-1}\right) \cdot f\left(s_t \middle| s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}\left(s_{t-1} \middle| Y_{t-1}\right) ds_{t-1}$$

- The goal is pursued via the principle of *efficient importance sampling* (EIS).

EIS (Richard and Zhang, 2007 JoE)

- Let $\varphi_t(\lambda_t)$ denote the integrand

$$f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}(s_{t-1}|Y_{t-1}),$$

with $\lambda_t = \left(s_{t-1}, s_t\right)$.

- Implementation of EIS begins with the pre-selection of a parametric class $K = \left\{k\left(\lambda_t; a_t\right); a_t \in A\right\}$ of analytically integrable auxiliary density kernels. The corresponding density functions (IS samplers) and IS ratios are given respectively by

$$\begin{aligned} g(\lambda_t|a_t) &= \frac{k(\lambda_t; a_t)}{\chi(a_t)}, \qquad \chi(a_t) = \int k(\lambda_t; a_t) d\lambda_t, \\ \omega_t(\lambda_t; a_t) &= \frac{\varphi_t(\lambda_t)}{g_t\left(\lambda_t|a_t\right)}. \end{aligned}$$

EIS, cont.

- Objective: select $\widehat{a}_t \in A$ to minimize the MC variance of the IS ratio over the full range of integration.
- A near-optimal value $\widehat{a}_t$ obtains as the solution to

$$(\widehat{a}_t, \widehat{c}_t) = \underset{(a_t, c_t)}{\arg\min} \int \left[\ln \varphi_t(\lambda_t) - c_t - \ln k(\lambda_t; a_t)\right]^2 g(\lambda_t | a_t) d\lambda_t,$$

  where c_t denotes an intercept meant to calibrate the ratio $\ln\left(\varphi_t / k\right)$.
- This represents a standard least squares problem, except that the auxiliary sampling density depends upon a_t. This is resolved via the specification of an initial value $\widehat{a}_t^0$, and the search for a fixed point solution via iterations on

$$(\widehat{a}_t^{l+1}, \widehat{c}_t^{l+1}) = \underset{(a_t, c_t)}{\arg\min} \sum_{i=1}^{R} \left[\ln \varphi_t(\lambda_{t,l}^i) - c_t - \ln k(\lambda_{t,l}^i; a_t)\right]^2.$$

The EIS Filter

- Having obtained the fixed-point solution $\widehat{a}_t$, the likelihood EIS estimate is given by

$$\widehat{f}\left(y_t | Y_{t-1}\right) = \frac{1}{S}\sum_{i=1}^{S}\omega_t\left(s_{t-1}^i, s_t^i; \widehat{a}_t\right),$$

$$\omega_t(\lambda_t; a_t) = \frac{\varphi_t(\lambda_t)}{g_t\left(\lambda_t | a_t\right)},$$

where $\{s_{t-1}^i, s_t^i\}_{i=1}^S$ denotes i.i.d. draws from the EIS sampler $g\left(s_{t-1}, s_t | \widehat{a}_t\right)$.

- A period-t filtering density approximation is then given by the marginal of g in $s_{t.}$:

$$\widehat{f}\left(s_t | Y_t\right) = \int g\left(s_{t-1}, s_t; \widehat{a}_t\right) ds_{t-1}.$$

Initialization

- The selection of a good initial sampler $g_t\left(s_t, s_{t-1}|\widehat{a}_t^0\right)$ is critical for achieving reliable convergence to an effective final sampler $g_t\left(s_t, s_{t-1}|\widehat{a}_t\right)$.
- We rely upon local Taylor Series expansions to construct initial Gaussian samplers. This is similar to the procedure proposed by Durbin and Koopman (1997) whereby (local) Gaussian approximations are used as importance samplers to evaluate the likelihood function of non-Gaussian state space models.
- Critical difference: we use these local approximations to construct starting values for fully iterated global EIS approximation.

Algorithmic Summary of the EIS Filter

- **Propagation:** Inheriting $\widehat{f}\left(s_{t-1} | Y_{t-1}\right)$ from period $(t-1)$, obtain the integrand

$$\varphi_t\left(s_{t-1}, s_t\right) = f\left(y_t | s_t, Y_{t-1}\right) \cdot f\left(s_t | s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}(s_{t-1} | Y_{t-1}).$$

- **EIS Optimization:** Construct an initialized sampler $g_t\left(s_{t-1}, s_t | \widehat{a}_t^0\right)$, and obtain the optimized parameterization $\widehat{a}_t$ as the solution to

$$(\widehat{a}_t^{l+1}, \widehat{c}_t^{l+1}) = \underset{(a_t, c_t)}{\arg\min} \sum_{i=1}^{R} \left[\ln \varphi_t(\lambda_{t,l}^i) - c_t - \ln k(\lambda_{t,l}^i; a_t)\right]^2.$$

- **Likelihood integral:** Obtain draws $\left\{s_{t-1}^{i}, s_{t}^{i}\right\}_{i=1}^{N}$ from $g_t\left(s_{t-1}, s_t \middle| \widehat{a}_t\right)$, and approximate $\widehat{f}\left(y_t \middle| Y_{t-1}\right)$ as

$$\widehat{f}\left(y_t \middle| Y_{t-1}\right) = \frac{1}{S}\sum_{i=1}^{S} \omega_t\left(s_{t-1}^{i}, s_t^{i}; \widehat{a}_t\right).$$

- **Filtering:** Approximate $\widehat{f}\left(s_t \middle| Y_t\right)$ as

$$\widehat{f}\left(s_t \middle| Y_t\right) = \int g\left(s_{t-1}, s_t; \widehat{a}_t\right) ds_{t-1}.$$

- **Continuation:** Pass $\widehat{f}\left(s_t \middle| Y_t\right)$ to the period-$(t+1)$ propagation step and proceed through period T.

Performance

- We demonstrate the performance of the EIS filter relative to the (unadapted) bootstrap particle filter of Gordon, Salmond, and Smith (1993 *IEEE Proceedings*).
- Application is to four data sets: two artificial/actual pairs associated with two DSGE models.
- Model 1: the two-state RBC model used by Fernandez-Villaverde and Rubio-Ramirez (2005 *J. Applied Econometrics*) to demonstrate the bootstrap particle filter.
- Model 2: a six-state version of the small open economy model fashioned from Mendoza (1991 *AER*), Schmitt-Grohe and Uribe (2003 *J. Int'l Economics*).

Performance, cont.

Each data set offers a unique challenge:

- RBC Model, Artificial Data: highly informative measurement densities
- RBC Model, Actual Data: outliers (1974:IV, 1980:II)
- SOE Model, both data sets: relatively high-dimensional state space, relatively significant departures from linearity in the state-transition equations.

Representative household's problem:

$$\max U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(c_t^{\varphi} l_t^{1-\varphi}\right)^{1-\phi}}{1-\phi},$$

subject to

$$\begin{aligned} y_t &= z_t k_t^{\alpha} n_t^{1-\alpha}, \\ 1 &= n_t + l_t, \\ y_t &= c_t + i_t, \\ k_{t+1} &= i_t + (1-\delta) k_t, \\ z_t &= z_0 e^{gt} e^{\omega_t}, \quad \omega_t = \rho \omega_{t-1} + \varepsilon_t. \end{aligned}$$

Example, cont.

State Transition Equations:

$$\begin{aligned}\left(1+\frac{g}{1-\alpha}\right)k'(k_t,z_t) &= i(k_t,z_t)+(1-\delta)k_t \\ \log z_t &= (1-\rho)\log(z_0)+\rho\log z_{t-1}+\varepsilon_t.\end{aligned}$$

Observation Equations:

$$\begin{aligned}x_t &= x(k_t,z_t)+u_{x,t}, \quad x=y,i,n, \\ u_{x,t} &\sim N(0,\sigma_x^2).\end{aligned}$$

SOE Model

Representative household's problem:

$$\begin{aligned} \max U &= E_0 \sum_{t=0}^{\infty} \theta_t \frac{\left[c_t - \varphi_t \omega^{-1} n_t^{\omega}\right]^{1-\gamma} - 1}{1-\gamma}, \qquad \omega > 0, \qquad \gamma \geqslant 0, \\ \theta_{t+1} &= \beta\left(\widetilde{c}_t, \widetilde{n}_t\right) \theta_t, \qquad \theta_0 = 1, \\ \beta\left(\widetilde{c}_t, \widetilde{n}_t\right) &= \left[1 + \widetilde{c}_t - \omega^{-1} \widetilde{n}_t^{\,\omega}\right]^{-\psi}, \qquad \psi > 0, \end{aligned}$$

where $(\widetilde{c}_t, \widetilde{n}_t)$ denote average per capita consumption and hours worked, subject to

SOE Model, cont.

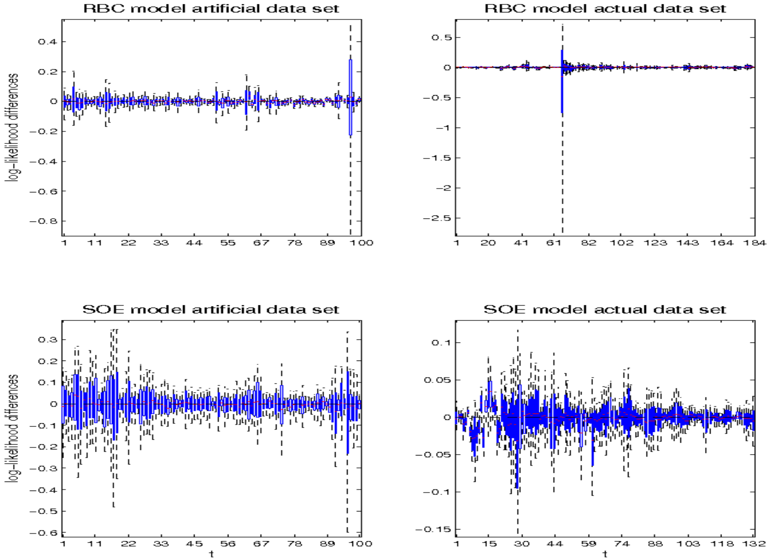
$$
\begin{aligned}
x_t &= A_t k_t^{\alpha} n_t^{1-a} \\
d_{t+1} &= (1 + r_t)\, d_t - x_t + c_t + i_t + \frac{\phi}{2}\left(k_{t+1} - k_t\right)^2 \\
k_{t+1} &= v_t^{-1} i_t + (1 - \delta)\, k_t \\
\ln A_{t+1} &= \rho_A \ln A_t + \varepsilon_{At+1}, \qquad \varepsilon_{At} \sim iidN(0, \sigma^2_{\varepsilon_A}) \\
\ln r_{t+1} &= (1 - \rho_r) \ln r_* + \rho_r \ln r_t + \varepsilon_{rt+1}, \qquad \varepsilon_{rt} \sim iidN(0, \sigma^2_{\varepsilon_r}) \\
\ln v_{t+1} &= \rho_v \ln v_t + \varepsilon_{vt+1}, \qquad \varepsilon_{vt} \sim iidN(0, \sigma^2_{\varepsilon_v}) \\
\ln \varphi_{t+1} &= \rho_\varphi \ln \varphi_t + \varepsilon_{\varphi t+1}, \qquad \varepsilon_{\varphi t} \sim iidN(0, \sigma^2_{\varepsilon_\varphi}).
\end{aligned}
$$

Observables: $y_t' = (x_t, \quad c_t, \quad i_t, \quad n_t)$.

Experiment 1: Bias in the EIS Filter?

- For each data set, generate 100 date-by-date log-likelihood approximations (using 100 different sets of random numbers) using the BP filter, $N = 1,000,000$. Given the unbiasedness of the BP filter, these approximations serve as a benchmark for judging the EIS filter.
- Next, generate 100 log-likelihood approximations using the EIS filter ($N = R = 100$ for the RBC model, $N = R = 200$ for the SOE model).
- Calculate the difference in approximations for each of the 10,000 possible combinations of likelihood values, and searched for instances in which differences were significantly different from zero.
- Result: in all instances, zero lies between the 5^{th} and 95^{th} percentiles of resulting boxplots. I.E., we cannot reject the null that differences between estimators merely reflect numerical error.

Example 1, cont.



Experiment 2: Comparison of Numerical Efficiency

- Once again, for each data set, generate 100 date-by-date log-likelihood approximations (using 100 different sets of random numbers) using the BP and EIS filters.
- Objective: compare the numerical standard errors associated with the filters.
- BP Filter: $N = 60,000$ for the RBC model (following F-V/R-R), $N = 150,000$ for the SOE model. Computational times range from 17.28 seconds (RBC, artificial) to 80.90 seconds (SOE, actual).
- EIS Filter: $N = R = 100$ for the RBC model, $N = R = 200$ for the SOE model. Computational times range from 0.55 seconds (RBC, artificial) to 2.18 seconds (SOE, actual).
- Result: Tremendous efficiency gains associated with the EIS filter.

Example 2, cont.

RBC Model

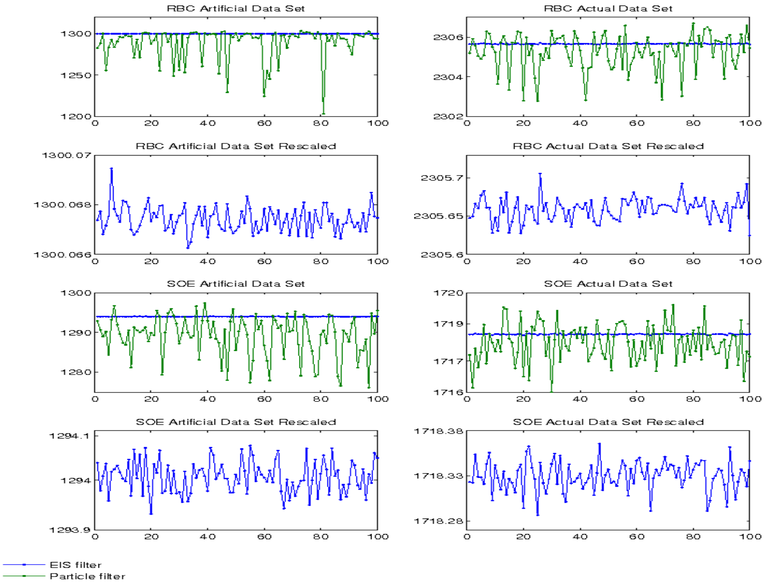
	BP Filter		EIS Filter		Initial Sampler	
	Mean	NSE	Mean	NSE	Mean	NSE
Art.	1289.4520	19.0234	1300.0524	5.1e-04	1283.8614	0.3486
Act.	2305.2687	0.9139	2305.6589	0.0151	2305.2988	4.0279

SOE Model

	BP Filter		EIS Filter		Initial Sampler	
	Mean	NSE	Mean	NSE	Mean	NSE
Art.	1289.1690	5.1659	1294.0069	0.0232	1253.0159	4.6041
Act.	1717.9816	0.7607	1718.3298	0.0166	1696.6785	3.7189

Table:

Example 2, cont.



Experiment 3: Are Results Data-Set Specific?

- Repeat Experiment 2 for 100 artificial data sets generated from the four model parameterizations represented in the previous experiments (EIS filter, only). This yields a distribution of NSEs, indicating whether the NSEs in the previous table are somehow unusual.
- In addition, we construct sampling (statistical) errors by calculating the standard deviation of likelihood estimates obtained across data sets using the EIS filter implemented with a single set of common random numbers.
- Result: the NSEs reported in Table 2 only appear unusual for the RBC model, actual data set, which features the two significant outliers.
- Also: statistical standard errors dominate NSEs (by two to five orders of magnitude).

Example 3, cont.

	RBC Model		
	SSE	NSE	
		Mean	Std. Dev.
Artificial Data	19.8978	4.9e-4	1.2e-4
Actual Data	1.1483	0.0113	6.9e-4

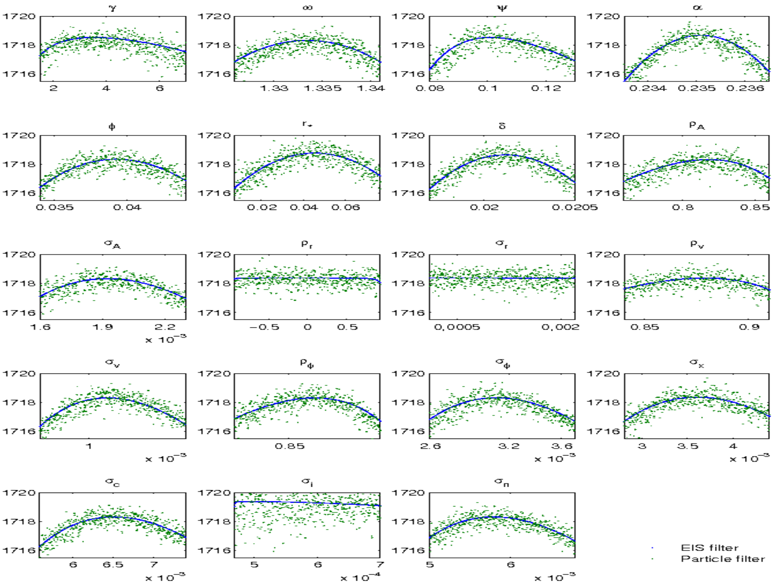
	SOE Model		
	SSE	NSE	
		Mean	Std. Dev.
Artificial Data	17.6041	0.0417	0.0365
Actual Data	15.6710	0.0134	0.0028

Notes: SSE stands for statistical standard errors, which were computed as standard deviations of log-likelihood values across 100 alternative data sets. NSE denotes numerical standard errors.

Table:

Experiment 4: Continuity of log-likelihood surfaces

- Generate log-likelihood surfaces by allowing each model parameter to vary individually above and below its ML estimate, holding all additional parameters fixed at their ML estimates. For each parameter combination, obtain log-likelihood approximations using the same set of CRNs, to eliminate numerical error.
- Result: surfaces associated with the BP filter are discontinuous; those associated with the EIS filter are continuous.
- Figure 3: SOE model, actual data set.


γ
ω
ψ
α
φ
r*
δ
ρA
σA
ρr
σr
ρv
σv
ρφ
σφ
σx
σc
σi
σn
1720
1718
1716
EIS filter
Particle filter

Experiment 5: Outliers, and Bias Redux

- For RBC model, artificial data set, generate 12 variations by inserting an outlier in the second observation of one of the observables, keeping the remaining variables fixed at their original values.
- Four outliers were generated for each variable: two deviated by ± 4 sample standard deviations from the sample mean, and two deviated ± 8 sample standard deviations from the sample mean.
- For each new data set (as well as for the original), log-likelihood values were calculated for periods 1 and 2 using the BP and EIS filters, and also the Gauss-Chebyshev quadrature method implemented with 250 nodes along all three dimensions of integration, for a total of $250^3 = 15,625,000$ nodes.
- By evaluating the first two periods only, implementation of the quadrature method is feasible, and provides a near-exact value of targeted log-likelihoods.
- Result: the EIS filter remains free of bias, and its associated NSEs are fairly uniform across data sets. The performance of the BP filter deteriorates in the presence of bias.

Example, 5 cont.

		BP Filter				EIS Filter		
		$t=1$		$t=2$		$t=1$		$t=2$
		Mean	NSE	Mean	NSE	Mean	NSE	Mean
	-8	10.7477	0.2922	13.1189	0.1550	10.8265	0.0007	13.1237
	-4	10.7477	0.2922	13.1777	0.1246	10.8265	0.0007	13.1981
x	0	10.7477	0.2922	13.1166	0.1327	10.8265	0.0007	13.1153
	4	10.7477	0.2922	12.8416	0.1681	10.8265	0.0007	12.8753
	8	10.7477	0.2922	12.4285	0.2377	10.8265	0.0007	12.4782
		Mean	NSE	Mean	NSE	Mean	NSE	Mean
	-8	10.7477	0.2922	-7.2901	2.2706	10.8265	0.0007	-4.8069
	-4	10.7477	0.2922	8.2613	0.6063	10.8265	0.0007	8.3088
i	0	10.7477	0.2922	13.1166	0.1327	10.8265	0.0007	13.1153
	4	10.7477	0.2922	9.5606	0.4855	10.8265	0.0007	9.5875
	8	10.7477	0.2922	-4.1055	1.8742	10.8265	0.0007	-2.2999
		Mean	NSE	Mean	NSE	Mean	NSE	Mean
	-8	10.7477	0.2922	-19.2097	0.2208	10.8265	0.0007	-19.1907
	-4	10.7477	0.2922	4.8902	0.1706	10.8265	0.0007	4.8973
n	0	10.7477	0.2922	13.1166	0.1327	10.8265	0.0007	13.1153

Summary

1. Particle-based filters are easy to implement and produce unbiased likelihood estimates.
2. However, they are prone to numerical inefficiency, induce spurious discontinuities in likelihood surfaces, and at best admit efforts towards approximating conditional adaptation.
3. In turn, the EIS filter implements continuous approximations of filtering densities, enabling the pursuit of unconditional adaption.
4. The EIS algorithm produces global approximations of targeted integrands, yields significant gains in numerical efficiency, and produces likelihood surfaces that are continuous in model parameters.

Extensions

- The foregoing results were obtained using Gaussian approximations of filtering densities.
- While such approximations proved sufficient in the applications to DSGE models we have considered, they are clearly not appropriate in general.
- We are currently working to develop operational EIS samplers that are more flexible than those drawn from the exponential family of distributions. One such extension entails the development of an EIS procedure to construct global mixtures of Gaussian samplers; under this approach, EIS optimization is pursued via non-linear least squares implemented using analytical derivatives.
- The goal is to facilitate EIS implementations using highly flexible samplers that will prove efficient in applications involving even the most challenging of targeted integrands.