Efficient Likelihood Evaluation of State-Space Representations

David N. DeJong*
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260, USA

Roman Liesenfeld
Department of Economics
Universität Kiel
24118 Kiel, Germany

Guilherme V. Moura
Department of Economics
Universidade Federal de Santa Catarina
88049970 Florianopolis, Brazil

Jean-François Richard
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260, USA

with: Hariharan Dharmarajan
Bates White, L.L.C.
Washington, DC 20005, USA

First Version: April 2007 This Revision: August 2011

Abstract

We develop a numerical procedure that facilitates efficient likelihood evaluation in applications involving non-linear and non-Gaussian state-space models. The procedure employs continuous approximations of filtering densities, and delivers unconditionally optimal global approximations of targeted integrands to achieve likelihood approximation. Optimized approximations of targeted integrands are constructed via efficient importance sampling. Resulting likelihood approximations are continuous functions of model parameters, greatly enhancing parameter estimation. We illustrate our procedure in applications to dynamic stochastic general equilibrium models.

Keywords: dynamic stochastic general equilibrium model; efficient importance sampling; kernel density approximation; particle filter.

*Contact Author: D.N. DeJong, Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260, USA; Telephone: 412-648-2242; Fax: 412-648-1793; E-mail: dejong@pitt.edu. Richard gratefully acknowledges research support provided by the National Science Foundation under grant SES-0516642; and DeJong and Richard do the same for NSF support provided under grant SES-0850448. For helpful comments, we thank anonymous referees, Chetan Dave, Jesus Fernandez-Villaverde, Hiroyuki Kasahara, Juan Rubio-Ramirez, Enrique Sentana, and seminar and conference participants at the Czech National Bank, the Universitites of Comenius, Di Tella, Texas (Dallas), Pennsylvania, the 2008 Econometric Society Summer Meetings (CMU), the 2008 Conference on Computations in Economics (Paris), and the 2008 Vienna Macroecnomics Workshop. We also thank Chetan Dave for assistance with compiling the Canadian data set we analyzed, and Albrecht Mengel for providing access to the grid computing facilities of the Institute for Statistics and Econometrics at Universität Kiel. GAUSS code and data sets used to demonstrate implementation of the EIS filter are available at `www.pitt.edu/~dejong/wp.htm`

1 Introduction

Likelihood evaluation and filtering in applications involving dynamic state-space models require the calculation of integrals over unobservable state variables. When models are linear and stochastic processes are Gaussian, required integrals can be calculated analytically via the Kalman filter. Departures entail integrals that must be approximated numerically. Here we introduce an efficient procedure for calculating such integrals: the Efficient Importance Sampling (EIS) filter.

State-space models are Markovian, thus the integrals they entail in period t are joint in the period-$(t-1)$ and period-t state variables s_{t-1} and s_t. The EIS filter we propose for approximating these integrals falls under the general classification of a Sequential Monte Carlo (SMC) method, which dates at least to the Sequential Importance Sampling (SIS) algorithm of Handschin and Mayne (1969) and Handschin (1970). Inheriting $\{s_{t-1}^i\}_{i=1}^N$ from the period-$(t-1)$ stage of the algorithm, SIS entails in period t the construction of conditional Importance Sampling (IS) densities $\{g\left(s_t|s_{t-1}^i\right)\}_{i=1}^N$ from which draws $\{s_t^i\}_{i=1}^N$ are obtained and passed to period $t+1$. An extension known as a Sampling Importance Resampling (SIR) algorithm obtains by resampling $\{s_t^i\}_{i=1}^N$ with probabilities determined by corresponding IS weights $\{\omega_t^i\}_{i=1}^N$ before passing to period-$(t+1)$.

SIS and SIR algorithms are intuitive and easy to implement, but are often numerically inefficient: they can produce large Monte Carlo (MC) standard deviations for filtered estimates (likelihood functions and/or expected values of state variables). Extensive efforts have been made to extend baseline SIS/SIR algorithms with the goal of enhancing numerical efficiency; for surveys, see Ristic et al. (2004), Cappe et al. (2007), the collection of papers in Doucet et al. (2001a), and the collection housed at http://www.sigproc.eng.cam.uk/smc/papers.html.

The root inefficiency of SIS/SIR algorithms stems from their use of conditional IS samplers to approximate joint integrals in (s_{t-1}, s_t), and from the fact that the conditioning draws $\{s_{t-1}^i\}_{i=1}^N$ are generated absent information available in period t. Significant efforts towards achieving efficiency enhancements entail the introduction of a resampling step for s_{t-1} that takes into account period-t information. A leading example is the Auxiliary Particle Filter (APF) of Pitt and Shephard (1999), which preserves the inherent tractability of SIR algorithms by imposing two key restrictions: resampling is limited to the initial swarm $\{s_{t-1}^i\}_{i=1}^N$ (generated absent time-t information); and period-t IS densities are constructed via local Taylor series approximations of the model's measurement density.

As discussed, e.g., in Geweke (1989, 2007), local approximations often exhibit thinner tails than targeted integrands, resulting in large (occasionally unbounded) MC variances.

Here we seek to attain maximum efficiency by constructing sequential IS densities that in period t are tailored to the joint integral in (s_t, s_{t-1}) associated with likelihood evaluation. The densities are continuous in both s_t and s_{t-1}, and are constructed using the EIS methodology developed by Richard and Zhang (2007). The objective of EIS is to produce global approximations to targeted integrands, thus minimizing MC variances (or equivalently, maximizing efficiency) over the full range of integration. Though the EIS principle is generic, its implementation is more demanding than that of particle filters; but as illustrated below, it is capable of producing dramatic efficiency gains in challenging applications.

Our focus here is on the achievement of near-optimal efficiency for likelihood evaluation. Example applications involve the analysis of Dynamic Stochastic General Equilibrium (DSGE) models, and are used to illustrate the relative performance of the particle and EIS filters. In a companion paper (DeJong et al., 2010) we focus on filtering, and present an application to the bearings-only tracking problem featured prominently, e.g., in the engineering literature.

As motivation for our focus on the analysis of DSGE models, a brief literature review is helpful. The pioneering work of Sargent (1989) demonstrated the mapping of DSGE models into linear/Gaussian state-space representations amenable to likelihood-based analysis achievable via the Kalman filter. DeJong et al. (2000) developed a Bayesian approach to analyzing these models. Subsequent work has involved the implementation of DSGE models towards a broad range of empirical objectives, including forecasting and guidance of the conduct of aggregate fiscal and monetary policy (following Smets and Wouters, 2003).

Prior to the work of Fernandez-Villaverde and Rubio-Ramirez (2005, 2009), likelihood-based implementation of DSGE models was conducted using linear/Gaussian representations. But their findings revealed an important caveat: approximation errors associated with linear representations of DSGE models can impart significant errors in corresponding likelihood representations. As a remedy, they demonstrated use of the bootstrap particle filter developed by Gordon, et al. (1993) for achieving likelihood evaluation for non-linear model representations. But as our examples illustrate, the numerical inefficiencies noted above suffered by the particle filter can be acute in applications involving DSGE models. By eliminating these inefficiencies, the EIS filter offers a significant advance

in the empirical analysis of DSGE models.

2 Likelihood Evaluation in State-Space Representations

Let s_t denote a $d \times 1$ vector of latent state variables, y_t denote an $n \times 1$ vector of observable variables, and denote $\{y_s\}_{s=1}^{t}$ as Y_t. State-space representations consist of a state-transition density $f(s_t|s_{t-1})$ coupled with a measurement density $f(y_t|s_t)$.[1] Since this representation is recursive, the likelihood function $f(Y_T)$ factors sequentially as

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

where $f(y_1|Y_0) = f(y_1)$. Evaluation of the period-t likelihood $f(y_t|Y_{t-1})$ takes as input the period-$(t-1)$ filtering density $f(s_{t-1}|Y_{t-1})$ (or an approximation thereof), and requires joint integration in s_t and s_{t-1}, conditionally on Y_t:

$$f\left(y_t|Y_{t-1}\right) = \int\int f\left(y_t|s_t\right) f(s_t|s_{t-1}) f\left(s_{t-1}|Y_{t-1}\right) ds_{t-1} ds_t. \tag{2}$$

The recursive updating of filtering densities follows from Bayes' theorem, and is given by

$$f\left(s_t|Y_t\right) = \frac{f\left(y_t|s_t\right)}{f\left(y_t|Y_{t-1}\right)} \cdot \int f\left(s_t|s_{t-1}\right) f\left(s_{t-1}|Y_{t-1}\right) ds_{t-1}, \tag{3}$$

initialized by $f(s_0|Y_0) = f(s_0)$. The integral in (3) represents the period-t predictive density $f(s_t|Y_{t-1})$, which appears explicitly in Kalman filter (KF) derivations. The updating of filtering densities constitutes the cornerstone of all filtering algorithms. Likelihood evaluation is achieved by evaluating (2) and (3) sequentially from period 1 to T. In turn, filtering is achieved via the calculation of

$$E[m(s_t)|Y_t] = \int m(s_t) f\left(s_t|Y_t\right) ds_t \tag{4}$$

for any function $m(s_t)$, including s_t itself.

[1] This formulation incorporates the implicit conditional independence assumptions that $s_t \perp\!\!\!\perp Y_{t-1}|s_{t-1}$ and $y_t \perp\!\!\!\perp (s_{t-1}, Y_{t-1})|s_t$, which validate subsequent density operations. These assumptions can be relaxed trivially, e.g., as in dynamic control environments.

3 Particle Filters

3.1 General Principle: Sequential Importance Sampling

Here we present a brief summary of particle filters in order to provide context for the EIS filter (for a detailed overview, see Ristic et al., 2004). Particle filters (PFs) are Sequential Importance Sampling (SIS) or Resampling (SIR) algorithms wherein filtering densities are approximated sequentially using mixtures of Dirac measures ($\delta(0) = 1; \delta(x) = 0$ for $x \neq 0$):

$$\hat{f}\left(s_t|Y_t\right) = \sum_{i=1}^{N} \omega_t^i \cdot \delta\left(s_t - s_t^i\right), \tag{5}$$

where $\{s_t^i\}_{i=1}^N$ denotes a swarm of particles drawn from a period-t Importance Sampling (IS) density, and $\{\omega_t^i\}_{i=1}^N$ denotes the corresponding IS weights (normalized such that $\sum_{i=1}^N \omega_t^i = 1$). The swarm $\{s_t^i\}_{i=1}^N$ constitutes the discrete support over which $\hat{f}\left(s_t|Y_t\right)$ is defined. Once established, this support remains fixed; in this sense approximations constructed in subsequent periods are defined as conditional on the supports established in the current and past periods. This conditionality is common to all PFs.

In period t, PFs propagate inherited particles s_{t-1}^i with counterparts s_t^i drawn from a (sequential) IS density $g_t(s_t|s_{t-1}^i)$. The latter is meant to approximate the integrand in the PF-likelihood integral, which obtains by substituting the period-$(t-1)$ filtering density approximation in (5) – lagged by one period – into the likelihood integral in (2):

$$\hat{f}\left(y_t|Y_{t-1}\right) = \sum_{i=1}^{N} \left[\omega_{t-1}^i \int f\left(y_t|s_t\right) f\left(s_t|s_{t-1}^i\right) ds_t\right]. \tag{6}$$

Corresponding IS weights are given by

$$\tilde{\omega}_t^i = \omega_{t-1}^i \frac{f(y_t|s_t^i) f(s_t^i|s_{t-1}^i)}{g_t(s_t^i|s_{t-1}^i)}, \tag{7}$$

and the period-t filtering density approximation is given by (5) with the normalized weights

$$\omega_t^i = \frac{\tilde{\omega}_t^i}{\sum_{j=1}^N \tilde{\omega}_t^j}, \tag{8}$$

whose denominator also represents the period-t likelihood estimate

$$\hat{f}\left(y_t|Y_{t-1}\right) = \sum_{j=1}^{N} \tilde{\omega}_t^j. \tag{9}$$

With reference to (3), (7) and (8) constitute the updating step of the filtering density of SIS and SIR algorithms.

3.2 Selection of Importance Densities

3.2.1 The Optimal Choice

It follows from (7) that the optimal IS density for $s_t|s_{t-1}$ – that which minimizes the MC variance of the corresponding IS likelihood estimate conditionally on s_{t-1} – obtains from the reverse factorization (when computable) of the numerator:

$$g_t^*\left(s_t|s_{t-1}\right) = f\left(s_t|s_{t-1}, Y_t\right) = \frac{f(y_t|s_t)f(s_t|s_{t-1})}{f(y_t|s_{t-1}, Y_{t-1})}. \tag{10}$$

In this case the period-t weights are given by

$$\tilde{\omega}_t^i = \omega_{t-1}^i f\left(y_t|s_{t-1}^i, Y_{t-1}\right), \tag{11}$$

and do not depend on $\{s_t^i\}_{i=1}^N$. It follows that the MC variance of the likelihood estimate in (9) would be zero, *conditionally* on $\{s_{t-1}^i\}_{i=1}^N$; the sampler $g_t^*\left(s_t|s_{t-1}\right)$ is referenced as the conditionally optimal particle filter (OPF). In all but special cases, the reverse factorization in (10) is not tractable analytically, rendering $g_t^*\left(s_t|s_{t-1}\right)$ as unavailable for MC simulation; for examples of special cases, see Doucet et al. (2000, 2001b).

3.2.2 Suboptimal Choices

If the transition density $f(s_t|s_{t-1})$ is amenable to MC simulation, it provides an operational specification of the IS sampler $g_t(s_t|s_{t-1})$. In this case, $\tilde{\omega}_t^i$ in (7) simplifies to

$$\tilde{\omega}_t^i = \omega_{t-1}^i f\left(y_t|s_t^i\right). \tag{12}$$

The Bootstrap Particle Filter (BPF) of Gordon et al. (1993) is a leading example of a PF implemented using $f(s_t|s_{t-1})$ as an importance sampler. But while straightforward to implement, $f(s_t|s_{t-1})$ tends to be inefficient, particularly if its variance is large relative to that of the measurement density $f(y_t|s_t)$, and also given the presence of outliers. Due to this shortcoming, considerable efforts have gone towards the development of importance samplers that approximate the conditionally optimal density g_t^* in (10); see Ristic et al. (2004, Section 3.4.2) for references. One such method is the auxiliary particle filter, discussed next.

3.3 Auxiliary Particle Filters

As shown in (2), period-t likelihood evaluation requires integration in s_t *and* s_{t-1}. Yet within the PF framework, attempts at achieving efficient (IS) integration in s_{t-1} are constrained by the fact that the discrete support $\{s_{t-1}^i\}_{i=1}^N$ of the period-$(t-1)$ filtering density approximation is kept fixed in period t. Subject to this operational constraint, the Auxiliary Particle Filter (APF) of Pitt and Shephard (1999) is designed to relocate the particles $\{s_{t-1}^i\}_{i=1}^N$ in regions of high period-t likelihood. To do so, the likelihood integral in (6) is reinterpreted as a mixed integral in (s_t, i), where $i \in \{1, \ldots, N\}$ denotes the index of the period-$(t-1)$ particles and follows the multinomial distribution $MN(N, \{\omega_{t-1}^j\}_{j=1}^N)$. The corresponding IS target integrand is given by

$$\varphi\left(s_t, i\right) = \omega_{t-1}^i f\left(y_t|s_t\right) f\left(s_t|s_{t-1}^i\right). \tag{13}$$

In its simplest form, APF employs a mixed IS kernel obtained from (13) replacing $f(y_t|s_t)$ by its (order zero Taylor series) local approximation $f(y_t|\mu(s_{t-1}^i))$, where $\mu(s_{t-1}^i) = \mathrm{E}(s_t|s_{t-1}^i)$. The resulting kernel is given by

$$k_t\left(s_t, i\right) = \omega_{t-1}^i f\left(y_t|\mu\left(s_{t-1}^i\right)\right) f\left(s_t|s_{t-1}^i\right), \tag{14}$$

and its integrating constant is

$$D_t = \sum_{j=1}^N \omega_{t-1}^j f\left(y_t|\mu\left(s_{t-1}^i\right)\right). \tag{15}$$

Corresponding IS weights are given by

$$\omega_t\left(s_t, i\right) = D_t \cdot \frac{f(y_t|s_t, Y_{t-1})}{f(y_t|\mu(s_{t-1}^i), Y_{t-1})}. \tag{16}$$

Let $\{s_t^j, i^j\}_{j=1}^N$ denote N i.i.d. draws from the IS sampler k_t/D_t. The APF likelihood estimate is then given by

$$\hat{f}\left(y_t|Y_{t-1}\right) = \frac{1}{N}\sum_{j=1}^{N}\tilde{\omega}_t^j, \qquad \tilde{\omega}_t^j = \omega_t(s_t^j, i^j), \tag{17}$$

and the period-t filtering weights obtain from (8).

As noted by Ristic et al. (2004, Section 3.5.2), if the variance of the transition density is small, the APF will often be less sensitive to outliers than the BPF; though if it is large, use of the APF can actually degrade performance. See also Johansen and Doucet (2008, p. 1503) who show that "...whilst the adaptation may be beneficial at the time which it is performed it may have a negative influence on the variance at a later point."

Under special circumstances, higher-order Taylor series approximations of $\ln f(y_t|s_t)$ around $s_t = \mu(s_{t-1}^i)$ can produce more efficient versions of the APF. A (theoretical) conditionally optimal version of the APF, known as the Fully Adapted Particle Filter (FAPF), obtains by substituting the reverse factorization (10) into (13) and using the latter as mixed IS kernel, in which case the IS weights in (16) no longer depend on s_t and have zero MC variance conditionally in $\{s_{t-1}^i\}_{i=1}^N$.

3.4 Illustration of Performance Tradeoffs

As noted, the BPF and APF differ fundamentally by the way in which they incorporate period-t information in constructing IS densities. To illustrate how this difference yields tradeoffs in efficiency and sensitivity to outliers in a context relevant to the applications presented below, we demonstrate their relative performance in a linear Gaussian version of a Real Business Cycle (RBC) model. Since the reverse factorization in (10) is available in this case, we can also demonstrate the performance of the OPF and FAPF, which respectively represent optimal versions of the BPF and APF.

Regarding the model, let (y_t, z_t, k_t) denote output, total factor productivity, and capital (measured as logged deviations from steady state). The state vector is $s_t' = (z_t, k_t)$, and state transition

is characterized by

$$k_t = \ln(1-\alpha\beta) + z_{t-1} + \alpha k_{t-1}, \quad 0 < \alpha < 1, \tag{18}$$

$$z_t = \rho z_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{i.i.d.} N\left(0, \sigma_\epsilon^2\right), \quad 0 < \rho < 1. \tag{19}$$

The measurement equation is

$$y_t = z_t + \alpha k_t + u_t \quad u_t \sim \text{i.i.d.} N\left(0, \sigma_y^2\right). \tag{20}$$

For the sake of simplicity, we focus on the approximation of the period-1 likelihood $f(y_1)$, pursued with the initial filtering density $f(s_0)$ represented as the stationary density of s_0 (thereby abstracting from small-sample complications arising from the use of a PF discrete support for s_0). The standard algebra of linear Gaussian models yields analytical expressions for period-1 IS densities, and for MC variation coefficients (standard derivation/mean) of PF estimates for $f(y_1)$. The derivations are straightforward but tedious; they are omitted here but available at http://www.pitt.edu/~dejong/wp.htm.

Using the generic notation f for true (model) densities and g for IS samplers, the IS densities for the four PF schemes under consideration are given by:

$$\text{BPF}: \quad g(s_1, s_0|y_1) = f(s_0) f(s_1|s_0) \tag{21}$$

$$\text{OPF}: \quad g(s_1, s_0|y_1) = f(s_0) f(s_1|s_0, y_1) \tag{22}$$

$$\text{APF}: \quad g(s_0, s_1|y_1) = g(s_0|y_1) f(s_1|s_0), \quad \text{with} \tag{23}$$

$$g(s_0|y_1) \propto f(s_0) f(y_1|s_1)\Big|_{s_1=\mu(s_0)} \tag{24}$$

$$\text{FAPF (KF)}: \quad g(s_0, s_1|y_1) = f(s_0|y_1) f(s_1|s_0, y_1). \tag{25}$$

Table 1 reports exact results for the MC variation coefficients of the PF estimates for $N = 1$ as a function of the standardized difference between y_1 and its predictive mean in units of its predictive standard deviation (results are to be divided by $\sqrt{N}$ for $N > 1$). The parameter values used to obtain these results are set equal to the posterior means obtained from the RBC model in Section 5 below, with a critical difference: our analytical results indicate that the MC variance

of the APF likelihood estimate is unbounded for $\sigma_y^2 < 2\sigma_\epsilon^2$ (reflecting Ristic et al.'s message that the performance of the APF degrades as the variance of measurement errors becomes small relative to the variance of stochastic innovations to the state). Thus we had to raise the value of σ_y and considered two cases differing by the proximity of σ_y and $\sigma_\epsilon\sqrt{2}$.

Beginning with the comparison of the BPF and OPF, we find that the efficiency gains realized by replacing the sampler $f(s_1|s_0)$ by $f(s_1|s_0, y_1)$ (leaving $f(s_0)$ unchanged) are relatively modest, reducing MC variation coefficients by less than 20%. Next, the comparison between the BPF and APF indicates that if σ_y is not too close to $\sigma_\epsilon\sqrt{2}$, the APF is less efficient than BPF for low standardized values of y_1, but significantly more robust against outliers (Table 1.A). However, the performance of APF degrades rapidly as σ_y approaches $\sigma_\epsilon\sqrt{2}$ (Table 1.B). This illustrates that the APF is vulnerable to thin-tail issues resulting from the use of local IS approximations. Note that for this linear Gaussian example, implementation of the APF using a second-order Taylor series approximation of $\ln f(y_1|s_1)$ produces an exact (global) approximation to the IS target integrand, which is Gaussian in both y_1 and s_1. Therefore, the second-order APF coincides with FAPF, which is also the Kalman filter: its corresponding likelihood estimate has zero MC variance.

We draw two lessons from this example. First, efficiency gains can be achieved by treating the likelihood integral (2) as a joint integral in s_t and s_{t-1}, and constructing an IS density for the pair using period-t information. Second, it is important to replace local with global approximations in order to avoid thin-tail issues. These lessons motivate implementation of the EIS filter, which applies global EIS approximation techniques to the full integrand in (2).

4 The EIS filter

4.1 Unconditional Optimality

Returning to the likelihood integral (2), consider the (theoretical) reverse factorization of the full integrand

$$f\left(y_t|s_t\right) f\left(s_t|s_{t-1}\right) f\left(s_{t-1}|Y_{t-1}\right) = f\left(s_t, s_{t-1}|Y_t\right) f\left(y_t|Y_{t-1}\right). \tag{26}$$

If this factorization were tractable analytically, $f(s_t, s_{t-1}|Y_t)$ would constitute the unconditionally optimal IS sampler for (2): a single draw would produce an estimate of $f(y_t|Y_{t-1})$ with zero MC

variance. The period-t filtering density would obtain by marginalization with respect to s_{t-1}:

$$f\left(s_t|Y_t\right) = \int f\left(s_t, s_{t-1}|Y_t\right) ds_{t-1}. \tag{27}$$

As illustrated in Section 3.4, these densities obtain analytically for linear Gaussian models and constitute the KF. They are generally unavailable for more general models, yet provide a useful point of reference for the EIS filter which targets unconditional optimality via global approximations of $f(s_t, s_{t-1}|Y_t)$.

4.2 General Principle

The target density kernel to be approximated in (2) is given by

$$\varphi_t\left(s_t, s_{t-1}\right) = f\left(y_t|s_t\right) f\left(s_t|s_{t-1}\right) f\left(s_{t-1}|Y_{t-1}\right). \tag{28}$$

Our objective is to construct a joint sampler $g_t(s_t, s_{t-1})$ that approximates φ_t, employing the EIS principle outlined in Section 4.3. Corresponding (E)IS ratios and likelihood estimates are given by

$$\omega_t\left(s_t, s_{t-1}\right) = \frac{\varphi_t(s_t, s_{t-1})}{g_t(s_t, s_{t-1})}, \tag{29}$$

$$\hat{f}\left(y_t|Y_{t-1}\right) = \frac{1}{N}\sum_{i=1}^{N} \omega_t\left(s_t^i, s_{t-1}^i\right), \tag{30}$$

where $\{(s_t^i, s_{t-1}^i)\}_{i=1}^N$ denotes N i.i.d. draws from g_t.

To derive the period-t filtering density approximation, we partition the EIS sampler as

$$g_t\left(s_t, s_{t-1}\right) = g_t\left(s_t\right) g_t\left(s_{t-1}|s_t\right). \tag{31}$$

The period-t filtering density in (3) is then rewritten as

$$f\left(s_t|Y_t\right) = \frac{g_t(s_t)}{f(y_t|Y_{t-1})} \int \omega_t\left(s_t, s_{t-1}\right) g_t\left(s_{t-1}|s_t\right) ds_{t-1}. \tag{32}$$

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

$$\bar{\omega}_t(s_t) = \frac{1}{N}\sum_{i=1}^{N} \omega_t\left(s_t, s_{t-1}^i(s_t)\right), \tag{33}$$

where $\{s_{t-1}^i(s_t)\}_{i=1}^N$ denotes draws from $g_t(s_{t-1}|s_t)$. It follows from (30) and (32) that an EIS estimate of the period-t filtering density obtains as

$$\hat{f}\left(s_t|Y_t\right) = g_t(s_t)\tilde{\omega}_t(s_t), \qquad \text{with} \tag{34}$$

$$\tilde{\omega}_t(s_t) = \frac{\sum_{i=1}^N \omega_t\left(s_t, s_{t-1}^i(s_t)\right)}{\sum_{i=1}^N \omega_t(s_t^i, s_{t-1}^i)}. \tag{35}$$

Implementation of this generic principle entails use of the EIS procedure to construct the joint sampler $g_t(s_t, s_{t-1})$; and the numerical evaluation of $\tilde{\omega}_t(s_t)$ in period $(t+1)$ for the construction of the period-$(t+1)$ EIS sampler. These steps are described in the following two subsections.

4.3 EIS integration

Here we sketch the application of EIS to the target kernel φ_t in (28) assuming an operational approximation to the filtering density $f(s_{t-1}|Y_{t-1})$; see Richard and Zhang (2007, hereafter RZ) for a detailed presentation.

Let $\varphi(\lambda)$ with $\lambda = (s, s_{-1})$ denote the target, where the subscript t is omitted for the ease of notation (since EIS is applied independently in each period). Implementation of EIS begins with the pre-selection of a parametric class $K = \{k(\lambda; a); a \in A\}$ of auxiliary IS density kernels amenable to MC simulation. The corresponding IS densities and ratios are given by

$$g\left(\lambda|a\right) = \frac{k(\lambda; a)}{\chi(a)}, \quad \text{with} \quad \chi(a) = \int k\left(\lambda; a\right) d\lambda, \tag{36}$$

$$\omega\left(\lambda; a\right) = \frac{\varphi(\lambda)}{g(\lambda|a)}. \tag{37}$$

The objective of EIS is to select a parameterization $\hat{a} \in A$ that minimizes the MC variance

of the IS ratio ω over the full support of φ. Following RZ a near-optimal solution obtains as the solution to the minimization problem

$$(\hat{a}, \hat{c}) = \arg\min_{(a,c)} \int \left[\ln \varphi(\lambda) - c - \ln k(\lambda; a)\right]^2 g(\lambda|a) d\lambda, \tag{38}$$

where c denotes an intercept meant to calibrate the ratio $\ln(\varphi/k)$. Equation (38) represents a standard least-squares problem, except that the auxiliary sampling density g itself depends on a. An operational MC version, implemented (typically) using $R \ll N$ draws, is as follows:

Step $(\ell + 1)$: Given $\hat{a}_\ell$, draw intermediate values $\{\lambda_\ell^i\}_{i=1}^R$ from $g(\lambda|\hat{a}_\ell)$ and solve:

$$(\hat{a}_{\ell+1}, \hat{c}_{\ell+1}) = \arg\min_{(a,c)} \sum_{i=1}^{R} \left[\ln \varphi\left(\lambda_\ell^i\right) - c - \ln k\left(\lambda_\ell^i; a\right)\right]^2. \tag{39}$$

Before we discuss the pre-selection of K, three points bear mentioning. First, the selection of the initial value $\hat{a}_0$ can be based upon an analytically tractable local approximation to $\varphi(\lambda)$. In Section 5, we rely upon Extended Kalman Filter (EKF) approximations. Such initial values generally suffice to achieve rapid fixed-point convergence (say less than five iterations). Second and most importantly, the EIS fixed-point iterations aim at replacing an initial local approximation of $\varphi(\lambda)$ by a global one over its full support. As we shall illustrate in Section 5, it is precisely the replacement of local approximations by a global EIS approximation that generates (often dramatic) efficiency gains resulting from EIS filtering. Third, to achieve rapid fixed-point convergence $\{\lambda_\ell^i\}_{i=1}^N$ must be obtained by transformation of a set of Common Random Numbers (CRNs) drawn from a canonical distribution, i.e. one that does not depend on a (e.g., uniform $(0, 1)$, or standardized normal draws if g is Gaussian). Note that if EIS is used to produce MC estimates of the likelihood function itself, then a new EIS sampler must be produced for each new parameter value. In such a case, reliance upon CRNs ensures continuity of the estimated likelihood function.

Continuing with convergence, it is important to distinguish between fixed-point convergence of the sequence $\{\hat{a}_\ell\}$, and MC convergence of the subsequent likelihood estimate as N increases. Fixed-point convergence is assessed by monitoring the $\{\hat{a}_\ell\}$ sequence and implementing a stopping rule, using e.g. a relative-change threshold. The threshold need not be tight as small changes in

$\hat{a}$ typically have minimal impact on the accuracy of the corresponding EIS estimates (by far the largest efficiency gains obtain through the first few iterations). Although there is no guarantee nor formal proof that the sequence $\{\hat{a}_\ell\}$ converges for a given pair (φ, k) in general, we have found repeatedly that it never fails to converge short of a major mismatch between the target φ and the kernel k, in which case failure to converge serves as a signal that the class K needs to be adjusted. Actually, fixed-point convergence is not critical, since all that matters is that the resulting EIS sampler delivers numerically efficient estimates of the likelihood integral. Regarding accuracy, or more precisely, convergence in distribution of the EIS MC estimate, this follows from standard regularity conditions for IS, as discussed e.g. by Geweke (1989). Final efficiency can be assessed using summary statistics relative to the dispersion of the EIS ratios $\omega(\lambda^i; \hat{a})$.

Finally, as noted above, the pre-selection of K is problem-specific. However, once EIS has been programmed for a particular class K, changes in model specification only require adjusting the dependent variable $\ln \varphi$ in the auxiliary regression (39). If K belongs to the exponential family of distributions, then there exists a natural parametrization in the sense of Lehmann (1986, Section 2.7) which follows from the representation of $\ln k(\lambda; a)$ as a linear function of a fixed-dimensional vector of sufficient statistics. If, e.g., $k(\lambda; a)$ denotes a density kernel for an n-dimensional normal $\lambda \sim N_n(\mu, \Omega)$, then λ together with the lower triangular part of $\lambda \cdot \lambda'$ constitutes a sufficient statistic of dimension $n + n(n+1)/2$ and the minimization problem in (39) takes the form of a linear (OLS) regression of $\ln \varphi(\lambda)$ on that sufficient statistic. Once $\hat{a}$ obtains, it is a trivial matter to transform it back into a Cholesky recursive factorization of $g(\lambda; \hat{a})$ as needed to produce the draws of λ based on CRNs. Details for the applications considered here are provided below.

4.4 Filtering Weights

The transfer of filtering density approximations across periods is the weak link of any filtering technique. Under PFs transfer takes the operational form of a swarm of particles $\{s_{t-1}^i\}_{i=1}^N$ together with period-$(t-1)$ weights, at the cost of restricting subsequent period-t IS as conditional on $\{s_{t-1}^i\}_{i=1}^N$. Under EIS filtering transfer consists of an analytical period-$(t-1)$ marginal sampler $g_{t-1}(s_{t-1})$ together with a weight functional $\tilde{\omega}_{t-1}(s_{t-1})$, as defined in (35) lagged by one period. The challenge in this case is that $\tilde{\omega}_{t-1}(s_{t-1})$ does not have an analytical form, nor can it be computed from (35) for any new IS draws of s_{t-1} in period t. There are two possible solutions to this challenge.

The first applies when the period-$(t-1)$ EIS sampler fits its target sufficiently well that the IS weights $\omega_{t-1}(s_{t-1}^i, s_{t-2}^i)$ exhibit relatively small MC coefficients of variation. In such cases, the solution is to set all weights equal to their arithmetic mean, which amounts to setting $\tilde{\omega}_{t-1}(s_{t-1})$ identically equal to one. Following (34), the resulting constant-weight approximation of $f(s_{t-1}|Y_{t-1})$ is given by $g_{t-1}(s_{t-1})$ itself. This approximation introduces a bias in corresponding EIS estimates but this can be a small price to pay if it results in large MC variance reduction. Note that the use of biased estimates has become increasingly common in situations where unbiased estimates exhibit large sampling variances. See, e.g., Judge et al. (1985, Chapter 3 and 22) for comprehensive discussions of biased estimation. The applications presented in Section 5 employ constant-weight approximations, which are validated by comparisons of the Root Mean Squared Errors (RMSE) associated with BPF and EIS filter estimates of log-likelihood functions (each of which, as discussed in Section 4.6, require adjustments for bias induced by the log transformation).

The second solution entails the use of numerical techniques for approximating $\tilde{\omega}_{t-1}(s_{t-1})$ (e.g., interpolation). This would be required in applications wherein differences in log-likelihood approximations across filters are greater than is attributable to MC error. The use of more flexible weighting procedures is discussed in Section 6.

4.5 Singular Transitions

Transition identities such as (18) are common across a wide range of state-space models, including the DSGE models analyzed in Section 5. Obviously, Dirac transitions are not revised by new observations, thus if $f(s_t|s_{t-1})$ is singular, then so too will be $f(s_t|s_{t-1}, Y_t)$, $f(s_t, s_{t-1}|Y_t)$ and $f(s_{t-1}|s_t, Y_t)$. As illustrated in Section 3.4, identities raise no computational issues for the Kalman filter, which trivially accommodates singular covariance matrices. Identities are also easily accommodated under PFs that employ the transition density $f(s_t|s_{t-1})$ for propagation. This being said, note that even under the Kalman filter, identities play a key implicit role. For the example discussed in Section 3.4, it is easy to verify that the density $f(s_{t-1}|s_t, Y_t)$ incorporates the identity

$$k_{t-1} + \frac{1}{\alpha} z_{t-1} = \frac{1}{\alpha} \left[k_t - \ln\left(1 - \alpha\beta\right) \right], \tag{40}$$

which obtains form (18) by inversion with respect to k_{t-1}.

Within EIS (as for any IS procedure) it is critical that the IS sampler shares a common support with its target density (applying a nonsingular IS sampler to a singular target would produce zero IS ratios with probability 1). Therefore, let s_t partition into $s_t = (p_t, q_t)$ such that the transition consists of a set of identities of the form

$$q_t = \phi\left(p_t, s_{t-1}\right), \tag{41}$$

in combination with a nonsingular transition density $f(p_t|s_{t-1})$ [2]. Such identities effectively reduce the dimensionality of the likelihood integral (2), which is rewritten as

$$f\left(y_t|Y_{t-1}\right) = \int\int f\left(y_t|s_t\right)\Big|_{q_t=\phi(p_t,s_{t-1})} \cdot f\left(p_t|s_{t-1}\right) f\left(s_{t-1}|Y_{t-1}\right) dp_t ds_{t-1}. \tag{42}$$

Note that computation of the period-t filtering density in (32) requires integration in $s_{t-1}|s_t, Y_t$, which is singular. This is the same issue noted in the context of (40). The remedy consists of the application of a transformation of variables from (p_t, s_{t-1}) into (s_t, p_{t-1}) to the integral in (42). Assuming that ϕ is one-to-one (on the relevant range), we denote its inverse and Jacobian by

$$q_{t-1} = \psi\left(s_t, p_{t-1}\right), \tag{43}$$

$$J\left(s_t, p_{t-1}\right) = \left|\left|\frac{\partial}{\partial q_t'}\psi\left(s_t, p_{t-1}\right)\right|\right|, \tag{44}$$

where $||\cdot||$ denotes the absolute value of a determinant. The likelihood integral in (42) is then rewritten as

$$f\left(y_t|Y_{t-1}\right) = \int\int f\left(y_t|s_t\right) J\left(s_t, p_{t-1}\right)\left[f\left(p_t|s_{t-1}\right) f\left(s_{t-1}|Y_{t-1}\right)\right]\Big|_{q_{t-1}=\psi(s_t,p_{t-1})} ds_t dp_{t-1}, \tag{45}$$

and the EIS target $\varphi_t(s_t, p_{t-1})$ is the integrand in (45). Application of the EIS procedure proceeds exactly as described above, and produces a sampler of the form $g_t(s_t, p_{t-1}) = g_t(s_t)g_t(p_{t-1}|s_t)$. The filtering density is now given by

$$f\left(s_t|Y_t\right) = \frac{g_t(s_t)}{f(y_t|Y_{t-1})}\int \omega_t\left(s_t, p_{t-1}\right) g_t\left(p_{t-1}|s_t\right) dp_{t-1}, \qquad \text{with} \tag{46}$$

[2] An alternative factorization of the form $q_t = \phi(s_{t-1})$ and $f(p_t|q_t, s_{t-1})$ would be treated in the same way.

$$\omega_t\left(s_t, p_{t-1}\right) = \frac{\varphi_t(s_t, p_{t-1})}{g_t(s_t, p_{t-1})}. \tag{47}$$

4.6 Bias Correction for Log-Likelihood Estimates

As proved, e.g., by Del Moral (2004) and Chopin (2004), BPF estimates of the likelihood function are unbiased under fairly weak assumptions. In contrast, EISF estimates of the likelihood obtained under the constant-weight approximation are biased if the IS weights in (29) or (47) are not constant. However, classical likelihood-based inference is based on estimates of the log-likelihood function, and since the log transformation is concave, it induces downward bias for log-likelihood estimates associated with any filtering technique. Moreover, the magnitude of these induced biases is an increasing function of the MC standard deviation of the likelihood estimates. Thus in order to meaningfully compare numerical efficiency (on the basis of Root Mean Squared Errors, hereafter RMSEs) across filtering techniques (as we shall do below), we must first correct for biases induced by the log transformation.

We use two alternative bias-correction methods: an asymptotic correction based upon a Central Limit Theorem (CLT); and a finite-sample correction calculated via a Monte Carlo method. The asymptotic correction employs the Taylor series expansion

$$\ln x = (x-1) - \frac{1}{2}(x-1)^2 - \left[\sum_{i=3}^{\infty} (-1)^{i-1} \frac{(x-1)^i}{i}\right], \quad |x-1| < \epsilon. \tag{48}$$

Let L and $\hat{L}_N$ denote the true (unknown) likelihood and its filtered estimate, respectively. Under a CLT for $\hat{L}_N/L$

$$\frac{\hat{L}_N}{L} \xrightarrow{d} N\left(1, \frac{\sigma^2}{N}\right), \tag{49}$$

it follows from (48) that

$$\ln\left(\frac{\hat{L}_N}{L}\right) \xrightarrow{d} N\left(-\frac{\sigma^2}{2N}, \frac{\sigma^2}{N}\right). \tag{50}$$

Thus an asymptotic first-order bias correction obtains by adding $\sigma^2/(2N)$ to the log of the likelihood estimates.

In order to validate this correction it is important to verify that N is sufficiently large to ensure

that the CLT in (50) applies and that σ^2/N is sufficiently small so that higher-order terms derived from the slow-converging series in (48) can be ignored. This requires producing i.i.d. draws of the log-likelihood estimates, say $\{\ln \hat{L}_{N,j}\}_{j=1}^{R}$ under different seeds and computing, for example, a Jarque-Bera test statistic for normality. Moreover, these draws can be used to compute a finite-sample bias correction as

$$BC = \frac{1}{R}\sum_{j=1}^{R} \ln\left(\frac{\hat{L}_{N,j}}{L_0}\right) - \ln\left(\frac{\bar{L}_R}{L_0}\right), \tag{51}$$

where $\bar{L}_R$ denotes the arithmetic mean of $\{\hat{L}_{N,j}\}_{j=1}^{R}$, and L_0 an arbitrary estimate of the (unknown) likelihood used to control for overflows.

5 Application to DSGE Models

As noted, Fernandez-Villaverde and Rubio-Ramirez (2005, 2009) showed that approximation errors associated with linear representations of DSGE models can impart significant errors in corresponding likelihood representations. As a remedy, they demonstrated use of the BPF developed by Gordon, et al. (1993) for achieving likelihood evaluation for non-linear model representations. Here we demonstrate the implementation of the EIS filter using two workhorse models. The first is the standard two-state real business cycle (RBC) model; the second is a small-open-economy (SOE) model patterned after those considered, e.g., by Mendoza (1991) and Schmitt-Grohe and Uribe (2003), but extended to include six state variables. Both models share a common statistical structure. The EIS implementation for that common structure is presented in Section 5.1. The specific models are then described in Section 5.2, and results are presented in Section 5.3.

5.1 A Generic DSGE Model

We use $f_N^n(x|\mu,\Omega)$ to denote the density of a normal random variable $x \in \mathbb{R}^n$ with mean vector μ and covariance matrix Ω. Let n, d, d_p, d_q denote the respective dimensions of y_t, s_t, p_t and q_t. The class of DSGE models we consider is characterized by the following state-space densities. The measurement density is given by

$$f\left(y_t|s_t\right) = f_N^n\left(y_t|\mu\left(s_t\right),\Omega_y\right), \tag{52}$$

where $\mu(s_t)$ denotes a non-linear vector function of s_t. The transition density consists of a non-singular density for $p_t|s_{t-1}$ and a non-linear transition identity for $q_t|s_{t-1}$:

$$f\left(p_t|s_{t-1}\right) = f_N^{d_p}\left(p_t|Rs_{t-1},\Omega_p\right), \tag{53}$$

$$q_t = \phi\left(s_{t-1}\right). \tag{54}$$

As in Section 4.5, ϕ is invertible, with inverse and Jacobian denoted by

$$q_{t-1} = \psi\left(q_t, p_{t-1}\right), \qquad J\left(q_t, p_{t-1}\right) = \left|\left|\frac{\partial}{\partial q_t'}\psi\left(q_t, p_{t-1}\right)\right|\right|. \tag{55}$$

We applied EIS using a class K of unconstrained Gaussian kernels for $\lambda_t = (s_t, p_{t-1})$, together with a constant-weight filtering density approximation, whereby (46) simplifies to

$$f\left(s_t|Y_t\right) = g_t(s_t) =_{(\text{say})} f_N^d\left(s_t|b_t, P_t\right). \tag{56}$$

According to (45), the period-t EIS target is given by

$$\begin{aligned}\varphi_t\left(s_t, p_{t-1}\right) &= f_N^n\left(y_t|\mu(s_t),\Omega_y\right) J\left(q_t, p_{t-1}\right) \\ &\quad \times \left[f_N^{d_p}\left(p_t|Rs_{t-1},\Omega_p\right) f_N^d\left(s_{t-1}|b_{t-1}, P_{t-1}\right)\right]\Big|_{q_{t-1}=\psi(q_t,p_{t-1})}.\end{aligned} \tag{57}$$

As discussed in Section 4.3, this selection of K produces an operational version of EIS whereby the auxiliary regressions in (39) are linear in λ_t and the lower triangular part of $\lambda_t\lambda_t'$. The only additional requirement consists of the selection of an initial sampler $g_t(\lambda_t|\hat{a}_t^0)$ for λ_t. As described next, we shall use for that purpose a Gaussian Extended Kalman Filter (EKF) density obtained by linearization of the non-linear components $\mu(s_t)$ and $\phi(s_{t-1})$ in (52) and (54). (For an example of the use of EKF as a final IS sampler, see Durbin and Koopman, 1997.) Selecting an EKF initial sampler also offers the advantage that it serves to illustrate the efficiency gains produced by the use of a global EIS Gaussian approximation in place of the local EKF approximation.

The construction of the initial EKF sampler follows the standard algebra of the Kalman filter;

this allows us to deal with the (linearized) transition identity through singular Gaussian covariance matrices. The period-$(t-1)$ filtering density is approximated by the constant-weight approximation in (56), lagged by one period. The EKF approximation of the measurement density (52) is

$$\hat{f}\left(y_t|s_t\right) = f_N^n\left(y_t|\kappa_t + H_t s_t, \Sigma\right), \tag{58}$$

where κ_t and H_t obtain by local linear approximation of $\mu(s_t)$ – specific details are provided below for each model. The EKF approximation of the singular transition density obtains by combining the transition density (53) with a local linear approximation of $\phi(s_{t-1})$:

$$\hat{f}\left(s_t|s_{t-1}\right) = f_N^d\left(s_t|c_t + F_t s_{t-1}, \Psi\right), \tag{59}$$

$$\hat{\phi}\left(s_{t-1}\right) = A_t s_{t-1} + d_t, \tag{60}$$

where

$$c_t = \begin{pmatrix} 0 \\ d_t \end{pmatrix}, \qquad F_t = \begin{pmatrix} R \\ A_t \end{pmatrix}, \qquad \Psi = \begin{pmatrix} \Omega_p & 0 \\ 0 & 0 \end{pmatrix}. \tag{61}$$

Following standard (E)KF algebra, we combine the approximations of the measurement density in (58), the transition density in (59), and the lagged filtering density in (56) into a joint (singular) normal density for $(y_t, s_t, s_{t-1}|Y_{t-1})$ with mean vector m_t and covariance matrix V_t:

$$m_t = \begin{pmatrix} \mu_{t|t-1} \\ \hline b_{t|t-1} \\ b_{t-1} \end{pmatrix}, \qquad V_t = \left(\begin{array}{c|cc} \Omega_{t|t-1} & H_t P_{t|t-1} & H_t F_t P_{t-1} \\ \hline P_{t|t-1} H_t' & P_{t|t-1} & F_t P_{t-1} \\ P_{t-1} F_t' H_t' & P_{t-1} F_t' & P_{t-1} \end{array}\right), \tag{62}$$

with

$$b_{t|t-1} = c_t + F_t b_{t-1}, \qquad P_{t|t-1} = \Psi + F_t P_{t-1} F_t' \tag{63}$$

$$\mu_{t|t-1} = \kappa_t + H_t b_{t|t-1}, \quad \Omega_{t|t-1} = \Sigma + H_t P_{t|t-1} H_t'. \tag{64}$$

The initial sampler $g_t(s_t, p_{t-1}|\hat{a}_t^0)$ then obtains via marginalization with respect to q_{t-1}, which

requires deleting the last d_q rows and columns in (m_t, V_t), and conditioning upon y_t. Since the resulting initial sampler is to be used to produce the initial draws $\{\lambda_0^i\}_{i=1}^R$ in (39), we require its Cholesky factorization. See Appendix A for details.

In summary the EIS algorithm (Gaussian kernels, constant weights) consists of six steps:

Step 1: Construct the EKF mean vector m_t and covariance matrix V_t as in (62).

Step 2: Delete from (m_t, V_t) the last d_q rows and columns corresponding to q_{t-1}.

Step 3: Condition upon y_t to obtain the Cholesky factorization of the initial sampler $g_t(s_t, p_{t-1}|\hat{a}_t^0)$, as described in Appendix A.

Step 4: Use CRN draws from the initial sampler as input, and compute a fully iterated global EIS kernel from the sequence of linear least-squares regressions in (39), under the natural parametrization introduced in Section 4.3, for a total of

$$n_k = (d + d_p) + \frac{1}{2}(d + d_p)(d + d_p + 1) \tag{65}$$

regressors plus an intercept. See Appendix B for details regarding this regression.

Step 5: Using CRN draws from $g_t(s_t, p_{t-1}|\hat{a}_t)$, where $\hat{a}_t$ denotes the EIS solution to (39), compute the IS ratios in (47) and the corresponding period-t likelihood estimate in (30).

Step 6: Marginalize $g_t(s_t, p_{t-1}|\hat{a}_t)$ with respect to p_{t-1} to obtain the constant-weight approximation of the period-t filtering density.

As a rule of thumb, we have found that starting from the initial EKF sampler and using a number of CRN draws R of the order of 3 to 5 times n_k secures fast convergence to a near optimal value $\hat{a}_t$, with typically 3 to 4 iterations for the convergence criterion

$$\left|\left|\frac{\hat{a}_{(\ell+1)}}{\hat{a}_{(\ell)}} - 1\right|\right|_2 < \varepsilon, \tag{66}$$

with ε of the order of 10^{-4}.

In conclusion, relative to EKF the EIS filter adds an intermediate EIS iterated linear least-squares auxiliary regression problem to replace the initial EKF local approximation of the target (57) by a global approximation thereof. As we shall see in Section 5.3, this additional step produces dramatic efficiency gains relative to EKF and BPF.

5.2 Example Models

As noted, implementation of the EIS filter is demonstrated using two specific DSGE models. The first is the RBC model used by Fernandez-Villaverde and Rubio-Ramirez (2005) to demonstrate implementation of the BPF. The model consists of a representative household that seeks to maximize the expected discounted stream of utility derived from consumption c and leisure l:

$$\max_{c_t, l_t} \quad U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(c_t^{\varphi} l_t^{1-\varphi}\right)^{1-\phi}}{1-\phi}, \tag{67}$$

where (β, ϕ, φ) represent the household's subjective discount factor, degree of relative risk aversion, and the relative importance assigned to c_t and l_t in determining period-t utility.

The household divides its available time per period (normalized to unity) between labor n_t and leisure. Labor combines with physical capital k_t and a stochastic productivity term z_t to produce a single good x_t, which may be consumed or invested (we use x in place of the usual representation for output – y – to avoid confusion with our use of y as representing the observable variables of a generic state-space model). Investment i_t combines with undepreciated capital to yield k_{t+1}, thus the opportunity cost of period-t consumption is period-$(t+1)$ capital. Collectively, the constraints faced by the household are given by

$$x_t = z_t k_t^{\alpha} n_t^{1-\alpha}, \tag{68}$$

$$1 = n_t + l_t, \tag{69}$$

$$x_t = c_t + i_t, \tag{70}$$

$$k_{t+1} = i_t + (1-\delta) k_t, \tag{71}$$

$$z_t = z_* e^{\omega_t}, \qquad \omega_t = \rho \omega_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d.} N(0, \sigma_\varepsilon^2), \tag{72}$$

where (α, δ, ρ) represent capital's share of output, the depreciation rate of capital, and the persistence of innovations to total factor productivity.

Optimal household behavior is represented as policy functions for (x_t, i_t, n_t) in terms of the state (z_t, k_t). The corresponding policy functions for (c_t, l_t) follow from identities (69) and (70). Policy functions are expressed as Chebyshev polynomials in the state variables (z_t, k_t), constructed using

the projection method described in DeJong and Dave (2007, Ch. 10.5.2). Given the form of (72), it is convenient to represent the state variables (z_t, k_t) as logged deviations from their steady state (z_*, k_*). Measurements for (x_t, i_t, n_t) are assumed to differ from their policy function values by i.i.d. Gaussian measurement errors. Thus the model is of the form given by (52) to (55) with $n = 3$, $d_p = d_q = 1$, and (with all variables presented as logged deviations from steady state):

- $s_t' = (p_t, q_t) = (z_t, k_t)$,
- $y_t' = (x_t, i_t, n_t)$,
- $\mu(s_t)$ denoting the policy functions for y_t,
- Ω_y the diagonal covariance matrix of the measurement errors, with diagonal elements $(\sigma_x^2, \sigma_i^2, \sigma_n^2)$,
- $R = (\rho, 0), \qquad \Omega_p = \sigma_\varepsilon^2$,
- $\phi(s_{t-1})$ denoting the degenerate transition obtained from identity (71) via substitution of the corresponding policy function, and application of the logged deviation transformation to (s_t, s_{t-1}). (Details regarding the construction and inversion of $\phi(s_{t-1})$ are presented in Appendix C.1.)

The second example is a model of a small open economy (SOE), patterned after those considered, e.g., by Mendoza (1991) and Schmitt-Grohe and Uribe (2003). The model consists of a representative household that seeks to maximize

$$U = E_0 \sum_{t=0}^{\infty} \theta_t \frac{\left[c_t - \varphi_t \omega^{-1} n_t^{\omega}\right]^{1-\gamma} - 1}{1-\gamma}, \qquad \omega > 0, \qquad \gamma \geqslant 0, \tag{73}$$

where φ_t is a preference shock that affects the disutility of labor effort (introduced, e.g., following Smets and Wouters, 2003). Following Uzawa (1968), the discount factor θ_t is endogenous and obeys

$$\begin{aligned} \theta_{t+1} &= \beta(\tilde{c}_t, \tilde{n}_t)\,\theta_t, \qquad \theta_0 = 1, \\ \beta(\tilde{c}_t, \tilde{n}_t) &= \left[1 + \tilde{c}_t - \omega^{-1}\tilde{n}_t^{\omega}\right]^{-\psi}, \qquad \psi > 0, \end{aligned} \tag{74}$$

where $(\tilde{c}_t, \tilde{n}_t)$ denote average per capita consumption and hours worked. The household takes these

as given; they equal (c_t, n_t) in equilibrium. The household's constraints are collectively

$$
\begin{aligned}
x_t &= A_t k_t^{\alpha} n_t^{1-a}, && (75)\\
d_{t+1} &= (1+r_t)\, d_t - x_t + c_t + i_t + \frac{\phi}{2}\left(k_{t+1} - k_t\right)^2, && (76)\\
k_{t+1} &= v_t^{-1} i_t + (1-\delta)\, k_t, && (77)\\
\ln A_{t+1} &= \rho_A \ln A_t + \varepsilon_{At+1}, \qquad \varepsilon_{At} \sim \text{i.i.d.} N(0, \sigma_A^2), && (78)\\
\ln r_{t+1} &= (1-\rho_r) \ln r_* + \rho_r \ln r_t + \varepsilon_{rt+1}, \qquad \varepsilon_{rt} \sim \text{i.i.d.} N(0, \sigma_r^2), && (79)\\
\ln v_{t+1} &= \rho_v \ln v_t + \varepsilon_{vt+1}, \qquad \varepsilon_{vt} \sim \text{i.i.d.} N(0, \sigma_v^2), && (80)\\
\ln \varphi_{t+1} &= \rho_\varphi \ln \varphi_t + \varepsilon_{\varphi t+1}, \qquad \varepsilon_{\varphi t} \sim \text{i.i.d.} N(0, \sigma_\varphi^2), && (81)
\end{aligned}
$$

where relative to the RBC model, the new variables are d_t, the stock of foreign debt, r_t, the exogenous interest rate at which domestic residents can borrow in international markets, v_t, an investment-specific productivity shock, and the preference shock φ_t.

The state variables of the model are $(d_t, k_t, A_t, r_t, v_t, \varphi_t)$; the controls are (x_t, c_t, i_t, n_t). In this application we achieve model approximation following Schmitt-Grohe and Uribe (2004). Specifically, we represent all variables as quadratic functions of the states, again expressed as logged deviations from steady states. Measurements for the logged controls are centered on their respective logged policy functions. The resulting model is of the form given by (52) to (55) with $n = 4$, $d_p = 4$, $d_q = 2$, and (with all variables once again presented as logged deviations from steady state):

- $p_t' = (A_t, r_t, v_t, \varphi_t)$,
- $q_t' = (d_t, k_t)$,
- $y_t' = (x_t, c_t, i_t, n_t)$,
- $\mu(s_t)$ denoting the logged policy functions for y_t,
- Ω_y a diagonal covariance matrix of the measurement errors, with diagonal elements $(\sigma_x^2, \sigma_c^2, \sigma_i^2, \sigma_n^2)$,
- $R = (R_p, 0)$, with R_p denoting a diagonal matrix, with diagonal elements $(\rho_A, \rho_r, \rho_v, \rho_\varphi)$,
- Ω_p a 4×4 diagonal covariance matrix, with diagonal elements $\left(\sigma_A^2, \sigma_r^2, \sigma_v^2, \sigma_\varphi^2\right)$,
- $\phi(s_{t-1})$ denoting the bivariate degenerate transition equations obtained directly from the Schmitt-Grohe/Uribe solution algorithm. (Details regarding the (recursive) inversion of $\phi(s_{t-1})$ are presented in Appendix C.2.)

The application of EIS to this six-state model accounting for the degenerate transition requires 10-dimensional EIS in (s_t, p_{t-1}) as described in Section 4.3.

5.3 Results

Here we present three MC experiments designed to illustrate: (i) Potential biases in EIS likelihood evaluation, and the relative numerical efficiency of EIS, BPF and EKF; (ii) Data robustness analysis; and (iii) numerical continuity and differentiability of EIS log-likelihood approximations. Each experiment involves four data sets, two for each model. For a given model, one data set consists of artificial data generated from an assumed set of parameter (true) values; the second consists of actual data that align with their theoretical counterparts. Using these actual data and as discussed further below, we computed posterior modes and standard deviations for both models. Parameter values were then set equal to these posterior modes to conduct our experiments under the real data sets. Both data sets are available at http://www.pitt.edu/~dejong/wp.htm. Additional model-specific details follow.

RBC Model: The artificial data set consists of 100 realizations of $\{x_t, i_t, n_t\}$, and was constructed by Fernandez-Villaverde and Rubio-Ramirez (2005, hereafter FVRR) using the parameter values reported in row 1 of the RBC panel in Table 2. The actual data set consists of 184 quarterly US observations on $\{x_t, i_t, n_t\}$, where i_t is defined as the sum of the consumption of durable goods and real gross fixed investment, x_t is the sum of i_t and the consumption of non-durable goods, and n_t is total non-farm employment. The data are quarterly, seasonally adjusted and span 1964:I – 2009:IV; output and investment are measured in real per-capita terms, and all series are HP filtered. This data set represents an updated version of the actual data set used by FVRR. The source is the Federal Reserve Bank of St. Louis. Posterior modes for the updated data sets were computed under the uniform priors chosen by FVRR. Prior intervals, posterior modes and (approximate) posterior standard deviations are reproduced in rows 2 to 5 of the RBC panel in Table 2.

SOE Model: The artificial data set consists of 100 realizations of $\{x_t, c_t, i_t, n_t\}$. It was generated using three complementary sets of parameter values: (i) aside from the parameters characterizing sources of uncertainty, we used the parameter values calibrated by Schmitt-Grohe and Uribe (2003, hereafter SGU) to match summary statistics on Canadian data as reported by Mendoza (1991); (ii) Additional parameters characterizing new sources of uncertainty in the model

were chosen as those that minimized the sum of squared differences between Mendoza's summary statistics (excluding trade balance) and the statistics implied by the model (standard deviations for (x_t, c_t, i_t, n_t), first-order serial correlations, and contemporaneous correlations with output); (iii) Finally, the standard derivations of all measurement errors were set at 0.5%. The corresponding parameter values are reported in row 1 (2 parts) of the SOE panel in Table 2.

The actual data set consists of 132 quarterly Canadian observations on $\{x_t, c_t, i_t, n_t\}$ spanning 1976:I – 2008:IV, and is an updated version of the data set used by Mendoza. Here, x_t is defined as GDP, c_t as personal expenditures on consumer goods and services, i_t as business gross fixed capital formation, and n_t is an index of man hours worked by paid workers. Once again, the data are quarterly, seasonally adjusted, and detrended using the HP filter. The source of the SOE data set is Statistics Canada.

Parameters for the actual data set were set at their posterior modes based upon independent prior normal distributions for each parameter. Aside from parameters that characterize stochastic uncertainty, prior means were set at the values specified by SGU, and prior standard deviations were set to reflect non-trivial uncertainty over these specifications. (Note that SGU priors for δ and r_* are appropriate for annual data and were transformed into priors for quarterly observations.) The prior means and standard derivations for the AR coefficients were set at 0.8 and 0.2, respectively. With two exceptions along ill-behaved dimensions (σ_r and σ_i), the prior means for the σ's were set at 0.5% with standard deviation 0.5%. The likelihood function implies a strong negative correlation between σ_r and ρ_r, thus the prior mean and standard deviation of ρ_r (0.1% and 0.22%, respectively) were set so that the posterior mode of ρ_r remained close to its prior mean. The posterior mode of σ_i was difficult to pin down, so its prior mean was set at 0.5% (like its counterparts) while its standard deviation was set at 0.05% to force its posterior mode to remain close to its prior mean. Prior means, prior standard deviations, posterior modes and (approximate) posterior standard deviations an reproduced in rows 2 to 5 of the SOE panel in Table 2.

Each data set poses a distinct challenge to efficient likelihood evaluation. In the RBC artificial data set, the standard deviations of the measurement errors (σ_x,σ_i,σ_n) are small relative to σ_ε, which as we have noted in Sections 3.2 and 3.4 can lead to increased sensitivity to outliers, likely to be further exacerbated by sample impoverishment. In the RBC actual data set, output and investment feature two sudden drops of more than 1.5 standard deviations between the third and fourth quarter

of 1974, and the first and second quarter of 1980. Such abrupt changes pose a challenge for filters that implement discrete and fixed-support importance samplers, as it is difficult for such samplers to generate proposals of states that appear as outliers.

In the SOE data sets, one challenge is the relatively high dimensionality of the state space (six versus two in the RBC model). A second is the additional non-linearities featured in the model (relative to the RBC model): e.g., the capital-adjustment cost term $\phi(k_{t+1} - k_t)^2/2$ appearing in (76). As opposed to the applications involving the RBC model, variances of measurement errors are closely comparable across data sets in this case. Instead, differences in data sets stem primarily from differences in the volatility and persistence of the model's structural shocks. In particular, with the model parameterization associated with the artificial data set calibrated to annual data, and the parameterization associated with the real data set estimated using quarterly observations, structural shocks are far less persistent, and generally more volatile, in the former case. The upshot is that in working with the actual data, the state variables are relatively easy to track, and in general the construction of likelihood approximations is less problematic.

Experiment 1: The first experiment is designed to analyze the MC properties of log-likelihood approximations generated by three filters: the BPF; the EKF used as the initial IS sampler for EIS iterations (EKFIS); and the EIS filter (EISF).[3] Setting $N = 1,000,000$ (1M), we generated 100 i.i.d. BPF log-likelihood estimates using 100 different seeds. Then setting $N = R = 100$ for the RBC model and $N = R = 200$ for the SOE model, we generated 100 i.i.d. EKFIS and EISF log-likelihood estimates using 100 different sets of CRNs. This was done for each of the four data sets listed in Table 2 (hereafter RBC-art, RBC-act, SOE-art, SOE-act). The MC means and numerical standard errors (NSEs) of log-likelihood estimates obtained for all four data sets using all three filters are reported in Table 3.A. In addition, the log-likelihood approximations obtained using the BPF and EISF are illustrated in Figure 1.

Note first from Table 3.A that the NSEs associated with the EKFIS are two to four orders of magnitude larger than those associated with the EISF. This illustrates that the most critical step in the EISF implementation is the replacement of an initial local (EKF) approximation to the target φ_t with a global (EIS) approximation. In light of this result, the remainder of our investigation

[3]We also attempted to analyze the APF. However, the tightly distributed measurement errors associated with the model parameterizations we examined prevented us from doing so: we frequently encountered numerical overflows for APF weights, symptomatic of the thin-tail problem discussed in Section 3.4.

focuses on the MC sampling properties of the BPF and EISF.

With the exception of RBC-art, NSEs associated with the BPF are between 10 and 60 times higher than those associated with the EISF; for RBC-art, NSEs differ by three orders of magnitude. The large difference in the latter case stems from the model's tightly distributed measurement errors, illustrating that efficiency gains from adaption are most significant in this situation. Regarding the MC means of log-likelihoods, these differ by at most 0.4 across the BPF and EISF.

As discussed in Section 4.6, a meaningful efficiency comparison of EISF and BPF log-likelihood estimates requires a correction for biases induced by the log-transformation. The NSEs reported in Table 3.A are estimates of $\sigma N^{-1/2}$ in equations (74) and (75), thus an asymptotic first-order bias correction obtains by adding $\sigma^2/(2N)$ to the corresponding means. Bias-corrected means for BPF and EISF are reported in columns 1 and 3 of Table 3.B. Note that the small NSEs that obtain for the EISF log-likelihood estimates imply that the biases induced by the log transformation are negligible in all four cases. As for BPF(1M), with the exception of SOE-art, NSEs above 0.5 raise concerns about the validity of asymptotic bias corrections. Moreover the RBC-BPF(1M) replications in Figure 1 exhibit left skewness and some kurtosis as confirmed by significant Jarque-Bera test statistics for normality (24.1 and 18.5 against a $\chi^2_{(2)}$ critical value of 6.9). Therefore, we also computed finite-sample bias corrections as in (51). They turn out to be closely aligned with their asymptotic counterparts. Finite-sample bias corrected means for BPF(1M) are reported in column 2 of Table 3.B.

Additional statistics reported in Table 3.B are the average variation coefficients (VARCOF) for the EIS ratios in (29), and t-statistics for tests of the equality of the EISF and BPF (asymptotic and finite-sample) bias-corrected means. The statistics indicate that potential biases induced by the constant-weight approximation are negligible. In the one case (SOE-act) in which the null of no bias is rejected at the 5% level, the difference between the two means (0.1341) represents less than 0.008% of the BPF log-likelihood estimate.

Returning to the consideration of numerical efficiency, we supplement the information conveyed by the NSEs reported in Table 3.A with Root Mean Squared Errors (RMSEs), reported in Table 3.C. These were constructed using the bias-corrected log-likelihood approximations of the BPF as pseudo-true values. Additional statistics provided in Table 3.C are CPU time (secs) per log-likelihood evaluation (calculated on a 2.9 GHz desktop computer using MATLAB); and the numbers

N (1M units) and CPU times (hours) required for the BPF to match the RMSEs associated with the EISF, assuming $N^{-1/2}$ convergence.[4]

Remembering that EISF estimates are based on 100 (RBC) and 200 (SOE) draws, these comparisons represent substantial efficiency gains, illustrating the key tradeoff between bias and variance reduction that has received considerable attention in the econometrics literature; e.g., see Judge et al. (1985 Chapters 3 and 22) for a comprehensive exposition of biased estimation.

Experiment 2: The second experiment is designed to assess whether the foregoing results are somehow particular to the specific data sets upon which they are based, and to provide context for the small NSEs associated with the EISF reported in Table 3.A. Note that the estimates produced in Experiment 1 are de facto functions of the data set Y and of a set U of CRNs, say $\ln \hat{L}(Y, U)$. In computing NSEs, $\ln \hat{L}$ is treated as a function of U for fixed Y. Here we compute Statistical Standard Error (SSEs), wherein $\ln \hat{L}$ is treated as a function of Y for a fixed U.

Specifically, here we compute SSEs for each of the four mean EISF estimates reported in Table 3.A, column 3. To do so, we first fix the sets of CRNs $\{U_i\}_{i=1}^{100}$ used for the 100 EISF individual estimates. Next for each model (parameterized as in Table 2), we generate 100 artificial data sets $\{Y_{j,k}\}_{j=1}^{100}$, where $k = 1$ to 4 indexes the models. For each data set, we replicate Experiment 1, and record the corresponding means $\{\text{MEAN}_{j,k}\}$ and NSEs $\{\text{NSE}_{j,k}\}$, thus producing 100 statistical replications of the EISF statistics reported in columns 3 and 4 of Table 3.A. Table 4 reports the following statistics for $k = 1$ to 4:

SSE_k:	Standard deviation of the 100 $\{\text{MEAN}_{j,k}\}_{j=1}^{100}$;
Mean NSE_k:	Arithmetic mean of the 100 $\{\text{NSE}_{j,k}\}_{j=1}^{100}$;
Std. dev. NSE_k:	Standard deviation of the 100 $\{\text{NSE}_{j,k}\}_{j=1}^{100}$.

Except for RBC-act, the NSEs reported in Table 3.A, column 4 all lie within 2 statistical standard deviations of their means in Table 4. The difference for RBC-act is 5.5 statistical standard derivations, and appears to result from the large outliers for 1974:IV and 1980:II noted above in the original data set: the probability of similarly large outliers in the replicated data sets is virtually

[4] Although we avoided the use of loops in programming BPF, differences in computing times are likely to be compressed using Fortran or C in place of MATLAB.

zero. Overall, the general message of Table 3 is not sensitive to specific features of the data sets upon which they are based.

Comparing NSEs and SSEs, Table 4 also indicates that statistical uncertainty swamps numerical inaccuracy. This is critical, because it indicates that numerical inaccuracy does not stand as a barrier in drawing likelihood-based inferences when using the EIS filter, a result further underscored in our final experiment.

Experiment 3: Here we analyze the numerical continuity and differentiability of log-likelihood approximations generated by BPF and EISF. We do so using SOE-act, which as reported in Table 3.A is the model for which BPF performs best. The experiment consists of producing log-likelihood sections for each parameter. Sections obtain by varying one parameter at a time along a 500-point grid, with all other parameters held fixed at their posterior modes. Thus this experiment requires $500 \times 19 = 9,500$ log-likelihood evaluations. Based on the CPU times reported in Table 3.C, this experiment requires approximately 5.75 hours under EISF(200); under BPF(1M) it would require approximately 2,470 hours, which is prohibitive. Thus we ran it under BPF with $N = 150,000$ (150K), which produces a log-likelihood function of 1717.9816 at the posterior mode with an NSE of 0.7607. This reduced total CPU time to 210 hours. Note that the NSE under BPF(150K) is about 4.5 times larger than that under BPF(1M).

Figure 2 illustrates all 19 log-likelihood sections for SOE. For both filters, log-likelihood values are indicated by dots; we do not connect the dots because doing so would obfuscate EISF contours. Clearly, the BPF log-likelihood contours are jagged; this result also holds under BPF(1M), except that the vertical scale is reduced by a factor of 4.5, ranging approximately from 1717.5 to 1718.5.

The discontinuity of the surface approximations generated by BPF poses challenges to the use of both classical and Bayesian techniques for obtaining parameter estimates. In addition, this discontinuity renders as problematic the use of derivative-based methods for computing covariance matrices associated with a given set of parameter estimates. For example, letting μ denote the parameter vector targeted for estimation, one means of estimating its posterior covariance matrix Σ_μ is via the Laplace approximation

$$\hat{\Sigma}_\mu^{-1} = \Sigma_p^{-1} - \left. \frac{\partial^2 \ln \hat{L}_N}{\partial \mu \partial \mu'} \right|_{\mu=\hat{\mu}}, \tag{82}$$

where Σ_p denotes the prior covariance matrix and $\hat{\mu}$ the posterior mode of μ. The posterior modes reported in Table 2 were produced under EISF(100/200) using a BFGS Quasi-Newton method with a cubic line search procedure (using the BFGS formula to update the Hessian). Corresponding posterior standard derivations were obtained by direct application of (82). In contrast, numerical approximations of $\hat{\Sigma}_\mu$ failed to be positive-definite when constructed under BPF.

6 Future Research

The simple (Gaussian constant-weight) EISF implementation demonstrated here produced dramatic efficiency gains in applications to workhorse DSGE models. The fact that these models (as is typical in the literature) feature Gaussian state-transition innovations and measurement errors makes them relatively amenable EIS implementation. However, the narrow measurement errors associated with their parameterization, coupled with the presence of outliers, renders them as challenging environments for achieving numerically accurate log-likelihood approximation. Moreover, the non-linearities embodied in these models are non-negligible, as highlighted by the inefficiency of the EKFIS initial sampler. This latter result underscores the findings of Fernandez-Villaverde and Rubio-Ramirez (2005, 2009), which caution that (local) linear representations of DSGE models can impart significant errors in likelihood estimation.

The efficiency of the EIS filter in these settings motivates ongoing research designed to increase flexibility on two critical counts: (i) allowing for the implementation of more flexible classes K of importance kernels; and (ii) designing interpolation methods to relax the constant-weight approximation of $\tilde{\omega}_t(s_t)$ in (35).

We are currently progressing on these two fronts, and one-dimensional pilot applications have yielded promising results. For the selection of K, we are investigating the use of mixtures of Gaussian kernels, which are highly flexible. The EIS auxiliary regressions in (39) become non-linear in a when using mixtures, but can be solved via Gauss-Newton algorithms implemented using analytical derivatives under an appropriate parametrization of the mixture. A pilot application to a univariate stochastic volatility model using 8-term mixtures has delivered high degrees of numerical accuracy. As for weight interpolation, we are exploring a variety of (potentially high-dimensional) interpolation techniques (surveyed, e.g., by Pflughaupt, 1993). A one-dimensional pilot application

to the approximation of a bimodal two-term mixture by a Gaussian kernel using a simpler Dirac interpolation technique has already proven successful.

A third frontier of exploration concerns the construction of the initial sampler. Since the initial sampler merely serves to generate a set of MC draws for computing the initial EIS iteration in (39), it need not itself be a member of pre-selected class K of auxiliary samplers. Here we used the EKF as an operational initial sampler (which happened to be in K), but more generally, a variety of local approximation techniques are available as alternatives (e.g., Laplace approximations, as in Tierney and Kadane, 1986). A potentially simpler approach consists of combining period-$(t-1)$ EIS draws $\{s_{t-1}^i\}_{i=1}^N$ with draws of s_t propagated through the transition density (as in BPF or APF).

7 Conclusion

In conducting likelihood analyses of state-space representations, particle-based filters offer two key advantages: they are easy to implement, and they produce unbiased likelihood estimates (under fairly weak conditions). However, they can be prone to numerical inefficiency, particularly in applications involving narrowly distributed measurement equations, and given the presence of outliers. Moreover, refinements designed to deliver improvements in numerical efficiency are constrained at best to generate conditionally adapted importance samplers, where conditionality in period t is with respect to the discrete density used to represent the period-$(t-1)$ filtering density. In addition, refinements themselves can be prone to efficiency problems, as conditional adaptation in early periods can have a negative impact on numerical efficiency in subsequent periods.

Here we have presented a filtering algorithm that targets unconditional optimality. The algorithm features two critical elements. First, in generating period-t approximations, it implements continuous rather than discrete approximations of filtering densities, thus enabling the pursuit of unconditional adaption with respect to (s_t, s_{t-1}). Second, adaption is achieved via implementation of the EIS algorithm, which produces global rather than local approximations of targeted integrands. As we have demonstrated, the efficiency of the resulting EIS filter owes considerably to this component of the algorithm.

Implementation of the EIS filter is relatively involved in comparison with particle-based filters. It relies upon three auxiliary steps in period t: (i) The construction of an initial sampler for (s_{t-1}, s_t);

(ii) Iterated auxiliary regressions in (39) in order to transform the initial sampler into a (globally) efficient one within a pre-selected class K; and (iii) the construction of a continuous approximation to the period-t filtering density. These three steps can be programmed as self-contained procedures. The particular EIS filter implementation used in the present paper relies upon three key simplifications. (i) The initial sampler obtains by application of the extended Kalman filter principle; (ii) The class K consists of Gaussian samplers so that the auxiliary EIS regressions in (39) are linear under the natural parametrization associated with Gaussian kernels; (iii) The (constant-weight) filtering density approximation is defined as the EIS marginal density of s_t. Through the implementation details provided in the paper, including annotated codes available at http://www.pitt.edu/~dejong/wp.htm, we have sought to reduce barriers to entry regarding its current implementation.

Fernandez-Villaverde and Rubio-Ramirez (2005, 2009) have demonstrated the importance of the preservation of non-linearities in the context of DSGE models. This finding is underscored here by the finding that the extended Kalman filter (used as our initial sampler) is numerically inefficient relative to the global adaption facilitated via EIS. This motivates our current research agenda, as outlined in Section 6, wherein we are seeking to develop EIS procedures based on more flexible families of distributions than that used here. The ultimate goal is to facilitate EIS implementations using highly flexible samplers that will prove efficient in applications involving even the most challenging of targeted integrands.

Appendix A: Cholesky factorization of the initial sampler $g_t(s_t, p_{t-1}|\hat{a}_t^0)$

Deleting time subscripts for the ease of notation, let $(\tilde{m}, \tilde{V})$ denote the mean vector and covariance matrix obtained by deleting the least d_q rows and columns from (m_t, V_t) in (62). Next, let $x' = (s', p'_{-1})$, and partition $(\tilde{m}, \tilde{V})$ conformably with (y, x) into

$$\tilde{m} = \begin{pmatrix} \tilde{m}_y \\ \tilde{m}_x \end{pmatrix}, \qquad \tilde{V} = \begin{pmatrix} \tilde{V}_{yy} & \tilde{V}_{yx} \\ \tilde{V}_{xy} & \tilde{V}_{xx} \end{pmatrix}. \tag{83}$$

It follows that

$$E\left(x|y\right) = \tilde{m}_{x\,\cdot\,y} + \tilde{\Delta}_{xy} y, \qquad \mathrm{Cov}\left(x|y\right) = \tilde{V}_{xx\,\cdot\,y}, \tag{84}$$

with

$$\tilde{\Delta}_{xy} = \tilde{V}_{xy}\tilde{V}_{yy}^{-1}, \qquad \tilde{m}_{x\cdot y} = \tilde{m}_x - \tilde{\Delta}_{xy}\tilde{m}_y, \qquad \tilde{V}_{xx\cdot y} = \tilde{V}_{xx} - \tilde{V}_{xy}\tilde{V}_{yy}^{-1}\tilde{V}_{yx}. \tag{85}$$

Next, let L denote the lower triangular Cholesky factorization of $\tilde{V}$, partitioned conformably

$$\tilde{V} = LL', \qquad L = \begin{pmatrix} L_{yy} & 0 \\ L_{xy} & L_{yy} \end{pmatrix}. \tag{86}$$

It immediately follows that

$$\tilde{\Delta}_{xy} = L_{xy}L_{yy}^{-1}, \qquad \tilde{V}_{xx\cdot y} = L_{yy}L_{yy}', \tag{87}$$

thus CRN draws of $x|y$ obtain as follows

$$x_i = \left(\tilde{m}_{x\cdot y} + \tilde{\Delta}_{xy}y\right) + L_{yy}u_i, \qquad u_i \sim N\left(0, I_{d+d_p}\right). \tag{88}$$

Appendix B: EIS Optimization for Multivariate Gaussian Kernels

Denote the targeted integrand as $\varphi(\lambda)$, with $\lambda \in \mathbb{R}^n$. Let K denote a class of Gaussian density kernels with mean vector m and covariance matrix H^{-1}. Under this initial parametrization, log kernels in K (ignoring all additive constants) are given by

$$\ln k(\lambda;\cdot) = -\frac{1}{2}\left(\lambda' H\lambda - 2\lambda' Hm\right). \tag{89}$$

This immediately suggests introducing the auxiliary parametrization

$$a' = \left(d', -\mathrm{vec}_*'(H)\right), \tag{90}$$

where $d = Hm$ and $\mathrm{vec}_*(H)$ denotes the column vector expansion of the lower triangular part of H. Accounting for symmetry, the transformation from (H, m) into a is one-to-one (positive definiteness of H is not an issue for the applications in Section 5). It follows from (89) that the corresponding

vector of sufficient statistics is given by

$$z' = T'(\lambda) = (\lambda', \frac{1}{2}\lambda_1^2, \lambda_2\lambda_1, \frac{1}{2}\lambda_2^2, \ldots, \lambda_n\lambda_1, \ldots, \lambda_n\lambda_{n-1}, \frac{1}{2}\lambda_n^2). \tag{91}$$

The EIS auxiliary regression in (39) is then linear in a and its solution obtains by OLS of $y = \ln \varphi(\lambda)$ on $(1, z')$.

Appendix C: Construction and Inversion of the Degenerate Transitions $\phi(s_{t-1})$

Here we characterize the inversion of (54), repeated here for convenience:

$$q_t = \phi\left(p_{t-1}, q_{t-1}\right).$$

Recall that the goal of inversion is to obtain

$$q_{t-1} = \psi\left(q_t, p_{t-1}\right), \qquad J\left(q_t, p_{t-1}\right) = \left|\left|\frac{\partial}{\partial q_t'}\psi\left(q_t, p_{t-1}\right)\right|\right|.$$

C.1 RBC Model

Under the RBC model, q_t specializes to $q_t = \ln k_t/k_*$, and in light of (71) the specific form of (54) is

$$e^{\ln k_t/k_*} = i\left(\ln k_t/k_*\,,\, \ln z_t/z_*\right) + (1-\delta)\, e^{\ln k_{t-1}/k_*}, \tag{92}$$

where the policy function $i\left(\ln k_t/k_*\,,\, \ln z_t/z_*\right)$ is a Chebyshev polynomial. We achieve inversion in $\ln k_{t-1}/k_*$ via a projection method wherein we postulate a third-order polynomial of the form

$$\ln k_{t-1}/k_* = \psi\left(\ln k_t/k_*\,,\, \ln z_{t-1}/z_*\right), \tag{93}$$

and specify the parameters of this polynomial such that

$$e^{\ln k_t/k_*} - i\left(\ln k_t/k_*\,,\, \ln z_t/z_*\right) + (1-\delta)\, e^{\psi(\ln k_t/k_*\,,\, \ln z_{t-1}/z_*)} = 0 \tag{94}$$

holds. Given the optimized specification

$$\ln k_{t-1}/k_* = \psi^* \left(\ln k_t/k_* \, , \, \ln z_{t-1}/z_* \right) , \tag{95}$$

the Jacobian

$$J \left(k_t, z_{t-1} \right) = \left\| \frac{\partial}{\partial k_t} \psi \left(k_t, z_{t-1} \right) \right\| \tag{96}$$

obtains analytically.

C.2 SOE Model

Under the SOE model, q_t specializes to $q_t' = (d_t, k_t)$, where here for ease of notation d_t and k_t are represented as logged deviations from steady state values. Since the model is solved using the second-order approximation technique of Schmitt-Grohe and Uribe (2004), (54) is quadratic in its arguments. Moreover, it turns out that for all parameterizations of the model we considered, the coefficient associated with d_{t-1} that appears in the identity for k_t is of the order 10^{-8}, and thus is safely set to zero. The upshot is that the quadratic system to be inverted in this case is triangular in q_{t-1}.

Let $\left(s_{t-1}^1\right)' = \left(k_{t-1}, p_{t-1}'\right)$; then the identity for k_t is given by

$$k_t = C_k + L_k s_{t-1}^1 + \frac{1}{2} s_{t-1}^{1\prime} Q_k s_{t-1}^1. \tag{97}$$

Partitioning L_k and Q_k conformably with $\left(k_{t-1}, p_{t-1}'\right)$ as

$$Q_k = \begin{bmatrix} Q_k^{11} & Q_k^{12} \\ Q_k^{21} & Q_k^{22} \end{bmatrix}, \qquad L_k = \begin{bmatrix} L_k^1 \\ L_k^2 \end{bmatrix}, \tag{98}$$

we note that $Q_k^{11} > 0$ for all parameterizations under consideration. Thus the inversion of (97) is given by the solution

$$k_{t-1}^* = \frac{-b_k + \sqrt{b_k^2 - 4a_k c_k}}{2a_k}, \tag{99}$$

with

$$a_k = \frac{1}{2}Q_k^{11}, \tag{100}$$

$$b_k = L_k^1 + Q_k^{12}p_{t-1}, \tag{101}$$

$$c_k = C_k + L_k^2 p_{t-1} + \frac{1}{2}p_{t-1}' Q_k^{22} p_{t-1} - k_t. \tag{102}$$

Next, replacing k_{t-1} by k_{t-1}^* in the identity for d_t, we obtain

$$d_t = C_d + L_d s_{t-1} + \frac{1}{2}s_{t-1}' Q_d s_{t-1}. \tag{103}$$

Inversion in d_{t-1} yields the solution

$$d_{t-1}^* = \frac{-b_d + \sqrt{b_d^2 - 4a_d c_d}}{2a_d}, \tag{104}$$

with

$$Q_d = \begin{bmatrix} Q_d^{11} & Q_d^{12} \\ Q_d^{21} & Q_d^{22} \end{bmatrix}, \qquad L_d = \begin{bmatrix} L_d^1 \\ L_d^2 \end{bmatrix}, \tag{105}$$

and

$$a_d = \frac{1}{2}Q_d^{11}, \qquad (>0), \tag{106}$$

$$b_d = L_d^2 + Q_d^{12}s_{t-1}^1, \tag{107}$$

$$c_d = C_d + L_d^2 s_{t-1}^1 + \frac{1}{2}s_{t-1}^{1\prime} Q_k^{22} s_{t-1}^1 - d_t. \tag{108}$$

Note that the solutions $\left(k_{t-1}^*, d_{t-1}^*\right)$ in (99) and (104) are in terms of the largest roots of their corresponding quadratic forms, since k_t is monotone and increasing in k_{t-1}, and d_t is monotone and increasing in d_{t-1}.

Finally, the Jacobian of this triangular inversion is given by

$$J\left(q_t, p_{t-1}\right) = \left(b_k^2 + 4a_k c_k\right)^{-\frac{1}{2}} \cdot \left(b_d^2 + 4a_d c_d\right)^{-\frac{1}{2}}. \tag{109}$$

References

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

[2] Chopin, N., 2004, "Central Limit Theorems for Sequential Monte Carlo and its Application to Bayesian Inference", *The Annals of Statistics, 32, 2385-2411.*

[3] DeJong, D.N. with C. Dave, 2007, *Structural Macroeconometrics.* Princeton: Princeton University Press.

[4] DeJong, D.N., H. Dharmarajan, R. Liesenfeld, and J.-F. Richard, 2010, "Efficient Filtering in State-Space Representations", University of Pittsburgh; working paper.

[5] DeJong, D.N., B.F. Ingram, and C.H. Whiteman, 2000, "A Bayesian Approach to Dynamic Macroeconomics", *Journal of Econometrics,* 98, 203-233.

[6] Del Moral, P., 2004, *Feynman-Kac Formulae Genealogicial and Interacting Particle Systems with Applications.* New York: Springer-Verlag.

[7] Doucet, A., S. Goodsill, and C. Andrieu 2000, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering", *Statistics and Computing*, 10, 197-208.

[8] Doucet, A., N. de Freitas and N. Gordon, 2001a, *Sequential Monte Carlo Methods in Practice.* New York: Springer.

[9] Doucet, A., N. Gordon, and V. Krishnamurthy, 2001b, "Particle Filters for State Estimation of Jump Markov Linear Systems", *IEEE Trans. Signal Processing*, 49, 613-624.

[10] Durbin, J., and S.J. Koopman, 1997, "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models", *Biometrika*, 84, 669-684.

[11] Fernandez-Villaverde, J. and J.F. Rubio-Ramirez, 2005, "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood", *Journal of Applied Econometrics* 20, 891-910.

[12] Fernandez-Villaverde, J. and J.F. Rubio-Ramirez, 2009, "Estimating Macroeconomic Models: A Likelihood Approach", *Review of Economic Studies,* 74, 1059-1087.

[13] Geweke, J., 1989, "Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica,* 57, 1317-1339.

[14] Geweke, J., 2007, "Monte Carlo Simulation and Numerical Integration", In Amman, H.M., Knedrik, D.A., and J. Rust (Eds.), *Handbook of Computational Economics*, Vol. 1. New York: North Holland.

[15] Gordon, N.J., D.J. Salmond and A.F.M. Smith, 1993, "A Novel Approach to Non-Linear and Non-Gaussian Bayesian State Estimation", *IEEE Proceedings F,* 140, 107-113.

[16] Handschin, J., 1970, "Monte Carlo Techniques for Prediction and Filtering of Non-Linear Stochastic Processes", *Automatica,* 6, 555-563.

[17] Handschin, J. and D. Mayne, 1969, "Monte Carlo Techniques to Estimate the Cponditional Expectation in Multi-Stage Non-Linear Filtering", *International Journal of Control,* 9, 547-559.

[18] Johansen, A. M. and A Doucet, 2008, A Note on Auxiliary Particle Filters, *Statistics and Probability Letters*, 78, 1498-1504.

[19] Judge, G.G., W.E. Griffiths, R. Carter Hill, H. Lütkepohl, and T.-C. Lee, 1985, *The Theory and Practice of Econometrics.* New York: John Wiley and Sons.

[20] Lehmann. E. L., 1986, *Testing Statistical Hypotheses.* New York: John Wiley & Sons.

[21] Mendoza, E., 1991, "Real Business Cycles in a Small-Open Economy", *American Economic Review,* 81, 797-818.

[22] Pitt, M.K. and N. Shephard, 1999, "Filtering via Simulation: Auxiliary Particle Filters", *Journal of the American Statistical Association*, 94, 590-599.

[23] Pflughaupt, R., 1993, "High-Dimensional Linear Data Interpolation", University of California, Berkely, technical report No. UCB/CSD-93-745.

[24] Richard, J.-F. and W. Zhang, 2007, "Efficient High-Dimensional Monte Carlo Importance Sampling", *Journal of Econometrics*, 141, 1385-1411.

[25] Ristic, B., S. Arulampalam, and N. Gordon, 2004, *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Boston: Artech House.

[26] Sargent, T.J., 1989, "Two Models of Measurements and the Investment Accelerator", *Journal of Political Economy*, 97, 251-287.

[27] Schmitt-Grohe, S. and M. Uribe, 2003, "Closing Small Open Economy Models", *Journal of International Economics,* 61, 163-185.

[28] Schmitt-Grohe, S. and M. Uribe, 2004, "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function", *Journal of Economic Dynamics and Control,* 28, 755-775.

[29] Smets, F. and R. Wouters, 2003, "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area", *Journal of the European Economic Association,* 1, 1123-1175.

[30] Tierney, L. and J.B. Kadane, 1986, "Accurate Approximations for Posterior Moments and Marginal Densities", *Journal of the American Statistical Association*, 81, 82-86.

[31] Uzawa, H., 1968, "Time Preference, the Consumption Function and Optimum Asset Holdings", In Wolfe, J.N. (Ed.), *Value, Capital and Growth: Papers in Honor of Sir John Hicks.* Edinburgh: The University of Edinburgh Press, 485-504.

Table 1. MC Variation Coefficients for RBC First-Period PF Likelihood Estimates ($N = 1$)

Standardized y_1	BPF	OPF	APF	FAPF & KF
1.A: $\alpha = 0.36$, $\beta = 0.99$, $\rho = 0.98$, $\sigma_y = 0.007$, $\sigma_\epsilon = 0.0004$				
0	1.51	1.36	1.72	0.0
0.6	1.70	1.54	1.75	0.0
1.2	2.36	2.17	1.83	0.0
1.8	3.85	3.55	1,97	0.0
2.4	7.29	6.72	2.17	0.0
3.0	16.2	14.8	2.45	0.0
1.B: $\alpha = 0.36$, $\beta = 0.99$, $\rho = 0.98$, $\sigma_y = 0.0057$, $\sigma_\epsilon = 0.0004$				
0	1.72	1.50	2.83	0.0
0.6	1.93	1.70	3.06	0.0
1.2	2.65	2.36	3.87	0.0
1.8	4.30	3.89	5.61	0.0
2.4	8.14	7.28	9.34	0.0
3.0	18.2	16.2	17.8	0.0

Table 2. Parameter Values

RBC Model

	α	β	ϕ	φ	δ	ρ	σ_ε	σ_x	σ_i	σ_n
Art. Data	0.4	0.99	2	0.357	0.01961	0.95	0.007	1.58e-04	8.66e-4	0.0011
Prior Max.	0	0.75	0	0	0	0	0	0	0	0
Prior Min.	1	1	100	1	0.05	1	0.1	0.1	0.1	0.1
Post. Mode	0.3561	0.9938	3.3631	0.2006	0.0109	0.9842	0.0053	0.0060	0.0007	0.0017
Post. S.D.	1.4e-04	1.9e-05	1.9e-02	4.4e-03	2.5e-05	3.7e-03	4.3e-04	3.8e-04	4.0e-04	9.0e-05

SOE Model

	γ	ω	ψ	α	ϕ	r_*	δ	ρ_A	σ_A
Art. Data	2	1.455	0.11135	0.32	0.028	0.04	0.1	0.53	0.0089
Prior Mean	2	1.455	0.11	0.32	0.028	0.007	0.025	0.8	0.005
Prior S.D.	1	0.2	0.001	0.05	0.01	0.025	0.025	0.2	0.005
Post. Mode	2.49	1.33	0.11	0.23	0.039	0.02	0.02	0.82	0.0019
Post. S.D.	0.0086	0.0213	0.0059	0.0047	0.0133	0.0010	0.0031	0.0177	0.0003

	ρ_r	σ_r	ρ_v	σ_v	ρ_φ	σ_φ	σ_x	σ_c	σ_i	σ_n
Art. Data	0.37	0.001	0.89	0.001	0.3	0.0152	0.005	0.005	0.005	0.005
Prior Mean	0.8	0.0022	0.8	0.005	0.8	0.005	0.005	0.005	0.005	0.005
Prior S.D.	0.2	0.0005	0.2	0.005	0.2	0.005	0.005	0.005	0.0005	0.005
Post. Mode	0.79	0.0022	0.87	0.001	0.86	0.0031	0.0038	0.0065	0.0046	0.0058
Post. S.D.	0.1099	0.0129	0.01329	0.0002	0.0145	0.0004	0.0006	0.0006	0.0010	0.0005

Table 3.A. MC Means and Standard Deviations of Log-Likelihood Estimates

	BPF(1M)		EISF(100/200)		EKFIS(100/200)	
	Mean	NSE	Mean	NSE	Mean	NSE
RBC art	1299.9492	0.5287	1300.0674	0.0005	1283.8614	4.0279
RBC act	2305.2687	0.9139	2305.6589	0.0151	2305.2988	0.3486
SOE art	1293.8399	0.7134	1294.1197	0.0229	1253.0159	4.6041
SOE act	1718.1814	0.1699	1718.3298	0.0165	1696.6785	3.7189

Table 3.B. Bias Corrected Means of Log-Likelihood Estimates and Auxiliary Statistics

	BPF(1M) asympt.	BFF(1M) fin.-sample	EISF(100/200) asympt.	VARCOF	t-stat. asympt.	t-stat. fin.-sample
RBC art	1300.0890	1300.0716	1300.0674	0.000051	0.4078	0.0794
RBC act	2305.6863	2305.5804	2305.6590	0.00063	0.2986	- 0.8588
SOE art	1294.0944	1294.0713	1294.1200	0.0043	- 0.3572	- 0.6781
SOE act	1718.1958	1718.1958	1718.3299	0.0029	- 7.8961	- 7.8961

Table 3.C. Relative Numerical Efficiencies

	RMSE		CPU (secs)		BPF-EISF equivalent	
	BPF(1M)	EISF(100/200)	BPF(1M)	EISF(100/200)	N(1M)	CPU(hours)
RBC art	0.5287	0.0216	375	0.55	599	62.4
RBC act	0.9139	0.0312	695	1.34	858	165.6
SOE art	0.7134	0.0307	678	0.65	540	101.7
SOE act	0.1699	0.0211	936	2.18	65	16.9

Table 4. EISF for Repeated Samples

		EISF-NSE	
	SSE	Mean NSE	Std. dev. NSE
RBC art	19.8978	4.9e-4	1.2e-4
RBC act	1.1483	0.0113	6.9e-4
SOE art	17.6041	0.0417	0.0365
SOE act	15.6710	0.0134	0.0028

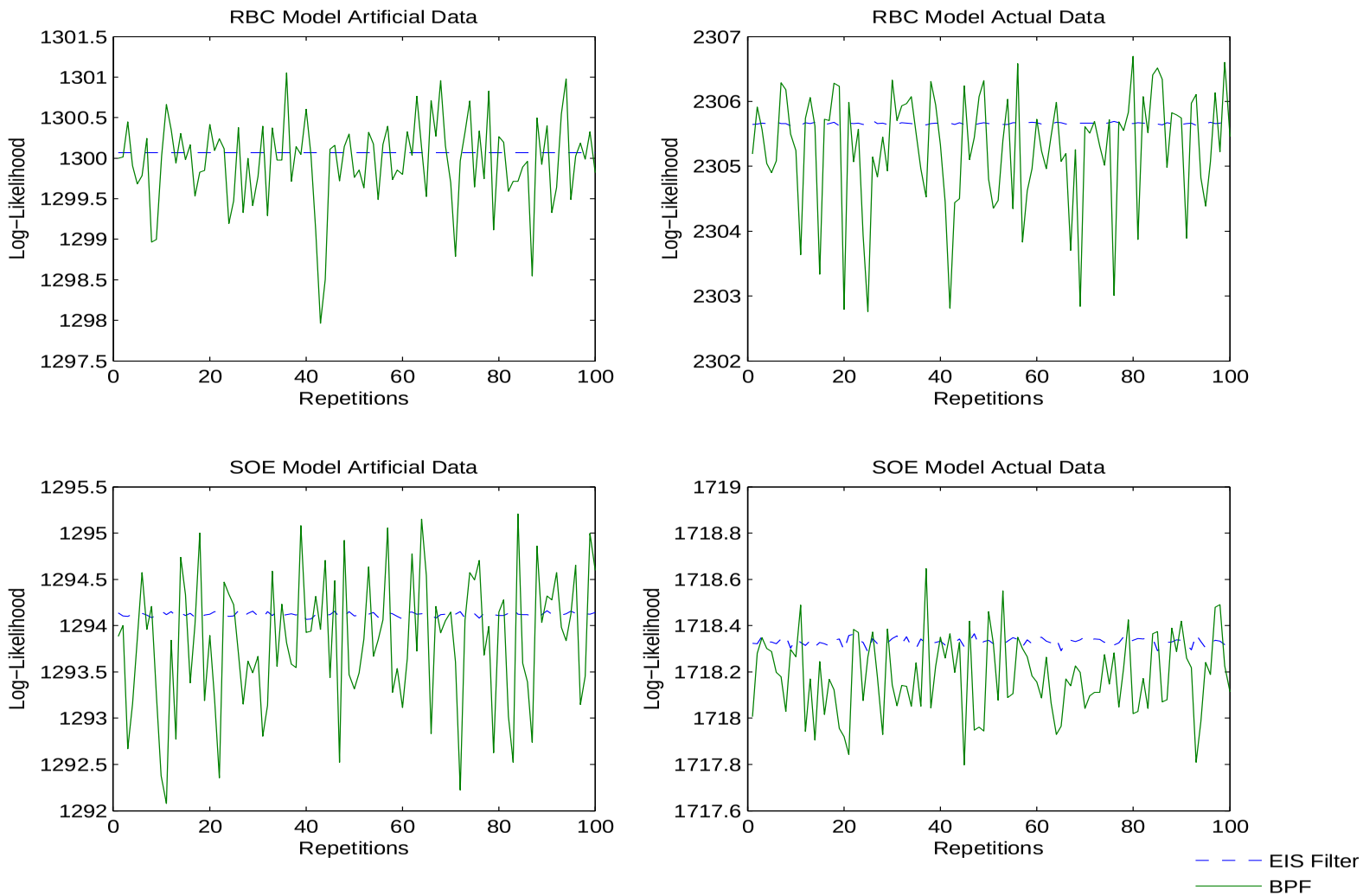

Figure 1. Log-likelihood approximations.

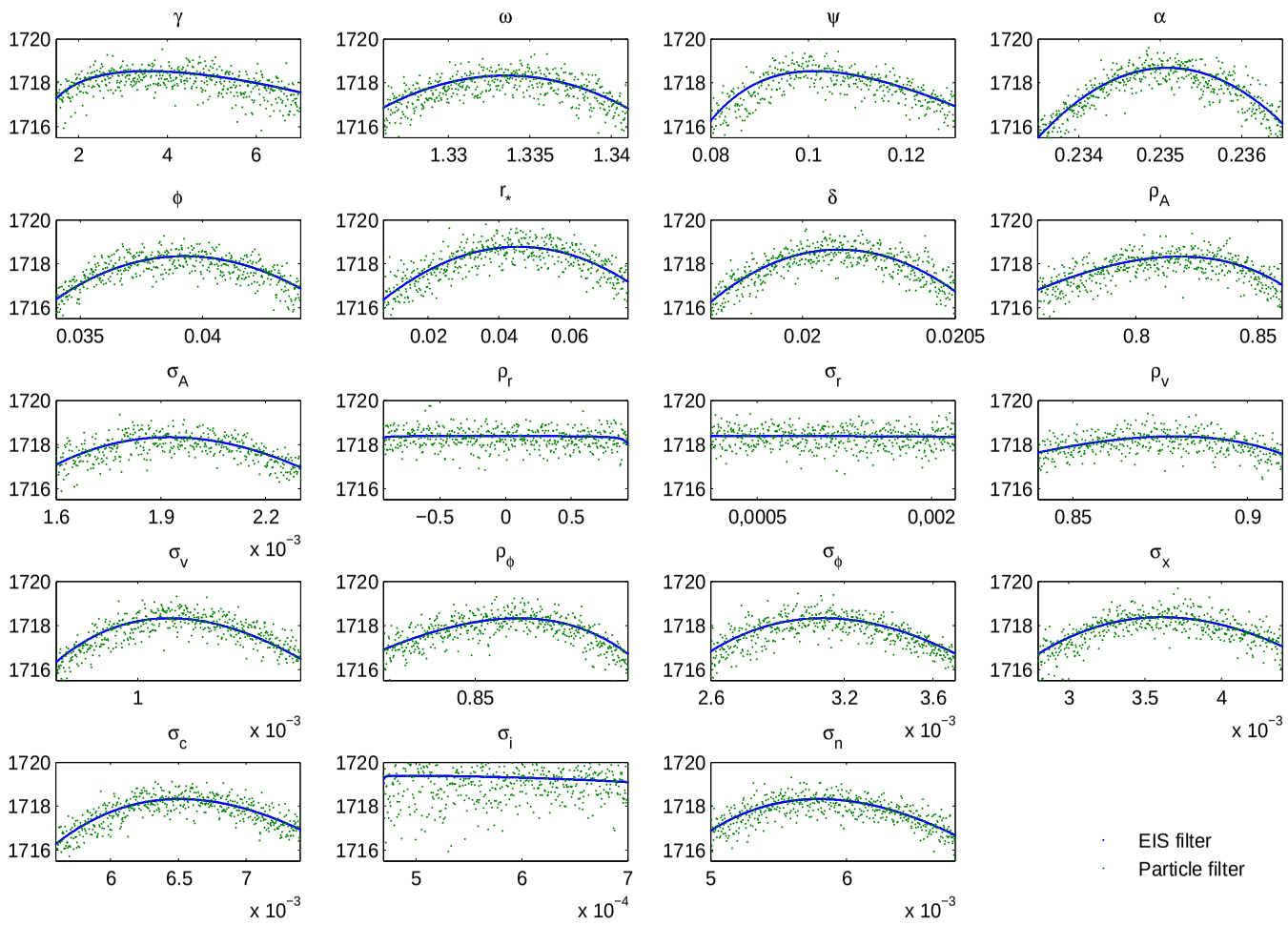

Figure 2. Log-likelihood sections of the SOE model for the actual data.