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 Econometrics
VU University
1081 HV Amsterdam, Netherlands

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

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

First Version: April 2007 This Revision: August 2010

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

*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. This paper circulated previously as "An Efficient Approach to Analyzing State-Space Representations". 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 state-space models requires 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.

The EIS filter falls under the general classification of a sequential Monte Carlo (SMC) method. Dating at least to the sequential importance sampling (SIS) algorithm of Handschin and Mayne (1969) and Handschin (1970), SMC methods have long served as workhorses for achieving likelihood evaluation and filtering. Moreover, extensive efforts have been made to develop refinements of the baseline SIS algorithm with the goal of enhancing accuracy and numerical efficiency in the face of challenging scenarios that give rise to problems known as degeneracy and sample impoverishment (for surveys, see Ristic, Arulampalan and Gordon, 2004; and Cappé, Godsill and Moulines, 2007). In the terminology of Pitt and Shephard (1999), the goal of these efforts is to achieve adaption.

The application of SMC methods to the evaluation of DSGE models has been demonstrated by Fernandez-Villaverde and Rubio-Ramirez (2005, 2009). The specific filtering method they implemented was developed by Gordon, Salmond and Smith (1993). Like the SIS algorithm, this filter employs discrete fixed-support approximations to unknown densities that appear in the predictive and updating stages of the filtering process. The discrete points that collectively provide density approximations are known as particles; the approach is known as the bootstrap particle filter, and is a leading example of a sampling importance resampling (SIR) algorithm.

While SIS and SIR algorithms are intuitive and straightforward to implement, they suffer two important shortcomings. First, because the density approximations they provide are discrete, associated likelihood approximations can feature spurious discontinuities (with respect to parameters), rendering as problematic the application of likelihood maximization procedures. Second, the supports upon which approximations are based are not adapted: period-t approximations are based on supports that incorporate information conveyed by values of the observable variables available in period $t-1$, but not period t. As characterized by Pitt and Shephard (1999), this renders the approximations as "blind". This problem gives rise to numerical inefficiencies that can be acute

when observable variables are highly informative with regard to state variables, particularly given the presence of outliers.

Numerous extensions of SIS and SIR algorithms have been proposed in attempts to address these problems. For examples, see Pitt and Shephard (1999); the collection of papers in Doucet, de Freitas and Gordon (2001); Pitt (2002); Ristic et al. (2004), and the collection housed at `http://www-sigproc.eng.cam.ac.uk/smc/papers.html`. While these efforts have produced algorithms that have proven to be effective in many applications, the use of discrete approximations to filtering densities entails a significant limitation: the support of the period-$(t-1)$ approximations carried over for the construction of period-t approximations can no longer be changed in period t. Only the weights attached to the period-$(t-1)$ discrete elements of the state space (i.e., particles) can be adjusted in period t. To be sure, there exist a wide range of algorithms designed to optimize (adapt) these weights in light of new information available at time t. Nevertheless, optimality is restricted as being conditional on the individual particles established in period $(t-1)$. That is, discrete approximations cannot achieve unconditionally optimality, or in other words, full adaption.

Here we seek to overcome this limitation by constructing sequential importance sampling densities that are tailored to targeted integrand using the Efficient Importance Sampling (EIS) methodology developed by Richard and Zhang (2007). Implementation entails the abandonment of discrete approximations to filtering densities. Instead, within a prespecified family of distributions, the goal is to construct an optimal importance sampling density based upon period-t information, and absent the fixed- and prespecified-support constraints: i.e., the goal is to achieve full adaption. Optimality is associated with the minimization of the numerical standard error of the weights associated with the chosen importance sampler, and typically takes the form of a sequence of least-squares regressions. While the algorithm is not as easy to implement as those based on discrete approximations, and the family of samplers successfully implemented to date is somewhat limited, the algorithm can produce dramatic efficiency gains (i.e., reductions of numerical standard errors) in challenging applications.

Here, our focus is on the achievement of near-optimal efficiency for likelihood evaluation. Example applications involve the analysis of 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, Ingram and Whiteman (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, Salmond and Smith (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 y_t be a $n \times 1$ vector of observable variables, and denote $\{y_j\}_{j=1}^t$ as Y_t. Likewise, let s_t be a $d \times 1$ vector of unobserved ('latent') state variables, and denote $\{s_j\}_{j=1}^t$ as S_t . State-space representations consist of a state-transition equation

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

where υ_t is a vector of innovations with respect to (s_{t-1}, Y_{t-1}), and a measurement equation

$$y_t = \delta\left(s_t, Y_{t-1}, u_t\right), \tag{2}$$

where u_t is a vector innovations with respect to (s_t, Y_{t-1}). Hereafter, we refer to v_t as structural shocks, and u_t as measurement errors.

The likelihood function $f(Y_T)$ is obtained by interpreting (1) and (2) in terms of the transition and measurement densities $f(s_t|s_{t-1}, Y_{t-1})$ and $f(y_t|s_t, Y_{t-1})$, respectively. Since the representation is recursive, $f(Y_T)$ factors sequentially as

$$f\left(Y_T\right) = \prod_{t=1}^{T} f\left(y_t|Y_{t-1}\right), \tag{3}$$

where $f\left(y_1|Y_0\right) \equiv f(y_1)$. The time-$t$ likelihood $f\left(y_t|Y_{t-1}\right)$ is obtained by marginalizing over s_t :

$$f\left(y_t|Y_{t-1}\right) = \int f\left(y_t|s_t, Y_{t-1}\right) f\left(s_t|Y_{t-1}\right) ds_t, \tag{4}$$

where the predictive density $f\left(s_t|Y_{t-1}\right)$ is given by

$$f\left(s_t|Y_{t-1}\right) = \int f\left(s_t|s_{t-1}, Y_{t-1}\right) f\left(s_{t-1}|Y_{t-1}\right) ds_{t-1}, \tag{5}$$

and $f\left(s_{t-1}|Y_{t-1}\right)$ is the time-$(t-1)$ filtering density. Advancing the time subscript by one period, from Bayes' theorem, $f\left(s_t|Y_t\right)$ is given by

$$f\left(s_t|Y_t\right) = \frac{f\left(y_t, s_t|Y_{t-1}\right)}{f\left(y_t|Y_{t-1}\right)} = \frac{f\left(y_t|s_t, Y_{t-1}\right) f\left(s_t|Y_{t-1}\right)}{f\left(y_t|Y_{t-1}\right)}. \tag{6}$$

Likelihood construction is achieved by calculating (4) and (5) sequentially from periods 1 to T, taking as an input in period t the filtering density constructed in period $(t-1)$. In period 1 the filtering density is the known marginal density $f(s_0)$, which can be degenerate as a special case.

3 Particle Filters

3.1 General Principle

Period-t computation inherently requires evaluating the following integral obtained by substituting (5) into the likelihood integral in (4)

$$f\left(y_t|Y_{t-1}\right) = \int\int f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}(s_{t-1}|Y_{t-1}) ds_{t-1} ds_t, \tag{7}$$

where $\widehat{f}(s_{t-1}|Y_{t-1})$ denotes an approximation to the period-$(t-1)$ filtering density. Particle filters rely upon approximations in the form of a mixture-of-Dirac measures associated with the period-$(t-1)$ swarm $\{s_{t-1}^i\}_{i=1}^N$ which is fixed in period-t :

$$\widehat{f}(s_{t-1}|Y_{t-1}) = \sum_{i=1}^{N} \omega_{t-1}^i \cdot \delta_{s_{t-1}^i}\left(s_{t-1}\right), \tag{8}$$

where $\delta_{s_{t-1}^i}(s)$ denotes the Dirac measure at point s_{t-1}^i, and ω_{t-1}^i the weight associated with particle s_{t-1}^i $\left(\text{with } \sum_{i=1}^{N} \omega_{t-1}^i = 1\right)$. If a resampling step took place in period-$(t-1)$, e.g., as in the bootstrap particle filter of Gordon, Salmond and Smith (1993), then $\omega_{t-1}^i = \frac{1}{N}$. For the purpose of comparing EIS filtering with existing SMC methods, resampling is unimportant and is omitted here for ease of presentation. Substituting the Dirac approximation (8) into the likelihood integral (7) effectively solves the (inner) integration in s_{t-1}, and produces the simplification

$$f\left(y_t|Y_{t-1}\right) = \sum_{i=1}^{N} \omega_{t-1}^i \int f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}^i, Y_{t-1}\right) ds_t. \tag{9}$$

Filtering algorithms differ by the construction of the importance sampler used to evaluate the integral in (9) and to construct the corresponding approximation for the period-t filtering density. Next we discuss first unadapted filtering, then characterize conditional adaption, and finally outline auxiliary particle filters, which aim at approximating conditional adaption.

3.2 Unadapted Filters

The baseline unadapted particle filter relies upon a propagation step whereby for each particle s_{t-1}^i one draws a particle s_t^i from the transition density $f\left(s_t|s_{t-1}^i, Y_{t-1}\right)$. The corresponding estimates of the likelihood integral and of the period-t filtering density are then given by

$$\widehat{f}\left(y_t|Y_{t-1}\right) = \sum_{i=1}^{N} \omega_{t-1}^i \cdot f\left(y_t|s_t^i, Y_{t-1}\right) \tag{10}$$

$$\widehat{f}\left(s_t|Y_t\right) = \sum_{i=1}^{N} \omega_t^i \delta_{s_t^i}\left(s_t\right), \tag{11}$$

where the (posterior) weights ω_t^i obtain from the (prior) weights ω_{t-1}^i by application of Bayes' theorem:

$$\omega_t^i = \omega_{t-1}^i \cdot \frac{f\left(y_t|s_t^i, Y_{t-1}\right)}{\widehat{f}\left(y_t|Y_{t-1}\right)}. \tag{12}$$

Unadapted filters are generally easy to implement as they only require that transition densities be amenable to (sequential) Monte Carlo simulation. Moreover, they provide unbiased estimates of likelihood integrals under weak though technically non-trivial conditions; e.g., see Del Moral (2004) or Chopin (2004). On the other hand, they are prone to degeneracy and sample impoverishment due to their lack of adaption.

3.3 Conditional Optimality

Note that the measurement density incorporates the critical assumption that y_t is independent of s_{t-1} given (s_t, Y_{t-1}). It follows that

$$f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) = f\left(s_t|s_{t-1}, Y_t\right) \cdot f\left(y_t|s_{t-1}, Y_{t-1}\right). \tag{13}$$

When this factorization is analytically tractable, it is possible to achieve conditionally optimal adaption. To see how, substitute this factorization into the likelihood integral (7) to obtain

$$\begin{aligned} f\left(y_t|Y_{t-1}\right) &= \int\int f\left(s_t|s_{t-1},Y_t\right)\cdot f\left(y_t|s_{t-1},Y_{t-1}\right)\cdot \widehat{f}\left(s_{t-1}|Y_{t-1}\right)ds_t ds_{t-1} \\ &= \int f\left(y_t|s_{t-1},Y_{t-1}\right)\cdot \widehat{f}\left(s_{t-1}|Y_{t-1}\right)ds_{t-1} \\ &= \sum_{i=1}^{N}\omega_{t-1}^i\cdot f\left(y_t|s_{t-1}^i,Y_{t-1}\right). \end{aligned} \tag{14}$$

The corresponding propagation stage then proceeds as follows: for each particle s_{t-1}^i, draw a particle s_t^i from $f\left(s_t|s_{t-1}^i,Y_t\right)$. The corresponding weights are given by

$$\omega_t^i = \omega_{t-1}^i\cdot\frac{f\left(y_t|s_{t-1}^i,Y_{t-1}\right)}{\widehat{f}\left(y_t|Y_{t-1}\right)}. \tag{15}$$

The key difference with the unadapted particle filter lies in the fact that the draws of s_t are now conditional on y_t also. Note that since ω_t^i does not depend on s_t^i, but only on s_{t-1}^i, its conditional variance is zero given $\{s_{t-1}^i\}_{i=1}^N$. This is referenced as the optimal sampler following Zaritskii et al. (1975) and Akaski and Kumamoto (1977). However, since the factorization in (13) is tractable only in special cases, this sampler represents a theoretical rather than an operational benchmark. But most importantly, this sampler is only optimal conditionally upon $\{s_{t-1}^i\}_{i=1}^N$: its unconditional MC variance is not zero. The limitation to conditional optimality is inherently linked to the use of mixtures of Dirac measures in representing filtering densities. As noted, the EIS implementation we propose below targets unconditional optimality.

3.4 Auxiliary Particle Filters

Since the conditionally optimal kernel is generally intractable, much effort has been devoted to its approximation under the restriction that the swarm $\{s_{t-1}^i\}_{i=1}^N$ is kept fixed in period t, so that only integration in s_t can be adapted; e.g., see Doucet (1998) or Vaswani (2008). The key to a number of such extensions lies in the interpretation of (9) as a mixed integral in (s_t, k_t), where k_t denotes the index of particles, and follows the multinomial distribution $MN\left(N,\{\omega_{t-1}^i\}_{i=1}^N\right)$. The likelihood integral may then be evaluated via importance sampling (IS), relying upon a mixed

density kernel of the form

$$\gamma_t(s,k) = \omega_{t-1}^k \cdot p_t(s,k) \cdot f\left(s_t | s_{t-1}^k, Y_{t-1}\right), \tag{16}$$

where $p_t(s,k)$ is introduced with the objective of minimizing the MC variance of the likelihood IS estimate. Numerical tractability requires that $p_t(s,k)$ be selected in such a way that the IS kernel $\gamma_t(s,k)$ can be normalized into an operational IS sampling density $g_t(s,k)$. As a prominent example, the Auxiliary Particle (AP) filter of Pitt and Shephard (1999) specifies $p_t(s,k)$ as

$$p_t(s,k) = f\left(y_t | \mu_t^k, Y_{t-1}\right), \qquad \mu_t^k = E(s_t | s_{t-1}^k, Y_{t-1}), \tag{17}$$

in which case the IS mixed sampler obtains as

$$g_t(s,k) = \pi_t^k \cdot f\left(s_t | s_{t-1}^k, Y_{t-1}\right), \tag{18}$$

with

$$\pi_t^k = D_t^{-1} \cdot \omega_{t-1}^k \cdot f(y_t | \mu_t^k, Y_{t-1}) \tag{19}$$

$$D_t = \sum_{j=1}^{N} \omega_{t-1}^j \cdot f(y_t | \mu_t^j, Y_{t-1}). \tag{20}$$

Let $\omega_t(s,k)$ denote the ratio between the integrand in (9) and the IS density in (18):

$$\omega_t(s,k) = D_t \frac{f(y_t|s, Y_{t-1})}{f(y_t|\mu_t^k, Y_{t-1})}. \tag{21}$$

Then the AP filter estimate of the likelihood integral is given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^{N} \widetilde{\omega}_t^i, \qquad \widetilde{\omega}_t^i = \omega_t\left(s_t^i, k_t^i\right), \tag{22}$$

and the approximation of the period-t filtering density is

$$\widehat{f}_N\left(s_t|Y_t\right) = \sum_{i=1}^{N} \omega_t^i \delta_{s_{t-1}^i}\left(s_t\right), \qquad \omega_t^i = \frac{\widetilde{\omega}_t^i}{\sum_{j=1}^{N} \widetilde{\omega}_t^j}. \tag{23}$$

Under special circumstances further adaption can be achieved. If the transition density $f\left(s_t|s_{t-1}, Y_{t-1}\right)$ belongs to a family of densities closed under multiplication, then selecting $p_t\left(s,k\right)$ in (16) from that same family can produce an operational and improved sampler. Its integrating constant obtains by first integrating the product $p_t\left(s,k\right) \cdot f\left(s|s_{t-1}^k, Y_{t-1}\right)$ with respect to s, then summing the remainder over k. Examples for the case in which $f\left(s_t|s_{t-1}, Y_{t-1}\right)$ is Gaussian in s_t are discussed in Pitt and Shephard (1999) and Smith and Santos (2006). Using for $p_t\left(s,k\right)$ a first-order Taylor series expansion of $\ln f\left(y_t|s_t, Y_{t-1}\right)$ in s_t around μ_t^j yields Pitt and Shephard's adapted particle filter; likewise, Smith and Santos demonstrate implementation of a second-order expansion.

Last but not least, we note that conditional adaption, even in the optimal case, can occasionally be outperformed by unadapted filters, as highlighted by Johansen and Doucet (2008) for the AP filter. Specifically the asymptotic variance decomposition derived by the authors contains terms propagated forward from all previous periods. It follows that "while the adaption may be beneficial at the time which it is performed it may have a negative influence on the variance at a later point". [p. 1503]

4 EIS Filters

4.1 Unconditional Optimality

The root inefficiency of particle-based filters lies in their reliance upon mixture-of-Dirac approximations of filtering densities. This effectively replaces the likelihood integral in (7) by that in (9), thereby eliminating the possibility of period-t adaption with respect to s_{t-1}. In order to further illustrate this critical distinction, let us revisit (7). Consider the (theoretical) factorization

$$f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}\left(s_{t-1}|Y_{t-1}\right) = f\left(s_t, s_{t-1}|Y_t\right) \cdot f\left(y_t|Y_{t-1}\right). \tag{24}$$

If analytically tractable, $f(s_t, s_{t-1}|Y_t)$ would clearly be the unconditionally optimal (fully adapted) IS sampler for the likelihood integral (7), as a single draw from it would produce an IS estimate of $f(y_t|Y_{t-1})$ with zero MC variance. The period-t filtering density would then obtain by marginalization with respect to s_{t-1} :

$$f(s_t|Y_t) = \int f(s_t, s_{t-1}|Y_t)\, ds_{t-1}. \tag{25}$$

Under full normality and linearity assumptions, such densities are immediately available and instrumental in the Kalman filter. They are generally intractable for more general dynamic state-space models, but nevertheless provide a useful point of reference for the EIS filter we propose below.

Our target is indeed that of constructing EIS samplers in (s_{t-1}, s_t) for the likelihood integral in (7). As discussed further below, this requires that approximations of filtering densities be continuous. Section 4.2 outlines the general principle behind EIS. Section 4.3 discusses implementation in the context of state-space representations. Finally, Section 4.4 discusses a special case that often characterizes state-space representations: degenerate transition densities.

4.2 EIS Integration

Let $\varphi_t(\lambda_t)$, with $\lambda_t = (s_{t-1}, s_t)$, denote the integrand in (7), where the subscript-t in φ_t replaces Y_t. Implementation of EIS begins with the pre-selection of a parametric class $K = \{k(\lambda_t; a_t); a_t \in A\}$ of analytically integrable auxiliary density kernels. The corresponding density functions (IS samplers) and IS ratios are given respectively by

$$g(\lambda_t|a_t) = \frac{k(\lambda_t; a_t)}{\chi(a_t)}, \qquad \chi(a_t) = \int k(\lambda_t; a_t) d\lambda_t, \tag{26}$$

$$\omega_t(\lambda_t; a_t) = \frac{\varphi_t(\lambda_t)}{g_t(\lambda_t|a_t)}. \tag{27}$$

The objective of EIS is to select a parameter value $\widehat{a}_t \in A$ that minimizes the MC variance of the IS ratio over the full range of integration. Following Richard and Zhang (2007), a near optimal value $\widehat{a}_t$ obtains as the solution to

$$(\widehat{a}_t, \widehat{c}_t) = \underset{(a_t, c_t)}{\arg\min} \int \left[\ln \varphi_t(\lambda_t) - c_t - \ln k(\lambda_t; a_t)\right]^2 g(\lambda_t|a_t) d\lambda_t, \tag{28}$$

where c_t denotes an intercept meant to calibrate the ratio $\ln(\varphi_t/k)$. Equation (28) represents a standard least squares problem, except that the auxiliary sampling density itself depends upon a_t. This is resolved by reinterpreting (28) as the search for a fixed-point solution. An operational MC version, implemented (typically) using $R << N$ draws, is as follows:

Step $l+1$: Given $\widehat{a}_t^l$, draw intermediate values $\{\lambda_{t,l}^i\}_{i=1}^R$ from the step-l EIS sampler $g(\lambda_t|\widehat{a}_t^l)$, and solve

$$(\widehat{a}_t^{l+1}, \widehat{c}_t^{l+1}) = \underset{(a_t,c_t)}{\arg\min} \sum_{i=1}^{R} \left[\ln \varphi_t(\lambda_{t,l}^i) - c_t - \ln k(\lambda_{t,l}^i; a_t)\right]^2. \tag{29}$$

Before we discuss the pre-selection of K, three points bear mentioning. First, the optimization problems in (28) and (29) are designed to provide global approximations of $\varphi_t(\lambda_t)$ over the full range of integration (in contrast with the selection of IS samplers from local approximations). See Section 3.4 in Richard and Zhang (2007) for further discussion of this critical issue in relation with the MC variance of IS estimates, and more generally, with their convergence. Second, the selection of the initial value $\widehat{a}_t^0$ can be important for achieving rapid fixed-point convergence (say less than 5 iterations). Section 5 below presents an effective algorithm for specifying $\widehat{a}_t^0$ in applications involving DSGE models. Third, to achieve rapid convergence, and foremost, to ensure continuity of the corresponding likelihood estimates, $\{\lambda_{t,l}^i\}$ must be obtained by transformation of a set of Common Random Numbers (CRNs) $\{u_t^i\}$ drawn from a canonical distribution associated with K (i.e., one that does not depend on a_t; e.g., standardized Normal draws when g is Gaussian).

Continuing with convergence, this can be assessed by monitoring $\widehat{a}_t^l$ across successive iterations l, and a stopping rule can be established using a relative-change threshold. Although there is no guarantee nor formal proof that the sequence $\left\{\widehat{a}_t^l\right\}$ converges for every possible pair $(\varphi(\cdot), k(\cdot))$, we have found repeatedly that it never fails to occur short of a complete mismatch between the target φ and the kernel k (e.g., φ is bimodal and k is unimodal), in which case failure to converge serves as a signal that the class K needs to be adjusted. Moreover, convergence is not critical: what matters is that the resulting sampler delivers accurate and numerically efficient approximations of the targeted integrand. Regarding accuracy, or more precisely, convergence of the MC estimate generated by the optimized importance sampler to the targeted integral, this is assured under the mild regularity conditions for importance samplers outlined, e.g., by Geweke (1989). Regarding efficiency, this can be assessed using diagnostic measures such as the R^2 statistic associated with (29), and summary

statistics regarding the dispersion of the importance sampling weights $\omega\left(\lambda^i, \widehat{a}(\theta)\right)$ (see Richard and Zhang, 2007, for further details, and the empirical applications presented in Section 5 for extensive examples of efficiency diagnostics).

Finally, as mentioned above, the preselection of K is bound to be problem-specific. (However, note that once EIS has been programmed for a particular class K, then changes in model specification only require adjusting the dependent variable $\ln \varphi_t$.) If K belongs to the exponential family of distributions, then there exists a natural parametrization in the sense of Lehmann (1986, Section 2.7) for which the auxiliary regressions in (29) are linear in a_t. Details regarding the implementation of Gaussian EIS samplers are provided in Appendix B.

4.3 The EIS Filter

4.3.1 Joint EIS

The simplest (one-step) implementation of EIS filter entails the application of EIS directly to the likelihood integral in (7) using a Gaussian EIS sampler in $\lambda_t = (s_{t-1}, s_t)$. The dependent variable in the auxiliary linear EIS regressors in (29) is then given by

$$\varphi_t\left(s_{t-1}, s_t\right) = f\left(y_t|s_t, Y_{t-1}\right) \cdot f\left(s_t|s_{t-1}, Y_{t-1}\right) \cdot \widehat{f}(s_{t-1}|Y_{t-1}). \tag{30}$$

In turn, the likelihood EIS estimate is given by

$$\widehat{f}\left(y_t|Y_{t-1}\right) = \frac{1}{S}\sum_{i=1}^{S} \omega_t\left(s_{t-1}^i, s_t^i; \widehat{a}_t\right), \tag{31}$$

where $\omega_t\left(\cdot\right)$ denotes the EIS ratio (27) and $\{s_{t-1}^i, s_t^i\}_{i=1}^S$ denotes i.i.d. draws from the EIS sampler $g\left(s_{t-1}, s_t|\widehat{a}_t\right)$. A period-$t$ filtering density approximation is then given by the marginal of g in s_t.:

$$\widehat{f}\left(s_t|Y_t\right) = \int g\left(s_{t-1}, s_t; \widehat{a}_t\right) ds_{t-1}. \tag{32}$$

As noted, particularly in the presence of outliers and/or tight measurements, the selection of a good initial sampler $g_t\left(s_t, s_{t-1}|\widehat{a}_t^0\right)$ can significantly accelerate convergence. For the applications discussed in Section 5, we rely upon local Taylor Series expansions to construct initial Gaussian

samplers. This is similar to the procedure proposed by Durbin and Koopman (1997) whereby (local) Gaussian approximations are used as importance samplers to evaluate the likelihood function of non-Gaussian state space models. However, our approach differs critically from theirs, in that such local approximations are used only to construct starting values for fully iterated global EIS approximation of the likelihood integrand (i.e. a fully adapted IS sampler within the pre-selected class K). As illustrated in Section 5, convergence to fully adapted EIS samplers produce substantial efficiency gains relative to initial local approximations.

Given a distributional family chosen for $g_t(\,)$; a systematic means of constructing $g_t\left(s_{t-1}, s_t|\widehat{a}_t^0\right)$; and with $\widehat{f}(s_0|Y_0)$ initialized using $f(s_0)$; a summary of the period-t algorithm of the EIS filter is as follows.

Propagation: Inheriting $\widehat{f}(s_{t-1}|Y_{t-1})$ from period $(t-1)$, obtain the integrand $\varphi_t(s_{t-1}, s_t)$ appearing in the likelihood integral (7), where $\varphi_t(s_{t-1}, s_t)$ is defined in (30).

EIS Optimization: Construct an initialized sampler $g_t\left(s_{t-1}, s_t|\widehat{a}_t^0\right)$, and obtain the optimized parameterization $\widehat{a}_t$ as the solution to (29). This yields $g_t(s_{t-1}, s_t|\widehat{a}_t)$.

Likelihood integral: Obtain draws $\left\{s_{t-1}^i, s_t^i\right\}_{i=1}^N$ from $g_t(s_{t-1}, s_t|\widehat{a}_t)$, and approximate $\widehat{f}(y_t|Y_{t-1})$ as in (31).

Filtering: Approximate $\widehat{f}(s_t|Y_t)$ as in (32).

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

4.3.2 Nested EIS

A more flexible EIS implementation relies upon the factorization used in constructing particle-based filters, wherein the likelihood integral (7) is rewritten as a nested integral

$$f(y_t|Y_{t-1}) = \int f(y_t|s_t, Y_{t-1}) \cdot \widehat{f}(s_t|Y_{t-1}) ds_t \tag{33}$$

$$\widehat{f}(s_t|Y_{t-1}) = \int f(s_t|s_{t-1}, Y_{t-1}) \cdot \widehat{f}(s_{t-1}|Y_{t-1}) ds_{t-1}, \tag{34}$$

and EIS is then applied separately to these two lower-dimensional integrals. However, the EIS exploitation of this factorization differs fundamentally from that of the particle filter. Specifically, preserving full adaption requires that (34) be evaluated under an "inner" (E)IS sampler individually

adapted to each value of s_t requested for "outer" EIS evaluation of (33). Such nesting provides greater flexibility in that it eliminates all analytical restrictions across the two nested EIS samplers, but adds significant computational burden since inner EIS must be repeated for each individual value of s_t requested by outer EIS. While intrinsically computationally more demanding, nested EIS can often be simplified. For example, if $f(s_t|s_{t-1}, Y_{t-1})$ is partially Gaussian, as in the case for the applications discussed in Section 5, and $\widehat{f}(s_{t-1}|Y_{t-1})$ is itself Gaussian, then the dimensionality of inner EIS can be reduced by partial analytical integration. Moreover, while the applications presented below were pursued via joint EIS, we also experimented at an earlier stage with an (initial) inner IS sampler produced by local Gaussian approximation of $f(s_t|s_{t-1}, Y_{t-1})$, which proved to be numerically reliable for achieving outer EIS evaluation. Finally, note that under nested EIS it is the outer EIS sampler $g_t(s_t|\widehat{a}_t)$ that provides a direct approximation of the filtering density $f(s_t|Y_t)$.

4.4 Degenerate Transitions

State transitions often include identities that effectively reduce the dimension of integration in (s_t, s_{t-1}). Let s_t partition into $s_t = (p_t, q_t)$ in such a way that the transition consists of a set of identities

$$q_t \equiv \phi(p_t, s_{t-1}), \tag{35}$$

in combination with a proper transition density $f(p_t|s_{t-1}, Y_{t-1})$. Such identities effectively reduce the dimensionality of the likelihood integral (7), which is rewritten as

$$f(y_t|Y_{t-1}) = \int\int f(y_t|s_t, Y_{t-1})\,|_{q_t=\phi(p_t,s_{t-1})} \cdot f(p_t|s_{t-1}, Y_{t-1}) \cdot \widehat{f}(s_{t-1}|Y_{t-1}) dp_t ds_{t-1}. \tag{36}$$

Application of EIS (whether joint or nested) to (36) produces an EIS sampler $g_t(p_t, s_{t-1}|\widehat{a}_t)$, which typically requires non-linear transformations and auxiliary EIS in order to produce a marginal approximation for the filtering density $f(s_t|Y_t)$.

Fortunately, a more direct approach avoids these problems. This consists of applying a transformation of variables from (p_t, s_{t-1}) into (s_t, p_{t-1}) to the integral in (36) Assuming that $\phi()$ is

one-to-one (on the relevant range), we denote its inverse and Jacobian by

$$q_{t-1} = \psi\left(s_t, p_{t-1}\right), \tag{37}$$

$$J\left(s_t, p_{t-1}\right) = ||\frac{\partial}{\partial q_t'}\psi(s_t, p_{t-1})||, \tag{38}$$

where $||\bullet||$ denotes the absolute value of a determinant. The likelihood integral in (36) is then rewritten as

$$f\left(y_t|Y_{t-1}\right) = \int\int f\left(y_t|s_t, Y_{t-1}\right)\cdot J\left(s_t, p_{t-1}\right)\left[f\left(p_t|s_{t-1}, Y_{t-1}\right)\cdot\widehat{f}(s_{t-1}|Y_{t-1})\right]|_{q_{t-1}=\psi(s_t,p_{t-1})}ds_t dp_{t-1}. \tag{39}$$

Application of the EIS filter to (39) rather than (7) then proceeds precisely as described above, producing an EIS sampler $g_t\left(s_t, p_{t-1}|\widehat{a}_t\right)$ for which an approximation of the filtering density $f(s_t|Y_t)$ obtains directly by marginalization with respect to p_{t-1}.

5 Application to DSGE Models

As noted, the work of Fernandez-Villaverde and Rubio-Ramirez (2005, 2009) revealed 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 bootstrap particle (BP) filter 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.

We analyze two data sets for both models: an artificial data set generated from a known model parameterization; and a corresponding real data set. Thus in total we consider four applications, each of which poses a significant challenge to the successful implementation of a numerical filtering algorithm. 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 presented in Section 5.2, and results are presented in Section 5.3.

In demonstrating the performance of the EIS filter, we provide context by comparing this performance with that of the BP filter. We also attempted to evaluate the performance of the AP filter of Pitt and Shephard (1999), but in all cases found that it could not be implemented successfully in the extensive battery of Monte Carlo experiments presented below. With reference to (21), its occasional failure traces to the denominator $f\left(y_t|\mu_t^k, Y_{t-1}\right)$, which occasionally takes on near-zero values. However, we do note that in conducting pilot experiments wherein the AP filter was implemented successfully, the gains in numerical efficiency it generated relative to the BP filter were dominated by those generated by the EIS filter (details are available upon request).

5.1 A Generic DSGE Model

The following functional notation for multivariate Gaussian densities proves useful for the following discussion:

$$f_N^n(x|\mu,\Omega) = (2\pi)^{-\frac{n}{2}}|\Omega|^{-\frac{1}{2}}\exp -\frac{1}{2}(x-\mu)'\Omega^{-1}(x-\mu). \tag{40}$$

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 here is characterized by the following state-space densities. The measurement density is given by

$$f\left(y_t|s_t, Y_{t-1}\right) = f_N^n\left(y_t|\mu\left(s_t\right), V_y\right), \tag{41}$$

where $\mu\left(s_t\right)$ denotes a non-linear vector function of s_t. The transition density consists of a non-degenerate density for p_t and a non-linear degenerate transition for q_t. The latter only depends on s_{t-1} for both models considered here:

$$\begin{aligned} f\left(p_t|s_{t-1}, Y_{t-1}\right) &= f_N^{d_p}\left(p_t|R\cdot s_{t-1}, \Sigma\right), & (42)\\ q_t &= \phi\left(p_{t-1}, q_{t-1}\right). & (43) \end{aligned}$$

As in Section 4.4, ϕ is assumed to be invertible (on the relevant range), and its inverse and Jacobian are denoted by

$$q_{t-1} = \psi\left(q_t, p_{t-1}\right), \qquad J\left(q_t, p_{t-1}\right) = ||\frac{\partial}{\partial q_t}\psi\left(q_t, p_{t-1}\right)||. \tag{44}$$

All computations in period t are based on an EIS Gaussian approximation for the period-$(t-1)$ filtering density, denoted by

$$\widehat{f}\left(s_{t-1}|Y_{t-1}\right) = f_N^d\left(s_{t-1}|\mu_{t-1}, \Omega_{t-1}\right). \tag{45}$$

The application of joint EIS to this model under a class K consisting of unconstrained Gaussian kernels for (s_t, s_{t-1}) is conceptually straightforward. The critical step for fast and numerically reliable EIS convergence consists of the derivation of a good initial value for the auxiliary EIS parameter $\widehat{a}_t^0$. Next, we outline the successive steps required for the computation of $\widehat{a}_t^0$. Details are presented in Appendix A.

Following (39) the period-t EIS target is given by

$$\varphi_t\left(s_t, p_{t-1}\right) = f_N^n\left(y_t|\mu\left(s_t\right), V_y\right) \cdot J\left(q_t, p_{t-1}\right) \cdot \left[f_N^{d_p}\left(p_t|Rs_{t-1}, \Sigma\right) \cdot f_N^d(s_{t-1}|\mu_{t-1}, \Omega_{t-1})\right]|_{q_{t-1}=\psi(q_t, p_{t-1})}. \tag{46}$$

The initial EIS sampler $g\left(s_t, p_{t-1}|\widehat{a}_t^0\right)$ employs the local linear approximations

$$\widehat{\mu}\left(s_t\right) = r_t + P_t s_t, \tag{47}$$

$$\widehat{\psi}\left(q_t, p_{t-1}\right) = A_t q_t + B_t p_{t-1} + c_t, \tag{48}$$

which obtain by first-order Taylor Series around the base points

$$\overline{s}_{t-1} = \mu_{t-1}, \qquad \overline{p}_t = R\mu_{t-1} \qquad \text{and} \qquad \overline{q}_t = \phi\left(\overline{s}_{t-1}\right). \tag{49}$$

For each of the four applications under consideration, these starting values produce fast and numerically reliable EIS convergence for all periods. The initial EIS kernel is defined as

$$k\left(s_t, p_{t-1}; \widehat{a}_t^0\right) = f_N^n\left(y_t|r_t + P_t s_t, V_y\right) \cdot \left[f_N^{d_p}\left(p_t|Rs_{t-1}, \Sigma\right) \cdot f_N^d(s_{t-1}|\mu_{t-1}, \Omega_{t-1})\right]|_{q_{t-1}=A_t q_t + B_t p_{t-1} + c_t}. \tag{50}$$

Being defined as a product of three linear Gaussian densities, $k\left(s_t, p_{t-1}; \widehat{a}_t^0\right)$ is itself a Gaussian kernel in (s_t, p_{t-1}). Its integrating constant $\chi\left(\widehat{a}_t^0\right)$ and first- and second-order moments obtain through a sequence of operations consisting of the following five steps:

Step 1. Combine the two Gaussian densities between brackets to produce a joint Gaussian density for (p_t, s_{t-1}).

Step 2. Transform the combined density into a Gaussian density for (s_t, p_{t-1}) by application of the linear transformation (48).

Step 3. Factorize the latter density into a marginal Gaussian density for s_t and a conditional Gaussian density for $p_{t-1}|s_t$, for which we use the short-hand notation $f_1^*(s_t)$ and $f_2^*(p_{t-1}|s_t)$, respectively.

Step 4. Combine $f_1^*(s_t)$ and the (linearized) measurement density $f_N^n(y_t|r_t + P_t s_t, V_y)$ into a joint Gaussian density for (y_t, s_t).

Step 5. Factorize the latter into a marginal Gaussian density for y_t and a conditional Gaussian density for $s_t|y_t$, for which we use the short-hand notation $f_3^*(y_t)$ and $f_4^*(s_t|y_t)$, respectively.

Given these steps, the integrating constant of the initial EIS kernel in (50) is given by

$$\chi\left(\widehat{a}_t^0\right) = f_3^*(y_t). \tag{51}$$

The corresponding initial sampler is given by

$$g\left(s_t, p_{t-1}|\widehat{a}_t^0\right) = f_2^*(p_{t-1}|s_t) \cdot f_4^*(s_t|y_t). \tag{52}$$

An initial IS estimate of the likelihood, constructed along the lines of Durbin and Koopman (1997), is then given by

$$\widehat{f}_t^0(y_t|Y_{t-1}) = f_3^*(y_t) \cdot \frac{1}{N}\sum_{i=1}^{N} \widetilde{\omega}_t^0\left(p_t^i, s_{t-1}^i; \widehat{a}_t^0\right), \tag{53}$$

where $\left\{p_t^i, s_{t-1}^i\right\}_{i=1}^N$ denotes N i.i.d. (CRN) draws from $g\left(s_t, p_{t-1}|\widehat{a}_t^0\right)$ and $\widetilde{\omega}_t^0(\cdot)$ denotes the corresponding IS ratios (with $f_3^*(y_t)$ factored out from $\widetilde{\omega}_t^0(\cdot)$ for ease of interpretation); detailed expressions are given in Appendix A.

Using the initialized sampler $g\left(s_t, p_{t-1}|\widehat{a}_t^0\right)$ as an input, we compute a fully iterated global EIS Gaussian kernel via the linear least-squares problem defined in (29). Under a natural parametrization in the sense of Lehmann (1986, Section 2.7), the regressors consist of $\left(s_t^i, p_{t-1}^i,\right)$ together with

squares and (lower-triangular) cross-product for a total of

$$n_k = (d + d_p) + \frac{1}{2}(d + d_p)(d + d_p + 1) \tag{54}$$

regressors, plus an intercept. Details are provided in Appendix B. The convergence criterion we use is given by

$$||\frac{\widehat{a}_t^{l+1} - \widehat{a}_t^l}{\widehat{a}_t^l}||_2 < \varepsilon, \tag{55}$$

with ε on the order of 10^{-4}. As a general rule of thumb we have found that starting from the initial sampler (52) and using a number of draws R of the order of 3 to 5 times n_k secures fast convergence to a (near) optimal value $\widehat{a}_t$ (typically 3 to 4 iterations for the applications described below).

5.2 Example Models

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

$$\max_{c_t, l_t} \qquad U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(c_t^{\varphi} l_t^{1-\varphi}\right)^{1-\phi}}{1-\phi} \quad , \tag{56}$$

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

$$
\begin{aligned}
x_t &= z_t k_t^{\alpha} n_t^{1-\alpha}, && (57)\\
1 &= n_t + l_t, && (58)\\
x_t &= c_t + i_t, && (59)\\
k_{t+1} &= i_t + (1-\delta)k_t, && (60)\\
z_t &= z_* e^{\omega_t}, \qquad \omega_t = \rho\omega_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2), && (61)
\end{aligned}
$$

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

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 (58) and (59). 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 (61), it proves 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. It follows that this two-state RBC model is of the form given by (41) to (43) with $n = 3$, $d_p = d_q = 1$, and (with all variables presented as logged deviations from steady state):

- $s_t' = (p_t, \quad q_t) = (z_t, \quad k_t)$,
- $y_t' = (x_t, i_t, n_t)$,
- $\mu(s_t)$ denoting the policy functions for y_t,
- V_y denoting the diagonal covariance matrix of the measurement error,
- $R = (\rho, \quad 0), \qquad \Sigma = \sigma_\varepsilon^2$,
- $\phi(s_{t-1})$ denoting the degenerate transition obtained from identity (60) 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 model is that 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

$$U = E_0 \sum_{t=0}^{\infty} \theta_t \frac{\left[c_t - \varphi_t \omega^{-1} n_t^{\omega}\right]^{1-\gamma} - 1}{1-\gamma}, \qquad \omega > 0, \qquad \gamma \geqslant 0, \tag{62}$$

where φ_t is a preference shock that affects the disutility generated by labor effort (introduced, e.g., following Smets and Wouters, 2002). Following Uzawa (1968), the discount factor θ_t is endogenous and obeys

$$\begin{aligned} \theta_{t+1} &= \beta\left(\widetilde{c}_t, \widetilde{n}_t\right)\theta_t, \qquad \theta_0 = 1, \\ \beta\left(\widetilde{c}_t, \widetilde{n}_t\right) &= \left[1 + \widetilde{c}_t - \omega^{-1}\widetilde{n}_t^{\omega}\right]^{-\psi}, \qquad \psi > 0, \end{aligned} \tag{63}$$

where $(\widetilde{c}_t, \widetilde{n}_t)$ 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

$$x_t = A_t k_t^{\alpha} n_t^{1-a} \tag{64}$$

$$d_{t+1} = (1 + r_t)\, d_t - x_t + c_t + i_t + \frac{\phi}{2}\left(k_{t+1} - k_t\right)^2 \tag{65}$$

$$k_{t+1} = v_t^{-1} i_t + (1 - \delta)\, k_t \tag{66}$$

$$\ln A_{t+1} = \rho_A \ln A_t + \varepsilon_{At+1}, \qquad \varepsilon_{At} \sim iidN(0, \sigma^2_{\varepsilon_A}) \tag{67}$$

$$\ln r_{t+1} = (1 - \rho_r)\ln r_* + \rho_r \ln r_t + \varepsilon_{rt+1}, \qquad \varepsilon_{rt} \sim iidN(0, \sigma^2_{\varepsilon_r}) \tag{68}$$

$$\ln v_{t+1} = \rho_v \ln v_t + \varepsilon_{vt+1}, \qquad \varepsilon_{vt} \sim iidN(0, \sigma^2_{\varepsilon_v}) \tag{69}$$

$$\ln \varphi_{t+1} = \rho_\varphi \ln \varphi_t + \varepsilon_{\varphi t+1}, \qquad \varepsilon_{\varphi t} \sim iidN(0, \sigma^2_{\varepsilon_\varphi}), \tag{70}$$

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 the full set of model 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 (41) to (43) with $n = 4$, $d_p = 4$, $d_q = 2$, and (with all variables once again presented as logged deviations from steady state):

- $p_t' = (A_t, \quad r_t, \quad v_t, \quad \varphi_t)$,
- $q_t' = (d_t, \quad k_t)$,
- $y_t' = (x_t, \quad c_t, \quad i_t, \quad n_t)$,
- $\mu(s_t)$ denoting the logged policy functions for y_t,
- $R = (R_p, \quad 0)$, with R_p denoting the 4×4 diagonal matrix with diagonal elements $(\rho_A, \rho_r, \rho_\gamma, \rho_p)$,
- Σ a 4×4 diagonal covariance matrix, with diagonal elements $\left(\sigma^2_{\varepsilon_A}, \sigma^2_{\varepsilon_r}, \sigma^2_{\varepsilon_v}, \sigma^2_{\varepsilon_\varphi}\right)$,
- $\phi(s_{t-1})$ 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 4.4.

5.3 Results

Here we present a series of five Monte Carlo (MC) experiments; each experiment involves four data sets, two for each model. For a given model, one data set consists of artificial data generated from a known parameterization of the model; the second consists of actual data that align with theoretical counterparts. For the 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). The actual data set consists of 184 quarterly 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 evaluated by Fernandez-Villaverde and Rubio-Ramirez (2005).

For the SOE model, the artificial data set consists of 100 realizations of $\{x_t, c_t, i_t, n_t\}$. The actual data set consists of Canadian data on $\{x_t, c_t, i_t, n_t\}$, spanning 1976:I-2008:IV (132 observations), and is an updated version of the data set evaluated by Mendoza (1991). 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 RBC data is the Federal Reserve Bank of St. Louis; and the source of the SOE data set is Statistics Canada. The data are available for downloading at `http://www.pitt.edu/~dejong/wp.htm`.

The MC experiments we conducted involve repeated approximations of the likelihood functions corresponding to each data set. Parameterizations of the likelihood functions are presented in Table 1. The parameters of the RBC model, artificial data set are those chosen by Fernandez-Villaverde and Rubio-Ramirez (2005); for the actual data set, the parameters are posterior modes associated with the non-informative priors chosen by Fernandez-Villaverde and Rubio-Ramirez.

Regarding the SOE model, aside from the parameters characterizing sources of uncertainty in the model, the artificial data set was generated using the parameter values calibrated by Schmitt-Grohe and Uribe (2003) to match the summary statistics on Canadian data reported by Mendoza (1991). The additional parameters characterizing new sources of stochastic uncertainty included 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. The statistics are the standard deviations of $\{x_t, c_t, i_t, n_t\}$, first-order serial correlations, and contemporaneous correlations with output. Finally, the standard deviations of all measurement errors were set at 0.5%.

For the actual data set of the SOE model, parameters were set at posterior modes estimated using the prior specification indicated in Table 1. The prior consists of independent normal distributions specified for each parameter. Aside from parameters that characterize stochastic uncertainty, prior means were set at the values specified by Schmitt-Grohe and Uribe (2003), and prior standard deviations were set to reflect non-trivial uncertainty over these specifications. (Note that the specifications of δ and r_* chosen by Schmitt-Grohe and Uribe are appropriate for annual data, and thus were translated under our prior into specifications appropriate for the quarterly observations we

employ.) The priors over AR parameters were centered at 0.8 (s.d. 0.2); and with two exceptions along ill-behaved dimensions (σ_r and σ_i), the priors over σ's were centered at 0.5% (s.d. 0.5%). The likelihood function implies strong negative correlation between σ_r and ρ_r, thus σ_r was set so that the posterior mode of ρ_r lied near its prior mean. Also, the posterior mode of σ_i was difficult to pin down, so its prior mean was centered at 0.5% like its counterparts, while its standard deviation was set to pin down the posterior mode at this value.

Table 1: Parameter Values

RBC Model

	α	β	ϕ	φ	δ	ρ	σ_ε	σ_y	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_y	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

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 can lead to 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 quarters of 1974, and the first and second quarters 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 $\frac{\phi}{2}(k_{t+1} - k_t)^2$ appearing in (65). 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.

The first experiment we conducted is designed to assess potential biases in likelihood approximations associated with the EIS filter. (As noted, following Del Moral, 2004, and Chopin, 2004, we know that likelihood estimates associated with the BP filter are unbiased.) Setting $N = 1,000,000$, we generated 100 log-likelihood approximations using 100 sets of CRNs generated under the BP filter. (With this very large specification of N, we hoped to capture as accurately as possible 'true' likelihood values.) Then setting $N = R = 100$ for the RBC model and $N = R = 200$ for the SOE model, we generated another 100 approximations using the EIS filter. Finally, we calculated the difference in approximations for each of the 10,000 possible combinations of values, and searched for instances in which differences were significantly different from zero. We did so on a date-by-date basis for all four data sets. Differences are illustrated in Figure 1, in the form of boxplots.

The red lines in the boxplots depict the median of the distribution of differences we obtained; the edges of the blue boxes depict the first (Q1) and third quartiles (Q3); and the black whiskers extend to the 5^{th} and 95^{th} percentiles. In all time periods and data sets, zero is contained in the interval of the empirical differences going from the 5^{th} to the 95^{th} percentiles, and in most cases zero is also contained in the blue box defined by Q1 and Q3. Thus we conclude that the EIS filter generates unbiased likelihood approximations in these applications. In the case of the RBC model, actual data set, note the large spike in the width of the distribution of differences associated with the second quarter of 1980, which coincides with the abrupt changes in output and investment noted above. This time period carries further implications discussed below.

The second experiment we conducted is designed to assess the relative numerical efficiency of the EIS and BP filters. Once again, we generated 100 approximations of the log-likelihood function using 100 sets of CRNs for both filters. Following Fernandez-Villaverde and Rubio-Ramirez (2005), the BP filter was implemented using $N = 60,000$ for the RBC model, requiring 17.28 seconds of CPU time per log-likelihood evaluation on a 2.9 GHz desktop computer using MATLAB for the artificial data set, and 25.72 seconds for the real data set. Given the relatively large dimensionality of s_t for the SOE model, we set $N = 150,000$ in this case, requiring 61.83 and 80.90 seconds per log-likelihood evaluation for the artificial and real data sets, respectively. In turn, the EIS filter was implemented using $N = R = 100$ for the RBC model, and 200 for the SOE model, with convergence criterion given by (55), and the maximum number of iterations set to 10. For the RBC model, this required 0.55 seconds per log-likelihood evaluation for the artificial data set and 0.6477 for the real data set; for the SOE model, this required 1.34 and 2.18 seconds per evaluation.[1] Results are presented in Table 2 and Figure 2.

Table 2: Monte Carlo Means and Standard Deviations

RBC Model

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

SOE Model

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

Notes: Monte Carlo means and standard deviations are based on 100 different sets of CRNs. The EIS filter and Initial Sampler are implemented using N=100 and N=200 for the RBC and SOE models respectively. The BP filter is implemented using N=60,000 and N=150,000 for the RBC and SOE models. NSE denotes numerical standard errors.

Table 2 presents means and standard deviations of log-likelihood approximations obtained across the 100 sets of CRNs for each data set. The standard deviations indicate numerical accuracy, and are

[1] Although we avoided the use of loops in programming the BP filter, the relative difference in computational times reported above are likely to be compressed using Fortran or C.

often referenced as numerical standard errors, or NSEs. For the EIS filter, we report results obtained using both initial and optimized samplers, where recall that initialized samplers are constructed as local linear Gaussian approximations of targeted densities, similar to the importance sampling procedure of Durbin and Koopman (1997), which is based on the Kalman smoother. Note that NSEs associated with the BP filter exceed those associated with the EIS filter by a factor ranging from 45.6 (SOE model, actual data) to 37,300 (RBC model, artificial data). As NSEs decay at the square-root of the rate of increase in N, the comparison associated with the SOE model, actual data set implies a required value of $N = 3.1e + 8$ in order for the BP filter to match the accuracy of the EIS filter with $N = R = 200$, and $N = 8.4e + 14$ for the RBC model, artificial data set. NSEs associated with the initial samplers are also inferior to those associated with the EIS filter. Specifically, for the RBC model the NSE associated with the initial sampler is 0.3486 rather than 0.00051 for the artificial data set, and 4.02 rather than 0.0151 for the actual data set. For the SOE model, the comparisons are 4.60 versus 0.0232 for the artificial data set, and 3.719 versus 0.0166 for the actual data set. Thus we observe substantial payoffs to the implementation of the optimization step of the EIS filtering algorithm.

Continuing with Table 2, note that the mean log-likelihoods associated with both artificial data sets differ relatively substantially across the BP and EIS filters. This discrepancy does not reflect bias in either filter. Rather, it reflects a tendency of the BP filter to provide very low log-likelihood estimates in certain cases, a problem that can be remedied by increasing N (as the boxplots in Figure 1 indicate). This tendency is illustrated in Figure 2, which plots each of the 100 log-likelihood approximations obtained for each filter and each data set. While the approximations obtained using the BP filter often lie near those generated using the EIS filter, there are many instances in which approximations fall far below. That is, the distribution of approximations obtained using the BP filter are skewed heavily leftward (though, as noted, unbiased). Finally, mean log-likelihoods obtained using the initial samplers also differ substantially from those obtained using the EIS filter, in part reflecting the caveats associated with the use of linearized approximations of DSGE models noted by Fernandez-Villaverde and Rubio-Ramirez (2005, 2009).

The third 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 NSEs reported in Table 2. Here, we repeated each of the four experiments summarized in Table 2 100 times using 100

Table 3: Repeated Samples

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

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

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

artificial data sets generated from each model, parameterized as in Table 1. Hereafter we refer to a specific parameterized model as a data generation process (DGP). So for each DGP, we generated an artificial data set, produced an NSE measure by obtaining 100 log-likelihood approximations using 100 sets of CRNs, and repeating for a total of 100 artificial data sets. We did this only for the EIS filter. Results are summarized in Table 3.

We report three sets of numbers for each DGP. The second and third columns of numbers are the mean and statistical standard deviation of NSEs obtained across realizations of artificial data sets. For three of the four DGPs, the NSEs reported in Table 2 lie within two standard deviations of the mean NSEs reported in Table 3. The exception is the DGP associated with RBC model, actual data set, for which the NSE reported in Table 2 is 5.5 standard deviations above the mean reported in Table 3. This discrepancy appears to reflect the challenge to numerical accuracy posed by the outlier associated with the second quarter of 1980. Despite this discrepancy, the general message of Table 2 does not appear overly sensitive to specific features of the data sets upon which they are based.

The first column of numbers reported in Table 3 are the standard deviations of log-likelihood estimates obtained using a single set of CRNs applied to the 100 realizations of artificial data generated using each DGP. These numbers are 'statistical moments' that indicate plausible ranges of log-likelihood values one would expect to obtain when confronted with realizations of alternative

data sets generated from a given DGP: we refer to them as statistical standard errors (SSEs). Note that SSEs exceed their numerical counterparts (the mean NSEs) by two to five orders of magnitude. This helps provide additional context for the NSEs: it indicates that statistical uncertainty swamps numerical inaccuracy. This is important, because among other things it implies that numerical inaccuracy does not stand as a barrier in obtaining information regarding statistical uncertainty, a result further underscored in the next experiment. In contrast, note that SSEs are roughly on par with the NSEs associated with the BP filter (as reported in Table 2) for three of the four DGPs we considered (the SOE model, actual data providing the exception).

The fourth experiment illustrates the continuity (with respect to parameters) of log-likelihood approximations generated by the EIS filter. Here we generated log-likelihood surfaces produced by allowing each model parameter to vary individually above and below its actual value, holding all additional parameters fixed at their true values. This was done using both filters, each implemented using CRNs. Figure 3 illustrates log-likelihood surfaces for the full set of parameters associated with the SOE model, constructed using 500 grid points. Once again, the BP filter was implemented using $N = 150,000$, and the EIS filter using $N = R = 200$. Results of applications to the additional models yield similar results. For both filters, surface heights are indicated by dots. We do not connect the dots using line segments because doing so obfuscates surfaces associated with the EIS filter. But clearly, connecting the dots with line segments yields highly jagged surfaces for the BP filter.

The discontinuity of the surface approximations generated by the BP filter clearly 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 collection of parameters associated with a given model targeted for estimation, and Σ_μ its associated posterior covariance matrix, one means of estimating Σ_μ is via the Laplace approximation

$$\widehat{\Sigma}_\mu^{-1} = \Sigma_p^{-1} - \frac{\partial^2 \ln L\left(\widehat{\mu}\right)}{\partial \mu \partial \mu'}, \tag{71}$$

where Σ_p denotes the prior covariance matrix, $L\left(\widehat{\mu}\right)$ represents the associated likelihood function, and $\widehat{\mu}$ denotes the posterior mode of μ. In the models considered here, approximations of $\widehat{\Sigma}_\mu$ fail

to be positive-definite when constructed using the BP filter, but are produced with no problems using the EIS filter: indeed, the posterior standard deviations reported in Table 1 were obtained via the application of the EIS filter to (71).

The final experiment we conducted offers further insights into challenges posed by outliers, as well as additional reassurance regarding potential bias associated with the EIS filter. Working with the RBC model, artificial data set, we generated 12 variations of this data set by inserting an outlier in the second observation of one of the observables, keeping the remaining variables fixed at their original values. We considered four different values for this outlier for each variable: two deviated by ± 4 sample standard deviations from the sample mean, and two deviated ± 8 sample standard deviations from the sample mean. For each of the 12 new data sets (as well as for the original data set), we calculated log-likelihood values for periods 1 and 2 using the BP filter implemented with $N = 60,000$, the EIS filter implemented with $N = R = 100$, and the Gauss-Chebyshev quadrature method implemented with 250 nodes along all three dimensions of integration, for a total of $250^3 = 15,625,000$ nodes. By evaluating the first two periods only, implementation of the quadrature method is feasible (since the dimensionality of the targeted integral is low), and provides a near-exact value of targeted log-likelihoods. Using each filter, we once again obtain 100 likelihood evaluations using 100 sets of CRNs; means and NSEs computed across sets of CRNs are reported in Table 4.

As Table 4 indicates, for the cases involving output and hours, both filters deal with outliers reasonable well: both typically provide accurate approximations of the targeted log-likelihood, and NSEs are fairly uniform across alternative data sets. This is not true for investment, which recall has a relatively tightly distributed measurement equation. In this case, while the performance of the EIS filter remains constant across data sets, that of the BP filter deteriorates dramatically: mean approximations diverge from targeted values, and NSEs jump by factors in the neighborhood of 4 and 15 for the ± 4 and ± 8 cases, respectively.

6 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

Table 4: Outliers Experiment

		BP Filter				EIS Filter				Quadrature	
		$t=1$		$t=2$		$t=1$		$t=2$		$t=1$	$t=2$
		Mean	NSE	Mean	NSE	Mean	NSE	Mean	NSE		
	-8	10.7477	0.2922	13.1189	0.1550	10.8265	0.0007	13.1237	0.0477	10.8266	13.1208
	-4	10.7477	0.2922	13.1777	0.1246	10.8265	0.0007	13.1981	0.0478	10.8266	13.1956
x	0	10.7477	0.2922	13.1166	0.1327	10.8265	0.0007	13.1153	0.0478	10.8266	13.1125
	4	10.7477	0.2922	12.8416	0.1681	10.8265	0.0007	12.8753	0.0478	10.8266	12.8726
	8	10.7477	0.2922	12.4285	0.2377	10.8265	0.0007	12.4782	0.0478	10.8266	12.4753
		Mean	NSE	Mean	NSE	Mean	NSE	Mean	NSE		
	-8	10.7477	0.2922	-7.2901	2.2706	10.8265	0.0007	-4.8069	0.0487	10.8266	-4.8434
	-4	10.7477	0.2922	8.2613	0.6063	10.8265	0.0007	8.3088	0.0478	10.8266	8.3022
i	0	10.7477	0.2922	13.1166	0.1327	10.8265	0.0007	13.1153	0.0478	10.8266	13.1125
	4	10.7477	0.2922	9.5606	0.4855	10.8265	0.0007	9.5875	0.0472	10.8266	9.5887
	8	10.7477	0.2922	-4.1055	1.8742	10.8265	0.0007	-2.2999	0.0474	10.8266	-2.3127
		Mean	NSE	Mean	NSE	Mean	NSE	Mean	NSE		
	-8	10.7477	0.2922	-19.2097	0.2208	10.8265	0.0007	-19.1907	0.0478	10.8266	-19.1937
	-4	10.7477	0.2922	4.8902	0.1706	10.8265	0.0007	4.8973	0.0478	10.8266	4.8944
n	0	10.7477	0.2922	13.1166	0.1327	10.8265	0.0007	13.1153	0.0478	10.8266	13.1125
	4	10.7477	0.2922	5.4680	0.1261	10.8265	0.0007	5.4635	0.0478	10.8266	5.4608
	8	10.7477	0.2922	-18.0546	0.1518	10.8265	0.0007	-18.0582	0.0477	10.8266	-18.0608

Notes: Monte Carlo means and NSEs are based on 100 different sets of CRNs. The EIS filter is implemented using S=R=100 and the BP filer using N=60,000.

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 filter owes considerably to this component of the algorithm.

Implementation of the EIS filter is relatively involved in comparison with particle-based filters. However, EIS iterations can be programmed as a self-contained procedure, and through the considerable details we have provided regarding implementation, including the annotated code that accompanies this paper, we have sought to reduce barriers to entry regarding its implementation. Moreover, given the importance of the preservation of non-linearities in the context of working with DSGE models, as Fernandez-Villaverde and Rubio-Ramirez (2005, 2009) have demonstrated, coupled with the outstanding performance of the EIS filter we have documented, the gains from its adaption are considerable.

The performance of the EIS filter in the applications we have presented motivates our current research agenda, wherein we are seeking to develop operational EIS samplers that are more flexible than those drawn from the exponential family of distributions. One such extension, which we have implemented successfully in a companion paper dedicated explicitly to filtering, entails the construction of highly flexible marginal densities specified along one or two dimensions (DeJong et al., 2010). A more promising approach entails the development of an EIS procedure to construct global mixtures of Gaussian samplers; under this approach, EIS optimization is pursued via non-linear least squares implemented using analytical derivatives. We have already successfully tested an initial univariate mixture implementation; high-dimensional extensions are under development. The goal is to facilitate EIS implementations using highly flexible samplers that will prove efficient in applications involving even the most challenging of targeted integrands.

7 Appendix A

Here we characterize derivation of the initial sampler (52).

- Step 1: The product of the two Gaussian densities between brackets in (50) is given by

$$f^*(p_t, s_{t-1}) = f_N^{d+d_p}\left(\left(\begin{array}{c} p_t \\ s_{t-1} \end{array}\right) \mid C\mu_{t-1}, V\right) \tag{72}$$

with $C = \left(\begin{array}{c} R \\ I_d \end{array}\right)$, $V = \left(\begin{array}{cc} \Sigma + R\Omega_{t-1}R' & R\Omega_{t-1} \\ \Omega_{t-1}R' & \Omega_{t-1} \end{array}\right) = \left(\begin{array}{cc} \Sigma & 0 \\ 0 & 0 \end{array}\right) + C\Omega_{t-1}C'$.

- Step 2: The linear approximation of ψ in (48) implies the following linear transformation of (s_t, p_{t-1}) into (p_t, s_{t-1}):

$$\left(\begin{array}{c} p_t \\ s_{t-1} \end{array}\right) = \delta_t + \Delta_t^{-1}\left(\begin{array}{c} s_t \\ p_{t-1} \end{array}\right) \tag{73}$$

with

$$\delta_t = \left(\begin{array}{c} 0 \\ 0 \\ c_t \end{array}\right), \qquad \Delta_t^{-1} = \left(\begin{array}{ccc} I_{d_p} & 0 & 0 \\ 0 & 0 & I_{d_p} \\ 0 & A_t & B_t \end{array}\right), \tag{74}$$

$$||\Delta_t|| = ||A_t||, \qquad \Delta_t = \left(\begin{array}{ccc} I_{d_p} & 0 & 0 \\ 0 & -A_t^{-1}B_t & A_t^{-1} \\ 0 & I_{d_p} & 0 \end{array}\right). \tag{75}$$

The quadratic form in the density (72) is rewritten as

$$\begin{aligned} \left[\left(\begin{array}{c} p_t \\ s_{t-1} \end{array}\right) - C\mu_{t-1}\right]' V^{-1} \left[\left(\begin{array}{c} p_t \\ s_t \end{array}\right) - C\mu_{t-1}\right] \;=\; & \left[\left(\begin{array}{c} s_t \\ p_{t-1} \end{array}\right) - \Delta_t\left(C\mu_{t-1} - \delta_t\right)\right]' \\ & \times(\Delta_t V \Delta_t')^{-1}\left[\left(\begin{array}{c} s_t \\ p_{t-1} \end{array}\right) - \Delta_t\left(C\mu_{t-1} - \delta_t\right)\right]. \end{aligned} \tag{76}$$

It follows that the initial kernel as defined in (50) is given by

$$k\left(s_t, p_{t-1}; \widehat{a}_t^0\right) = ||A_t||^{-1} \cdot f_N^n\left(y_t | r_t + P_t s_t, V_y\right) \cdot f_N^{d+d_p}\left(\left(\begin{array}{c} s_t \\ p_{t-1} \end{array}\right) \mid \widetilde{m}_t, \widetilde{\Sigma}_t\right) \tag{77}$$

with

$$\widetilde{m}_t = \Delta_t\left(C\mu_{t-1} - \delta_t\right), \qquad \widetilde{\Sigma}_t = \Delta_t V \Delta_t'. \tag{78}$$

- Step 3: $\widetilde{m}_t$ and $\widetilde{\Sigma}_t$ are partitioned conformably with (s_t, p_{t-1}) into

$$\widetilde{m}_t = \left(\begin{array}{c} \widetilde{m}_1^t \\ \widetilde{m}_2^t \end{array}\right), \qquad \widetilde{\Sigma}_t = \left(\begin{array}{cc} \widetilde{\Sigma}_{11}^t & \widetilde{\Sigma}_{12}^t \\ \widetilde{\Sigma}_{21}^t & \widetilde{\Sigma}_{22}^t \end{array}\right). \tag{79}$$

It immediately follows that

$$f_1^*(s_t) = f_N^d\left(s_t | \widetilde{m}_1^t, \widetilde{\Sigma}_{11}^t\right) \tag{80}$$

$$f_2^*(p_{t-1}|s_t) = f_N^{d_p}\left(p_{t-1} | \widetilde{m}_{2-1}^t + \widetilde{\Delta}_{21}^t s_t, \widetilde{\Sigma}_{22-1}^t\right) \tag{81}$$

with

$$\widetilde{m}_{2-1}^t = \widetilde{m}_2^t - \widetilde{\Delta}_{21}^t \widetilde{m}_1^t, \qquad \widetilde{\Delta}_{21}^t = \widetilde{\Sigma}_{21}^t \left(\widetilde{\Sigma}_{11}^t\right)^{-1}, \qquad \widetilde{\Sigma}_{22-1}^t = \widetilde{\Sigma}_{22}^t - \widetilde{\Delta}_{21}^t \widetilde{\Sigma}_{12}^t. \tag{82}$$

- Step 4: The product of the measurement density by $f_1^*(s_t)$ produces the following joint density for (y_t, s_t) :

$$f(y_t, s_t | Y_{t-1}) = f_N^{n+d}\left(\left(\begin{array}{c} y_t \\ s_t \end{array}\right) \mid \left(\begin{array}{c} r_t + P_t \widetilde{m}_1^t \\ \widetilde{m}_1^t \end{array}\right), \left(\begin{array}{cc} V_y + P_t \widetilde{\Sigma}_{11}^t P_t' & P_t \widetilde{\Sigma}_{11}^t \\ \widetilde{\Sigma}_{11}^t P_t' & \widetilde{\Sigma}_{11}^t \end{array}\right)\right). \tag{83}$$

Note that

$$\left(\begin{array}{cc} V_y + P_t \widetilde{\Sigma}_{11}^t P_t' & P_t \widetilde{\Sigma}_{11}^t \\ \widetilde{\Sigma}_{11}^t P_t' & \widetilde{\Sigma}_{11}^t \end{array}\right)^{-1} = \left(\begin{array}{cc} Q & -QP_t \\ -P_t'Q & H_t + P_t'QP_t \end{array}\right) \tag{84}$$

with

$$Q = V_y^{-1}, \qquad H_t = \left(\widetilde{\Sigma}_{11}^t\right)^{-1}. \tag{85}$$

- Step 5: The joint density in (83) factorizes into

$$\begin{aligned} f_3^*(y_t) &= f_N^n(y_t|r_t + P_t\widetilde{m}_1^t, V_y + P_t\widetilde{\Sigma}_{11}^t P_t') && (86)\\ f_4^*(s_t|y_t) &= f_N^d(s_t|\mu_t^0, \Omega_t^0) && (87) \end{aligned}$$

with

$$\begin{aligned} \mu_t^0 &= \widetilde{m}_1^t + \left(H_t + P_t'QP_t\right)^{-1} P_t'Q\left(y_t - r_t - P_t\widetilde{m}_1^t\right) && (88)\\ \Omega_t^0 &= \left(H_t + P_t'QP_t\right)^{-1} && (89) \end{aligned}$$

and the initial IS sampler is given by

$$g\left(s_t, p_{t-1}|\widehat{a}_t^0\right) = f_2^*\left(p_{t-1}|s_t\right) \cdot f_4^*\left(s_t|y_t\right). \tag{90}$$

Note that the integrating constant of the kernel $k\left(s_t, p_{t-1}; \widehat{a}_t^0\right)$ is given by $\chi\left(\widehat{a}_t^0\right) = ||A_t||$. All together the IS ratio associated with the initial sampler is given by

$$\omega_t^0\left(s_t, p_{t-1}; \widehat{a}_t^0\right) = \frac{\varphi_t\left(s_t, p_{t-1}\right)}{f_2^*\left(p_{t-1}|s_t\right) \cdot f_4^*\left(s_t\right)} = f_3^*\left(y_t\right) \cdot \widetilde{\omega}_t^0\left(s_t, p_{t-1}; \widehat{a}_t^0\right) \tag{91}$$

with

$$\begin{aligned}
\widetilde{\omega}_t^0\left(s_t, p_{t-1}; \widehat{a}_t^0\right) &= R_1\left(q_t, p_{t-1}\right) \cdot R_2\left(q_t, p_{t-1}\right) \cdot R_3\left(s_t\right) \\
R_1\left(q_t, p_{t-1}\right) &= \frac{J\left(q_t, p_{t-1}\right)}{||A_t||} \\
R_2\left(q_t, p_{t-1}\right) &= \frac{f_N^{d+d_p}\left(\begin{pmatrix} p_t \\ s_{t-1} \end{pmatrix} \mid C\mu_{t-1}, V\right)|_{q_{t-1}=\psi(q_t,p_{t-1})}}{f_N^{d+d_p}\left(\begin{pmatrix} p_t \\ s_{t-1} \end{pmatrix} \mid C\mu_{t-1}, V\right)|_{q_{t-1}=\widehat{\psi}(q_t,p_{t-1})}} \\
R_3\left(s_t\right) &= \frac{f_N^n\left(y_t|\mu\left(s_t\right), V_y\right)}{f_N^n\left(y_t|\widehat{\mu}\left(s_t\right), V_y\right)},
\end{aligned}$$

where $\widehat{\mu}\left(s_t\right)$ and $\widehat{\psi}\left(q_t, p_{t-1}\right)$ denote the linear Taylor Series approximations defined in (47) and (48), respectively. This factorization allows one to analyze which of the two linear approximations is most critical.

8 Appendix B

Here we characterize the construction of EIS samplers belonging to the exponential family of distributions, highlighting the Gaussian samplers used in the applications presented in Section 5.

The exponential family of distributions is defined as haven kernels of the form

$$\ln k\left(\lambda; a\right) = b\left(\lambda\right) + a' T\left(\lambda\right), \tag{92}$$

where $T\left(\lambda\right)$ denotes a fixed-dimensional vector of sufficient statistics and the corresponding natural parameterizations (for details, e.g., see Lehmann, 1986, Section 2.7). For $\lambda \sim N_k\left(m, H^{-1}\right)$, we have (up to an additive constant)

$$-2 \ln k\left(\lambda; a\right) \propto \left(\lambda - m\right)' H \left(\lambda - m\right), \tag{93}$$

where a denotes a one-to-one transformation of (m, H) characterized as follows.

Note that

$$\begin{aligned} \lambda' H \lambda &= \sum_{i=1}^{k} h_{ii}\lambda_i^2 + \sum_{i=2}^{k}\sum_{j=1}^{i-1} (2h_{ij})\,\lambda_j\lambda_i \\ &= a_2' vech\left(\lambda\lambda'\right), \end{aligned} \tag{94}$$

where $vech\left(\lambda\lambda'\right)$ denotes the lower-triangular column expansion of $\lambda\lambda'$, and a_2 is the vector

$$a_2' = (h_{11}; 2h_{21}, h_{22}; 2h_{31}, 2h_{32}, h_{33}; ...)\,. \tag{95}$$

Since we are working with kernels we can ignore the additive constant $m'Hm$ in (93) (which will be accounted for by the intercept of the EIS regression), and thus

$$-2\ln k\left(\lambda; a\right) \propto a'T(\lambda), \tag{96}$$

with $a' = (a_1', a_2')$, and

$$a_1 = -2Hm, \qquad T'(\lambda) = \left(\lambda, vech(\lambda\lambda')\right). \tag{97}$$

Under this parameterization, the EIS regression problem in (29) is linear in a_t :

$$(\widehat{a}_t^{l+1}, \widehat{c}_t^{l+1}) = \underset{(a_t, c_t)}{\arg\min} \sum_{i=1}^{R} \left[y_{t,l}^i - c_t - a_t' T\left(\lambda_{t,l}^i\right)\right]^2, \tag{98}$$

with $y_{t,l}^i = -2\ln\varphi_t\left(\lambda_{t,l}^i\right)$. Given the fixed-point solution $\widehat{a}_t$, the optimized information matrix $\widehat{H}_t$ is constructed as in (95), and the corresponding optimized mean vector is obtained using

$$\widehat{m}_t = -\frac{1}{2}\widehat{H}^{-1}\widehat{a}_{1,t}. \tag{99}$$

We conclude by noting that for pathological problems, $\widehat{a}_t$ may not transform into positive-definite $\widehat{H}_t$'s (although this never occurs in the example applications presented in Section 5). In such cases, the construction of $\widehat{a}_t$ may be pursued via shrinkage estimation or constrained optimization, rather than unconstrained OLS; see Richard and Zhang (2007) for an example.

9 Appendix C

Here we characterize the inversion of (43), repeated here for convenience:

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

Recall that the goal of inversion is to obtain

$$q_{t-1} = \psi\left(q_t, p_{t-1}\right), \qquad J\left(q_t, p_{t-1}\right) = ||\frac{\partial}{\partial q_t}\psi\left(q_t, p_{t-1}\right)||.$$

9.1 C.1 RBC Model

Under the RBC model, q_t specializes to $q_t = \ln k_t/k_*$, and in light of (60), the specific form of (43) is

$$e^{\ln k_t/k_*} = i\left(\ln k_t/k_* \ln z_t/z_*\right) + (1-\delta)\, e^{\ln k_{t-1}/k_*}, \tag{100}$$

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

$$\ln k_{t-1}/k_* = \psi\left(\ln k_t/k_*, \ln z_{t-1}/z_*\right), \tag{101}$$

and specify the parameters of this polynomial such that

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

holds. Given the optimized specification

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

the Jacobian

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

obtains analytically.

9.2 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), (43) is quadratic in its arguments. Moreover, it turns out that for all paramterizations 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

$$k_t = C_k + L_k s_{t-1}^1 + \frac{1}{2} s_{t-1}^{1\prime} Q_k s_{t-1}^1. \tag{102}$$

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

$$Q_k = \begin{bmatrix} Q_k^{11} & Q_k^{12} \\ Q_k^{21} & Q_k^{22} \end{bmatrix} \quad , \quad L_k = \begin{bmatrix} L_k^1 \\ L_k^2 \end{bmatrix} \ , \tag{103}$$

we note that $Q_k^{11} > 0$ for all parameterizations under consideration. Thus the inversion of (102) is given by the solution

$$k_{t-1}^* = \frac{-b_k + \sqrt{b_k^2 - 4a_k c_k}}{2a_k}, \tag{104}$$

with

$$a_k = \frac{1}{2} Q_k^{11}, \tag{105}$$

$$b_k = L_k^1 + Q_k^{12} p_{t-1}, \tag{106}$$

$$c_k = C_k + L_k^2 p_{t-1} + \frac{1}{2} p_{t-1}' Q_k^{22} p_{t-1} - k_t. \tag{107}$$

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

$$d_t = C_d + L_d s_{t-1} + \frac{1}{2} s_{t-1\prime} Q_d s_{t-1}. \tag{108}$$

Inversion in d_{t-1} yields the solution

$$d_{t-1}^{*} = \frac{-b_d + \sqrt{b_d^2 - 4a_d c_d}}{2a_d}, \tag{109}$$

with

$$Q_d = \begin{bmatrix} Q_d^{11} & Q_d^{12} \\ Q_d^{21} & Q_d^{22} \end{bmatrix} \quad , \quad L_d = \begin{bmatrix} L_d^1 \\ L_d^2 \end{bmatrix} \ , \tag{110}$$

and

$$a_d = \frac{1}{2} Q_d^{11}, \qquad (>0)\,, \tag{111}$$

$$b_d = L_d^2 + Q_d^{12} s_{t-1}^1, \tag{112}$$

$$c_d = C_d + L_d^2 s_{t-1}^1 + \frac{1}{2} s_{t-1}^{1\prime} Q_k^{22} s_{t-1}^1 - d_t. \tag{113}$$

Note that the solutions $\left(k_{t-1}^{*}, d_{t-1}^{*}\right)$ in (104) and (109) 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

$$J\left(q_t, p_{t-1}\right) = \left(b_k^2 - 4a_k c_k\right)^{-\frac{1}{2}} \cdot \left(b_d^2 - 4a_d c_d\right)^{-\frac{1}{2}}. \tag{114}$$

References

[1] Akashi, H. and H. Kumamoto, 1977, "Random Sampling Approoach to State Estimation in Switching Environments", *Automatica*, 13, 429-434.

[2] Cappe, 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.

[3] Carpenter, J.R., P. Clifford and P. Fernhead, 1999, "An Improved Particle Filter for Non-Linear Problems", *IEE Proceedings-Radar, Sonar and Navigation*, 146, 1, 2-7.

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

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

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

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

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

[9] Doucet, A., 1998, On Dequential Monte Carlo Sampling Methods for Bayesian Filterins, Cambridge University, Dept. of Engineering, Cambridge, UK.

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

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

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

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

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

[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] Hendry, D.F., 1994, "Monte Carlo Experimentation in Econometrics", in R.F. Engle and D.L. McFadden, Eds. *The Handbook of Econometrics*, Vol. IV. New York: North Holland.

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

[20] Kitagawa, G., 1996, "Monte Carlo Filter and Smoother for Non-Gaussian Non-Linear State-Space Models", *Journal of Computational and Graphical Statistics*, 5, 1-25.

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

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

[23] Pitt, M.K., 2002, "Smooth Particle Filters for Likelihood Evaluation and Maximisation", University of Warwick working paper.

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

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

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

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

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

[29] 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.

[30] 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.

[31] Smith, J.Q. and A.A.F. Santos, 2006, "Second-Order Filter Distribution Approximations for Financial Time Series with Extreme Outliers", *Journal of Business and Economic Statistics*, 24, 329-337.

[32] Uzawa, H., 1968, "Time Preference, the Consumption FUnction adn 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.

[33] Vaswani, N., 2008, "Particle Filtering for Large-Dimensional State Spaces with Multimodal Observation Likelihoods", *IEEE Transactions on Signal Processing*, 56, 4583-4597.

[34] Zaritskii, V., V. Svetnik, and L. Shimelevich, 1975, "Monte Carlo Techniques in Problems of Optimal Data Processing", *Automation and Remote Control*, 36 (3), 2015-2022.

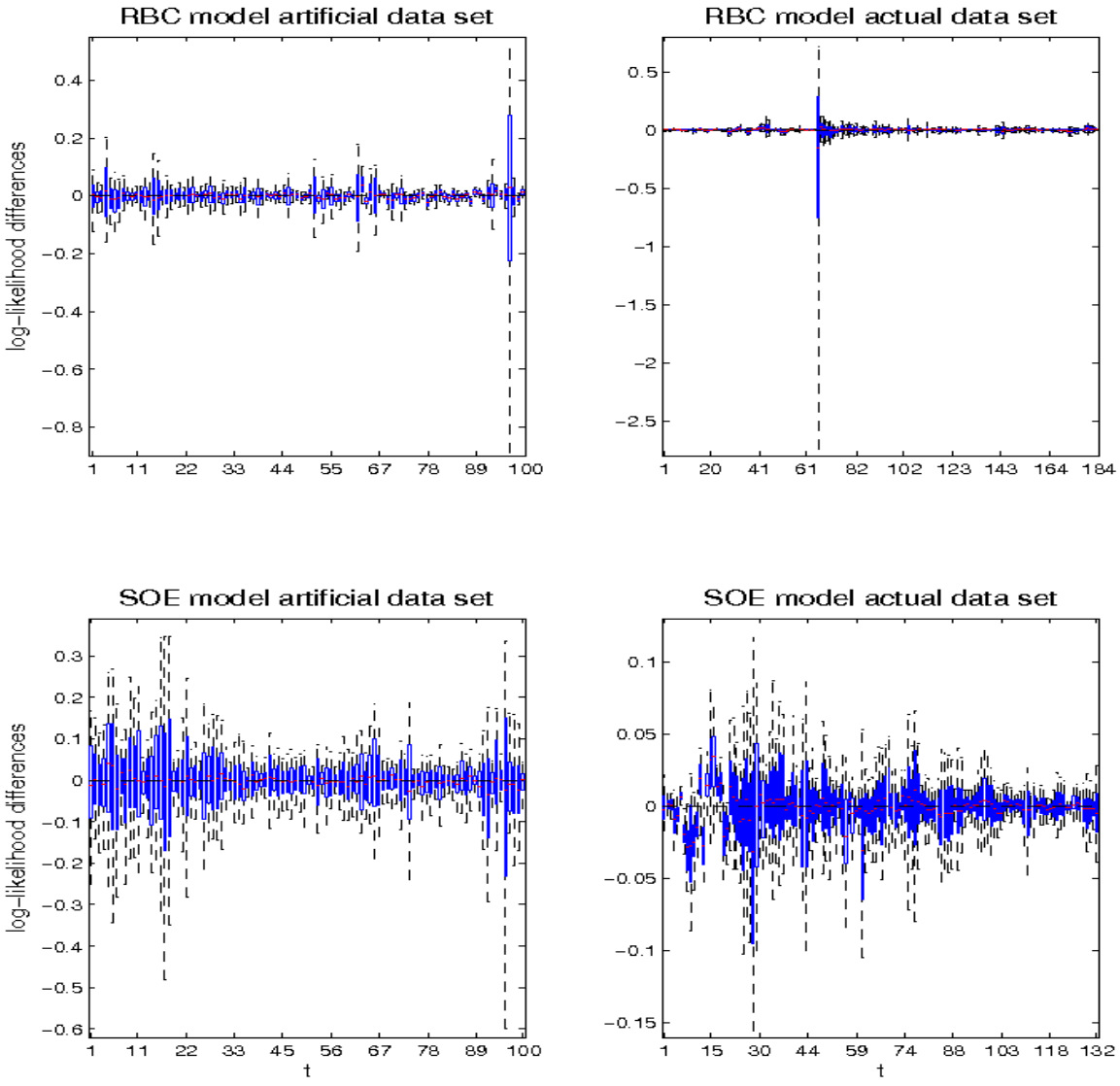

Figure 1. Box Plots

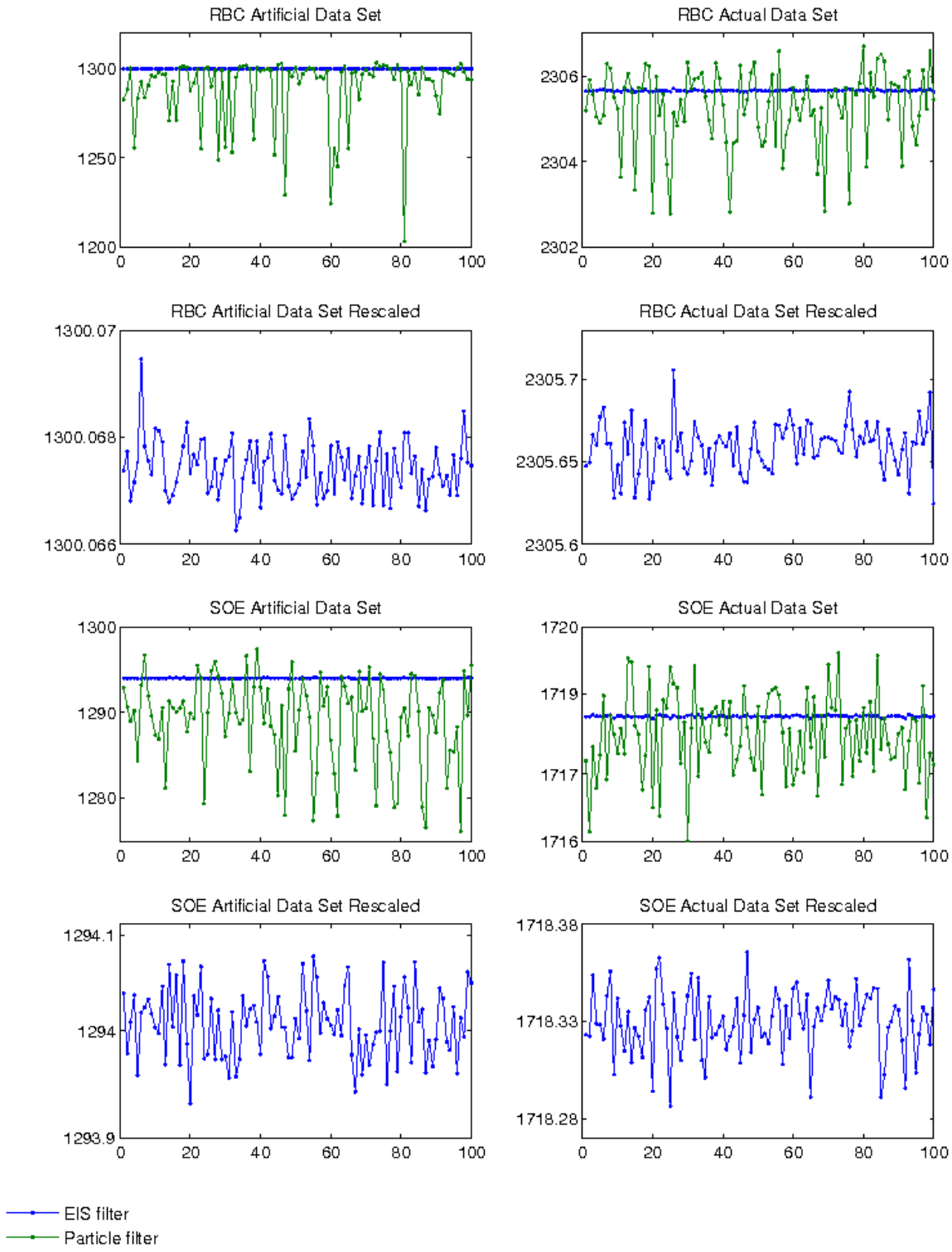

Figure 2. Log-Likelihood Approximations

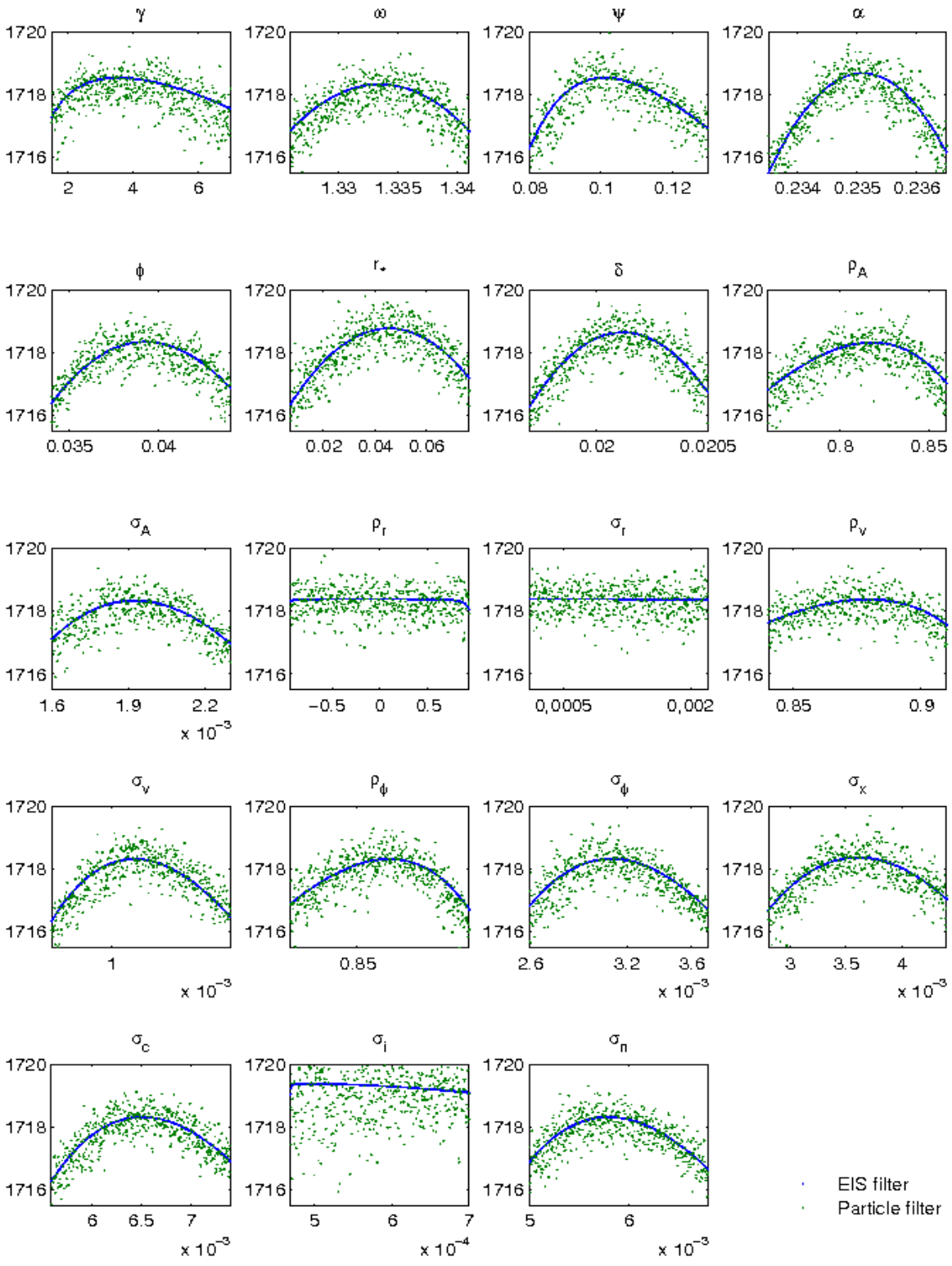

Figure 3. Log-likelihood Surface Plots, SOE Model, Actual Data Set