

Efficient Likelihood Evaluation of State-Space Representations

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

Hariharan Dharmarajan
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260, USA

Roman Liesenfeld
Department of Economics
Universität Kiel
24118 Kiel, Germany

Guilherme V. Moura
Department of Economics
Universität Kiel
24118 Kiel, Germany

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

First Version: April 2007 This Revision: May 2009

Abstract

We develop a numerical procedure that facilitates efficient likelihood evaluation in applications involving non-linear and non-Gaussian state-space models. The procedure approximates necessary integrals using continuous approximations of target densities. Construction is achieved via efficient importance sampling, and approximating densities are adapted to fully incorporate current information. 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. Richard gratefully acknowledges research support provided by the National Science Foundation under grant SES-0516642. For helpful comments, we thank Chetan Dave, Jesus Fernandez-Villaverde, Hiroyuki Kasahara, Juan Rubio-Ramirez, Enrique Sentana, and seminar and conference participants at 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 Macroeconomics Workshop. We also thank Chetan Dave for assistance with compiling the Canadian data set we analyzed. Pseudo-code, GAUSS and MATLAB 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 procedure takes as a building block the pioneering approach to likelihood evaluation and filtering developed by Gordon, Salmond and Smith (1993) and Kitagawa (1996). Their approach 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 particle filter. Examples of its use are becoming widespread; in economics, e.g., see Kim, Shephard and Chib (1998) for an application involving stochastic volatility models; and Fernandez-Villaverde and Rubio-Ramirez (2005, 2009) for applications involving dynamic stochastic general equilibrium (DSGE) models.

While conceptually simple and easy to program, the particle filter suffers two shortcomings. First, because the density approximations it provides are discrete, associated likelihood approximations can feature spurious discontinuities (with respect to parameters), rendering as problematic the application of likelihood maximization procedures (e.g., see Pitt, 2002). 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 (e.g., see Pitt and Shephard, 1999). This 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 the particle filter 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>. Typically, efficiency gains are sought through attempts at adapting period- t densities via the use of information available through period t . However, with the exception of the extension proposed by Pitt (2002), once period- t supports are established they

remain fixed over a discrete collection of points as the filter advances forward through the sample, thus failing to address the problem of spurious likelihood discontinuity. (Pitt employs a bootstrap-smoothing approximation designed to address this problem for the specialized case in which the state space is unidimensional.) Moreover, as far as we are aware, no existing extension pursues adaption in a manner that is designed to achieve optimal efficiency.

Here we propose an extension that constructs adapted period- t approximations, but that features a unique combination of two characteristics. In working with continuous targeted likelihoods, corresponding approximations are also continuous; and period- t supports are adjusted using a method designed to produce approximations that achieve near-optimal efficiency at the adaption stage. The approximations are constructed using the efficient importance sampling (EIS) methodology developed by Richard and Zhang (RZ, 2007). Construction is facilitated using an optimization procedure designed to minimize numerical standard errors associated with the approximated integral.

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., 2008) 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 particle filter 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 $m \times 1$ 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

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

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

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

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 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

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

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

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

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

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

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

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

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; i.e., $f(s_0|Y_0) \equiv f(s_0)$.

In turn, filtering entails the approximation of the conditional (upon Y_t) expectation of some function $h(s_t)$ (including s_t itself). In light of (6) and (4), this can be written as

$$E_t(h(s_t)|Y_t) = \frac{\int h(s_t)f(y_t|s_t, Y_{t-1})f(s_t|Y_{t-1})ds_t}{\int f(y_t|s_t, Y_{t-1})f(s_t|Y_{t-1})ds_t}. \quad (7)$$

3 The Particle Filter and Leading Extensions

Since our procedure is an extension of the particle filter developed by Gordon, Salmond and Smith (1993) and Kitagawa (1996), we provide a brief overview here. The particle filter is an algorithm that recursively generates random numbers approximately distributed as $f(s_t|Y_t)$. To characterize its implementation, let $s_t^{r,i}$ denote the i^{th} draw of s_t obtained from the conditional density $f(s_t|Y_{t-r})$ for $r = 0, 1$. A single draw $s_t^{r,i}$ is a particle, and a set of draws $\{s_t^{r,i}\}_{i=1}^N$ is a swarm of particles. The object of filtration is that of transforming a swarm $\{s_{t-1}^{0,i}\}_{i=1}^N$ to $\{s_t^{0,i}\}_{i=1}^N$. The filter is initialized by a swarm $\{s_0^{0,i}\}_{i=1}^N$ drawn from $f(s_0|Y_0) \equiv f(s_0)$.

Period- t filtration takes as input a swarm $\{s_{t-1}^{0,i}\}_{i=1}^N$. The predictive step consists of transforming this swarm into a second swarm $\{s_t^{1,i}\}_{i=1}^N$ according to (5). This is done by drawing $s_t^{1,i}$ from the conditional density $f(s_t|s_{t-1}^{0,i}, Y_{t-1})$, $i = 1, \dots, N$. Note that $\{s_t^{1,i}\}_{i=1}^N$ can be used to produce an MC estimate of $f(y_t|Y_{t-1})$, which according to (4) is given by

$$\hat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N f(y_t|s_t^{1,i}, Y_{t-1}). \quad (8)$$

Next, $f(s_t|Y_t)$ is approximated by re-weighting $\{s_t^{1,i}\}_{i=1}^N$ in accordance with (6) (the updating step): a particle $s_t^{1,i}$ with prior weight $\frac{1}{N}$ is assigned the posterior weight

$$w_t^{0,i} = \frac{f(y_t|s_t^{1,i}, Y_{t-1})}{\sum_{j=1}^N f(y_t|s_t^{1,j}, Y_{t-1})}. \quad (9)$$

The filtered swarm $\{s_t^{0,i}\}_{i=1}^N$ is then obtained by drawing with replacement from the swarm $\{s_t^{1,i}\}_{i=1}^N$ with probabilities $\{w_t^{0,i}\}_{i=1}^N$ (i.e., bootstrapping).

Having characterized the particle filter, its strengths and weaknesses (well documented in previous studies) can be pinpointed. Its strength lies in its simplicity: the algorithm described above is straightforward and universally applicable.

Its weaknesses are twofold. First, it provides discrete approximations of $f(s_t|Y_{t-1})$ and $f(s_t|Y_t)$, which moreover are discontinuous functions of the model parameters. The associated likelihood approximation is therefore also discontinuous, rendering the application of maximization routines problematic (a point raised previously, e.g., by Pitt, 2002).

Second, as the filter enters period t , the discrete approximation of $f(s_{t-1}|Y_{t-1})$ is set. Hence the swarm $\{s_t^{1,i}\}_{i=1}^N$ produced in the augmentation stage ignores information provided by y_t . (Pitt and Shephard, 1999, refer to these augmenting draws as “blind”.) It follows that if $f(y_t|s_t, Y_{t-1})$ - treated as a function of s_t given Y_t - is sharply peaked in the tails of $f(s_t|Y_{t-1})$, $\{s_t^{1,i}\}_{i=1}^N$ will contain few elements in the relevant range of $f(y_t|s_t, Y_{t-1})$. Thus $\{s_t^{1,i}\}_{i=1}^N$ represents draws from an inefficient sampler: relatively few of its elements will be assigned appreciable weight in the updating stage in the following period. This is known as “sample impoverishment”: it entails a reduction in the effective size of the particle swarm.

Extensions of the particle filter employ adaption techniques to generate gains in efficiency. An extension proposed by Gordon et al. (1993) and Kitagawa (1996) consists simply of making $N' \gg N$ blind proposals $\{s_t^{1,j}\}_{j=1}^{N'}$ as with the particle filter, and then obtaining the swarm $\{s_t^{0,i}\}_{i=1}^N$ by sampling with replacement, using weights computed from the N' blind proposals. This is the sampling-importance resampling filter; it seeks to overcome the problem of sample impoverishment by brute force, and can be computationally expensive.

Carpenter, Clifford and Fearnhead (1999) sought to overcome sample impoverishment using

a stratified sampling approach to approximate the prediction density. This is accomplished by defining a partition consisting of K subintervals in the state space, and constructing the prediction density approximation by sampling (with replacement) N_k particles from among the particles in each subinterval. Here N_k is proportional to a weight defined for the entire k^{th} interval; also, $\sum_{k=1}^K N_k = N$. This produces wider variation in re-sampled particles, but if the swarm of proposals $\{s_t^{1,i}\}_{i=1}^N$ are tightly clustered in the tails of $f(s_t|Y_{t-1})$, so too will be the re-sampled particles.

Pitt and Shephard (1999) developed an extension that ours perhaps most closely resembles. They tackle adaption using an Importance Sampling (IS) procedure. Consider as an example the marginalization step. Faced with the problem of calculating $f(y_t|Y_{t-1})$ in (4), but with $f(s_t|Y_{t-1})$ unknown, importance sampling achieves approximation via the introduction into the integral of an importance density $g(s_t|Y_t)$:

$$f(y_t|Y_{t-1}) = \int \frac{f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})}{g(s_t|Y_t)} g(s_t|Y_t) ds_t. \quad (10)$$

Obtaining drawings $s_t^{0,i}$ from $g(s_t|Y_t)$, this integral is approximated as

$$\hat{f}(y_t|Y_{t-1}) \approx \frac{1}{N} \sum_{i=1}^N \frac{f(y_t|s_t^{0,i}, Y_{t-1}) f(s_t^{0,i}|Y_{t-1})}{g(s_t^{0,i}|Y_t)}. \quad (11)$$

Pitt and Shephard referred to the introduction of $g(s_t|Y_t)$ as adaption. Full adaption is achieved when $g(s_t|Y_t)$ is constructed as being proportional to $f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})$, rendering the ratios in (11) as constants. Pitt and Shephard viewed adaption as computationally infeasible, due to the requirement of computing $f(s_t^{0,i}|Y_{t-1})$ for every value of $s_t^{0,i}$ produced by the sampler. Instead they developed samplers designed to yield partial adaption.

The samplers result from Taylor series approximations of $f(y_t|s_t, Y_{t-1})$ around $s_t = \mu_t^k = E(s_t|s_{t-1}^{0,k}, Y_{t-1})$. A zero-order expansion yields their auxiliary particle filter; a first-order expansion yields their adapted particle filter. (Smith and Santos, 2006, study examples under which it is possible to construct samplers using second-order expansions.)

These samplers help alleviate blind sampling by reweighting $\{s_{t-1}^{0,i}\}$ to account for information conveyed by y_t . However, sample impoverishment can remain an issue, since the algorithm does not allow adjustment of the support of $\{s_{t-1}^{0,i}\}$. Moreover, the samplers are suboptimal, since μ_t^k

is incapable of fully capturing the characteristics of $f(y_t|s_t, Y_{t-1})$. Finally, these samplers remain prone to the discontinuity problem.

Pitt (2002) addressed the discontinuity problem for the special case in which the state space is unidimensional by replacing the weights in (9) associated with the particle filter (or comparable weights associated with the auxiliary particle filter) with smoothed versions constructed via a piecewise linear approximation of the empirical c.d.f. associated with the swarm $\left\{s_t^{0,i}\right\}_{i=1}^N$. This enables the use of common random numbers (CRNs) to produce likelihood estimates that are continuous functions of model parameters (Hendry, 1994).

4 The EIS Filter

EIS is an automated procedure for constructing continuous importance samplers fully adapted as global approximations to targeted integrands. Section 4.1 outlines the general principle behind EIS, in the context of evaluating (4). Section 4.2 then discusses a key contribution of this paper: the computation of $f(s_t|Y_{t-1})$ in (4) at auxiliary values of s_t generated under period- t EIS optimization. Section 4.3 discusses a special case that often characterizes state-space representations: degenerate transition densities. Elaboration on the pseudo-code presented below is available at www.pitt.edu/~dejong/wp.htm.

4.1 EIS integration

Let $\varphi_t(s_t) = f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$ in (4), where the subscript t in φ_t replaces (y_t, Y_{t-1}) . Implementation of EIS begins with the preselection of a parametric class $K = \{k(s_t; a_t); a_t \in A\}$ of auxiliary density kernels. Corresponding density functions g are

$$g(s_t; a_t) = \frac{k(s_t; a_t)}{\chi(a_t)}, \quad \chi(a_t) = \int k(s_t; a_t) ds_t. \tag{12}$$

The selection of K is problem-specific; here we discuss Gaussian specifications; DeJong et al. (2008) discusses an extension to piecewise-continuous specifications. The objective of EIS is to select the parameter value $\hat{a}_t \in A$ that minimizes the variance of the ratio $\frac{\varphi_t(s_t)}{g(s_t|a_t)}$ over the range of integration.

Following RZ, a (near) optimal value \hat{a}_t is obtained as the solution to

$$(\hat{a}_t, \hat{c}_t) = \arg \min_{a_t, c_t} \int [\ln \varphi_t(s_t) - c_t - \ln k(s_t; a_t)]^2 g(s_t; a_t) ds_t, \quad (13)$$

where c_t is an intercept meant to calibrate $\ln(\varphi_t/k)$. Equation (13) is a standard least squares problem, except that the auxiliary sampling density itself depends upon a_t . This is resolved by reinterpreting (13) as the search for a fixed-point solution. An operational MC version implemented (typically) using $R \ll N$ draws, is as follows:

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

$$(\hat{a}_t^{l+1}, \hat{c}_t^{l+1}) = \arg \min_{a_t, c_t} \sum_{i=1}^R [\ln \varphi_t(s_{t,l}^i) - c_t - \ln k(s_{t,l}^i; a_t)]^2. \quad (14)$$

If K belongs to the exponential family of distributions, there exists a parameterization a_t such that the auxiliary problems in (14) are linear. Full details on their solution are provided in the posted psuedo-code noted above.

Three technical points bear mentioning here. First, the evaluation of $\varphi_t(s_t)$ entails the evaluation of $f(s_t|Y_{t-1})$, which is unavailable analytically and must be approximated; this is discussed below in Section 4.2. Second, the selection of the initial value \hat{a}_t^1 is important for achieving rapid convergence; Section 5 presents an effective algorithm for specifying \hat{a}_t^1 in applications involving DSGE models (one step in each of the examples we consider). Third, to achieve rapid convergence, and to ensure continuity of corresponding likelihood estimates, $\{s_{t,j}^i\}$ must be obtained by a transformation of a set of common random numbers (CRNs) $\{u_t^i\}$ drawn from a canonical distribution (i.e., one that does not depend on a_t ; e.g., standardized Normal draws when g is Gaussian).

An additional substantive point also bears mentioning. At convergence to \hat{a}_t , the EIS sampler $g(s_t; \hat{a}_t)$ not only provides the optimal global approximation to the targeted integrand $\varphi_t(s_t) = f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$, but also serves as the optimized approximation to the time- t filtering density $f(s_t|Y_t)$. Thus as with the particle filter, the EIS filter facilitates likelihood evaluation and filtering simultaneously.

At convergence to \widehat{a}_t , the EIS filter approximation of $f(y_t|Y_{t-1})$ in (4) is given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N \omega(s_t^i; \widehat{a}_t), \quad (15)$$

$$\omega(s_t^i; \widehat{a}_t) = \frac{f(y_t|s_t^i, Y_{t-1}) f(s_t^i|Y_{t-1})}{g(s_t^i; \widehat{a}_t)}, \quad (16)$$

where $\{s_t^i\}_{i=1}^N$ are drawn from the (final) EIS sampler $g(s_t; \widehat{a}_t)$. This estimate converges almost surely towards $f(y_t|Y_{t-1})$ under weak regularity conditions (outlined, e.g., by Geweke, 1989). Violations of these conditions typically result from the use of samplers with thinner tails than those of φ_t . RZ offer a diagnostic measure that is adept at detecting this problem. The measure compares the MC sampling variances of the ratio $\frac{\varphi_t}{g}$ under two values of a_t : the optimal \widehat{a}_t , and one that inflates the variance of the s_t draws by a factor of 3 to 5.

Pseudo-code for implementing the EIS filter is as follows:

- At period t , we inherit the sampler $g(s_{t-1}; \widehat{a}_{t-1})$, and corresponding draws and weights $\{s_{t-1}^i, \omega(s_{t-1}^i; \widehat{a}_{t-1})\}_{i=1}^N$ from period $t-1$, where in period 0 $g(s_0; \widehat{a}_0) \equiv f(s_0)$.
- Using an initial value \widehat{a}_t^1 , obtain R draws $\{s_t^{i,l}\}_{i=1}^R$ from $g(s_t; \widehat{a}_t^1)$, and solve (14) to obtain \widehat{a}_t^2 . Repeat until convergence, yielding \widehat{a}_t .
- Obtain N values $\{s_t^i\}_{i=1}^N$ from the optimized sampling density $g(s_t; \widehat{a}_t)$, and calculate (15).
- Pass $g(s_t; \widehat{a}_t)$ and $\{s_t^i, \omega(s_t^i; \widehat{a}_t)\}_{i=1}^N$ to period $t+1$. Repeat until period T is reached.

As we shall now explain, $\{s_t^i, \omega(s_t^i; \widehat{a}_t)\}_{i=1}^N$ are passed from period t to $t+1$ to facilitate the approximation of the unknown $f(s_t|Y_{t-1})$ appearing in (14) and (16).

4.2 Continuous approximations of $f(s_t|Y_{t-1})$

As noted, the EIS filter requires the evaluation of $f(s_t|Y_{t-1})$ at any value of s_t needed for EIS iterations. Here we discuss three operational alternatives for overcoming this hurdle (a fourth, involving mixture approximations, is under development). Below, S denotes the number of points used for each individual evaluation of $f(s_t|Y_{t-1})$.

Weighted-sum approximations

Combining (5) and (6), we can rewrite $f(s_t|Y_{t-1})$ as a ratio of integrals:

$$f(s_t|Y_{t-1}) = \frac{\int f(s_t|s_{t-1}, Y_{t-1})f(y_{t-1}|s_{t-1}, Y_{t-2})f(s_{t-1}|Y_{t-2})ds_{t-1}}{\int f(y_{t-1}|s_{t-1}, Y_{t-2})f(s_{t-1}|Y_{t-2})ds_{t-1}}, \quad (17)$$

where the denominator represents the likelihood integral for which an EIS sampler has been constructed in period $t - 1$. A direct MC estimate of $f(s_t|Y_{t-1})$ is given by

$$\hat{f}_S(s_t|Y_{t-1}) = \frac{\sum_{i=1}^S f(s_t|s_{t-1}^i, Y_{t-1}) \cdot \omega(s_{t-1}^i; \hat{a}_{t-1})}{\sum_{i=1}^S \omega(s_{t-1}^i; \hat{a}_{t-1})}, \quad (18)$$

where $\{s_{t-1}^i\}_{i=1}^S$ denotes EIS draws from $g(s_{t-1}|\hat{a}_{t-1})$, and $\{\omega(s_{t-1}^i; \hat{a}_{t-1})\}_{i=1}^S$ denotes associated weights (both of which are carried over from period- $t - 1$).

Obviously $g(s_{t-1}|\hat{a}_{t-1})$ is not an EIS sampler for the numerator in (17). This can impart a potential loss of numerical accuracy if the MC variance of $f(s_t|s_{t-1}, Y_{t-1})$ is large over the support of $g(s_{t-1}|\hat{a}_{t-1})$. This would be the case if the conditional variance of $s_t|s_{t-1}, Y_{t-1}$ were significantly smaller than that of $s_{t-1}|Y_{t-1}$. But the fact that we are using the same set of draws for the numerator and the denominator typically creates positive correlation between their respective MC estimators, thus reducing the variance of their ratio.

A constant weight approximation

When EIS delivers a close global approximation to $f(s_{t-1}|Y_{t-1})$, the weights $\omega(s_{t-1}; \hat{a}_{t-1})$ will be near constants over the range of integration. Replacing these weights by their arithmetic means $\bar{\omega}(\hat{a}_{t-1})$ in (17) and (18), we obtain the following simplification:

$$f(s_t|Y_{t-1}) \simeq \int f(s_t|s_{t-1}, Y_{t-1}) \cdot g(s_{t-1}; \hat{a}_{t-1})ds_{t-1}. \quad (19)$$

This substitution yields rapid implementation if additionally the integral in (19) has an analytical solution. This will be the case if, e.g., $f(s_t|s_{t-1}, Y_{t-1})$ is a conditional normal density for $s_t|s_{t-1}$, and g is also normal. In cases for which we lack an analytical solution, we can use the standard

MC approximation

$$\widehat{f}_S(s_t|Y_{t-1}) \simeq \frac{1}{S} \sum_{i=1}^S f(s_t|s_{t-1}^i, Y_{t-1}). \quad (20)$$

EIS evaluation

Evaluation of $f(s_t|Y_{t-1})$ can sometimes be delicate, including situations prone to sample impoverishment (such as when working with degenerate transitions, discussed below). Under such circumstances, one might consider applying EIS not only to the likelihood integral (“outer EIS”), but also to the evaluation of $f(s_t|Y_{t-1})$ itself (“inner EIS”).

While outer EIS is applied only once per period, inner EIS must be applied for every value of s_t generated by the former. Also, application of EIS to (5) requires the construction of a continuous approximation to $f(s_{t-1}|Y_{t-1})$. Two obvious candidates are as follows. The first is the period- $(t-1)$ EIS sampler $g(s_{t-1}; \widehat{a}_{t-1})$, under the implicit assumption that the corresponding weights $\omega(s_{t-1}; \widehat{a}_{t-1})$ are near-constant, at least over the range of integration. The second is the use of a more flexible sampler, such as a mixture of Gaussian densities.

4.3 Degenerate transitions

When state transition equations include identities, corresponding transition densities are degenerate (or Dirac) in some of their components; this requires an adjustment to EIS implementation. Let s_t partition into $s_t = (p_t, q_t)$, such that we have a proper transition density $f(p_t|s_{t-1}, Y_{t-1})$ for p_t , and an identity for $q_t|p_t, s_{t-1}$ (which could also depend on Y_{t-1} , omitted here for ease of notation):

$$q_t \equiv \phi(p_t, p_{t-1}, q_{t-1}) = \phi(p_t, s_{t-1}). \quad (21)$$

The evaluation of $f(s_t|Y_{t-1})$ in (5) now requires special attention, since its evaluation at a given s_t (as selected by the EIS algorithm) requires integration in the strict subspace associated with identity (21). Note in particular that the presence of identities raises a conditioning issue known as the Borel-Kolmogorov paradox (e.g., see DeGroot, 1975, Section 3.10). We resolve this issue here by reinterpreting (21) as the limit of a uniform density for $q_t|p_t, s_{t-1}$ on the interval $[\phi(p_t, s_{t-1}) - \varepsilon, \phi(p_t, s_{t-1}) + \varepsilon]$.

Assuming that $\phi(p_t, s_{t-1})$ is differentiable and strictly monotone in q_{t-1} , with inverse

$$q_{t-1} = \psi(p_t, q_t, p_{t-1}) = \psi(s_t, p_{t-1}) \quad (22)$$

we can take the limit of the integral in (5) as ε tends to zero, producing

$$f(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|s_{t-1}, Y_{t-1}) f(p_{t-1}, q_{t-1}|Y_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}, \quad (23)$$

where with $\|\cdot\|$ denoting the absolute value of a determinant,

$$J(s_t, p_{t-1}) = \left\| \frac{\partial}{\partial q_t} \psi(s_t, p_{t-1}) \right\|. \quad (24)$$

Note that (23) requires that for any s_t , $f(s_{t-1}|Y_{t-1})$ must be evaluated along the zero-measure subspace $q_{t-1} = \psi(s_t, p_{t-1})$. This rules out use of the weighted-sum approximation introduced above, since the probability that any of the particles $s_{t-1}^{0,i}$ lies in that subspace is zero. Instead, we can approximate (23) by replacing $f(s_{t-1}|Y_{t-1})$ by $\bar{w}(\hat{a}_{t-1}) g(s_{t-1}|\hat{a}_{t-1})$:

$$\hat{f}(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|q_{t-1}, Y_{t-1}) g(p_{t-1}, q_{t-1}|\hat{a}_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}. \quad (25)$$

In this case, since $g(p_{t-1}, \psi(s_t, p_{t-1})|\hat{a}_{t-1})$ is not a sampler for $p_{t-1}|s_t$, we must evaluate (25) either by quadrature or its own EIS sampler.

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 particle filter 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. Details follow.

5.1 Example 1: Two-State RBC Model

The first application is to the simple DSGE 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 :

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

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 ζ_t , which may be consumed or invested (we use ζ 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

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

where $(\alpha, \delta, g, \rho)$ represent capital's share of output, the depreciation rate of capital, the growth

rate of total factor productivity (TFP), and the persistence of innovations to TFP.

Optimal household behavior is represented as policy functions for $(\zeta_t, c_t, n_t, l_t, i_t)$ in terms of the state (k_t, z_t) . Given the policy function $i(k_t, z_t)$, the state-transitions equations reduce to

$$\left(1 + \frac{g}{1 - \alpha}\right) k_t = i(k_{t-1}, z_{t-1}) + (1 - \delta)k_{t-1} \quad (26)$$

$$\ln z_t = (1 - \rho) \ln(z_0) + \rho \ln z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (27)$$

and the observation equations are

$$x_t = x(k_t, z_t) + u_{x,t}, \quad x = \zeta, i, n, \quad u_{x,t} \sim N(0, \sigma_x^2). \quad (28)$$

Policy functions are expressed as Chebyshev polynomials in the state variables (k_t, z_t) , constructed using the projection method described in DeJong and Dave (2007, Ch. 10.5.2).

Given the form of (27), it will be useful represent state variables as logged deviations from steady state: $s_t = [\ln(k_t/k^*) \quad \ln(z_t/z^*)]'$. For ease of notation, hereafter we will denote $\ln(k_t/k^*)$ as k_t , and $\ln(z_t/z^*)$ as z_t . In addition, given the form of (28), y_t is defined as $y_t = [\zeta_t \quad i_t \quad n_t]'$. All subsequent formulas should be read in accordance with these representations.

To obtain the predictive density associated with (26) and (27), note that since (26) is an identity, the transition density of the system is degenerate in k_t . Thus we invert (26) to obtain

$$k_{t-1} = \psi(k_t, z_{t-1}), \quad (29)$$

$$J(k_t, z_{t-1}) = \left| \frac{\partial}{\partial k_t} \psi(k_t, z_{t-1}) \right|, \quad (30)$$

and express the predictive density as

$$f(s_t | Y_{t-1}) = \int J(k_t, z_{t-1}) f(z_t | s_{t-1}) f(s_{t-1} | Y_{t-1}) |_{k_{t-1} = \psi(k_t, z_{t-1})} dz_{t-1}. \quad (31)$$

From (27), note that

$$f(z_t | s_{t-1}) \sim N_1 \left(\left[\begin{array}{c} 0 \\ \rho \end{array} \right] s_{t-1}, \sigma_\varepsilon^2 \right),$$

with $N_r(\cdot)$ denoting an r -dimensional normal distribution. Finally, as the inversion of Chebyshev

polynomials is awkward, we approximate (29) and (30) using third-order polynomials in (k_t, z_{t-1}) .

With the predictive density established, the time- t likelihood is standard:

$$f(y_t|Y_{t-1}) = \int f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t, \quad (32)$$

where from (??)-(28),

$$f(y_t|s_t, Y_{t-1}) \sim N_3(\mu(s_t), V), \quad (33)$$

$$\mu(s_t) = \begin{bmatrix} \zeta(s_t) \\ i(s_t) \\ n(s_t) \end{bmatrix}, \quad V = \begin{bmatrix} \sigma_y^2 & 0 & 0 \\ 0 & \sigma_i^2 & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}.$$

To achieve likelihood evaluation in period t , our approach is to construct a normally distributed EIS sampler $g(s_t; \widehat{a}_t)$ for the integrand $f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$ in (32). In so doing, $f(s_t|Y_{t-1})$ is represented using the constant-weight approach to approximation described above. That is, we use the time- $(t-1)$ sampler $g(s_{t-1}; \widehat{a}_{t-1})$ as a stand-in for $f(s_{t-1}|Y_{t-1})$, yielding

$$f(s_t|Y_{t-1}) \simeq \int J(k_t, z_{t-1}) f(z_t|s_{t-1}) g(s_{t-1}; \widehat{a}_{t-1})|_{k_{t-1}=\psi(k_t, z_{t-1})} dz_{t-1}. \quad (34)$$

We initialize the process by constructing $f(s_0)$ as the unconditional distribution of the Kalman filter associated with a linear approximation of the model. We proceed via forward recursion, taking $g(s_{t-1}; \widehat{a}_{t-1})$ as an input, and passing $g(s_t; \widehat{a}_t)$ to the subsequent period. Full details follow.

Consider first the evaluation of (34). Representing $g(s_t; \widehat{a}_t)$ as

$$g(s_t; \widehat{a}_t) \sim N_2(\mu, \Omega),$$

and with $f(z_t|s_{t-1})$ distributed as

$$f(z_t|s_{t-1}) \sim N_1\left(\begin{bmatrix} 0 & \rho \end{bmatrix} s_{t-1}, \sigma_\varepsilon^2\right),$$

$f(z_t|s_{t-1})g(s_{t-1}; \widehat{a_{t-1}})$ combines to form the joint density

$$\begin{pmatrix} z_t \\ s_{t-1} \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 0 & \rho \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \mu, \quad V \right), \quad (35)$$

$$V = \sigma_\varepsilon^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & \rho \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \Omega \begin{pmatrix} 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix}.$$

As (34) must be evaluated for each candidate s_t used to calculate (32), we must transform (35) into a distribution over $(s_t, z_{t-1})'$. Approximating (29) linearly as $k_{t-1} = a_k k_t + b_k z_{t-1}$, which implies

$$\begin{pmatrix} s_t \\ z_{t-1} \end{pmatrix} = A \begin{pmatrix} z_t \\ s_{t-1} \end{pmatrix},$$

$$A = \begin{pmatrix} 0 & 1/a_k & -b_k/a_k \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad |A| = 1/|a_k|,$$

we can express $(z_t \quad s_{t-1})'$ as a function of $(s_t \quad z_{t-1})'$. This expression, coupled with (35), yields the joint density

$$f^*(s_t, z_{t-1}) \sim N_3(m, \Sigma), \quad \Sigma = AVA', \quad (36)$$

$$m = B\mu,$$

$$B = A \begin{bmatrix} 0 & \rho \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Finally, we partition (36) into a product of two densities, one for s_t and one for $z_{t-1}|s_t$:

$$f_1^*(s_t) \sim N_2(m_1, \Sigma_{11}), \quad f_2^*(z_{t-1}|s_t) \sim N_1(m_{2.1} + \Delta_{21}s_t, \sigma_{22.1}),$$

where

$$m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \sigma_{22} \end{pmatrix},$$

with Σ_{11} 2×2 , Σ_{12} 2×1 , Σ_{21} 1×2 , and σ_{22} 1×1 , and

$$\Delta_{21} = \Sigma_{21} (\Sigma_{11})^{-1}, \quad m_{2,1} = m_2 - \Delta_{21} m_1, \quad \sigma_{22,1} = \sigma_{22} - \Sigma_{21} (\Sigma_{11})^{-1} \Sigma_{12}.$$

Having accomplished these steps, (34) is approximately

$$f(s_t | Y_{t-1}) \approx f_1^*(s_t) \int \frac{J(s_t, z_{t-1})}{|a_k|} f_2^*(z_{t-1} | s_t) dz_{t-1}, \quad (37)$$

where since

$$|\Sigma|^{-1/2} = \frac{1}{|a_k|} |V|^{-1/2},$$

the term $|a_k|$ enters via the usual change-of-variables formula. For each candidate s_t that enters into the approximation of (32), we use $f_2^*(z_{t-1} | s_t)$ as an EIS sampler, and approximate (37) as

$$f(s_t | Y_{t-1}) \approx \left(\frac{f_1^*(s_t)}{|a_k|} \right) \frac{1}{S} \sum_{i=1}^S J(s_t, z_{t-1}^i), \quad (38)$$

where $\{z_{t-1}^i\}_{i=1}^S$ are simulated drawings from $f_2^*(z_{t-1} | s_t)$.

Turning to the approximation of (32), this is straightforward once a reliable initial EIS sampler is constructed. To construct this initial sampler, we seek a close approximation of the integrand $f(y_t | s_t, Y_{t-1}) \cdot f(s_t | Y_{t-1})$. Towards this end, we use $f_1^*(s_t)$ in place of $f(s_t | Y_{t-1})$, and construct a linear Gaussian approximation of $f(y_t | s_t, Y_{t-1})$. Recall from (33) that $f(y_t | s_t, Y_{t-1})$ is already Gaussian, but with a non-linear mean function, say $\mu(s_t)$. Approximating this function as $\mu(s_t) \approx r + P s_t$, we obtain

$$f^*(y_t | s_t, Y_{t-1}) \sim N_3(r + P s_t, \quad V).$$

Combining these approximations, and letting $Q = V^{-1}$, $H = \Sigma_{11}^{-1}$, we obtain the initial sampler

$$\begin{aligned}
g(s_t; a_t^0) &= f_1^*(s_t) f^*(y_t | s_t, Y_{t-1}) \\
&\sim N(\mu^0, \Sigma^0), \\
\mu^0 &= (H + P'QP)^{-1} [Hm_1 + P'Q(y_t - r)], \\
\Sigma^0 &= (H + P'QP)^{-1}.
\end{aligned} \tag{39}$$

To summarize, EIS implementation is achieved for the two-state RBC model as follows.

Model Representation

- Policy functions $x(k_t, z_t)$, $x = (\zeta, c, i, n, l)$, expressed as Chebyshev polynomials in $s_t = [k_t \ z_t]'$, are constructed via the projection method.
- With the law of motion for capital given by $k_t = i(k_{t-1}, z_{t-1}) + (1 - \delta)k_{t-1}$, we solve for k_{t-1} to obtain

$$k_{t-1} = \psi(k_t, z_{t-1}), \quad J(k_t, z_{t-1}) = \frac{\partial}{\partial k_t} \psi(k_t, z_{t-1}),$$

represented as third-order polynomials in (k_t, z_{t-1}) . We also construct the linear approximation $k_{t-1} = a_k k_t + b_k z_{t-1}$.

Likelihood Evaluation

- The EIS sampler $g(s_{t-1}; \widehat{a}_{t-1})$ serves as an input in the construction of the period- t likelihood function. In period 1, $g(s_0; \widehat{a}_0) \equiv \widehat{f}(s_0)$ is constructed as the unconditional distribution of the Kalman filter associated with a linear approximation of the model.
- To approximate the integrand $f(y_t | s_t, Y_{t-1}) \cdot f(s_t | Y_{t-1})$ in (32), we construct the initial sampler $g(s_t; a_t^0)$ as in (39):

$$g(s_t; a_t^0) = f_1^*(s_t) f^*(y_t | s_t, Y_{t-1})$$

- Using drawings $\{s_{t,0}^i\}_{i=1}^R$ obtained from $g(s_t; a_t^0)$, we construct \widehat{a}_t as the solution to (14).

This entails the computation of

$$\varphi_t(s_{t,l}^i) = f(y_t | s_{t,l}^i, Y_{t-1}) f(s_{t,l}^i | Y_{t-1}),$$

where $f(y_t | s_t, Y_{t-1})$ is given in (33), and $f(s_t | Y_{t-1})$ is approximated as indicated in (38).

- Having constructed $g(s_t; \hat{a}_t)$, $\hat{f}(y_t | Y_{t-1})$ is approximated as indicated in (15).
- The sampler $g(s_t; \hat{a}_t)$ is passed to the period- $(t+1)$ step of the algorithm. The algorithm concludes with the completion of the period- T step.

To demonstrate the performance of the EIS filter in this setting, we conducted Monte Carlo experiments using two data sets. The first is an artificial data set consisting of 100 realizations of $\{\zeta_t, i_t, n_t\}$ generated from the RBC model. This was constructed by Fernandez-Villaverde and Rubio-Ramirez (2005) under the model parameterization presented in the second row of Table 1. The second consists of actual quarterly observations on $\{\zeta_t, i_t, n_t\}$ used by Fernandez-Villaverde and Rubio-Ramirez to estimate the RBC model using the particle filter. The data are quarterly, span 1964:I-2003:II (158 observations), and were detrended using the Hodrick-Prescott filter. Posterior means of the estimates they obtained using this data set are presented in the third row of Table 1. Both data sets are available for downloading at

<http://qed.econ.queensu.ca/jae/2005-v20.7/fernandez-rubio/>

Table 1. Parameter Values, RBC Model

| | α | β | ϕ | φ | δ | g | ρ | σ_ε | σ_y | σ_i | σ_n |
|------------|----------|---------|--------|-----------|----------|-----|--------|----------------------|------------|------------|------------|
| Artificial | 0.4 | 0.99 | 2 | 0.357 | 0.02 | 0 | 0.95 | 0.007 | 0.000158 | 0.000866 | 0.0011 |
| Actual | 0.324 | 0.997 | 1.717 | 0.390 | 0.006 | 0 | 0.978 | 0.020 | 0.045 | 0.038 | 0.015 |

Each data set poses a distinct challenge to efficient filtering. In the artificial data set, note from Table 1 that the standard deviations of the measurement errors ($\sigma_y, \sigma_i, \sigma_n$) are small relative to σ_ε , which as noted above can lead to problems associated with sample impoverishment. In the real data set, the investment series contains two outliers: the values at 1976:III and 1984:IV, which lie 7.7 and 4.7 standard deviations above the sample mean. Outliers can induce bias in likelihood estimates associated with the particle filter. Both of these challenges are overcome via implementation of the EIS filter, as we now demonstrate.

Using both data sets, we conducted a Monte Carlo experiment under which we produced 1,000 approximations of the likelihood function (evaluated at Table-1 parameter values) for both the particle and EIS filters using 1,000 different sets of random numbers. Differences in likelihood approximations across sets of random numbers are due to numerical approximation errors. Following Fernandez-Villaverde and Rubio-Ramirez, the particle filter was implemented using $N = 60,000$, requiring 40.6 seconds of CPU time per likelihood evaluation on a 1.8 GHz desktop computer using GAUSS for the artificial data set, and 63 seconds for the real data set. The EIS filter was implemented using $N = R = 20$, $S = 10$, with one iteration used to construct \hat{a}_t ; this required 0.22 seconds per likelihood evaluation for the artificial data set, and 0.328 seconds for the real data set.

Considering first the artificial data set, the mean and standard deviation of the 1,000 log-likelihood approximations obtained using the particle filter are (1, 285.51, 33.48), and (1, 299.81, 0.00177) using the EIS filter (the likelihood value obtained using the Kalman filter is 1, 300.045). Thus the EIS filter reduces numerical approximation errors by four orders of magnitude in this application. Figure 1 plots the first 200 likelihood approximations obtained using both filters (in order to enhance visibility). Note that the particle-filter approximations (top panel) often fall far below the EIS sample mean of 1, 299.81 (by more than 50 on the log scale in twenty instances, and by more than 100 in eight instances); this largely accounts for the distinct difference in sample means obtained across methods. But as the figure indicates, even abstracting from the occasional large likelihood crashes suffered by the particle filter, the EIS filter is extremely precise: the maximum difference in log-likelihood values it generates is less than 0.012 (bottom panel), while differences of 10 are routinely observed for the particle filter.

Hereafter, we shall refer to differences observed between sample means of log-likelihood values obtained using the particle and EIS filters as reflecting bias associated with the particle filter. This presumes that the values associated with the EIS filter closely represent “truth”. This presumption is justified in a number of ways, in this experiment and each of those that follow. First, the small numerical approximation errors associated with the EIS filter indicate virtual replication across sets of random numbers. Second, as we increase the values of (N, R) used to implement the EIS filter, resulting mean log-likelihood approximations remain virtually unchanged, while numerical errors are inversely proportional to $N^{-1/2}$, as expected. Finally, when we implement the EIS filter using linear model approximations, the log-likelihood values we obtain match those produced by

the Kalman filter almost exactly (virtually to the limits of numerical precision).

Figure 2 provides an illustration and diagnosis of the problems faced by the particle filter in this application. The focus here is on a representative Monte Carlo replication generated as described above. The figure contains two panels, each of which represents a distinct scenario observed routinely across time periods within this replication. The top panel corresponds with $t = 53$, the bottom with $t = 18$.

Each panel contains two graphs, both of which depict z_t on the vertical axis and k_t on the horizontal axis. The measurement density $f(y_t|s_t, Y_{t-1})$ is the large thin ellipse depicted in both graphs (differences in vertical scales across graphs account for differences in its appearance). In the bottom graph, the swarm of dots comprises the particle-filter representation of $f(s_t|Y_{t-1})$, and the wide ellipse comprises the EIS representation of $f(s_t|Y_{t-1})$. In the upper graph, the swarm of dots comprises the particle-filter representation of $f(s_t|Y_t)$; particles in the upper swarm were obtained by sampling repeatedly from the bottom swarm, with probabilities assigned by the measurement density. The upper graph also depicts the EIS representation of $f(s_t|Y_t)$ (small ellipse).

Beginning with period 53, note that the vast majority of particles in the bottom graph are assigned negligible weight by the measurement density, and are thus discarded in the resampling step. Specifically, only 407 particles, or 0.68% of the total candidates, were re-sampled at least once in this instance. The average (across time periods) number of re-sampled particles is 350, or 0.58% of the total. This phenomenon reflects the sample impoverishment problem noted above. It results from the ‘blindness’ of proposals generated under the particle filter algorithm, and accounts for its numerical inaccuracy.

As noted, the small ellipse depicted in the upper graph is the EIS representation of $f(s_t|Y_t)$. The difference between this and the corresponding particle-filter representation reflects a second problem suffered by the particle filter in this application: there is non-trivial bias in the filtered values of the state it produces. This also reflects the ‘blindness’ problem, coupled with the fact that alternative proposals for s_t cannot be re-generated in light of information embodied in y_t . (Note from the figure that this bias is not easily eliminated through an increase in the number of particles included in the proposal swarm, since the probability that the proposal density will generate particles centered on the EIS representation of $f(s_t|Y_t)$ is clearly miniscule.) As described above, under suitable initialization the EIS filter avoids these issues by generating proposals from an importance density

tailored as the optimal global approximation of the targeted integrand $f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$.

Regarding period 18, note that the representations of both $f(s_t|Y_{t-1})$ and $f(s_t|Y_t)$ generated using the particle filter are discontinuous in k_t , a spurious phenomenon that occurs frequently through the sample. This exacerbates the bias associated with filtered values of the state, and contributes to a final problem associated with the particle filter illustrated in Figure 3.

Like its predecessor, Figure 3 was produced using a representative Monte Carlo replication. It depicts an approximation of the log-likelihood surface over α obtained by holding the remaining parameters fixed at their true values, and varying α above and below its true value. Three surfaces are depicted: those associated with the particle, EIS, and Kalman filters (the latter obtained using a linear model representation). The particle and EIS surfaces were produced with common random numbers, so that changes in α serve as the lone source of variation in log-likelihoods.

Note that while the surfaces associated with the EIS and Kalman filters are continuous and peak at the true value of 0.4, the surface associated with the particle filter is discontinuous and has a slightly rightshifted peak. Thus in addition to being numerically inefficient and producing biased filtered values of the state, the particle filter generates likelihood surfaces that are spuriously discontinuous in the underlying parameters of the model, rendering as problematic the attainment of likelihood-based model estimates.

Turning to the experiment conducted using the actual data set, we begin by noting that in this case there are non-trivial differences between the log-likelihood values associated with the Kalman and EIS filters. The mean and standard deviation of the 1,000 log-likelihood approximations obtained using the EIS filter are (921.56, 0.056), while the log-likelihood value associated with the Kalman filter is 928.06. The explanation for this difference is as follows. Since there are no outliers in the artificial data set, deviations from steady state are relatively small, thus the linear model approximations employed in implementing the Kalman filter are relatively accurate. However, accuracy breaks down when deviations from steady state are large (i.e., in the presence of outliers). Indeed, when the EIS filter is implemented using linear model approximations, differences in likelihoods produced by the EIS and Kalman filters virtually disappear (becoming at most $1.58e-9$ in 1976:IV).

Second, compared with the first example, differences between the particle and EIS filters are relatively modest in this case. The mean and standard deviation of the 1,000 log-likelihood ap-

proximations obtained using the particle filter are (926.84, 0.1958). Thus the difference in sample means is $5\frac{1}{4}$ in this case, compared with more than 14 in the first example, while the difference in sample means is by a factor of 3.5, as opposed to four orders of magnitude in the first example. The explanation for this is that measurement densities are relatively diffuse in this case, thus the sample impoverishment problem is less acute in general.

As the primary interest in this example is on the impact of outliers, Figure 4 illustrates the relative accuracy and precision of the Kalman, particle and EIS filters on a date-by-date basis. The top panel of the figure illustrates average (across CRNs) deviations from pseudo-true values of log-likelihood values associated with all three filters; the bottom panel illustrates MC standard deviations (across CRNs) of log-likelihood approximations associated with the EIS and particle filters. Thus the top panel provides a characterization of bias, and the bottom panel numerical precision. To generate pseudo-true values, we obtained log-likelihood values using the EIS filter implemented by setting (N, R) as (2000, 1000).

Absent the two outlier periods, the Kalman and particle filters exhibit non-trivial but relatively moderate bias. The average and total (across dates) absolute deviations from pseudo-true values of log-likelihoods obtained using the EIS filter are (0.002, 0.358), compared with (0.057, 8.955) using the Kalman filter and (0.043, 6.721) using the EIS filter. However, bias is distinctly more pronounced in the presence of the outliers, particularly for that observed in 1976:III. In this period absolute deviations in log-likelihoods are 0.041, 1.321 and 0.682 for the EIS, Kalman, and particle filters.

The outliers also have a distinct impact on the numerical precision of the particle filter, as the bottom panel of Figure 4 illustrates. In particular, while the average (across dates) MC standard deviation calculated for the particle filter is 0.0077 absent the two outliers, the standard deviation jumps to 0.41 in 1976:III, and 0.04 in 1984:IV. Associated values for the EIS filter are 0.0041, 0.03, and 0.01.

An illustration of the source of these biases and numerical inaccuracies is provided in Figures 5 and 6. The former focusses on the Kalman filter, the latter on the EIS filter; both pertain to the period 1976:III, and were generated using a single set of CRNs.

Figure 5 illustrates the measurement density $f(y_t|s_t, Y_{t-1})$ and predictive density $f(s_t|Y_{t-1})$ associated with both the Kalman and EIS filters. Recall that the integral of these densities over

the sample space yields the likelihood function $f(y_t|Y_{t-1})$. Note first that the predictive densities associated with the two filters are virtually indiscernible, and lie near the steady state location of the diagram (indicated by values of zero for both k_t and z_t). This reflects the fact that the time- $(t-1)$ observations y_{t-1} lie relatively close to their steady state values. In contrast, due to the realization of the time- t outlier, the measurement densities lie far from steady state, which accounts for the distinct difference in their shapes and locations: the quality of the linear model approximation associated with the Kalman filter deteriorates as the state variables deviate further from their steady state values, thus its characterization of $f(y_t|s_t, Y_{t-1})$ deviates from that of the EIS filter. Since the Kalman-filter representation of $f(y_t|s_t, Y_{t-1})$ is relatively tightly distributed, its height is much lower than that of the EIS filter at the location of $f(s_t|Y_{t-1})$, thus accounting for the negative bias associated with the Kalman filter.

Turning to the particle filter, the construction of Figure 6 mirrors that of Figure 2, with one small exception. In the top panel, the swarm comprising the particle-filter representation of $f(s_t|Y_t)$ has individual elements that vary by size in direct proportion to the number of times they were resampled from the swarm $f(s_t|Y_{t-1})$ in the bottom diagram. This is done to more clearly illustrate how the particle filter yields a biased approximation of $f(s_t|Y_t)$ in this case. The largest particle in the figure is more than 10,000 times larger than the smallest, thus it has 10,000 times more weight in representing $f(s_t|Y_t)$. The uneven size of the particles in the swarm illustrates the sample impoverishment that results from the outlier.

From the bottom panel of Figure 6, note that the particle-filter representation of $f(s_t|Y_{t-1})$ is left-shifted relative to the EIS-filter representation. Since the particle-filter representation has a discrete and fixed support, this left-shift persists in the re-sampling step under which the filtering density $f(s_t|Y_t)$ is obtained. The upshot is that the particle-filter representation of $f(s_t|Y_t)$ provides insufficient (virtually non-existent) coverage of the north-east portion of the upper diagram: its representation of $f(s_t|Y_t)$ is biased.

The left-shift in $f(s_t|Y_{t-1})$ also induces bias in the approximation of the likelihood function produced by the particle filter. To see why, recall that $f(s_t|Y_{t-1})$ serves as the importance-sampling distribution used by the particle filter: its approximation of the likelihood function $f(y_t|Y_{t-1})$ is given by the average value of the measurement density $f(y_t|s_t, Y_{t-1})$ evaluated at each particle in the swarm $f(s_t|Y_{t-1})$. Since the particle-filter representation of $f(s_t|Y_{t-1})$ is left-shifted, and

lies spuriously close to $f(y_t|s_t, Y_{t-1})$, the resulting likelihood approximation it produces is biased upwards.

As a final note on Figure 6, the filtering density $f(s_t|Y_t)$ associated with the particle filter depicted in the upper diagram helps illustrate sources of numerical imprecision that plague the particle filter given the realization of an outlier. Only a handful of particles are assigned appreciable weight in this representation. Moreover, the exact location of these particles owes much to random chance: alternative sets of CRNs give rise to subtle locational shifts that are magnified due to the imbalanced weight assigned to a select few particles. Thus the spikes in numerical error associated with the outlier dates evident in Figure 4 are not surprising.

As emphasized by RZ, it is important to distinguish between the numerical error associated with a given approximation technique (quantified using the MC standard errors described above), and the sampling error associated with the statistic being approximated (in this case, the log-likelihood function). To characterize sampling error, we conducted two additional experiments. In both, we constructed a data generation process (DGP) using a parameterization of the RBC model, and generated 100 artificial data sets consisting of time-series observations of (ζ, i, n) of length T . For each artificial data set, we used the EIS filter implemented using $(N, R) = (200, 100)$ to obtain 100 approximations of the log-likelihood function. The standard deviation of the log-likelihoods calculated in this manner serves as an estimate of the statistical sampling error associated with this summary statistic.

The DGPs employed in the two experiments were tailored to the empirical applications described above. The first was constructed using the parameters reported in the second row of Table 1, with $T = 100$; the second using the parameters reported in the third row of Table 1, with $T = 158$. The first yielded an estimated sampling error of 16.48; the second 17.99. For comparison, recall that the corresponding MC standard errors associated with the particle filter are 33.48 and 0.1958, while those associated with the EIS filter are 0.00177 and 0.0557. This comparison indicates that the particle filter is an unreliable tool for assessing statistical uncertainty in the context of the first example, since its associated numerical errors are first-order comparable to the associated statistical errors targeted for approximation.

To conclude, tightly-distributed measurement distributions and sample outliers are troublesome sources of numerical error and bias that can plague applications of the particle filter, but that can

be overcome via application of the EIS filter. We now demonstrate application of the EIS filter in a second example model featuring an expanded state space.

5.2 Example 2: Six-State Small Open Economy Model

This application is to a small-open-economy (SOE) model 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

$$U = E_0 \sum_{t=0}^{\infty} \theta_t \frac{[c_t - \varphi_t \omega^{-1} n_t^\omega]^{1-\gamma} - 1}{1-\gamma}, \quad \omega > 0, \quad \gamma \geq 0,$$

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

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

where $(\tilde{c}_t, \tilde{n}_t)$ denote average per capita consumption and hours worked. The household takes these as given; they equal (c_t, n_t) in equilibrium. The household's constraints are collectively

$$\begin{aligned} d_{t+1} &= (1 + r_t) d_t - \zeta_t + c_t + i_t + \frac{\phi}{2} (k_{t+1} - k_t)^2 \\ \zeta_t &= A_t k_t^\alpha n_t^{1-\alpha} \\ k_{t+1} &= \nu_t^{-1} i_t + (1 - \delta) k_t \\ \ln A_{t+1} &= \rho_A \ln A_t + \varepsilon_{At+1} \\ \ln r_{t+1} &= (1 - \rho_r) \ln \bar{r} + \rho_r \ln r_t + \varepsilon_{rt+1} \\ \ln \nu_{t+1} &= \rho_\nu \ln \nu_t + \varepsilon_{\nu t+1} \\ \ln \varphi_{t+1} &= \rho_\varphi \ln \varphi_t + \varepsilon_{\varphi t+1}, \end{aligned}$$

where relative to the RBC model, the new variables are d_t , the stock of foreign debt, r_t , the exogenous interest rate at which domestic residents can borrow in international markets, ν_t , an

investment-specific productivity shock, and the preference shock φ_t .

The state variables of the model are $(d_t, k_t, A_t, r_t, \nu_t, \varphi_t)$; the controls are (ζ_t, c_t, i_t, n_t) . In this application we obtain non-linear policy functions

$$x_t = x(s_t), \quad x_t = (\zeta_t, c_t, i_t, n_t), \quad s_t = (d_t, k_t, A_t, r_t, \nu_t, \varphi_t)$$

using a second-order Taylor Series approximation of the system of expectational difference equations associated with the model, following Schmitt-Grohe and Uribe (2004). Given these policy functions, the state-transitions equations reduce to

$$d_{t+1} = (1 + r_t) d_t - \zeta(s_t) + c(s_t) + i(s_t) + \frac{\phi}{2} (k_{t+1} - k_t)^2 \quad (40)$$

$$k_{t+1} = \nu_t^{-1} i(s_t) + (1 - \delta) k_t \quad (41)$$

$$\ln A_{t+1} = \rho_A \ln A_t + \varepsilon_{At+1} \quad (42)$$

$$\ln r_{t+1} = (1 - \rho_r) \ln \bar{r} + \rho_r \ln r_t + \varepsilon_{rt+1} \quad (43)$$

$$\ln \nu_{t+1} = \rho_\nu \ln \nu_t + \varepsilon_{\nu t+1} \quad (44)$$

$$\ln \varphi_{t+1} = \rho_\varphi \ln \varphi_t + \varepsilon_{\varphi t+1}, \quad (45)$$

and the observation equations are

$$\ln(x_t/x(s_t)) = u_{x,t}, \quad x = \zeta, c, i, n, \quad (46)$$

$$u_{x,t} \sim N(0, \sigma_x^2). \quad (47)$$

As with the RBC model, hereafter we represent state variables as logged deviations from steady state. In addition, given the form of (46), y_t is defined as $y_t = [\ln \zeta_t \quad \ln c_t \quad \ln i_t \quad \ln n_t]'$. All subsequent formulas should be read in accordance with these representations.

Notice that (40) and (41) characterize a bivariate degenerate transition of the form

$$q_t = \phi(p_{t-1}, q_{t-1}), \quad (48)$$

where following the notation of Section 4.3, $p_t = (A_t, r_t, \nu_t, \varphi)$, and $q_t = (d_t, k_t)$. Its inverse and

corresponding linear approximation are denoted respectively as

$$q_{t-1} = \psi(q_t, p_{t-1}), \quad \tilde{q}_{t-1} = \tilde{\psi}(q_t, p_{t-1}).$$

The Jacobian associated with ψ is given by

$$J(q_t, p_{t-1}) = \left\| \frac{\partial}{\partial q_t} \psi(q_t, p_{t-1}) \right\|. \quad (49)$$

We achieve the inversion of (40) and (41) as follows. It turns out that the transition equation for k_t is independent of d_t , thus we exploit the triangular structure of the system by first solving for k_{t-1} , and then using this result to find d_{t-1} . Defining $s_{t-1}^1 = [k_{t-1}, p_{t-1}]'$, the second-order approximation to the law of motion of k_t is given by

$$k_t = C_k + L_k s_{t-1}^1 + \frac{1}{2} s_{t-1}^{1'} Q_k s_{t-1}^1, \quad (50)$$

which is a quadratic equation in k_{t-1} with solutions

$$k_{t-1} = \frac{-b_k \pm \sqrt{b_k^2 - 4a_k c_k}}{2a_k}, \quad (51)$$

$$a_k = \frac{1}{2} Q_k^{11}, \quad (52)$$

$$b_k = L_k^1 + Q_k^{12} p_{t-1}, \quad (53)$$

$$c_k = C_k + L_k^2 p_{t-1} + \frac{1}{2} p_{t-1}' Q_k^{22} p_{t-1} - k_t, \quad (54)$$

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

As the capital stock evolves slowly, the solution to (50) we seek is chosen as

$$k_{t-1}^* = \arg \min [|(k_{t-1}^1 - k_t)|, |(k_{t-1}^2 - k_t)|], \quad (55)$$

where k_{t-1}^1 and k_{t-1}^2 are the roots (51).

Having obtained k_{t-1}^* , the solution of d_{t-1} proceeds as follows. Substituting (41) for k_{t+1} in (40), the second-order approximation to the law of motion of d_t is given by:

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

which is a quadratic equation in d_{t-1} with solutions

$$d_{t-1} = \frac{-b_d \pm \sqrt{b_d^2 - 4a_d c_d}}{2a_d},$$

$$a_d = \frac{1}{2} Q_d^{11}, \quad (56)$$

$$b_d = L_d^2 + Q_d^{12} s_{t-1}^1, \quad (57)$$

$$c_d = C_d + L_d^2 s_{t-1}^1 + \frac{1}{2} s_{t-1}^{1'} Q_k^{22} s_{t-1}^1 - d_t, \quad (58)$$

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

Again we select the solution d_{t-1}^* following (55).

The sequence of operations just described effectively transforms a triangular inverse transformation into a diagonal transformation. The corresponding Jacobian is given by

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

Having achieved inversion, implementation of the EIS filter proceeds precisely as with the RBC model, with the following straightforward modifications:

- $J(k_t, z_{t-1})$ in (30) is replaced by $J(d_t, k_t, p_{t-1})$ in (49).
- The predictive density $f(s_t | Y_{t-1})$ in (34) becomes

$$f(s_t | Y_{t-1}) \simeq \int J(q_t, p_{t-1}) f(p_t | s_{t-1}) g(s_{t-1}; \widehat{a_{t-1}}) |_{q_{t-1} = \psi(q_t, p_{t-1})} dp_{t-1},$$

where

$$f(p_t|s_{t-1}) = N_4 \left(\begin{bmatrix} 0 & 0 & \rho_A & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_r & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_v & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_\varphi \end{bmatrix} p_{t-1}, \begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 \\ 0 & \sigma_r^2 & 0 & 0 \\ 0 & 0 & \sigma_v^2 & 0 \\ 0 & 0 & 0 & \sigma_\varphi^2 \end{bmatrix} \right).$$

- With y_t now defined as $y_t = [\ln \zeta_t \quad \ln c_t \quad \ln i_t \quad \ln n_t]'$, the measurement density becomes

$$f(y_t|s_t, Y_{t-1}) \sim N_3(\mu(s_t), V),$$

$$\mu(s_t) = \begin{bmatrix} \ln \zeta(s_t) \\ \ln c(s_t) \\ \ln i(s_t) \\ \ln n(s_t) \end{bmatrix}, \quad V = \begin{bmatrix} \sigma_y^2 & 0 & 0 & 0 \\ 0 & \sigma_c^2 & 0 & 0 \\ 0 & 0 & \sigma_i^2 & 0 \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix}.$$

We demonstrate the performance of the EIS filter with two Monte Carlo experiments patterned exactly after those used in working with the RBC model. We again work with two data sets: an artificial data set consisting of 100 realizations of $\{\zeta_t, c_t, i_t, n_t\}$ generated using the parameterization of the model given in Table 2; and a Canadian data set consisting of quarterly real per capita observations on $\{\zeta_t, c_t, i_t, n_t\}$, spanning 1976:I-2008:IV (132 observations), and detrended using the Hodrick-Prescott filter. The latter was obtained from Statistics Canada; both are available for downloading at www.pitt.edu/~dejong/wp.htm.

Aside from the parameters that characterize sources of stochastic uncertainty in the model, the artificial data were generated using the parameter values calibrated by Schmitt-Grohe and Uribe (2003) to match the summary statistics on Canadian data reported by Mendoza (1991): parameter values are listed in their Table 1 (and our Table 3), and the summary statistics in their Table 3. The parameters that characterize sources of stochastic uncertainty in the model were chosen as those that minimized the sum of squared differences between Mendoza's summary statistics (excluding the trade balance) and the statistics implied by the model; the statistics are standard deviations of $\{\zeta_t, c_t, i_t, n_t\}$, first-order serial correlations, and contemporaneous correlations with output. Finally, the standard deviations of all measurement errors were set at 0.5%. The same

parameters used to generate the data were also used to evaluate the likelihood function in the MC experiment.

The parameters used to evaluate the likelihood function associated with the actual data are posterior modes estimated using the prior specification indicated in Table 2. 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, and prior standard deviations were set to reflect non-trivial uncertainty over these specifications. (Note that the specifications of δ and \bar{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 2. Parameter Values, SOE Model

| | γ | ω | ψ | α | ϕ | \bar{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 Std. Dev. | 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 | | |
| | 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 Std. Dev. | 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 | |

Results from the two MC experiments are presented in Table 3. Due to the increased dimensionality of the state space, we set N to 150,000 in working with the particle filter (requiring 128.59 and 169.66 seconds per function evaluation in the artificial and real data sets), and $N = R = 50$,

$S = 30$ in working with the EIS filter (requiring 5.02 and 6.65 seconds per function evaluation).

Table 3. Monte Carlo Means and Standard Deviations, SOE Model

| | Particle Filter | | EIS Filter | | Kalman |
|-----------------|-----------------|-----------|------------|-----------|-----------|
| | Mean | Std. Dev. | Mean | Std. Dev. | Filter |
| Artificial Data | 1292.8274 | 2.0391 | 1283.6767 | 0.0721 | 1271.9214 |
| Actual Data | 1718.1382 | 0.4884 | 1713.6243 | 0.04771 | 1719.4209 |

To begin, we note that neither data set contains an outlier observation: in both data sets and across all variables, the largest deviation observed from sample means is 2.7 sample standard deviations. Despite this absence of outliers, there are significant differences between the likelihood values produced by the EIS and Kalman filters in both data sets. Relative to the RBC model, this reflects the added sources of non-linearity featured in the SOE model: e.g., the capital-adjustment cost term $\frac{\phi}{2} (k_{t+1} - k_t)^2$ in (40), and the endogenous discount factor θ_t featured in the household's objective function. Once again, when the EIS filter is implemented using linear model approximations, differences in log-likelihoods produced by the EIS and Kalman filters virtually disappear (becoming at most 6.3e-11 across data sets and all time periods). Thus there is clearly a significant payoff in the implementation of a non-linear model representation in this application.

It is also interesting to note that in this application the artificial data set is the more challenging of the two. This is evident along two dimensions. First, MC standard deviations obtained using the artificial data set are relatively high for both filters. Second, the bias suffered by the particle filter is more substantial in the application involving the artificial data set. Specifically, the difference in mean log-likelihood approximations generated by the particle and EIS filters is more than 9 in working with the artificial data set, compared to less than 5 in working with the actual data set.

As opposed to the applications involving the RBC model, the explanation for these differences across data sets does not lie in the behavior of associated measurement errors: variances of measurement errors are closely comparable across data sets in this case. Instead, differences 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 up-

shot 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.

Comparing the EIS and particle filters, as noted, the particle filter once again suffers non-trivial bias, on scales similar to those observed in working with the RBC model. Regarding MC standard errors, these differ by two orders of magnitude in the artificial data set, but by only one order of magnitude in the actual data set. These results indicate that increases in the dimensionality of the state space do not necessarily amplify the numerical problems suffered by the particle filter: outliers and narrow measurement densities are far more important sources of difficulty.

We conclude our analysis of the SOE model by reporting sampling errors associated with the log-likelihood estimates reported in Table 4. Following the procedure described above in working with the RBC model, we estimate these errors to be 25.55 using the parameterization associated with the artificial data set, and 17.92 using the parameterization associated with the actual data set. Comparing these estimates with the MC standard errors reported in Table 4, we see that the particle filter serves as a better potential gauge of statistical uncertainty than was the case in the applications involving the RBC model. In particular, its MC standard errors are only 1/13th and 1/36th the size of their associated sampling errors in this case, while recall that in working with the RBC model, these ratios were roughly 2 and 1/92nd. The ratios associated with the EIS filter are 1/354th and 1/323rd in this case, compared with roughly 1/10,000 and 1/200 in working with the RBC model.

5.3 Repeated Samples

To this point our comparisons of the bias and numerical errors associated with the EIS and particle filters have been based on four data sets. This raises the question of whether the results we have reported are somehow sensitive to special features of these data sets. Thus we conducted a large set of additional experiments designed to address this issue.

Specifically, using each of the four models described above as data generating processes (the two parameterizations of the RBC and SOE models reported in Table 1 and 2), we generated 100 artificial data sets of length T (with T specified to match the corresponding data set associated with the parameterized model). For each realized data set, we obtained 100 sets of log-likelihood estimates using the EIS and particle filters on a date-by-date basis using 100 sets of CRNs. For each

data set, we also obtained pseudo-true log-likelihood estimates using the EIS filter implemented by setting (N, R) as $(2000, 1000)$. We then calculated the average (across CRNs) of the deviations from pseudo-true values of log-likelihood approximations associated with the EIS and particle filter for each data set and each time period. We also calculated the standard deviation (over CRNs) of log-likelihood approximations. The former provides a measure of bias; the latter a measure of numerical accuracy.

Summing the bias measure over time periods, and averaging the numerical accuracy measure over time periods, we obtained a single measure of bias and numerical accuracy for each data set. Respectively, these measures are given by

$$\sum_{t=1}^T \frac{1}{M} \sum_{m=1}^M (l_{t,m} - l_t^*), \quad \frac{1}{T} \sum_{t=1}^T \sqrt{\frac{1}{M} \sum_{m=1}^M (l_{t,m} - \bar{l}_t^M)^2},$$

where $l_{t,m}$ denotes time- t log-likelihood calculated using the m^{th} set of CRNs, l_t^* denotes its pseudo-true value, \bar{l}_t^M denotes the sample average of $l_{t,m}$ obtained across CRNs, and M denotes the total number of CRNs obtained in the experiment. Assessing the mean and standard deviation of both measures calculated over artificial data sets, and comparing these with corresponding measures obtained using the four data sets analyzed above, we obtain context for interpreting the specific results reported above for the four original data sets. Results are reported in Table 4.

Table 4. Repeated Samples

| DGP | Orig. DS | Bias | | Num. Acc. | | |
|-----------|----------|-----------|-----------|------------|-----------|-----------|
| | | Exp. Avg. | Exp. Stdv | Orig. Smpl | Exp. Avg. | Exp. Stdv |
| RBC, Art. | 0.1369 | 0.1364 | 0.0481 | 0.0015 | 0.0017 | 0.000589 |
| RBC, Act. | 0.8226 | 1.0619 | 0.0950 | 0.0044 | 0.0056 | 0.000383 |
| SOE, Art. | 0.5838 | 1.8279 | 1.3934 | 0.0053 | 0.0037 | 0.000786 |
| SOE, Act. | 0.8381 | 1.2202 | 0.1651 | 0.0026 | 0.0026 | 0.000173 |

Notes: Orig. DS denotes original data set; Exp. Avg. (Stdv) denotes average (standard deviation) of indicated statistic obtained across experimental data sets. RBC, Art. (Act.) denotes the parameterization of the RBC model associated with the artificial (actual) data set; likewise for the SOE model.

Note first that for the two cases in which the original data sets were generated artificially, bias and summary statistics lie within two standard deviations of the average values obtained in the repeated-sample experiments. This is as expected, since in these cases all 101 data sets (the original and those generated in the experiment) come from the same DGP.

For the actual data associated with the SOE model, the measure of numerical accuracy exactly equals its corresponding experimental average, while the measure of bias is only 69% of its corresponding experimental average and lies 0.06 below the ± 2 standard deviation range around the average. For the actual data set associated with the RBC model, the measures of bias and numerical accuracy are only 77% and 79% of their corresponding experimental averages, and lie 0.05 and 0.0004 below their ± 2 standard deviation ranges. Given the two large outliers present in the original RBC data set, it is perhaps surprising that these measures are so close.

The differences noted for the actual data sets indicate that the theoretical models used as DGPs in these experiments do not produce data sets that closely mimic the actual series. To explore this possibility, we repeated the experiments using as alternative DGPs unrestricted vector autoregressions estimated using the original series. In this case all summary statistics lied well within their associated ± 2 standard deviation ranges.

The bottom line we take from these experiments is that the results we have detailed above for the four original data sets are largely representative of those one would expect to obtain in working with repeated samples from appropriate data generating processes. Moreover, the algorithm we have presented for implementing the EIS filter is remarkably reliable, having succeeded in quickly producing likelihood estimates for each of the many hundreds of data sets it confronted.

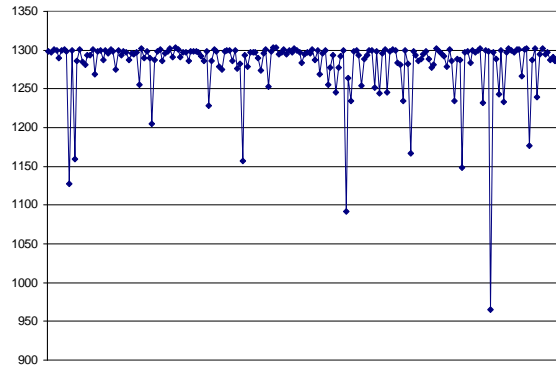
6 Conclusion

We have proposed an efficient means of facilitating likelihood evaluation in applications involving non-linear and/or non-Gaussian state space representations: the EIS filter. The filter is adapted using an optimization procedure designed to minimize numerical standard errors associated with targeted integrals. Resulting likelihood approximations are continuous in underlying likelihood parameters, greatly facilitating the implementation of ML estimation procedures. Implementation of the filter is straightforward, and the payoff of adoption can be substantial.

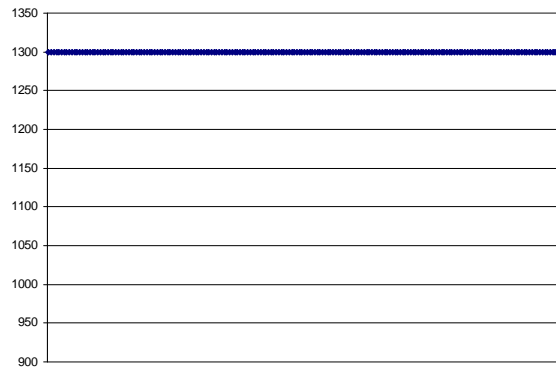
References

- [1] Carpenter, J.R., P. Clifford and P. Fernhead, 1999, “An Improved Particle Filter for Non-Linear Problems”, *IEEE Proceedings-Radar, Sonar and Navigation*, 146, 1, 2-7.
- [2] DeJong, D.N. with C. Dave, 2007, *Structural Macroeconometrics*. Princeton: Princeton University Press.
- [3] DeJong, D.N., H. Dharmarajan, R. Liesenfeld, and J.-F. Richard, 2008, “Efficient Filtering in State-Space Representations”, University of Pittsburgh Working Paper.
- [4] DeJong, D.N., B.F. Ingram, and C.H. Whiteman, 2000, “A Bayesian Approach to Dynamic Macroeconomics”, *Journal of Econometrics*. 98, 203-233.
- [5] Devroye, L., 1986, *Non-Uniform Random Variate Generation*. New York: Springer.
- [6] Doucet, A., N. de Freitas and N. Gordon, 2001, *Sequential Monte Carlo Methods in Practice*. New York: Springer.
- [7] Fernandez-Villaverde, J. and J.F. Rubio-Ramirez, 2005, “Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood”, *Journal of Applied Econometrics* 20, 891-910.
- [8] Fernandez-Villaverde, J. and J.F. Rubio-Ramirez, 2009, “Estimating Macroeconomic Models: A Likelihood Approach”, *Review of Economic Studies*. Forthcoming.
- [9] Geweke, J., 1989, “Bayesian Inference in Econometric Models Using Monte Carlo Integration”, *Econometrica*. 57, 1317-1339.
- [10] 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.
- [11] 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.
- [12] Kim, S., N. Shephard, and S. Chib, 1998, “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models”, *Review of Economic Studies*. 65, 361-393.
- [13] 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.
- [14] Mendoza, E., 1991, “Real Business Cycles in a Small-Open Economy”, *American Economic Review*. 81, 797-818.
- [15] Pitt, M.K., 2002, “Smooth Particle Filters for Likelihood Evaluation and Maximisation”, University of Warwick Working Paper.

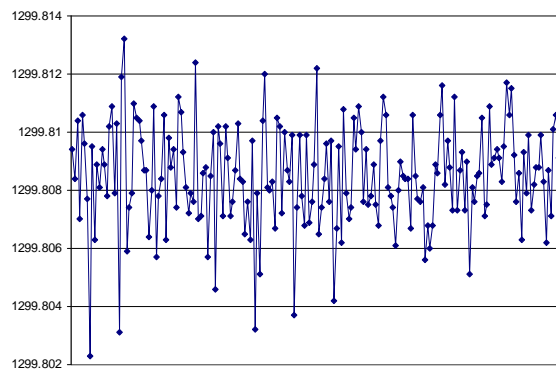
- [16] Pitt, M.K. and N. Shephard, 1999, "Filtering via Simulation: Auxiliary Particle Filters", *Journal of the American Statistical Association*. 94, 590-599.
- [17] Richard, J.-F. and W. Zhang, 2007, "Efficient High-Dimensional Monte Carlo Importance Sampling", *Journal of Econometrics*. 141, 1385-1411.
- [18] Ristic, B., S. Arulampalam, and N. Gordon, 1984, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston: Artech Hous Publishers.
- [19] Sargent, T.J., 1989, "Two Models of Measurements and the Investment Accelerator", *Journal of Political Economy*. 97, 251-287.
- [20] Schmitt-Grohe, S. and M. Uribe, 2003, "Closing Small Open Economy Models", *Journal of International Economics*. 61, 163-185.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Uzawa, H., 1968, "Time Preference, the Consumption Function and Optimum Asset Holdings", In Wolfe, J.N. (Ed.), *Value, Capital and Growth: Papers in Honor of Sir John Hicks*. Edinburgh: The University of Edinburgh Press, 485-504.



1a. Particle Filter

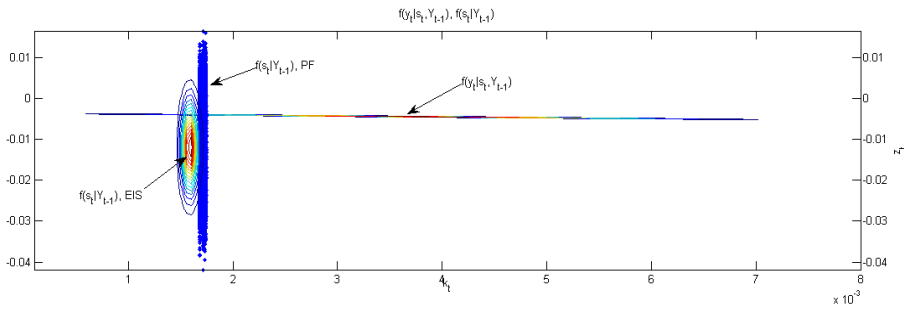
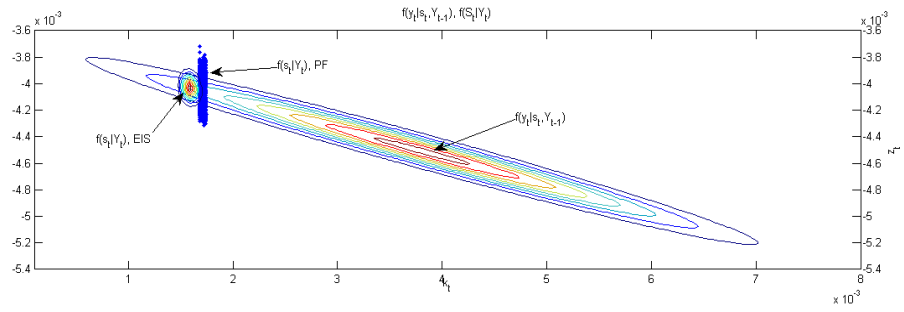


1b. EIS Filter

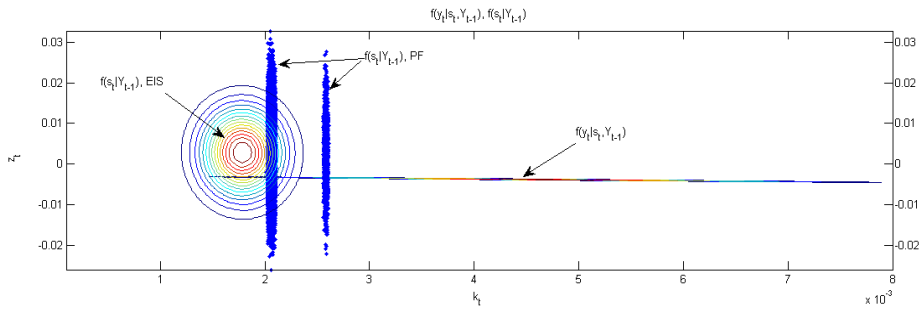
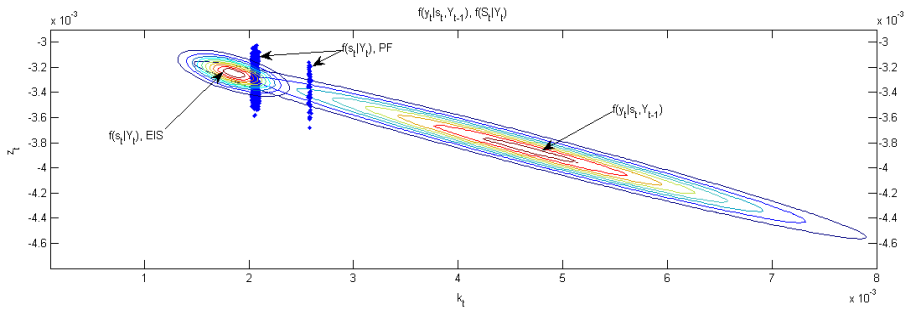


1c. EIS Filter, Rescaled

Figure 1. Log-Likelihood Approximations



2a. Time Period 53



2b. Time Period 18

Figure 2. Sample Impoverishment

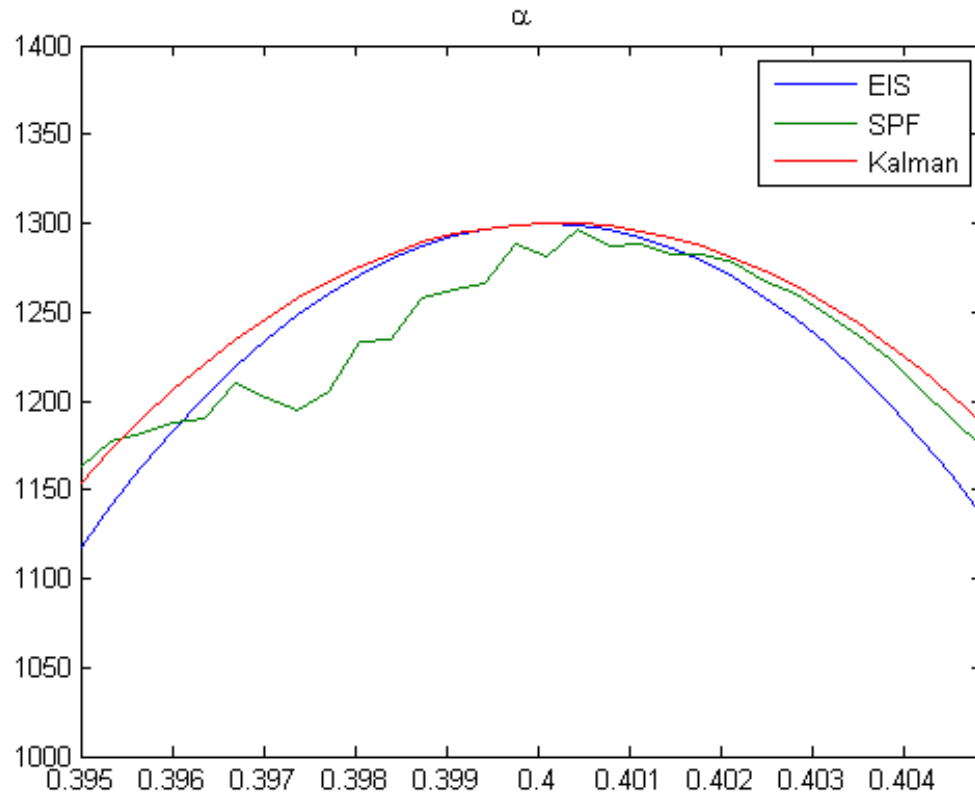
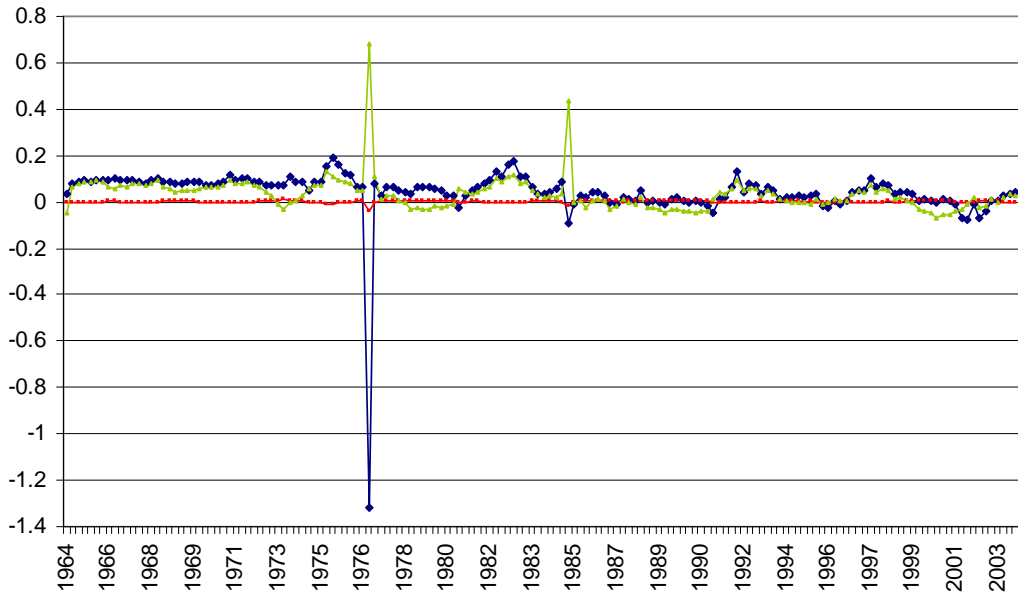
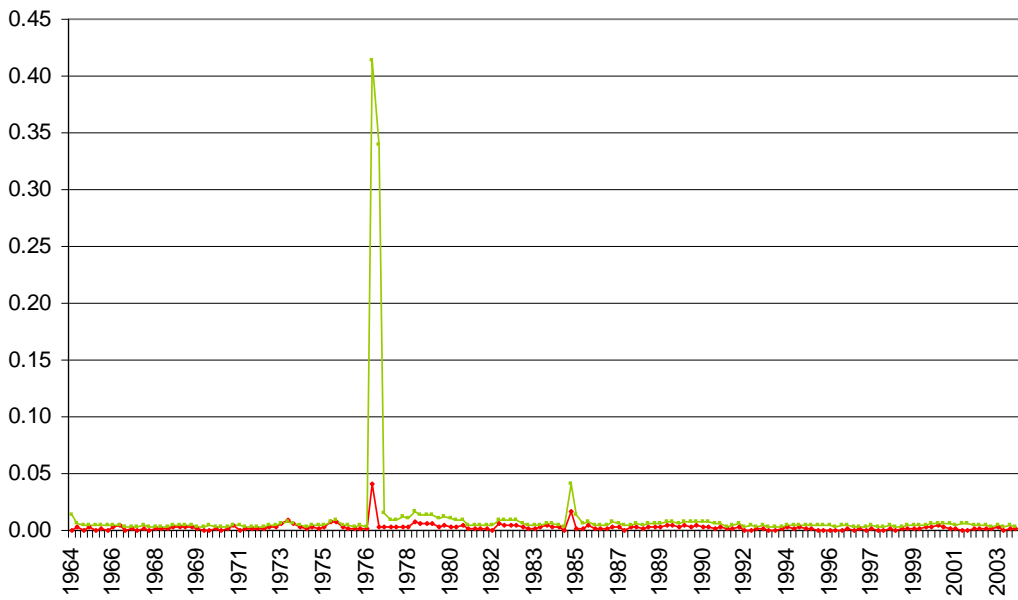


Figure 3. Likelihood Contours

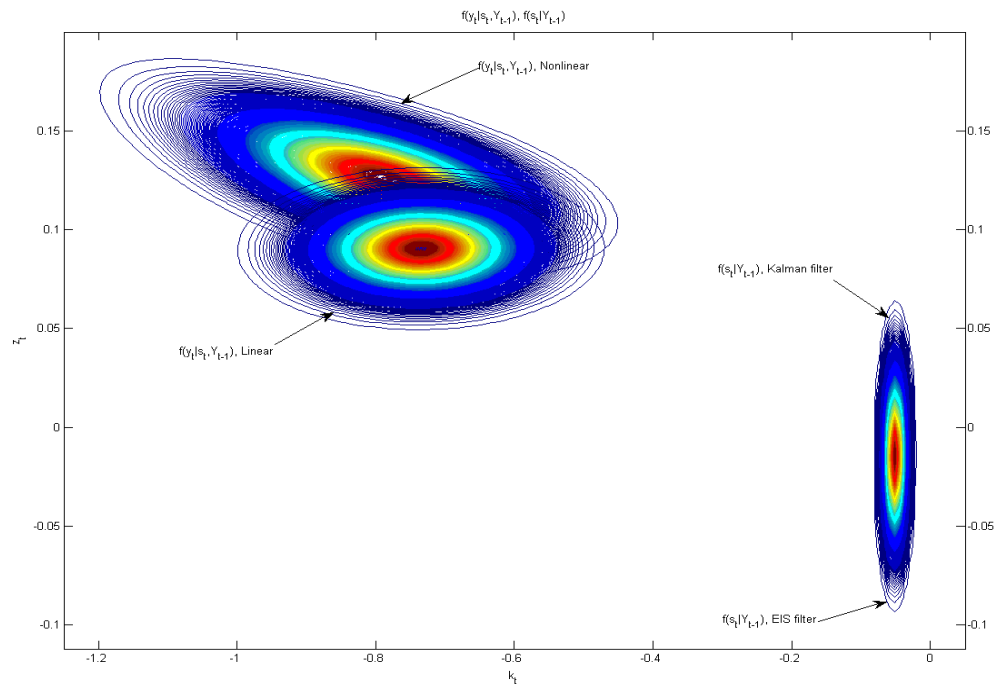


4a. Deviations of Log-Likelihoods from Pseudo-True Values



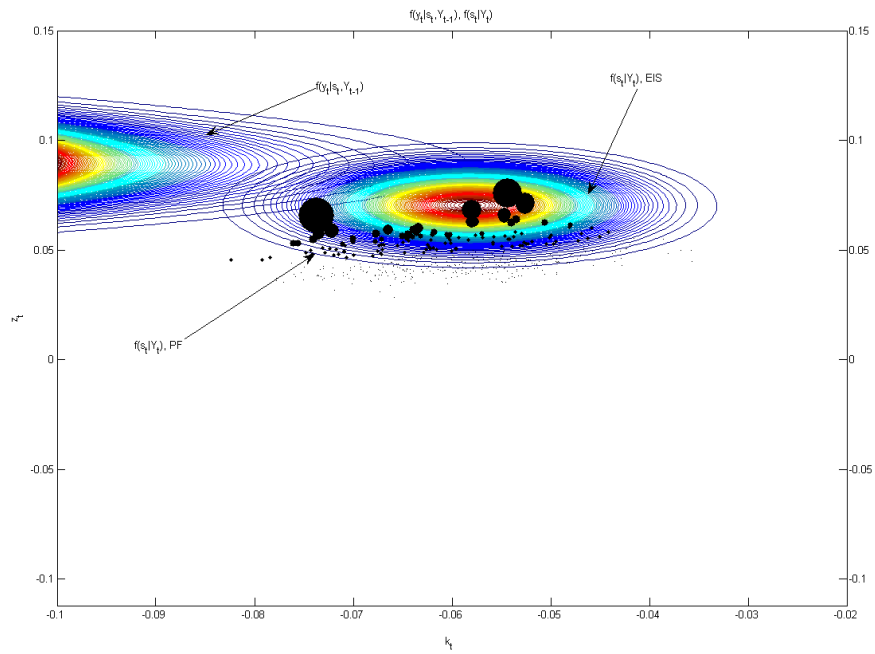
4b. MC Standard Deviations of Log-Likelihoods

Figure 4. Date-by-Date Filter Comparisons, RBC Model, Actual Data
 Blue: Kalman Filter; Green: Particle Filter; Red: EIS Filter

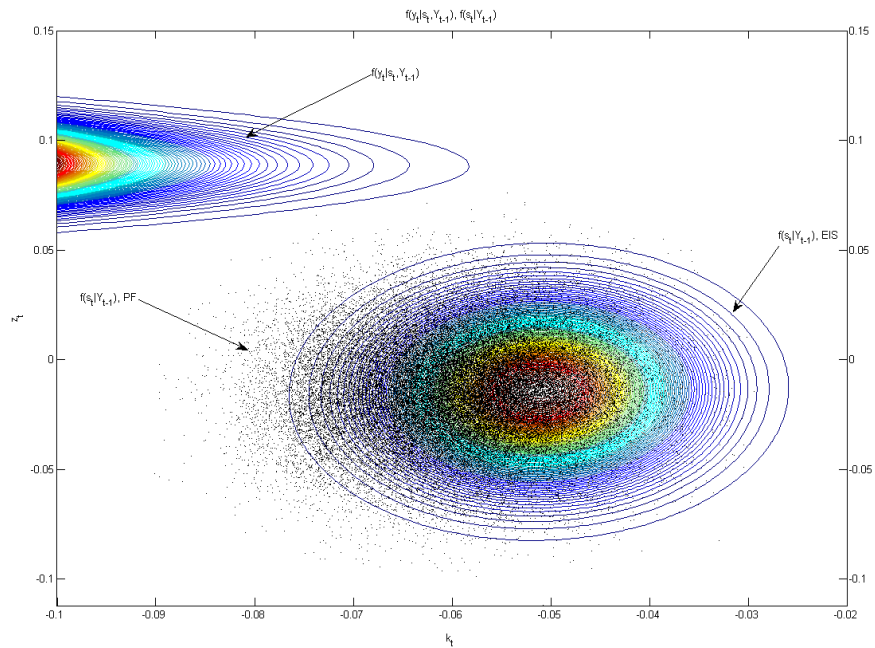


Measurement Densities

Figure 5. Kalman Versus EIS Filter, Actual Data, 1976:III



6a. Measurement and Filtering Densities



6b. Measurement and Predictive Densities

Figure 6. Particle Versus EIS Filter, Actual Data, 1976:III