DeJong, David Neil; Dharmarajan, Hariharan; Liesenfeld, Roman; Richard, Jean-François

Working Paper

An Efficient Filtering Approach to Likelihood Approximation for State-Space Representations

Economics Working Paper, No. 2007-25

Provided in Cooperation with:
Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: DeJong, David Neil; Dharmarajan, Hariharan; Liesenfeld, Roman; Richard, Jean-François (2007) : An Efficient Filtering Approach to Likelihood Approximation for State-Space Representations, Economics Working Paper, No. 2007-25, Kiel University, Department of Economics, Kiel

This Version is available at:
http://hdl.handle.net/10419/22041

An Efficient Filtering Approach to Likelihood Approximation for State-Space Representations

by David N. DeJong, Hariharan Dharmarajan, Roman Liesenfeld and Jean-François Richard

Christian-Albrechts-Universität Kiel

Department of Economics

Economics Working Paper

No 2007-25

An Efficient Filtering Approach to Likelihood Approximation for 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

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

April 2007

Abstract

We develop a numerical filtering procedure that facilitates efficient likelihood evaluation in applications involving non-linear and non-gaussian state-space models. The procedure approximates necessary integrals using continuous or piecewise-continuous approximations of target densities. Construction is achieved via efficient importance sampling, and approximating densities are adapted to fully incorporate current information.

Keywords: particle filter; adaption, efficient importance sampling; kernel density approximation.

*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. All reported computational times are based on runs executed on a 1.86 GHz Opteron 165 processor, using Compaq Visual Fortran 6.6. Documentation and code used to execute 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 possible realizations of unobservable state variables. When models are linear and stochastic processes are gaussian, required integrals can be calculated analytically via the Kalman filter. Departures from these situations entail integrals that must be approximated numerically. Here we introduce an efficient procedure for calculating such integrals: the 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 (1997). Their approach employs discrete fixed-support approximations to unknown densities that appear in the marginalization 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, 2007) for applications involving dynamic stochastic general equilibrium models.

While conceptually simple and easy to program, the particle filter suffers two shortcomings. First, the density approximations it provides are discrete. As a result, associated likelihood approximations can feature spurious discontinuities, 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 problem gives rise to numerical inefficiencies that can be acute when the model's observable variables are highly informative with regard to the 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); 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 density supports 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. First, the approximations are continuous or piecewise-continuous. Second, period-t supports are adjusted using a method designed to produce approximations that achieve optimal efficiency at the adaption stage. Specifically, 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 likelihood. The approach is demonstrated through a series of example applications.

While the focus here is on optimizing the efficiency of likelihood approximations, filtering (i.e., inferences regarding unobservable state variables) obtains as a natural byproduct. However, optimization of the filtering process itself is an alternative objective to that of optimizing likelihood approximations. In subsequent work, we will demonstrate how the EIS filter can be used to optimize the filtering process.

2 State-Space Representations and Likelihood Evaluation

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 . The state-space representations we analyze consists of two equations. The first is a state-transition equation:

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

where $\gamma(s_{t-1}, Y_{t-1}) = E(s_t|s_{t-1}, Y_{t-1})$, and thus υ_t is a vector of innovations with respect to (s_{t-1}, Y_{t-1}). The second equation is an observation (or measurement) equation:

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

where $\delta(s_t, Y_{t-1}) = E(y_t|s_t, Y_{t-1})$, and thus u_t is a vector innovations with respect to (s_t, Y_{t-1}). Hereafter, we refer to υ_t as structural shocks, and u_t as measurement errors.

Often, conditional independence assumptions are imposed to yield the exclusion of lagged components of y in $\gamma(\cdot)$ and $\delta(\cdot)$. Nevertheless, validation of the algorithms discussed here requires that $\gamma(\cdot)$ and $\delta(\cdot)$ be interpreted as being conditional on Y_{t-1}. Implementation of the particle filter is facilitated by interpreting (1) and (2) respectively in terms of the densities $f(s_t|s_{t-1},Y_{t-1})$ and $f(y_t|s_t,Y_{t-1})$. The state-space process is initialized with a marginal density $f(s_0)$, which can be degenerate as a special case.

The likelihood function $f(Y_T)$ is factorized sequentially as

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

where $f(y_1|Y_0) \equiv f(y_1)$. The individual factors are evaluated via marginalization of the measurement densities with respect to $s_t|Y_{t-1}$:

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

Marginalization requires the recursive evaluation of densities $f(s_t|Y_{t-1})$, which is accomplished via Bayes' theorem:

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

where

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

Note that the conditioning sets in (1) - (6) do not contain S_{t-1}. The common interpretation of this omission we follow here is that (u_t, v_{t+1}) are also innovations relative to S_{t-1}, which is omitted for ease of notation.

The particle filter and related extensions employ (4) - (6) to achieve likelihood evaluation and filtering via forward recursion. Note that (4) and (5) respectively require integration with respect to s_t and s_{t-1}. The integral in (4) is conditional on Y_{t-1} alone, while the integral in (5) is conditional on (Y_{t-1}, s_t).

To highlight the conditioning sets in (1) - (6), and to illustrate their internal consistency, it is useful to represent the sequence of operations they entail as a flowchart (as in Figure 1). Densities associated with the model specification are included in rectangles. The term filtration traditionally refers to the transformation of $f(s_{t-1}|Y_{t-1})$ into $f(s_t|Y_t)$, both of which are included in ovals. The numbers included in the figure correspond with equation

numbers associated with filtering, likelihood evaluation and prediction. While filtration focusses on inference on the latent process itself, Figure 1 illustrates that filtration and likelihood evaluation are intertwined.

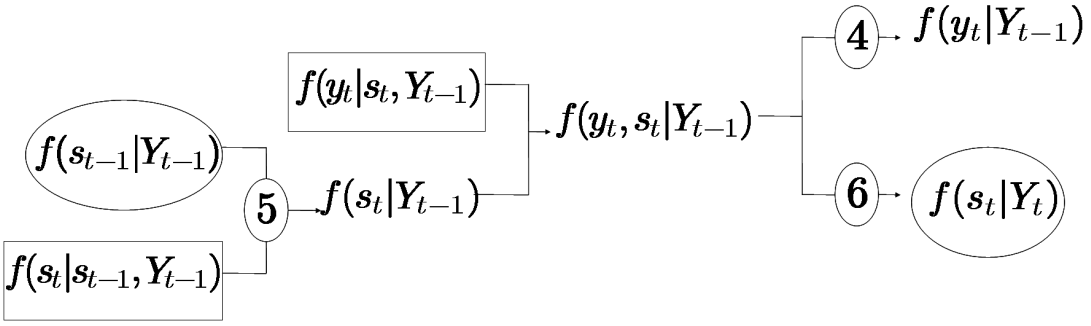

Figure 1: Sequence of operations in the particle filter.

For cases in which the state-space representation is linear and the structural shocks and observation errors are gaussian (or at least zero mean and i.i.d.), the integrals needed to accomplish likelihood evaluation can be calculated analytically via the Kalman filter. The particle filter and associated extensions provide alternative evaluation procedures given departures from these cases. Following is a brief characterization.

3 The Particle Filter

3.1 Standard Filter

The particle filter represents the continuous densities $f(s_t|Y_{t-1})$ and $f(s_t|Y_t)$ using discrete approximations, which are used to produce a Monte Carlo (MC) estimate of $f(y_t|Y_{t-1})$ in (4). 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 $\{s_t^{0,i}\}_{i=1}^N$ through the sequence of operations highlighted in Figure 1. 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 first operation consists of transforming this swarm into a second swarm $\{s_t^{1,i}\}_{i=1}^N$ according to (5). This is accomplished by drawing $s_t^{1,i}$ from the conditional density $f\left(s_t|s_{t-1}^{0,i}, Y_{t-1}\right)$, $i = 1, ..., N$. This

swarm is then used to produce an MC estimate of $f(y_t|Y_{t-1})$, which according to (4) is given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N}\sum_{i=1}^{N} f(y_t|s_t^{1,i}, Y_{t-1}). \tag{7}$$

Next, the filter approximation of $f(s_t|Y_t)$ is obtained by preserving the discrete support associated with the swarm $\{s_t^{1,i}\}_{i=1}^N$, but re-weighting its particles in accordance with (6). Specifically, a particle $s_t^{1,i}$ with initial (prior) weight $\frac{1}{N}$ is assigned the new (posterior) weight

$$w_t^{0,i} = \frac{f(y_t|s_t^{1,i}, Y_{t-1})}{\sum_{j=1}^{N} f(y_t|s_t^{1,j}, Y_{t-1})}. \tag{8}$$

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 weaknesses (well documented in previous studies) can be pinpointed. First, it provides discrete approximations of $f(s_t|Y_{t-1})$ and $f(s_t|Y_t)$, which respectively enable completion of the marginalization and updating stages. Such approximations, particularly the discrete bootstrap associated with the weights in (8), 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 (as indicated by (5) in Figure 1) ignores critical 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 $f(s_t|Y_{t-1})$, the discrete approximation $\{s_t^{1,i}\}_{i=1}^N$ 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 for likelihood evaluation: relatively few of its elements will be assigned appreciable weight in the updating stage in the following period. This phenomenon is known as "sample impoverishment": it entails a reduction in the effective size of the particle swarm. In extreme cases, the size of the effective swarm can be reduced to a handful of particle sequences.

Extensions of the particle filter seek to employ adaption techniques to generate gains in efficiency. Following is a characterization of some prominent examples.

3.2 Extensions of the Particle Filter

Many extensions have been proposed in efforts to address the problems noted above. An extension proposed by Gordon et al. (1993) and Kitagawa (1996) consists simply of making $N' >> N$ blind proposals $\{s_t^{1,j}\}_{j=1}^{N'}$ as with the standard particle filter, and then obtaining the swarm $\left\{s_t^{0,i}\right\}_{i=1}^{N}$ by sampling with replacement, using the weights

$$w^i = \frac{f\left(y_t|s_t^{1,i}, Y_{t-1}\right)}{\sum_{j=1}^{N'} f\left(y_t|s_t^{1,j}, Y_{t-1}\right)}. \tag{9}$$

This is the sampling-importance resampling (SIR) particle filter, and is widely used in literature (see Doucet et al., 2001). Setting $N' = N$ yields the standard particle filter. This extension 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 similar to (9), but 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 the problem of adaption through the use of an Importance Sampling (IS) procedure. Consider as an example the marginalization step of the filtering process. Faced with the problem of calculating

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

but with $f(s_t|Y_{t-1})$ unknown, importance sampling achieves approximation via the introduction into the integral of an importance density $g(s_t|Y_t)$:

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

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

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

Pitt and Shephard referred to the introduction of $g(s_t|Y_t)$ in this context 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. Our procedure is designed to achieve as nearly as possible full adaption. Pitt and Shephard viewed adaption as being computationally infeasible, due to the cost of having to compute $f\left(s_t^{0,i}|Y_{t-1}\right)$ for every value of $s_t^{0,i}$ produced by the sampler. Instead they introduced the concept of an auxiliary particle filter designed to yield partial adaption.

Their algorithm takes as input in period t a period-$(t-1)$ approximation of $f(s_t|Y_{t-1})$ of the form

$$\widehat{f_N}(s_t|Y_{t-1}) = \sum_{i=1}^{N} \pi_{t-1}^i f(s_t|s_{t-1}^{0,i},Y_{t-1}), \tag{12}$$

where $\left\{s_{t-1}^{0,i}\right\}_{i=1}^{N}$ denotes the period-$(t-1)$ swarm of particles, and $\left\{\pi_{t-1}^i\right\}_{i=1}^{N}$ a vector of probabilities initialized by $\pi_0^i = 1/N$, and recursively updated as described below. The corresponding approximation of $f(y_t,s_t|Y_{t-1})$ is then interpreted as the marginal of the mixed density

$$\widehat{f}(y_t,s_t,k|Y_{t-1}) = \pi_{t-1}^k f(y_t|s_t,Y_{t-1}) f(s_t|s_{t-1}^{0,k},Y_{t-1}), \tag{13}$$

with an auxiliary discrete random variable $k \in \{1,...,N\}$ (omitting a subscript t for ease of notation).

The adaption step consists of constructing an importance sampler $g(s_t,k|Y_t)$. In its simplest form this obtains by replacing s_t in $f(y_t|s_t,Y_{t-1})$ with its conditional expectation $\mu_t^k = E\left(s_t|s_{t-1}^{0,k},Y_{t-1}\right)$. The corresponding (normalized) importance sampler is then given by

$$g(s_t,k|Y_t) = \lambda_k f(s_t|s_{t-1}^{0,k},Y_{t-1}), \tag{14}$$

with

$$\lambda_k = \frac{1}{D}\pi_{t-1}^k f\left(y_t|\mu_t^k,Y_{t-1}\right), \tag{15}$$

$$D = \sum_{j=1}^{N} \pi_{t-1}^j f\left(y_t|\mu_t^j,Y_{t-1}\right). \tag{16}$$

Draws from $g(s_t,k|Y_t)$ are obtained as follows: first draw $k^i \in \{1,...,N\}$ with probabilities $\{\lambda_k\}_{k=1}^N$; next, conditionally upon k^i, draw $s_t^{0,i}$ from $f(s_t|s_{t-1}^{0,k^i},Y_{t-1})$. The IS estimate of the period-t likelihood is then given by

$$\begin{aligned}\widehat{f}_N(y_t|Y_{t-1}) &= \frac{1}{N}\sum_{i=1}^{N}\frac{\widehat{f}_N\left(y_t,s_t^{0,i},k^i|Y_{t-1}\right)}{g\left(s_t^{0,i},k^i|Y_t\right)} \\ &= \frac{D}{N}\sum_{i=1}^{N}\omega_t^i,\end{aligned} \tag{17}$$

with

$$\omega_t^i = \frac{f\left(y_t|s_t^{0,i},Y_{t-1}\right)}{f\left(y_t|\mu_t^{k^i},Y_{t-1}\right)},$$

and $\left\{s_t^{0,i},k^i\right\}_{i=1}^N$ denoting i.i.d. draws from $g(s_t,k|Y_t)$.

Similarly, the density $f(s_{t+1}|Y_t)$ is approximated by

$$\widehat{f}_N(s_{t+1}|Y_t) = \frac{1}{\widehat{f}_N(y_t|Y_{t-1})}\int f(s_{t+1}|s_t,Y_t)\,\widehat{f}_N(y_t,s_t|Y_{t-1})\,ds_t,$$

whose IS estimate under $g()$ is given by

$$\widehat{f}_N(s_{t+1}|Y_t) = \sum_{i=1}^{N}\pi_t^i f\left(s_{t+1}|\widehat{s}_t^{0,i},Y_t\right),$$

with

$$\pi_t^i = \frac{\omega_t^i}{\sum_{j=1}^{N}\omega_t^j}.$$

Note from (17) that $g(s_t,k|Y_t)$ is based implicitly on a zero-order Taylor series expansion of $\ln f(y_t|s_t,Y_{t-1})$ around $s_t=\mu_t^k$. Further adaption obtains if a higher-order expansion can be implemented feasibly.

Let the exponential version of such an approximation be given by

$$f(y_t|s_t,Y_{t-1}) \simeq f\left(y_t|\mu_t^k,Y_{t-1}\right)h\left(s_t;\mu_t^k,Y_t\right). \tag{18}$$

Combining $h(\cdot)$ with $f(s_t|s_{t-1}^{0,k}, Y_{t-1})$, we obtain

$$f_*\left(s_t|\mu_{*t}^k, Y_t\right) = \frac{f\left(s_t|s_{t-1}^{0,k}, Y_{t-1}\right) h\left(s_t;\mu_t^k, Y_t\right)}{\chi\left(\mu_{*t}^k; Y_t\right)},$$

where $\chi(\cdot)$ and μ_{*t}^k denote the integrating constant and reparameterization associated with the replacement of $f(\cdot)$ by $f_*(\cdot)$.

Two requirements must be met in order for $f_*(\cdot)$ to be implemented feasibly. First, it must be possible to draw from $f_*(\cdot)$; second, $\chi(\cdot)$ must be an analytical expression, since as shown below, it is used to construct adapted resampling weights. As explained, e.g., in RZ, these conditions are met in working within the exponential family of distributions, which are closed under multiplication. (Pitt and Shephard, 1999, present examples involving first-order approximations; and Smith and Santos, 2006, present examples involving second-order expansions.)

Given the use of $f_*(\cdot)$, the adapted importance sampler obtains by replacing $f(y_t|s_t, Y_{t-1})$ in (13) by its approximation in (18):

$$g_*(s_t, k|Y_t) = \lambda_*^k f_*(s_t|\mu_{*t}^k, Y_t), \tag{19}$$

with

$$\lambda_*^k = \frac{1}{D_*}\pi_{t-1}^k f\left(y_t|\mu_t^k, Y_{t-1}\right)\chi\left(\mu_{*t}^k; Y_t\right), \tag{20}$$

$$D_* = \sum_{j=1}^{N}\pi_{t-1}^j f\left(y_t|\mu_t^j, Y_{t-1}\right)\chi\left(\mu_{*t}^k; Y_t\right). \tag{21}$$

The IS weight in (17) is replaced by

$$\omega_{*t}^i = \frac{f\left(y_t|s_t^{0,i}, Y_{t-1}\right)}{f\left(y_t|\mu_t^{k^i}, Y_{t-1}\right) h\left(s_t;\mu_t^k, Y_t\right)}. \tag{22}$$

Relative to ω_t^i in (17), ω_{*t}^i has a smaller MC sampling variance, since it is based upon a higher-order Taylor series expansion of $f(y_t|s_t, Y_{t-1})$. The corresponding likelihood estimate is

$$\widehat{f}_{*N}(y_t|Y_{t-1}) = \frac{D_*}{N}\sum_{i=1}^{N}\omega_{*t}^i.$$

Note that this sampler implicitly reweights $\left\{s_{t-1}^{0,i}\right\}$ to account for the new information conveyed by y_t. It is therefore expected to alleviate the problem of blind sampling. However, sample impoverishment potentially remains an issue, since the algorithm does not allow the support of $\left\{s_{t-1}^{0,i}\right\}$ to be adjusted. Moreover, the sampler remains suboptimal, to the extent that $\mu_t(s_{t-1}^{0,i})$ is incapable of fully capturing the characteristics of $f(y_t|s_t, Y_{t-1})$. Finally, this extension does not address the discontinuity problem.

Pitt (2002) addressed the latter problem for the special case in which the state space is unidimensional by replacing the weights in (8) associated with the particle filter, or the weights in (15) 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 associated likelihood estimates that are continuous functions of model parameters (e.g., see Hendry, 1984; or RZ).

As noted, the approach to which we now turn extends Pitt and Shephard (1999) and Pitt (2002) by directly facilitating the process of adaption. Instead of employing discrete filters to approximate necessary integrals, we instead use the EIS methodology developed by RZ. This yields sampling densities that are (nearly) fully adapted, along with continuous approximations of the likelihood function. We do so by developing operational procedures for evaluating $f(s_t|Y_{t-1})$ at any value of s_t required by the EIS algorithm, not merely those in the swarm $\left\{s_t^{1,i}\right\}_{i=1}^{N}$. Details follow.

4 The EIS Filter

Efficient Importance Sampling (EIS) is an automated procedure for constructing numerically efficient importance samplers for analytically intractable integrals. EIS samplers are continuous and 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 introduces a class of piecewise-continuous samplers for dealing with pathological cases. Section 4.3 then discusses a complication for implementing EIS that arises in this context, and outlines alternative approaches for overcoming this complication. Briefly, implementation of any procedure requires that one account explicitly for the unknown density $f(s_t|Y_{t-1})$ in (4). Recall that the particle filter does so by representing $f(s_t|Y_{t-1})$ with swarm $\{s_t^{1,i}\}_{i=1}^{N}$, thus eliminating the need to recompute $f(s_t|Y_{t-1})$ in period t. In contrast,

implementation of the EIS filter requires that $f(s_t|Y_{t-1})$ be computed at auxiliary values of s_t generated under period-t EIS optimization. This requirement is the focus of Section 4.3. Section 4.4 discusses two special cases that often characterize state-space representations: partial observability of the state space; and degenerate transition densities. Section 4.5 concludes with a gaussian example that admits an analytical comparison of the standard, auxiliary, and EIS particle filters.

4.1 EIS integration

For ease of notation, let $\varphi_t(s_t)$ represent the integrand in (4):

$$\varphi_t(s_t) = f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}); \tag{23}$$

the subscript t in φ_t replaces $Y_t = (Y_{t-1}, y_t)$. Implementation of EIS begins with the pre-selection of a parametric class $K = \{k(s_t; a_t); a_t \in A\}$ of auxiliary density kernels. Corresponding density functions g are given by

$$g(s_t; a_t) = \frac{k(s_t; a_t)}{\chi(a_t)},\ \chi(a_t) = \int k(s_t; a_t) ds_t. \tag{24}$$

The selection of K is inherently problem-specific; below we discuss gaussian and piecewise-continuous alternatives. The objective of EIS is to select the parameter value $\widehat{a}_t \in A$ that minimizes the variance of the ratio $\frac{\varphi_t(s_t)}{g(s_t|a_t)}$ over the range of integration. Thus EIS is a global approximation technique. Following RZ, a (near) optimal value $\widehat{a}_t$ is obtained as the solution to the least-squares problem

$$(\widehat{a}_t, \widehat{c}_t) = \arg\min_{a_t, c_t} \int \left[\ln \varphi_t(s_t) - c_t - \ln k(s_t; a_t)\right]^2 g(s_t; a_t) ds_t, \tag{25}$$

where c_t denotes an intercept meant to calibrate $\ln(\varphi_t/k)$. Equation (25) has the form of a standard least squares problem, except that the auxiliary sampling density itself depends upon a_t. This can be resolved by reinterpreting (25) as the search for a fixed-point solution. An operational MC version of this analytically intractable problem, implemented using $R << N$ draws, is as follows:

Step $l+1$: Given $\widehat{a}_t^l$, draw intermediate values $\{s_{t,l}^i\}_{i=1}^R$ from the step-l EIS sampler

$g(s_t;\widehat{a}_t^l)$, and solve

$$(\widehat{a}_t^{l+1},\widehat{c}_t^{l+1}) = \arg\min_{a_t,c_t} \sum_{i=1}^{R} \left[\ln \varphi_t(s_{t,l}^i) - c_t - \ln k(s_{t,l}^i;a_t)\right]^2. \tag{26}$$

The initial value $\widehat{a}_t^1$ can be chosen in a variety of ways, with minimal impact on convergence. To avoid potential problems involving sample impoverishment, we employ a crude grid search over the unconditional range of s_t. In the case of evaluating (4), the initial EIS iteration can be based on draws from $f(s_t|Y_{t-1})$ itself (which then implicitly augments K). If K belongs to the exponential family of distributions, the auxiliary problems in (26) are linear in a_t. For well-posed problems, convergence to a fixed point is typically achieved within five to ten iterations.

In order to guarantee fast (and smooth) fixed-point convergence, and to ensure continuity of corresponding likelihood estimates, it is critical that all draws $\{s_{t,j}^i\}$ 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). Obvious examples are standardized Normal draws when g is gaussian, or more generally, uniform draws transformed into draws from g by the inverse c.d.f technique (e.g., see Devroye, 1986). See RZ for CRN application to EIS; and Pitt (2002) for the use of CRNs as a technique for smoothing the particle filter bootstrap.

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

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N}\sum_{i=1}^{N} \frac{f\left(y_t|s_t^i,Y_{t-1}\right) f\left(s_t^i|Y_{t-1}\right)}{g(s_t^i;a_t)}, \tag{27}$$

where $\left\{s_t^i\right\}_{i=1}^N$ are drawn from the (final) EIS sampler.

The estimate $\widehat{f}_N(y_t|Y_{t-1})$ 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 multiplicative factor of 3 to 5.

4.2 A piecewise-continuous class of samplers

As noted, kernels within the exponential family of distributions offer the computational advantage of yielding auxiliary EIS regressions in (26) that are linear in a_t. Nevertheless, there exist potential pathologies of the integrand in (4) that cannot be replicated efficiently within the exponential family. Examples include skewness, thick tails, and bimodality.

Here we propose to deal with these pathologies using an adaptable class of auxiliary samplers that provide piecewise log-linear approximations to the integrand φ_t. The parameters of such approximations are the grid points $a' = (a_0, ..., a_R)$, with $a_0 < a_1 < ... < a_R$. As we shall see, $\ln k(.;a)$ then depends non-linearly on a. Furthermore, R must be sufficiently large for good approximation. This prevents application of the EIS least-squares optimization step as described in (26). Instead we implement (essentially) equal probability division of the domain of integration, which guarantees excellent adaption even for pathological integrands.

We first describe the approximating kernel $k(s;a)$ for a preassigned grid a. The interval $[a_0, a_R]$ is understood as being sufficiently wide to cover the support of the density kernel $\varphi(s)$. Note that while R represents the number of grid-points here, and the number of auxiliary draws used to construct the EIS sampler $g(s_t;\widehat{a}_t^l)$ in (26), this does not represent an abuse of notation. Indeed, for the piecewise-continuous sampler, use of R grid-points translates precisely into the use of R auxiliary draws.

The kernel $k(s;a)$ is given by

$$\ln k_j(s;a) = \alpha_j + \beta_j s \quad \forall s \in [a_{j-1}, a_j], \tag{28}$$

with

$$\beta_j = \frac{\ln \varphi(a_j) - \ln \varphi(a_{j-1})}{a_j - a_{j-1}}, \quad \alpha_j = \ln \varphi(a_j) - \beta_j a_j. \tag{29}$$

Since k is piecewise integrable, its distribution function can be written as

$$K_j(s;a) = \frac{\chi_j(s;a)}{\chi_n(a)}, \quad \forall s \in [a_{j-1}, a_j], \tag{30}$$

with

$$\chi_j(s;a) = \chi_{j-1}(a) + \frac{1}{\beta_j}\left[k_j(s;a) - k_j(a_{j-1};a)\right], \tag{31}$$

$$\chi_0(a) = 0, \quad \chi_j(a) = \chi_j(a_j;a). \tag{32}$$

Its inverse c.d.f. is given by

$$s = \frac{1}{\beta_j}\left\{\ln\left[k_j(a_{j-1};a) + \beta_j\left(u\chi_R(a) - \chi_{j-1}(a)\right)\right] - \alpha_j\right\} \tag{33}$$

$$u \in]0,1[\quad \text{and} \quad \chi_{j-1}(a) < u\chi_R(a) < \chi_j(a).$$

Note that with $\beta_1 < 0$ and $\beta_R > 0$, we have

$$\lim_{s\to-\infty} k_1(s;a) = \lim_{s\to+\infty} k_R(s;a) = 0. \tag{34}$$

Thus trivial modification of the boundary conditions extends $k(s;a)$ to the real line.

The recursive construction of an equal-probability-division kernel $k(s;\widehat{a})$ is based upon the non-random equal division of $[0,1]$ with $u_i = \frac{i}{R}$ for $i = 1,...,R-1$. It proceeds as follows.

Step $l+1$: Given the step-l grid $\widehat{a}^l$, construct the density kernel k and its c.d.f K as described above. The step-$l+1$ grid is then computed as

$$\widehat{a}_i^{l+1} = K^{-1}(u_i), \quad i = 1,...,R-1. \tag{35}$$

The algorithm iterates until (approximate) convergence.

Extra care may be required for the tail areas $i = 1$ and $i = R-1$, since a combination of wide support $[a_0, a_R]$ and thin tails can result in wide tail intervals and poor approximations. We address this issue here by reducing the probability of these tail intervals. Further refinement would rely upon tail approximations for $\varphi(\cdot)$, a common practice in random number generation (e.g., see Devroye, 1986).

The resulting approximation is highly adapted and computationally inexpensive. Given a sufficiently large number of division points, it will dominate standard EIS, which is based typically on (much) lower-dimensional parametric classes of samplers. The piecewise-continuous class of samplers can be generalized to higher-dimensional state spaces, though the curse of dimensionality can rapidly become acute. Thus in working with multi-dimensional

state spaces, it is advisable to begin with standard parametric families of distributions, and reserve the use of piecewise continuous approximations for those dimensions along which the integrand appears to be ill-behaved.

4.3 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. This requirement is absent under the particle filter, wherein the discrete approximation of $f(s_t|Y_{t-1})$ by the swarm $\{s_t^{1,i}\}_{i=1}^N$ is not revisited or adapted in period t. Here we propose four operational alternatives for overcoming this hurdle. Weighted-sum approximations, which include constant weights in its simplest version, are the easiest to implement. Nonparameteric and full EIS approximations can deliver higher accuracy, though at additional computational cost, and thus may be advisable to implement in pathological cases. Below, S denotes the number of points used for each individual evaluation of $f(s_t|Y_{t-1})$.

4.3.1 Weighted-sum approximations

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

$$f(s_t|Y_{t-1}) = \frac{\int f(s_t|s_{t-1},Y_{t-1})f(y_{t-1}|s_{t-1},Y_{t-2})f(s_{t-1}|Y_{t-2})ds_{t-1}}{\int f(y_{t-1}|s_{t-1},Y_{t-2})f(s_{t-1}|Y_{t-2})ds_{t-1}}, \tag{36}$$

where the denominator represents the likelihood integral for which an EIS sampler has been constructed in period $t-1$. Following, e.g., Geweke (1989), a direct MC estimate of $f(s_t|Y_{t-1})$ is given by

$$\widehat{f}_N(s_t|Y_{t-1}) = \frac{\sum_{i=1}^S f(s_t|s_{t-1}^{0,i},Y_{t-1})\cdot\omega(s_{t-1}^{0,i};\widehat{a}_{t-1})}{\sum_{i=1}^S \omega(s_{t-1}^{0,i};\widehat{a}_{t-1})}, \tag{37}$$

where $\{s_{t-1}^{0,i}\}_{i=1}^S$ denotes EIS draws from $g(s_{t-1}|\widehat{a}_{t-1})$, and $\omega(\cdot)$ denotes associated weights (both of which are carried over from period-$t-1$):

$$\omega(s_{t-1};\widehat{a}_{t-1}) = \frac{f(y_{t-1}|s_{t-1},Y_{t-2})f(s_{t-1}|Y_{t-2})}{g(s_{t-1}|\widehat{a}_{t-1})}. \tag{38}$$

Obviously $g(s_{t-1}|\widehat{a}_{t-1})$ is not an EIS sampler for the numerator in (36). 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 $g(s_{t-1}|\widehat{a}_{t-1})$. This would be the case if the conditional variance of $s_t|s_{t-1},Y_{t-1}$ were significantly smaller than that of $s_{t-1}|Y_{t-1}$. On the other hand, 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. This weighted-sum approximation is the simplest of the three methods to implement, and thus provides a good starting point.

4.3.2 A constant weight approximation

When EIS delivers a close global approximation to $f(s_{t-1}|Y_{t-1})$, the weights $\omega(s_{t-1};\widehat{a}_{t-1})$ will be near constants over the range of integration. Replacing these weights by their arithmetic means $\overline{\omega}(\widehat{a}_{t-1})$ in (36) and (37), we obtain the following simplification:

$$f(s_t|Y_{t-1}) \simeq \frac{\overline{\omega}(\widehat{a}_{t-1})}{f(y_{t-1}|Y_{t-2})} \int f(s_t|s_{t-1},Y_{t-1}) \cdot g(s_{t-1};\widehat{a}_{t-1}) ds_{t-1}. \tag{39}$$

This substitution yields rapid implementation if additionally the integral in (39) 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 either normal or piecewise continuous as described in Section 4.2. Examples are provided in Section 5. In cases for which we lack an analytical solution, we can use the standard MC approximation

$$\widehat{f}_S(s_t|Y_{t-1}) \simeq \frac{\overline{\omega}(\widehat{a}_{t-1})}{f(y_{t-1}|Y_{t-2})} \frac{1}{S} \sum_{i=1}^{S} f(s_t|s_{t-1}^{0,i},Y_{t-1}). \tag{40}$$

4.3.3 Non-parametric approximations

An obvious candidate for smoothing $f(s_t|Y_{t-1})$ consists of replacing the discrete swarm $\{s_t^{1,i}\}_{i=1}^N$ with a non-parametric approximation. In this case, care must be taken to account for the fact that under EIS, the swarm $\{s_{t-1}^{0,i}\}_{i=1}^N$, which is transformed into $\{s_t^{1,i}\}_{i=1}^N$ through the transition $f(s_t|s_{t-1},Y_{t-1})$, is drawn from the EIS auxiliary sampler $g(s_{t-1}|\widehat{a}_{t-1})$, not from $f(s_{t-1}|Y_{t-1})$ itself. This requires re-weighting accordingly the indi-

vidual factors in the non-parametric approximation of $f(s_t|Y_{t-1})$, given by

$$\widehat{f}_S(s_t|Y_{t-1}) = \frac{1}{h}\sum_{i=1}^{S} p_{i,t}\ \kappa\left(\frac{s_t - s_t^{1,i}}{h}\right), \tag{41}$$

where $\kappa(\cdot)$ denotes an auxiliary density, h the bandwidth, and $\{p_{i,t}\}$ probabilities associated with $s_t^{1,i}$. Regarding these probabilities, for $t = 1$ the swarm $\{s_1^{1,i}\}_{i=1}^S$ is drawn from $f(s_1|Y_0)$, and thus $p_{i,0} = \frac{1}{S}$. Beyond period 1, the EIS algorithm, including the calculation of $p_{i,t}$, is as follows.

Period t: Apply the EIS algorithm as described in Section 4.1, replacing $f(s_t|Y_{t-1})$ with $\widehat{f}_S(s_t|Y_{t-1})$. Drawing from $\widehat{f}_S(s_t|Y_{t-1})$ in order to initialize EIS iteration proceeds as follows: draw an index j from the discrete distribution $\{1,2,...,S\}$ with probabilities $\{p_{i,t}\}_{i=1}^S$; draw e from the density $\frac{1}{h}\kappa(\frac{e}{h})$; then combine to obtain the desired draw $s_t^{1,j} + e$.

The EIS swarm $\{s_t^{0,i}\}_{i=1}^S$ is transformed into a swarm $\{s_{t+1}^{1,i}\}_{i=1}^S$ through the transition density $f(s_t|s_{t-1},Y_{t-1})$ in the usual way. Since $\{s_t^{0,i}\}_{i=1}^S$ was obtained using the EIS sampler $g(s_t|\widehat{a}_t)$, and not from $f(s_t|Y_t)$, we must re-weight accordingly the non-parametric factors, using the EIS weights $\omega(s_t;\widehat{a}_t)$. Thus $p_{i,t+1}$ is given by

$$p_{i,t+1} = \frac{\widehat{\omega}_S(s_t^{0,i};\widehat{a}_t)}{\sum_{j=1}^{S}\widehat{\omega}_S(s_t^{0,j};\widehat{a}_t)}, \tag{42}$$

with

$$\widehat{\omega}_S(s_t;\widehat{a}_t) = \frac{f(y_t|s_t,Y_{t-1})\widehat{f}_S(s_t|Y_{t-1})}{g(s_t|\widehat{a}_t)}. \tag{43}$$

4.3.4 EIS evaluation

Accurate numerical evaluation of $f(s_t|Y_{t-1})$ can sometimes be delicate, especially in situations prone to sample impoverishment (such as when working with degenerate transitions, discussed in Section 4.4.2 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 $f(s_{t-1}|Y_{t-1})$. Given the preceding discussion, two

obvious candidates are as follows. The first is a non-parametric approximation based upon a swarm $\{s_{t-1}^{0,i}\}_{i=1}^{S}$:

$$\widehat{f}_S(s_{t-1}|Y_{t-1}) = \frac{1}{Sh}\sum_{i=1}^{S}\kappa\left(\frac{s_{t-1}-s_{t-1}^{0,i}}{h}\right).$$

The second 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. It is expected that in pathological cases, significant gains in accuracy resulting from inner EIS will far outweigh approximation errors in $f(s_{t-1}|Y_{t-1})$.

4.4 Special cases

4.4.1 Partial observability

Consider the case in which only a subset of the state variables is measured, thus information on the remaining state variables is indirect. Let s_t partition into $s_t = (p_t, q_t)$, and assume

$$f(y_t|s_t Y_{t-1}) \equiv f(y_t|p_t, Y_{t-1}). \tag{44}$$

In this case, likelihood evaluation requires integration only with respect to p_t:

$$f(y_t|Y_{t-1}) = \int f(y_t|p_t, Y_{t-1})\, f(p_t|Y_{t-1})\, dp_t, \tag{45}$$

and the updating equation (6) factorizes into the product of the following two densities:

$$f(p_t|Y_t) = \frac{f(y_t|p_t, Y_{t-1})\, f(p_t|Y_{t-1})}{f(y_t|Y_{t-1})}; \tag{46}$$

$$f(q_t|p_t, Y_t) = f(q_t|p_t, Y_{t-1}). \tag{47}$$

Stronger conditional independence assumptions are required in order to produce factorizations in (5). In particular, if p_t is independent of q_t given (p_{t-1}, Y_{t-1}), so that

$$f(p_t|s_{t-1}, Y_{t-1}) \equiv f(p_t|p_{t-1}, Y_{t-1}), \tag{48}$$

then

$$f(p_t|Y_{t-1}) = \int f(p_t|p_{t-1}, Y_{t-1})\, f(p_{t-1}|Y_{t-1})\, dp_{t-1}. \tag{49}$$

Note that under conditions (44) and (48), likelihood evaluation does not require processing

sample information on $\{q_t\}$. The latter is required only if inference on $\{q_t\}$ is itself of interest.

4.4.2 Degenerate transitions

When state transition equations include identities, corresponding transition densities are degenerate (or Dirac) in some of their components. This situation requires an adjustment to EIS implementation. Again, let s_t partition into $s_t = (p_t, q_t)$, and assume that the transition equations consist of two parts: 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):

$$q_t \equiv \phi(p_t, p_{t-1}, q_{t-1}) = \phi(p_t, s_{t-1}). \tag{50}$$

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 (50). 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 (50) 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

$$q_{t-1} = \psi(p_t, q_t, p_{t-1}) = \psi(s_t, p_{t-1}) \tag{51}$$

and Jacobian

$$J(s_t, p_{t-1}) = \frac{\partial}{\partial q_t}\psi(s_t, p_{t-1}), \tag{52}$$

we can take the limit of the integral in (50) as ε tends to zero, producing the following integral in p_t only:

$$f(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|q_{t-1}, Y_{t-1}) f(p_{t-1}, q_{t-1}|Y_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}. \tag{53}$$

Note that (53) 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 in Section 4.3.1, since the probability that any of the particles $s_{t-1}^{0,i}$ lies in that subspace is zero. The non-parametric approximation in Section 4.3.2 remains applicable, since it does not require that the particles $\left\{s_{t-1}^{0,i}\right\}$in (41) lie in the subspace. As in

Section 4.3.3, we can also approximate the integral in (53) by replacing $f(s_{t-1}|Y_{t-1})$ by $\overline{\omega}(\widehat{a}_{t-1})\, g(s_{t-1}|\widehat{a}_{t-1})$:

$$\widehat{f}(s_t|Y_{t-1}) = \frac{\overline{\omega}(\widehat{a}_{t-1})}{f(y_{t-1}|Y_{t-2})} \int J(s_t, p_{t-1})\, f(p_t|q_{t-1}, Y_{t-1})\, g(p_{t-1}, q_{t-1}|\widehat{a}_{t-1})\,|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}. \tag{54}$$

In this case, since $g(.|\widehat{a}_{t-1})$is not a sampler for $p_{t-1}|s_t$, we must evaluate (54) either by quadrature or its own EIS sampler.

One might infer from this discussion that the EIS filter is particularly tedious to implement under degenerate transitions, while the standard particle filter handles such degeneracy trivially in the transition from $\left\{s_{t-1}^{0,i}\right\}$ to $\left\{s_t^{1,i}\right\}$. While this is true, it is also true that these situations are prone to significant sample impoverishment problems, as illustrated in example 2 of Section 5.

4.5 A heuristic comparison between the standard, auxiliary and EIS particle filters

The standard, auxiliary and EIS particle filters essentially differ by their choice of importance samplers for the likelihood integral in (4):

- Standard: $f(s_t|Y_{t-1})$.
- Auxiliary:

$$h(s_t, Y_t) \propto \int f(y_t|\mu_t(s_{t-1}), Y_{t-1}) f(s_t|s_{t-1}, Y_{t-1}) f(s_{t-1}|Y_{t-1}) ds_{t-1}. \tag{55}$$

- EIS: $g(s_t|\widehat{a}_t)$.

Note that $g(s_t|\widehat{a}_t)$ approximates the actual posterior density $f(s_t|Y_t)$:

$$f(s_t|Y_t) \propto f(y_t|s_t, Y_{t-1}) \cdot \int f(s_t|s_{t-1}, Y_{t-1}) f(s_{t-1}|Y_{t-1}) ds_{t-1}. \tag{56}$$

Thus the auxiliary particle filter essentially differs from the EIS filter in that s_t is replaced by $\mu_t(s_{t-1})$ in $f(y_t|s_t, Y_{t-1})$. Assuming that $\mu_t(s_{t-1})$ represents the conditional expectation of $s_t|s_{t-1}, Y_{t-1}$, the following variance decomposition comes into play:

$$Var(s_t|Y_{t-1}) = E_{s_{t-1}|Y_{t-1}}\left[Var(s_t|s_{t-1}, Y_{t-1})\right] + Var_{s_{t-1}|Y_{t-1}}\left[\mu_t(s_{t-1})\right]. \tag{57}$$

This suggests that in order for the auxiliary particle filter to approach the EIS filter in terms of numerical accuracy, the variance of $s_t|s_{t-1}, Y_{t-1}$ must be negligible relative to that of $\mu_t(s_{t-1})$. The following analytical example provides heuristic support to our interpretation. Assume

$$s_{t-1}|Y_{t-1} \sim N(0, \upsilon^2), \quad s_t|s_{t-1}, Y_{t-1} \sim N(\alpha s_{t-1}, s^2), \quad y_t|s_t, Y_{t-1} \sim N(s_t, \sigma^2).$$

It then follows that

$$s_t|Y_t \sim N\left(\frac{\omega^2}{\omega^2+\sigma^2} \cdot y_t \, , \, \frac{\omega^2\sigma^2}{\omega^2+\sigma^2}\right), \tag{58}$$

$$s_t|Y_{t-1} \sim N(0, \omega^2), \tag{59}$$

$$h(s_t|Y_{t-1}) \sim N\left(\frac{\alpha^2\upsilon^2}{\alpha^2\upsilon^2+\sigma^2} \cdot y_t \, , \, \frac{\alpha^2 s^2\upsilon^2+\omega^2\sigma^2}{\alpha^2\upsilon^2+\sigma^2}\right), \tag{60}$$

with $\omega^2 = s^2 + \alpha^2\upsilon^2$. Furthermore,

$$\omega^2 \geq_{(1)} \frac{\alpha^2 s^2\upsilon^2+\omega^2\sigma^2}{\alpha^2\upsilon^2+\sigma^2} \geq_{(2)} \frac{\omega^2\sigma^2}{\omega^2+\sigma^2} \tag{61}$$

and

$$0 \leq_{(1)} \frac{\alpha^2\upsilon^2}{\alpha^2\upsilon^2+\sigma^2} \leq_{(2)} \frac{\omega^2}{\omega^2+\sigma^2}, \tag{62}$$

with equalities (1) obtaining i.f.f. $s^2 = 0$, and (2) i.f.f. $\alpha^2\upsilon^2 = 0$. Note that the auxiliary particle filter lies "between" the standard particle filter and the actual posterior density (which is also the EIS sampler under normality). If $s^2 = 0$, the auxiliary filter coincides with EIS filter; and if $\alpha^2\upsilon^2 = 0$, with the particle filter. Note that the terms s^2 and $\alpha^2\upsilon^2$ are precisely those appearing in the variance decomposition (57).

5 Examples

Here we present two numerical examples that illustrate the relative performance of the standard, auxiliary (zero-order Taylor series expansion), adapted (first-order Taylor series expansion), and EIS particle filters. We begin with some lessons gleaned through these examples regarding the selection of the three auxiliary sample sizes employed under the EIS filter: N, the number of draws used for likelihood evaluation (e.g., see (27)); R, the number of draws used to construct EIS samplers (e.g., see (26)); and S, the number of

draws used to evaluate $f(s_t|Y_{t-1})$ (e.g., see (37)).

First, the efficiency of the EIS filter typically translates into substantial reductions (relative to the particle filter) in the number of draws N needed to reach given levels of numerical accuracy: often by two to three orders of magnitude. In all but the most well-behaved cases, this translates into efficiency gains that more than compensate for the additional calculations required to implement the EIS filter. More importantly, the EIS filter is far more reliable in generating numerically stable and accurate results when confronted with ill-behaved problems (e.g., involving outliers).

Second, in every case we have considered, EIS samplers can be constructed reliably using small values for R (e.g., 100 has sufficed for univariate applications).

Third, as with any filter, the range $s_t|Y_{t-1}$ must be sufficiently wide to accommodate period-t surprises (outliers in s_t and/or y_t). At the same time, the approximation grid must be sufficiently fine to accommodate the realization of highly informative realizations of y_t, which generate significant tightening of the distribution of $s_t|Y_t$ relative to that of $s_t|Y_{t-1}$. Both considerations push towards relatively large values for S. The standard particle filter implicitly sets $N = S$. However, repeated evaluations of $f(s_t|Y_{t-1})$ constitute a substantial portion of EIS computing time, thus in certain cases setting $S << N$ can yield significant gains in overall efficiency. Indeed, in the first example we set $S = 100$. (In the second example, we have an analytical approximation for $f(s_t|Y_{t-1})$, thus the need to set S is eliminated.) Note that it is trivial to rerun an EIS algorithm under different values for S, thus it is advisable to experiment with alternative values of S in trial runs before launching full-scale analyses in complex applications.

5.1 Example 1: Univariate model with frequent outliers

This example is from Fernandez-Villaverde and Rubio-Ramirez (2004). The state-transition and measurement equations are given by

$$s_{t+1} = \alpha + \beta \frac{s_t}{1+s_t^2} + \upsilon_{t+1} \tag{63}$$

$$y_t = s_t + u_t, \tag{64}$$

where $\upsilon_{t+1} \sim N\left(0, \sigma_\upsilon^2\right)$ and u_t is t-distributed with ν degrees of freedom:

$$f(u_t) \sim \left(\nu + u_t^2\right)^{-0.5(\nu+1)}, \qquad Var(u_t) = \frac{\nu}{\nu-2} \text{ for } \nu > 2.$$

In all cases, the parameters α and β are both set to 0.5; adjustments to these settings have minimal impact on our results. Note that the expectation of $s_{t+1}|s_t$ is highly non-linear around $s_t = 0$, and becomes virtually constant for $|s_t| > 10$.

We consider two values for ν: 2 and 50. For $\nu = 2$, the variance of u_t is infinite and the model generates frequent outliers: e.g., $\Pr(|u_t| > 4.303) = 0.05$. For $\nu = 50$, u_t is virtually gaussian: its variance is 1.42, and $\Pr(|u_t| > 2.010) = 0.05$. We consider four values for σ_υ: $(1/3, 1, 3, 10)$. Thus the parameterizations we consider cover a wide range of scenarios, ranging from well-behaved $(\nu = 50, \sigma_\upsilon = 1/3)$ to ill-behaved $(\nu = 2, \sigma_\upsilon = 10)$.

We compare the relative numerical efficiency of five algorithms. The first three are the standard, auxiliary, and adapted particle filters. These are implemented using values of N ranging from 100 to 200,000. Recall that for each index k, the auxiliary particle filter is based on a zero-order approximation of $f(y_t|s_t, Y_{t-1})$,which is given by

$$f\left(y_t|\mu_t^k, Y_{t-1}\right) \propto \left[\nu + \left(y_t - \mu_t^k\right)^2\right]^{-\frac{1}{2}(\nu+1)},$$

where

$$\mu_t^k = \alpha + \beta \frac{s_{t-1}^{0,k}}{1 + \left(s_{t-1}^{0,k}\right)^2}.$$

The adapted particle filter is based on the first-order approximation defined in (18), where $h(\cdot)$ in this case is given by

$$h\left(s_t; \mu_t^k, Y_t\right) = \exp\left[(\nu+1)\frac{y_t - \mu_t^k}{\nu + \left(y_t - \mu_t^k\right)^2}\left(s_t - \mu_t^k\right)\right].$$

This implies the following conditional expectation for the adapted sampler in (19):

$$\mu_{*t}^k = \mu_t^k + \sigma_\nu^2 (\nu+1)\frac{y_t - \mu_t^k}{\nu + \left(y_t - \mu_t^k\right)^2}.$$

The remaining algorithms are the gaussian EIS filter and the piecewise-linear EIS filter. These are implemented using N ranging from 100 to 1,000. Evaluation of $f(s_t|Y_{t-1})$ is based on the weighted-sum approximation introduced in Section 4.3.1 – see (37).

Results obtained using artificial data sets of size $T = 100$ are presented in Figure 2 and Tables 1 $(\nu = 2)$ and 2 $(\nu = 50)$. Numerical accuracy is assessed by running 100

i.i.d. likelihood evaluations obtained under different seeds. Means of these likelihood evaluations are interpreted as 'final' likelihood estimates; MC standard deviations of these means provide a direct measure of the stochastic numerical accuracy of the final likelihood estimates.

Figure 2 illustrates relationships between MC standard deviations and computing time obtained using the five algorithms. The tables also report this information, along with MC means of likelihood evaluations. In addition, the tables report a convenient measure of the relative time efficiency of filters i and j:

$$RTE_{i,j} = \frac{T_i V_i}{T_j V_j},$$

where T_i represents computing time per function evaluation, and V_i the MC variance associated with filter i. In the tables, i represents the standard particle filter in all comparisons, thus for ratios less than one, the standard particle filter is the relatively efficient estimator. Reported ratios are based on $N = 200,000$ for the particle, auxiliary and adapted filters, and $N = 1,000$ for the EIS filters.

Note first that RTEs obtained for the auxiliary particle filter range from 0.7 to 1.1 across all cases considered. Thus roughly speaking, regardless of whether the model is well- or ill-behaved, the efficiency gains it generates are offset by associated increases in required computing time, which are on the order of 40%.

Next, for well-behaved cases, RTEs of the adapted particle filter are good; e.g., for $\sigma_\upsilon = 1/3$, efficiency ratios are 8.2 for $\nu = 2$ and 11.6 for $\nu = 50$. However, its performance deteriorates dramatically as σ_υ increases. Indeed, results are not reported for $(\nu = 2, \sigma_\upsilon = 10; \nu = 50, \sigma_\upsilon = 3; \nu = 50, \sigma_\upsilon = 10)$, since in these cases estimated likelihood values diverge pathologically. This reflects the general inability of *local* approximations to provide reliable global approximations of $f(y_t|s_t, Y_{t-1})$ when relevant ranges for s_t become too large. In the present case, problems become critical for Taylor expansions around inflection points of the non-log-concave Student-t density ($y_t = \mu_t^k \pm \sqrt{\nu}$). Note that these are precisely points where second derivatives with respect to s_t are zero, which implies that the use of second-order approximations (e.g., as advocated by Smith and Santos, 2006) would fail to provide an effective remedy in this application.

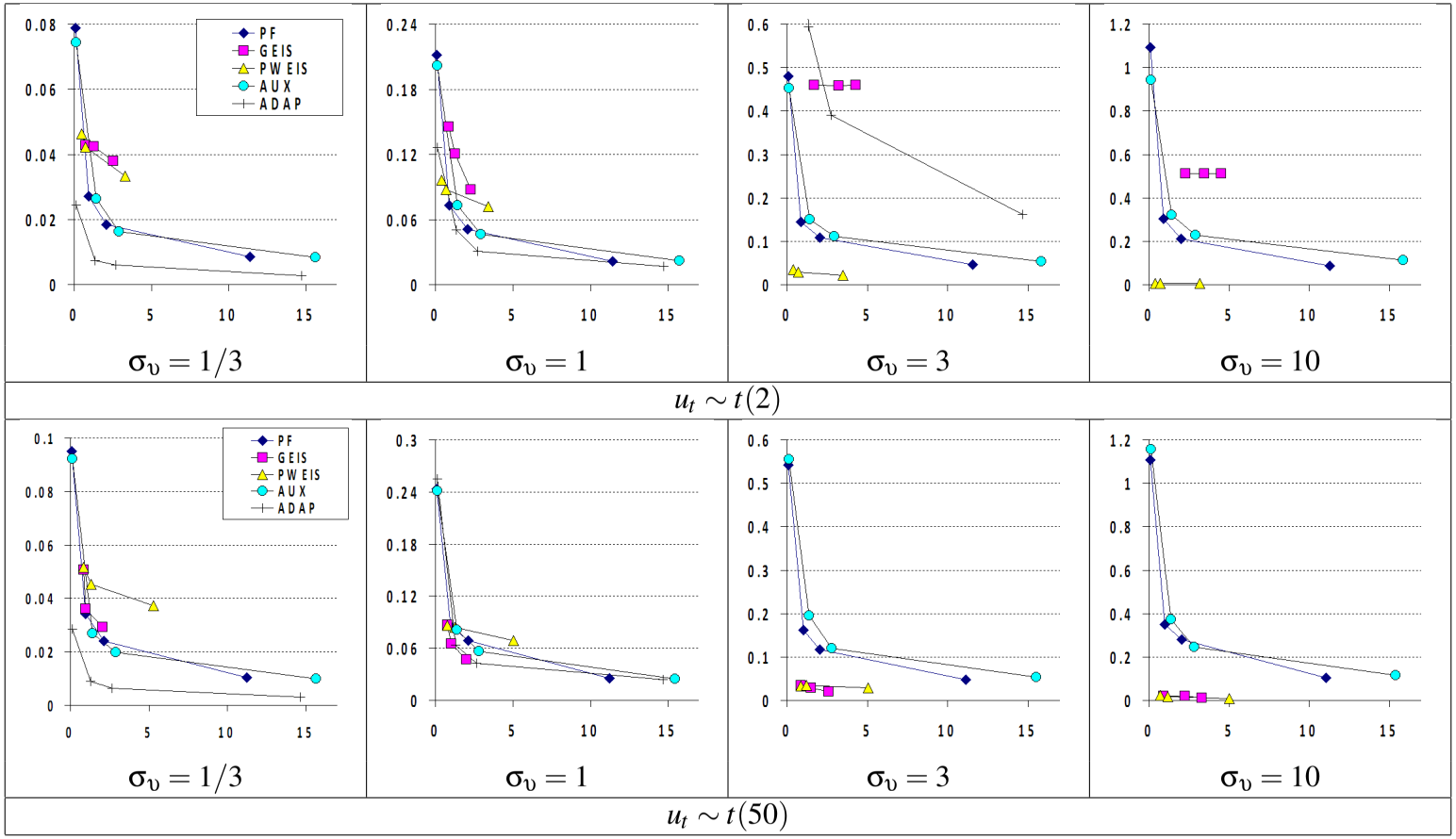

Figure 2: Plot of MC Standard Deviation of Likelihood vs. Time per Function Evaluation

Legend: From **Left** to **Right**, N=10K, 20K, 100K, 200K for the points on all PF curves. From **Left** to **Right**, N=100, 200, 1000 for the points on all EIS curves. R=100 and S=100 for all EIS Curves. Piecewise EIS uses a grid of R intervals. Two intervals of 0.001 probability weight each (at either end of the support) use constant weights. The domain of approximation is chosen using an estimated $\widehat{f}(s_t|y_{t-1})$ with its mean serving as the center and $\pm 5 \times Std.Deviation$ as the upper and lower limits.

Table 1: Univariate Model with Frequent Outliers, $u_t \sim t(2)$.

STANDARD PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-195.874	0.2441	0.010	-213.866	0.6750	0.008	-264.819	1.5108	0.008	-374.168	3.5771	0.008
200	-195.854	0.1740	0.013	-213.704	0.4755	0.015	-264.269	1.1253	0.016	-371.737	2.5684	0.016
1000	-195.845	0.0788	0.071	-213.606	0.2114	0.073	-263.578	0.4806	0.073	-368.935	1.0939	0.074
10000	-195.847	0.0271	0.924	-213.596	0.0729	0.902	-263.534	0.1442	0.861	-368.624	0.3035	0.890
20000	-195.846	0.0183	2.065	-213.593	0.0512	2.083	-263.541	0.1089	2.054	-368.542	0.2117	1.995
100000	-195.844	0.0085	11.391	-213.589	0.0212	11.432	-263.525	0.0458	11.542	-368.519	0.0869	11.286
200000	-195.843	0.0057	22.119	-213.591	0.0159	22.593	-263.534	0.0339	24.269	-368.532	0.0700	22.798
AUXILIARY PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-195.842	0.2427	0.012	-213.747	0.6804	0.012	-264.614	1.7226	0.012	-373.722	3.6753	0.012
200	-195.859	0.1700	0.024	-213.778	0.4470	0.025	-264.330	1.0785	0.024	-371.512	2.4577	0.024
1000	-195.826	0.0744	0.125	-213.576	0.2015	0.126	-263.674	0.4532	0.128	-369.028	0.9425	0.127
10000	-195.839	0.0263	1.411	-213.604	0.0731	1.414	-263.537	0.1507	1.416	-368.567	0.3214	1.401
20000	-195.839	0.0162	2.907	-213.605	0.0461	2.912	-263.531	0.1117	2.930	-368.545	0.2285	2.906
100000	-195.840	0.0084	15.630	-213.603	0.0222	15.728	-263.537	0.0537	15.838	-368.511	0.1122	15.858
200000	-195.840	0.0055	31.811	-213.601	0.0159	32.343	-263.526	0.0335	32.075	-368.520	0.0697	32.285
Rel. Time Efficiency			0.748			0.699			0.775			0.713
ADAPTED PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-195.836	0.0792	0.011	-213.750	0.4513	0.012	-285.835	8.0919	0.012			
200	-195.829	0.0525	0.024	-213.563	0.3039	0.024	-272.899	4.7383	0.024			
1000	-195.837	0.0244	0.123	-213.618	0.1264	0.125	-265.598	1.6661	0.123			
10000	-195.840	0.0074	1.337	-213.600	0.0506	1.329	-263.730	0.5948	1.334		N/A	
20000	-195.839	0.0059	2.710	-213.600	0.0306	2.730	-263.635	0.3902	2.740			
100000	-195.840	0.0026	14.730	-213.601	0.0169	14.726	-263.544	0.1625	14.649			
200000	-195.840	0.0017	29.945	-213.604	0.0113	29.935	-263.543	0.1312	29.778			
Rel. Time Efficiency			8.182			1.497			0.0543			

Note: Table continues on next page.

Table 1, continued. Univariate Model with Frequent Outliers, $u_t \sim t(2)$.

PIECEWISE-EIS PARTICLE FILTER $(R = 100, S = 100)$												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-194.866	0.046	0.493	-212.564	0.097	0.391	-262.489	0.035	0.407	-367.091	0.0076	0.409
200	-195.925	0.042	0.721	-214.317	0.088	0.701	-263.555	0.029	0.716	-367.904	0.0063	0.719
1000	-199.556	0.033	3.300	-217.810	0.072	3.410	-264.419	0.022	3.480	-368.449	0.0048	3.180
Rel. Time Efficiency			0.197			0.322			16.998			1545.116
GAUSSIAN-EIS PARTICLE FILTER $(R = 100, S = 100)$												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-195.825	0.043	0.717	-213.815	0.145	0.860	-266.016	0.459	1.699	-372.937	0.5117	2.278
200	-195.820	0.042	1.248	-213.714	0.121	1.248	-265.189	0.459	3.218	-372.020	0.5120	3.462
1000	-195.819	0.038	2.530	-213.601	0.087	2.287	-264.156	0.459	4.265	-370.858	0.5110	4.491
Rel. Time Efficiency			0.199			0.325			0.031			0.095

Note: Means and standard deviations are based on 100 Monte Carlo replications; Relative Time Efficiency is based on N=200,000 for the SPF and N=1,000 for the EIS filter.

Table 2: Univariate Model with Frequent Outliers $u_t \sim t(50)$.

STANDARD PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-141.209	0.2809	0.007	-166.495	0.8473	0.007	-242.466	1.5374	0.007	-366.869	6.4435	0.007
200	-141.199	0.2017	0.015	-166.317	0.5482	0.014	-241.804	1.0940	0.015	-361.076	3.7848	0.015
1000	-141.166	0.0951	0.078	-166.166	0.2437	0.079	-241.207	0.5425	0.076	-357.374	1.1088	0.076
10000	-141.168	0.0342	0.962	-166.159	0.0852	0.971	-241.070	0.1621	1.005	-356.852	0.3493	0.992
20000	-141.169	0.0240	2.127	-166.151	0.0687	2.134	-241.043	0.1166	2.053	-356.860	0.2797	2.026
100000	-141.165	0.0104	11.237	-166.149	0.0256	11.235	-241.035	0.0473	11.118	-356.801	0.1033	11.039
200000	-141.166	0.0077	23.131	-166.149	0.0195	22.915	-241.038	0.0350	22.485	-356.821	0.0750	22.533
AUXILIARY PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-141.198	0.3184	0.012	-166.544	0.8582	0.012	-242.784	1.6599	0.012	-367.118	6.4728	0.012
200	-141.160	0.2327	0.023	-166.184	0.6184	0.024	-241.491	1.0960	0.024	-360.795	3.7327	0.024
1000	-141.159	0.0921	0.124	-166.166	0.2410	0.124	-241.297	0.5557	0.125	-357.560	1.1562	0.125
10000	-141.135	0.0268	1.386	-166.143	0.0808	1.375	-241.067	0.1951	1.358	-356.906	0.3725	1.354
20000	-141.138	0.0199	2.878	-166.153	0.0564	2.806	-241.052	0.1203	2.794	-356.879	0.2464	2.803
100000	-141.137	0.0100	15.632	-166.141	0.0246	15.455	-241.042	0.0534	15.503	-356.831	0.1142	15.401
200000	-141.139	0.0064	31.423	-166.147	0.0172	31.752	-241.037	0.0350	31.806	-356.811	0.0763	31.234
Rel. Time Efficiency			1.067			0.929			0.707			0.697
ADAPTED PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-141.169	0.0958	0.012	-166.391	0.6827	0.012						
200	-141.162	0.0603	0.023	-166.280	0.5327	0.023						
1000	-141.168	0.0286	0.120	-166.189	0.2551	0.120						
10000	-141.171	0.0090	1.282	-166.146	0.0634	1.288		N/A			N/A	
20000	-141.168	0.0063	2.629	-166.144	0.0425	2.643						
100000	-141.170	0.0032	14.635	-166.147	0.0237	14.671						
200000	-141.170	0.0020	29.892	-166.145	0.0159	30.015						
Rel. Time Efficiency			11.590			1.157						

Note: Table continues on next page.

Table 2, continued. Univariate Model with Frequent Outliers, $u_t \sim t(50)$.

PIECEWISE-EIS PARTICLE FILTER $(R = 100, S = 100)$												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-140.212	0.052	0.840	-165.123	0.087	0.791	-240.888	0.034	0.865	-355.866	0.0230	0.720
200	-140.759	0.045	1.310	-165.748	0.084	1.250	-241.062	0.035	1.199	-356.372	0.0181	1.174
1000	-141.125	0.037	5.277	-166.341	0.069	5.025	-241.148	0.028	5.055	-356.812	0.0096	4.996
Rel. Time Efficiency			0.187			0.370			6.804			273.146
GAUSSIAN-EIS PARTICLE FILTER $(R = 100, S = 100)$												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**	**Mean**	**Stdev**	**Time**
100	-141.171	0.051	0.833	-166.160	0.087	0.791	-241.088	0.034	0.865	-356.868	0.0215	0.896
200	-141.173	0.036	0.972	-166.159	0.065	1.012	-241.058	0.029	1.506	-356.837	0.0212	2.252
1000	-141.175	0.029	2.049	-166.159	0.047	1.996	-241.046	0.020	2.576	-356.812	0.0116	3.308
Rel. Time Efficiency			0.782			1.996			25.493			284.197

Note: Means and standard deviations are based on 100 Monte Carlo replications; Relative Time Efficiency is based on N=200,000 for the SPF and N=1,000 for the EIS filter.

As expected, RTEs of the gaussian EIS filter are also poor given $\nu = 2$, especially when σ_υ is large. However, the gaussian EIS filter does not fail as dramatically as the adapted filter given $\sigma_\upsilon = 10$. This reflects the fact that EIS approximations are *global*, and thus are more robust than local ones (see RZ for further discussion). For $\nu = 50$, the gaussian EIS filter performs well, with impressive RTEs for large values of σ_υ (reaching 273 for $\sigma_\upsilon = 10$).

Due once again to relative computational expense, the piecewise-linear EIS filter is also inefficient relative to the standard particle filter in well-behaved cases, but the payoff of its adoption is dramatic in the challenging cases. For $\nu = 2$, its RTE ranges from 0.197 to $1,545$ as σ_υ increases from 1/3 to 10; for $\nu = 50$, its RTE ranges from 0.187 to 273. For context in interpreting these results, an RTE of 1,545 implies that the standard particle filter would require approximately 1 hour and 22 minutes (the time required to process approximately 42.5 million particles) to match the numerical accuracy of the piecewise-linear filter with $N = 1,000$ (which requires 3.18 seconds). These results clearly reflect the payoffs associated with the flexibility, in addition to the global nature, of approximations provided by the piecewise-linear filter.

In sum, the standard, auxiliary, and adapted particle filters perform relatively well under well-behaved scenarios. In these cases, their relative numerical inaccuracy is more than offset by their relative speed. However, expansions in the range of s_t, along with the presence of outliers, can lead to dramatic reductions in RTEs, and in the case of the auxiliary and adapted filters, can also lead to unreliable likelihood estimates. In contrast, the EIS filters provide insurance against these problems and exhibit superior RTEs in all but the most well-behaved cases.

However, while relative numerical efficiency is an important feature of any likelihood approximation procedure, it is not the only important feature. In pursuing maximum-likelihood parameter estimates, continuity with respect to model parameters is also critical. The next example highlights this feature.

5.2 Example 2: A dynamic stochastic general equilibrium model

Following Sargent (1989), likelihood-based analyses of dynamic stochastic general equilibrium (DSGE) models have long involved the application of the Kalman filter to log-linear model approximations (e.g., see DeJong, Ingram and Whiteman, 2000; Otrok, 2001; Ireland, 2004; and the survey by An and Schorfheide, 2007). However, Fernandez-

Villaverde, Rubio-Ramirez and Santos (2006) have shown that second-order approximation errors in model solutions map into first-order effects on the corresponding likelihood function, due to the accumulation of approximation errors on a period-by-period basis. Fernandez-Villaverde and Rubio-Ramirez (2005) document the quantitative relevance of this phenomenon in an empirical analysis involving estimates of a neoclassical growth model obtained using the particle filter.

Here we demonstrate the performance of the EIS filter in an exercise constructed following DeJong with Dave (2007). The objective is to estimate the structural parameters of a simple growth model via maximum likelihood. Regarding the model, let q_t, k_t, c_t, i_t, and a_t represent output, capital, consumption, investment, and total factor productivity (TFP). Labor is supplied inelastically and fixed at unity. The model is of a representative agent who seeks to maximize the expected value of lifetime utility

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \ln(c_t),$$

subject to

$$\begin{aligned} q_t &= a_t k_t^{\alpha} && (65)\\ q_t &= c_t + i_t && (66)\\ k_{t+1} &= i_t + (1-\delta)k_t && (67)\\ \ln(a_{t+1}) &= \rho \ln(a_t) + \varepsilon_t. && (68) \end{aligned}$$

Regarding parameters, α is capital's share of output, δ is capital depreciation, ρ determines the persistence of innovations to TFP, and the innovations $\varepsilon_t \sim N(0, \sigma^2)$. The state variables (a_t, k_t) are unobserved, and the distribution of (a_0, k_0) is known.

The solution of this problem can be represented as a policy function for consumption of the form $c(a_t, k_t)$. For the special case in which $\delta = 1$, $c(a_t, k_t) = (1 - \alpha\beta) a_t k_t^{\alpha}$. This is the case studied here.

We take q_t and i_t as observable, subject to measurement error. Combining equations,

the measurement equations are

$$\begin{aligned} q_t &= a_t k_t^{\alpha} + u_{q_t}, && (69) \\ i_t &= a_t k_t^{\alpha} - c(a_t, k_t) + u_{i_t} && (70) \\ &= \alpha\beta a_t k_t^{\alpha} + u_{i_t}, \end{aligned}$$

and the state-transition equations are (68) and

$$\begin{aligned} k_t &= a_{t-1} k_{t-1}^{\alpha} - c(a_{t-1}, k_{t-1}) && (71) \\ &= \alpha\beta a_t k_t^{\alpha}. \end{aligned}$$

Closer examination of (68) to (71) suggests reparameterizing the state variables as

$$z_t = \ln(a_t) \quad \text{and} \quad l_t = e^{z_t} k_t^{\alpha}, \tag{72}$$

where l_t denotes (unobserved) output, and $s_t = [l_t \quad z_t]'$ denotes the state vector. The transition process (68) then takes the form of a gaussian AR(1) in z_t, and the identity (71) can be rewritten as

$$l_t = e^{z_t} \left(\alpha\beta l_{t-1}\right)^{\alpha}. \tag{73}$$

Note that this example combines the two special cases discussed in Section 4.4. First, there is partial observability, in that y_t is independent of z_t conditionally on l_t (and Y_{t-1}):

$$y_t | s_t, Y_{t-1} \equiv y_t | l_t, Y_{t-1} \sim N_2 \left(\begin{pmatrix} 1 \\ \alpha\beta \end{pmatrix} l_t, \begin{bmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_i^2 \end{bmatrix} \right). \tag{74}$$

Second, (73) represents a degenerate Dirac transition, with inverse

$$l_{t-1} = \psi(s_t) = \frac{1}{\alpha\beta} \left(l_t e^{-z_t} \right)^{\frac{1}{\alpha}} \tag{75}$$

and Jacobian

$$J(s_t) = \frac{\partial \psi(s_t)}{\partial l_t} = \frac{1}{\alpha^2\beta} \left(l_t^{1-\alpha} e^{-z_t} \right)^{\frac{1}{\alpha}}. \tag{76}$$

In view of (74), and as discussed in Section 4.4.1, the likelihood integral simplifies into a univariate integral in l_t – see (45) – whose efficient evaluation requires only an EIS sampler for $l_t | Y_t$. Nevertheless, in period $t+1$, we still need to approximate $f(z_t | l_t, Y_t)$ in

order to compute $\widehat{f}(y_t|Y_{t-1})$. To efficiently capture the dependence between z_t and l_t given Y_t, it proves convenient to construct directly a single bivariate EIS sampler for $z_t, l_t|Y_t$. Whence the likelihood integral

$$f(y_t|Y_{t-1}) = \int f(y_t|l_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t \tag{77}$$

is evaluated under a bivariate gaussian EIS sampler $g(s_t|\hat{a}_t)$. Next, $f(s_t|Y_{t-1})$ is approximated according to (54), where we can exploit the fact that the Jacobian $J(s_t)$ does not depend on z_{t-1}:

$$\widehat{f}(s_t|Y_{t-1}) = J(s_t) \frac{\overline{\omega}(\widehat{a}_{t-1})}{f(y_{t-1}|Y_{t-2})} \int f(z_t|z_{t-1}) g(\psi(s_t), z_{t-1}|\hat{a}_{t-1}) dz_{t-1}. \tag{78}$$

Note that the integrand is quadratic in $z_{t-1}|s_t$, thus standard algebraic operations amounting to the completion of a quadratic form in z_{t-1} yield an analytical solution for $\widehat{f}(s_t|Y_{t-1})$. Thus under the implicit assumption that the EIS weights $\omega(s_t;\hat{a}_t)$ are near constant (to be verified empirically), we have derived a particularly fast and efficient EIS implementation based on a bivariate gaussian outer EIS, and an inner analytical approximation for $f(s_t|Y_{t-1})$.

Model estimates are based on artificial data simulated from the model. Parameter values used to simulate the data are as follows: $\alpha = 0.33$, $\beta = 0.96$, $\rho = 0.8$, $\sigma = 0.05$, $\sigma_q = 0.014$, $\sigma_i = 0.02$. The first four values are typical of this model calibrated to annual data; and given σ, the latter two values represent approximately 5% and 20% of the unconditional standard deviations of q_t and i_t. The unconditional mean and standard deviation of a_t implied by ρ and σ equal 1.0035 and 0.08378. The following results are based on a single data set of sample size $T = 100$.

To begin, we compute likelihood values at the true parameter values using from 100 to 100,000 particles for the particle filter, and from 100 to 1,000 auxiliary draws for the EIS filter (with R held fixed at 100). We do so for 100 MC replications. Results are reported in Table 3.

RTEs computed using as a numeraire the particle filter with $N = 100,000$ range from 6.825 (for $N = 100$ under the EIS filter) to 217.728 (for $N = 1,000$). That is, the time required for the particle filter to attain the same standard of numerical accuracy exceeds the time required by the EIS filter with $N = 1,000$ by a factor of approximately 217 (the time required to process approximately 8.7 million particles). This difference in efficiency

is due to the fact that the bivariate gaussian EIS samplers $g(s_t|\widehat{a}_t)$ provide close (global) approximations of the densities $f(s_t|Y_{t-1})$. Indeed, on a period-by-period basis, ratios of standard deviations to the means of the weights $\left\{\omega\left(s_t^{0,i};\widehat{a}_t\right)\right\}_{i=1}^{N}$ range from 1.14e-8 to 3.38e-3. Such small variations validate our reliance on (78) to approximate $f(s_t|Y_{t-1})$.

Table 3: DSGE Model

STANDARD PARTICLE FILTER				
N	**Mean**	**Stdev**	**Time**	**Rel. Time Efficiency**
100	434.416	3.382	0.0239	0.903
200	436.560	2.133	0.0481	1.127
1000	438.180	0.972	0.2512	1.039
10000	438.479	0.256	2.7846	1.344
20000	438.561	0.206	5.7651	1.000
100000	438.545	0.0774	29.036	1.417
GAUSSIAN-EIS PARTICLE FILTER $(R = 100)$				
N	**Mean**	**Stdev**	**Time**	**Rel. Time Efficiency**
100	438.313	0.1027	3.425	6.825
200	438.621	0.0278	5.731	55.440
1000	438.633	0.0083	16.414	217.728

Note: Means and standard deviations are based on 100 Monte Carlo replications; Relative Time Efficiency is based on N=20,000 for the SPF.

Next, we apply both the particle filter and the gaussian EIS filter to compute maximum likelihood estimates (MLEs) for $\theta = (\alpha, \beta, \rho, \sigma, \sigma_q, \sigma_i)$, under simulated samples of size T=40, 100 and 500. Using (70), the stepwise MLE of β given α is given by

$$\widehat{\beta} = \frac{\bar{i}}{\alpha \bar{l}},$$

where $\bar{i}$ and $\bar{l}$ denote sample means of i_t and l_t. MLEs for the remaining parameters are obtained via maximization of the concentrated log-likelihood function.

Results for the particle filter are based on $N = 20,000$; results for the EIS filter are based on $N = 200$ and $R = 100$. According to Table 3, computing times for a single likelihood evaluation are approximately the same under both methods (on the order of 5.5 seconds for $T = 100$), while MC estimates of the log-likelihood function are more accurate under the EIS filter (which has an RTE of approximately 55 given these settings for N and R). In addition, a graphical characterization of the relative accuracy of the EIS filter is provided in

Figure 3, which plots estimated log-likelihood functions along the α dimension for $T = 100$ (all other parameters being set at their ML values) for both the standard and EIS filters. In the figure, numerical inaccuracy is manifested by choppiness in the likelihood surface; note that the surface associated with the standard particle filter is particularly rough.

In order to minimize the impact of numerical inaccuracies, we employ the Nelder-Meade (1965) simplex optimization routine for all MLE computations, implemented following Lagarias et al. (1988). Unsurprisingly, convergence towards a solution is relatively fast under EIS, but only by a factor of approximately 1.4 to 2 depending on sample size. Following RZ, we use i.i.d replications (30 in the present set-up) of the complete ML algorithm in order to produce two sets of means and standard deviations for MLEs:

1. *Statistical* means and standard deviations are obtained from 30 different samples $\{y_t\}_{t=1}^T$ under a single set of auxiliary draws $\{u_i\}_{i=1}^N$. These statistics characterize the finite sample distribution of the MLEs. Under the EIS filter, we also compute the asymptotic standard deviations obtained by inversion of a numerical Hessian. As in Fernandez-Villaverde and Rubio-Ramirez (2007), we find that Hessians computed under the standard particle filter are unreliable and often fail to be positive definite.

2. *Numerical* means and standard deviations are obtained under 30 different sets of CRNs for a fixed sample $\{y_t\}_{t=1}^T$. Such means constitute our most accurate MC estimates of the MLEs and accordingly, the numerical standard deviations we report are those for the means.

Results of this experiment are given in Table 4. Highlights are as follows:

1. Log-likelihood functions are generally tightly peaked, as attested by the statistical standard deviations of the MLEs. This explains why the numerical inefficiency of the standard particle filter relative to the EIS filter is attenuated considerably for MLEs (accounting for computing time, RTEs range from 2 to 3, instead of 55 for the log-likelihood function MC estimates).

2. For $T = 40$, MLEs of α are significantly upward biased (by about 4 standard deviations), which is why we also report root mean-squared errors.

3. Under the EIS filter, there is remarkably close agreement between finite sample (MC) and asymptotic (Hessian) statistical standard deviations, especially as T increases.

While conceptually not surprising in view of the well-behaved log-likelihood functions, such concordance highlights the high numerical accuracy and reliability of EIS filter computations (including Hessians).

4. As T increases, the numerical standard deviations of the individual MLEs (which are $\sqrt{30}$ larger than those reported for the mean MLEs in Table 4) get closer to the corresponding statistical standard deviations. This does not appear to create a problem for the EIS filter (which employs CRNs), but clearly contaminates the computation of statistical standard deviations under the standard particle filter. For this example at least, the number of particles would need to be increased dramatically in order for the particle filter to provide reliable estimates of statistical standard deviations, whether finite sample or (moreso) asymptotic.

In sum, MLEs derived using the EIS filter ($N = 200$, $R = 100$) are numerically and statistically significantly more reliable than those derived under the standard particle filter ($N = 20,000$). They are also obtained relatively more rapidly (by a factor of 25% to 50%).

Table 4: MLE Comparisons

STANDARD PARTICLE FILTER								
	True	**Stat. Moments**				**Num. Moments**		
T=40	True	Mean	*S.D.*[a]	*S.D.*[b]	RMSE	Mean	*S.D.*[c]	*S.D.*[d]
α	0.33	0.34559	5.888E-03		1.667E-02	0.34826	3.158E-03	5.765E-04
β	0.96	0.91989	1.843E-02		4.414E-02	0.93891	8.490E-03	1.550E-03
ρ	0.8	0.82840	2.899E-02		4.058E-02	0.81292	1.735E-02	3.168E-03
σ	0.05	0.04802	4.363E-03	n/a	4.792E-03	0.05184	1.384E-03	2.527E-04
σ_l	0.014	0.01462	4.118E-03		4.165E-03	0.01588	6.831E-04	1.247E-04
σ_i	0.02	0.01955	2.030E-03		2.080E-03	0.02149	4.275E-04	7.804E-05
T=100								
α	0.33	0.33404	5.701E-03		6.987E-03	0.33922	3.027E-03	5.527E-04
β	0.96	0.91995	1.568E-02		4.301E-02	0.94024	9.035E-03	1.650E-03
ρ	0.8	0.79697	1.961E-02		1.984E-02	0.82860	1.698E-02	3.100E-03
σ	0.05	0.05413	3.348E-03	n/a	5.316E-03	0.05085	1.246E-03	2.275E-04
σ_l	0.014	0.01350	2.958E-03		3.001E-03	0.01416	6.052E-04	1.105E-04
σ_i	0.02	0.01991	1.259E-03		1.262E-03	0.02030	3.544E-04	6.470E-05
T=500								
α	0.33	0.33162	5.440E-03		5.739E-03	0.33399	3.498E-03	6.386E-04
β	0.96	0.95501	1.523E-02		1.611E-02	0.95049	9.930E-03	1.813E-03
ρ	0.8	0.81776	1.599E-02		2.459E-02	0.80170	1.759E-02	3.212E-03
σ	0.05	0.05238	2.579E-03	n/a	3.717E-03	0.05365	1.899E-03	3.468E-04
σ_l	0.014	0.01361	1.717E-03		1.849E-03	0.01400	4.898E-04	8.942E-05
σ_i	0.02	0.01971	5.413E-04		6.368E-04	0.01922	2.604E-04	4.754E-05
GAUSSIAN-EIS PARTICLE FILTER								
T=40	True	Mean	*S.D.*[a]	*S.D.*[b]	RMSE	Mean	*S.D.*[c]	*S.D.*[d]
α	0.33	0.34071	2.389E-03	1.974E-03	1.097E-02	0.34868	1.652E-03	3.016E-04
β	0.96	0.93364	1.085E-02	n/a	2.851E-02	0.94006	4.442E-03	8.110E-04
ρ	0.8	0.81669	1.562E-02	1.145E-02	2.286E-02	0.81811	9.791E-03	1.788E-03
σ	0.05	0.04879	3.193E-03	3.009E-03	3.416E-03	0.05425	9.340E-03	1.705E-03
σ_l	0.014	0.01535	2.276E-03	2.107E-03	2.646E-03	0.01594	2.415E-04	4.410E-05
σ_i	0.02	0.01984	1.586E-03	1.544E-03	1.594E-03	0.02155	2.349E-04	4.289E-05
T=100								
α	0.33	0.33380	4.635E-03	3.977E-03	5.996E-03	0.33601	1.577E-03	2.880E-04
β	0.96	0.92038	1.557E-02	n/a	4.337E-02	0.93960	4.441E-03	8.109E-04
ρ	0.8	0.79867	1.764E-02	1.432E-02	1.769E-02	0.82182	9.667E-03	1.765E-03
σ	0.05	0.05083	3.284E-03	3.312E-03	3.388E-03	0.05041	9.492E-03	1.733E-03
σ_l	0.014	0.01407	2.998E-03	2.919E-03	2.999E-03	0.01448	2.471E-04	4.512E-05
σ_i	0.02	0.01990	1.163E-03	1.214E-03	1.167E-03	0.02183	2.144E-04	3.914E-05
T=500								
α	0.33	0.33032	2.235E-03	2.243E-03	2.385E-03	0.33095	1.395E-03	2.547E-04
β	0.96	0.95610	7.478E-03	n/a	8.830E-03	0.95744	7.988E-03	1.458E-03
ρ	0.8	0.81082	6.408E-03	6.199E-03	1.275E-02	0.80102	7.142E-03	1.304E-03
σ	0.05	0.05108	1.064E-03	1.088E-03	1.559E-03	0.05031	4.068E-03	7.428E-04
σ_l	0.014	0.01400	7.206E-04	6.825E-04	7.623E-04	0.01400	2.193E-04	4.003E-05
σ_i	0.02	0.01997	2.295E-04	2.262E-04	2.444E-04	0.01998	1.891E-04	3.453E-05

a. Finite Sample S.D., b. Asymptotic S.D., c. S.D. of a single Draw, d. S.D. of the mean.

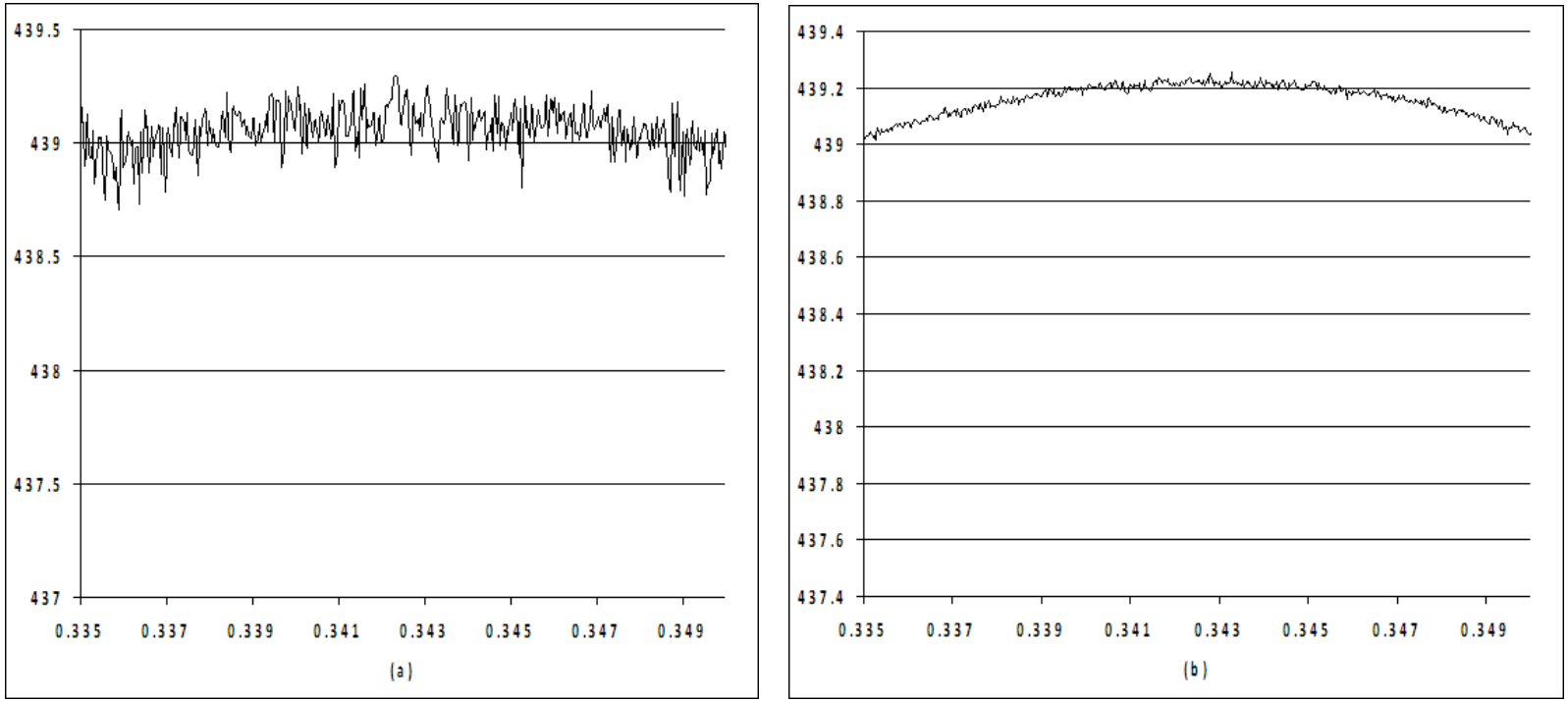

Figure 3: Plot of Log-Likelihood Function vs. α

(a) - Standard Particle Filter with N=20000. (b) - Gaussian-EIS Particle Filter (R=100 and N=200)

The likelihood function was computed with only α being changed while all the other parameters were kept constant at the values listed in Section 5.2.

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 constructs likelihood approximations using an optimization procedure designed to minimize numerical standard errors associated with the approximated likelihood. Resulting approximations are continuous in underlying likelihood parameters, greatly facilitating the implementation of ML estimation procedures. Implementation of the filter is straightforward, and as the examples we have presented demonstrate, the payoff of adoption can be substantial.

As noted in the introduction, an important byproduct of filtering, if not the primary focus in many applications, are estimates of filtered moments of the state variables. The efficient numerical estimation of these moments is the subject of future research.

References

[1] An, S., and F. Schorfheide, 2007, "Bayesian Estimation of DSGE Models", *Econometric Reviews.* Forthcoming.

[2] 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.

[3] DeGroot, M.H., 1984, *Probability and Statistics*. Reading PA: Addison-Wesley.

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

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

[6] Devroye, L., 1986, *Non-Uniform Random Variate Generation*. New York: Springer.

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

[8] Fernandez-Villaverde, J. and J.F. Rubio-Ramirez, 2004, "Sequential Monte Carlo Filtering: An Example", University of Pennsylvania Working Paper.

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

[10] Fernandez-Villaverde, J. and J.F. Rubio-Ramirez, 2007, "Estimating Macroeconomic Models: A Likelihood Approach", *Review of Economic Studies.* Forthcoming.

[11] Fernandez-Villaverde, J., J.F. Rubio-Ramirez, and M.S. Santos, 2006, "Convergence Properties of the Likelihood of Computed Dynamic Models", *Econometrica* 74, 93-119.

[12] Geweke, J., 1989, "Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica.* 24, 1037-1399.

[13] 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.

[14] Granger, C.W.J., 1969, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica.* 37, 424-438.

[15] Hendry, D.F., 2004, "Monte Carlo Experimentation in Econometrics", in R.F. Engle and D.L. McFadden, Eds. *The Handbook of Econmetrics*, Vol. IV. New York: North Holland.

[16] Ireland, P., 2004, "A Method for Taking Models to Data", *Journal of Economic Dynamics and Control.* 28, 1205-1226.

[17] Kim, S., N. Shephard, and S. Chib, 1998, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*. 65, 361-393.

[18] Kitagawa, G., 1987, "Non-Gaussian State-Space Modeling of Non-Stationary Time Series", *Journal of the American Statistical Association*. 82, 503-514.

[19] Otrok, C., 2001, "On Measuring the Welfare Cost of Business Cycles", *Journal of Monetary Economics.* 47, 61-92.

[20] Pitt, M.K., 2002, "Smooth Particle Filters for Likelihood Evaluation and Maximisation", University of Warwick Working Paper.

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

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

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

[24] 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.