An Efficient Approach to Analyzing 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

First Version: April 2007 This Revision: October 2007

Abstract

We develop a numerical procedure that facilitates efficient likelihood evaluation and filtering 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.

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, pseudo-code and Fortran 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 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 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 (1987). 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, 2007) for applications involving dynamic stochastic general equilibrium 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, 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); 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. The approximations are 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.

2 Overview of the Filtering Problem

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 objective of filtering is to infer the behavior of s_t given Y_t, and an assumed state-space representation; likelihood evaluation obtains as a by-product of the filtering process.

State-space representations consist of a state-transition equation

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

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

surement) equation

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

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

Filtering is facilitated 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. The process is initialized with a marginal density $f(s_0)$, which can be degenerate as a special case. From these densities, the goal is to construct $f\left(s_t|Y_t\right)$, which can then be used to calculate, e.g., $E_t\left(s_t|Y_t\right)$.

From Bayes' theorem, $f\left(s_t|Y_t\right)$ is given by

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

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

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

and $f\left(y_t|Y_{t-1}\right)$ is given by

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

Note that the recursive structure of $f\left(s_t|Y_t\right)$ evident in (3) and (4) indicates that filtering can be implemented via forward recursion, beginning with the known density $f(s_0) \equiv f\left(s_0|Y_0\right)$. Note also that since the likelihood function $f(Y_T)$ factors sequentially as

$$f\left(Y_T\right) = \prod_{t=1}^{T} f\left(y_t|Y_{t-1}\right), \tag{6}$$

where $f\left(y_1|Y_0\right) \equiv f(y_1)$, likelihood evaluation obtains as a by-product of the filtering process.

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 (3) and (5), this can be written as

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

3 The Particle Filter and Leading Extensions

Since our procedure is an extension of the particle filter, we provide a brief overview here. The particle filter is an algorithm that recursively generates random numbers approximately distributed as $f\left(s_t|Y_t\right)$. To characterize its implementation, let $s_t^{r,i}$ denote the i^{th} draw of s_t obtained from the conditional density $f\left(s_t|Y_{t-r}\right)$ 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 (4). This is done by drawing $s_t^{1,i}$ from the conditional density $f\left(s_t|s_{t-1}^{0,i}, Y_{t-1}\right)$, $i = 1, ..., N$. Note that $\{s_t^{1,i}\}_{i=1}^N$ can be used to produce an MC estimate of $f\left(y_t|Y_{t-1}\right)$, which according to (5) is given by

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

Next, $f\left(s_t|Y_t\right)$ is approximated by re-weighting $\{s_t^{1,i}\}_{i=1}^N$ in accordance with (3) (the updating step): a particle $s_t^{1,i}$ with prior weight $\frac{1}{N}$ is assigned the posterior weight

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

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 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 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 (1987) consists simply of making $N' >> N$ blind proposals $\{s_t^{1,j}\}_{j=1}^{N'}$ as with the particle filter, and then obtaining the swarm $\left\{s_t^{0,i}\right\}_{i=1}^N$ 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 $f\left(y_t|Y_{t-1}\right)$ in (5), but with $f\left(s_t|Y_{t-1}\right)$ unknown, importance sampling achieves approximation via the introduction into the integral of an importance density $g(s_t|Y_t)$:

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

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

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

Pitt and Shephard referred to the introduction of $g(s_t|Y_t)$ as adaption. Full adaption is achieved when $g(s_t|Y_t)$ is constructed as being proportional to $f\left(y_t|s_t, Y_{t-1}\right) f\left(s_t|Y_{t-1}\right)$, rendering the ratios in (11) as constants. Pitt and Shephard viewed adaption as computationally infeasible, due to the requirement of computing $f\left(s_t^{0,i}|Y_{t-1}\right)$ 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 $f\left(y_t|s_t, Y_{t-1}\right)$ around $s_t = \mu_t^k = E\left(s_t|s_{t-1}^{0,k}, Y_{t-1}\right)$. 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 $\left\{s_{t-1}^{0,i}\right\}$ 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 $\left\{s_{t-1}^{0,i}\right\}$. Moreover, the samplers are suboptimal, since μ_t^k is incapable of fully capturing the characteristics of $f\left(y_t|s_t, Y_{t-1}\right)$. 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 (5). Section 4.2 introduces a class of piecewise-continuous samplers for dealing with pathological cases. Section 4.3 then discusses a key contribution of this paper: the computation of $f(s_t|Y_{t-1})$ in (5) at auxiliary values of s_t generated under period-t EIS optimization. Section 4.4 discusses two special cases that often characterize state-space representations: partial measurement of the state space; and degenerate transition densities. Pseudo-code is available at `www.pitt.edu/~dejong/wp.htm`.

4.1 EIS integration

Let $\varphi_t(s_t) = f(y_t|s_t, Y_{t-1})f(s_t|Y_{t-1})$ in (5), 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

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

The selection of K is 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. A (near) optimal value $\widehat{a}_t$ is

obtained as the solution to

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

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 $R << N$ 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

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

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 to locate the mode of $\varphi_t(s_t)$. If K belongs to the exponential family of distributions, there exists a parameterization a_t such that the auxiliary problems in (14) are linear. Convergence to a fixed point is typically achieved within five to ten iterations.

To guarantee fast fixed-point 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). Examples are standardized Normal draws when g is Gaussian, or uniform draws transformed into draws from g by the inverse c.d.f technique (e.g., see Devroye, 1986).

At convergence, the EIS filter approximation of $f(y_t|Y_{t-1})$ in (5) is given by

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

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.

4.2 A piecewise-continuous class of samplers

While kernels within the exponential family of distributions yield EIS regressions that are linear in a_t, there exist potential pathologies of the integrand in (5) that they cannot replicate efficiently (e.g., skewness, thick tails, and bimodality). Here we propose an approach that provides high flexibility along one or two pathological dimensions, and as illustrated in Example 4 below, can be combined with (conditional) Gaussian samplers along additional better-behaved dimensions. It entails the use of samplers that provide piecewise log-linear approximations to the integrand φ_t; their parameters are the grid points $a' = (a_0, ..., a_R)$, with $a_0 < a_1 < ... < a_R$ (the index t is suppressed for ease of notation). 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 least-squares optimization step (14). Instead we implement near equal probability division of the domain of integration.

We first describe $k(s; a)$ for a preassigned grid a, where the interval $[a_0, a_R]$ covers the support of $\varphi(s)$. Note that while $R + 1$ represents the number of grid-points here, and R the number of auxiliary draws used to construct $g(s_t; \widehat{a}_t^l)$ in (14), this does not represent an abuse of notation: for the piecewise-continuous sampler, use of $R + 1$ grid-points translates precisely into the use of R auxiliary draws.

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

$$\ln k_j(s; a) = \alpha_j + \beta_j s \qquad \forall s \in [a_{j-1}, a_j], \tag{16}$$

$$\beta_j = \frac{\ln \varphi(a_j) - \ln \varphi(a_{j-1})}{a_j - a_{j-1}}, \quad \alpha_j = \ln \varphi(a_j) - \beta_j a_j. \tag{17}$$

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

$$K_j(s;a) = \frac{\chi_j(s;a)}{\chi_n(a)}, \quad \forall s \in [a_{j-1}, a_j], \tag{18}$$

$$\chi_j(s;a) = \chi_{j-1}(a) + \frac{1}{\beta_j}\left[k_j(s;a) - k_j(a_{j-1};a)\right], \tag{19}$$

$$\chi_0(a) = 0, \quad \chi_j(a) = \chi_j(a_j;a). \tag{20}$$

Its inverse c.d.f. is given by

$$s = \frac{1}{\beta_j}\left\{\ln\left[k_j(a_{j-1};a) + \beta_j\left(u \cdot \chi_R(a) - \chi_{j-1}(a)\right)\right] - \alpha_j\right\}, \tag{21}$$

$$u \in]0,1[\quad \text{and} \quad \chi_{j-1}(a) < u \cdot \chi_R(a) < \chi_j(a). \tag{22}$$

The recursive construction of an equal-probability-division kernel $k(s;\widehat{a})$ is based upon the non-random equal division of $[\varepsilon, 1-\varepsilon]$ with $u_i = \varepsilon + (2-\varepsilon)\frac{i}{R}$ for $i = 1, ..., R-1$, with ε sufficiently small (typically $\varepsilon = 10^{-4}$) to avoid tail intervals of excessive length. 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

$$\widehat{a}_i^{l+1} = K^{-1}(u_i), \qquad i = 1, ..., R-1. \tag{23}$$

The algorithm iterates until (approximate) convergence.

The resulting approximation is highly adapted and computationally inexpensive. Given a sufficiently large number of division points, it will outperform lower-dimensional parametric classes of samplers. Piecewise-continuous 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 parametric families of distributions, and reserve the use of log-linear piecewise continuous approximations for those dimensions along which the integrand is 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. Here we discuss three operational alternatives for overcoming this hurdle (a fourth, involving non-parametric approximations, is also possible but omitted here). Below, S denotes the number of points used for each individual evaluation of $f(s_t|Y_{t-1})$.

Weighted-sum approximations

Combining (4) and (3), we can rewrite $f(s_t|Y_{t-1})$ as a ratio of integrals:

$$f(s_t|Y_{t-1}) = \frac{\int f(s_t|s_{t-1}, Y_{t-1}) f(y_{t-1}|s_{t-1}, Y_{t-2}) f(s_{t-1}|Y_{t-2}) ds_{t-1}}{\int f(y_{t-1}|s_{t-1}, Y_{t-2}) f(s_{t-1}|Y_{t-2}) ds_{t-1}}, \tag{24}$$

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

$$\widehat{f}_S(s_t|Y_{t-1}) = \frac{\sum_{i=1}^{S} f(s_t|s_{t-1}^{0,i}, Y_{t-1}) \cdot \omega(s_{t-1}^{0,i}; \widehat{a}_{t-1})}{\sum_{i=1}^{S} \omega(s_{t-1}^{0,i}; \widehat{a}_{t-1})}, \tag{25}$$

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

$$\omega(s_{t-1}; \widehat{a}_{t-1}) = \frac{f(y_{t-1}|s_{t-1}, Y_{t-2}) f(s_{t-1}|Y_{t-2})}{g(s_{t-1}|\widehat{a}_{t-1})}. \tag{26}$$

Obviously $g(s_{t-1}|\widehat{a}_{t-1})$ is not an EIS sampler for the numerator in (24). 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 $s_{t-1}|Y_{t-1}$. 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 $\omega(s_{t-1};\widehat{a}_{t-1})$ will be near constants over the range of integration. Replacing these weights by their arithmetic means $\overline{\omega}\,(\widehat{a}_{t-1})$ in (24) and (25), we obtain the following simplification:

$$f(s_t|Y_{t-1}) \simeq \int f(s_t|s_{t-1},Y_{t-1}) \cdot g(s_{t-1};\widehat{a}_{t-1})ds_{t-1}. \tag{27}$$

This substitution yields rapid implementation if additionally the integral in (27) 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

$$\widehat{f}_S(s_t|Y_{t-1}) \simeq \frac{1}{S}\sum_{i=1}^{S} f(s_t|s_{t-1}^{0,i},Y_{t-1}). \tag{28}$$

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 (4) requires the construction of a continuous approximation to $f(s_{t-1}|Y_{t-1})$. 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

Partial measurement

Partial measurement refers to cases (e.g., see Examples 2 and 4) in which s_t can be partitioned (possibly after transformation) into $s_t = (p_t, q_t)$, so that

$$f\left(y_t|s_t, Y_{t-1}\right) \equiv f\left(y_t|p_t, Y_{t-1}\right). \tag{29}$$

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

$$f\left(y_t|Y_{t-1}\right) = \int f\left(y_t|p_t, Y_{t-1}\right) f\left(p_t|Y_{t-1}\right) dp_t, \tag{30}$$

and the updating equation (3) factorizes into the product of the following two densities:

$$f\left(p_t|Y_t\right) = \frac{f\left(y_t|p_t, Y_{t-1}\right) f\left(p_t|Y_{t-1}\right)}{f\left(y_t|Y_{t-1}\right)}; \tag{31}$$

$$f\left(q_t|p_t, Y_t\right) = f\left(q_t|p_t, Y_{t-1}\right). \tag{32}$$

Stronger conditional independence assumptions are required in order to produce factorizations in (4). In particular, if p_t is independent of q_t given (p_{t-1}, Y_{t-1}), so that

$$f\left(p_t|s_{t-1}, Y_{t-1}\right) \equiv f\left(p_t|p_{t-1}, Y_{t-1}\right), \tag{33}$$

then

$$f\left(p_t|Y_{t-1}\right) = \int f\left(p_t|p_{t-1}, Y_{t-1}\right) f\left(p_{t-1}|Y_{t-1}\right) dp_{t-1}. \tag{34}$$

Note that under conditions (29) and (33), 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.

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):

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

The evaluation of $f(s_t|Y_{t-1})$ in (4) 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 (35). 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 (35) 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

$$q_{t-1} = \psi(p_t, q_t, p_{t-1}) = \psi(s_t, p_{t-1}) \tag{36}$$

and Jacobian

$$J(s_t, p_{t-1}) = \frac{\partial}{\partial q_t}\psi(s_t, p_{t-1}), \tag{37}$$

we can take the limit of the integral in (35) as ε tends to zero, producing

$$f(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|s_{t-1}, Y_{t-1}) f(p_{t-1}, q_{t-1}|Y_{t-1})|_{q_{t-1}=\psi(s_t,p_{t-1})} dp_{t-1}. \tag{38}$$

Note that (38) 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. We can also approximate the integral in (38) by replacing $f(s_{t-1}|Y_{t-1})$ by $\overline{\omega}(\widehat{a}_{t-1})\, g(s_{t-1}|\widehat{a}_{t-1})$:

$$\widehat{f}(s_t|Y_{t-1}) = \int J(s_t, p_{t-1})\, f(p_t|q_{t-1}, Y_{t-1})\, g(p_{t-1}, q_{t-1}|\widehat{a}_{t-1})\,|_{q_{t-1}=\psi(s_t,p_{t-1})} dp_{t-1}. \tag{39}$$

In this case, since $g(.|\widehat{a}_{t-1})$ is not a sampler for $p_{t-1}|s_t$, we must evaluate (39) either by quadrature or its own EIS sampler.

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

5 Examples

Here we present four examples that illustrate the relative performance of the particle, auxiliary, adapted, and EIS filters. The first two focus on likelihood evaluation; the last two on filtering. 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 (15)); R, the number of draws used to construct EIS samplers (e.g., see (14)); and S, the number of draws used to evaluate $f(s_t|Y_{t-1})$.

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 the applications we have considered).

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 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 setting $S << N$ can yield significant gains in overall efficiency. Indeed, we typically we set $S = 100$. 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). Let

$$s_{t+1} = \alpha + \beta \frac{s_t}{1 + s_t^2} + \upsilon_{t+1} \tag{40}$$

$$y_t = s_t + u_t, \tag{41}$$

where $\upsilon_{t+1} \sim N(0, \sigma_\upsilon^2)$ 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 1.042, and $\Pr(|u_t| > 2.010) = 0.05$. We consider four values for σ_v: $(1/3, 1, 3, 10)$. Thus the parameterizations we consider cover a wide range of scenarios, ranging from well-behaved ($\nu = 50, \sigma_v = 1/3$) to ill-behaved ($\nu = 2, \sigma_v = 10$).

We compare the relative numerical efficiency of five algorithms. The first three are the particle, auxiliary, and adapted filters. These are implemented using $N = 20,000$ and $N = 200,000$. The remaining algorithms are the Gaussian and piecewise-linear EIS filters. These are implemented using $N = 100$ and $N = 1,000$. Evaluation of $f(s_t|Y_{t-1})$ is based on the weighted-sum approximation introduced in Section 4.3.1 – see (25).

Results obtained using artificial data sets of size $T = 100$ are presented in 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.

Table 1 reports relationships between MC standard deviations and computing time, along with MC means of likelihood evaluations. 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 particle filter; for ratios less than one, the 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.

As expected, RTEs of the Gaussian EIS filter are also poor given $\nu = 2$, especially when σ_υ is large. For $\nu = 50$, the Gaussian EIS filter performs well, with impressive RTEs for large values of σ_υ (reaching 284 for $\sigma_\upsilon = 10$).

The piecewise-linear EIS filter outperforms the particle filter in all cases, with the payoff to adoption increasing with σ_υ. For $\nu = 2$, its RTE ranges from 1.6 to 2,001 as σ_υ increases from 1/3 to 10; for $\nu = 50$, its RTE ranges from 2.3 to 1,401. An RTE of 1,400 implies that the particle filter requires approximately 1 hour and 15 minutes (the time required to process approximately 38.4 million particles) to match the numerical accuracy of the piecewise-linear filter with $N = 1,000$ (which requires 3.18 seconds). These results reflect the payoffs associated with the flexibility, in addition to the global nature, of approximations provided by the piecewise-linear filter.

In sum, the particle, auxiliary, and adapted filters perform well under well-behaved scenarios. In these cases, their relative numerical inaccuracy is often 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. The EIS filters provide insurance against these problems and exhibit superior RTEs in all but the most well-behaved cases. But while numerical efficiency is an important feature of likelihood approximation procedures, it is not the only important feature. In pursuing ML estimates, continuity with respect to 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 over time of approximation errors. 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 by estimating 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

$$\begin{aligned} q_t &= a_t k_t^{\alpha} && (42)\\ q_t &= c_t + i_t && (43)\\ k_{t+1} &= i_t + (1-\delta)k_t && (44)\\ \ln(a_{t+1}) &= \rho \ln(a_t) + \upsilon_{t+1}. && (45) \end{aligned}$$

Regarding parameters, α is capital's share of output, δ is capital depreciation, ρ determines the persistence of innovations to TFP, and the innovations $\upsilon_t \sim N(0, \sigma^2)$. 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

$$\begin{aligned} q_t &= a_t k_t^{\alpha} + u_{q_t}, && (46)\\ i_t &= a_t k_t^{\alpha} - c(a_t, k_t) + u_{i_t} && (47)\\ &= \alpha\beta a_t k_t^{\alpha} + u_{i_t}, \end{aligned}$$

and the state-transition equations are (45) and

$$\begin{aligned} k_{t+1} &= a_t k_t^{\alpha} - c(a_t, k_t) && (48)\\ &= \alpha\beta a_t k_t^{\alpha}. \end{aligned}$$

Examination of (45) to (48) suggests reparameterizing the state variables as $z_t = \ln(a_t)$ and $l_t = e^{z_t} k_t^{\alpha}$, where l_t denotes (unobserved) output, and $s_t = [l_t \quad z_t]'$ denotes the state vector. The transition process (45) then takes the form of a Gaussian AR(1) in z_t, and the

identity (48) can be rewritten as

$$l_t = e^{z_t} \left(\alpha\beta l_{t-1}\right)^{\alpha}. \tag{49}$$

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

$$y_t|s_t, Y_{t-1} \equiv y_t|l_t, Y_{t-1} \sim N_2\left(\begin{pmatrix} 1 \\ \alpha\beta \end{pmatrix} l_t, \begin{bmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_i^2 \end{bmatrix}\right). \tag{50}$$

Second, (49) represents a degenerate Dirac transition, with inverse

$$l_{t-1} = \psi(s_t) = \frac{1}{\alpha\beta}\left(l_t e^{-z_t}\right)^{\frac{1}{\alpha}} \tag{51}$$

and Jacobian

$$J(s_t) = \frac{\partial\psi(s_t)}{\partial l_t} = \frac{1}{\alpha^2\beta}\left(l_t^{1-\alpha} e^{-z_t}\right)^{\frac{1}{\alpha}}. \tag{52}$$

In view of (50), the likelihood integral simplifies into a univariate integral in l_t whose evaluation requires only an EIS sampler for $l_t|Y_t$. Nevertheless, in period $t+1$ we still need to approximate $f\left(z_t|l_t, Y_t\right)$ in order to compute $\widehat{f}(y_t|Y_{t-1})$. To capture the dependence between z_t and l_t given Y_t, it proves convenient to construct a single bivariate EIS sampler for $z_t, l_t|Y_t$. Whence the likelihood integral

$$f(y_t|Y_{t-1}) = \int f(y_t|l_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t \tag{53}$$

is evaluated under a Gaussian EIS sampler $g(s_t|\hat{a}_t)$. Next, $f(s_t|Y_{t-1})$ is approximated according to (39), where we exploit the fact that the Jacobian $J(s_t)$ does not depend on z_{t-1}:

$$\widehat{f}(s_t|Y_{t-1}) = J(s_t) \int f(z_t|z_{t-1}) g(\psi(s_t), z_{t-1}|\hat{a}_{t-1}) dz_{t-1}. \tag{54}$$

Note that the integrand is quadratic in $z_{t-1}|s_t$, so 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.

To begin, we compute likelihood values at the true parameter values using 20,000 and 100,000 particles for the particle filter, and 100 and 1,000 auxiliary draws for the EIS filter (with R held fixed at 100). We do so for 100 MC replications, applied to a single data set of sample size $T = 100$. Results are reported in Table 3.

RTEs computed using as a numeraire the particle filter with $N = 20,000$ are 55.44 (for $N = 100$ under the EIS filter) and 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 $g\left(s_t|\widehat{a}_t\right)$ 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 (54) to approximate $f(s_t|Y_{t-1})$.

Next, we apply both the particle and EIS filters 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

(47), the stepwise MLE of β given α is given by $\widehat{\beta} = \bar{i}/\alpha\bar{l}$, 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 $R = 100$. Note from Table 3 that 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 much more accurate under the EIS filter (which has an RTE of approximately 55 given these settings for N and R).

Figure 1 plots estimated log-likelihoods for a representative data set along the α dimension for $T = 100$ for both the particle and EIS filters; all other parameters are set at their ML values. Note that the surface associated with the particle filter is particularly rough relative to that associated with EIS.

We employ the Nelder-Meade simplex optimization routine for all MLE computations. 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. The first are *statistical* means and standard deviations, obtained from 30 different samples $\{y_t\}_{t=1}^T$ under a single set of auxiliary draws $\{u_i\}_{i=1}^N$. These 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 particle filter are unreliable and often fail to be positive definite. The second are *numerical* means and standard deviations, obtained under 30 different sets of CRNs for a fixed sample $\{y_t\}_{t=1}^T$. These constitute our most accurate MC estimates of the MLEs and accordingly, the numerical standard deviations we report are those for the means.

Results are given in Table 4. Highlights are as follows. (1) Log-likelihood functions are tightly peaked, as statistical standard deviations attest. (2) For $T = 40$, MLEs of α are upward biased (by about 4 standard deviations), thus we also report root mean-

squared errors. (3) Under the EIS filter, there is close agreement between finite sample (MC) and asymptotic (Hessian) statistical standard deviations, especially as T increases. This highlights the numerical accuracy and reliability of EIS filter computations (including Hessians). (4) As T increases, numerical standard deviations (which are $\sqrt{30}$ larger than those reported for the mean MLEs) approach corresponding statistical standard deviations. This does not create a problem for the EIS filter (which employs CRNs), but contaminates the computation of statistical standard deviations under the particle filter. For this example, N would need to be increased dramatically in order for the particle filter to provide reliable estimates of statistical standard deviations.

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

5.3 Example 3: Stochastic Volatility

The stochastic volatility (SV) model is given by

$$y_t = u_t \beta \exp(s_t/2) \tag{55}$$

$$s_{t+1} = \phi s_t + \upsilon_t, \tag{56}$$

where u_t and υ_t are independent Gaussian random variables with variances 1 and σ^2; β represents modal volatility; and ϕ and σ^2 determine the persistence and variance of volatility shocks. This model was introduced by Taylor (1982, 1986) in attempts to account for the time-varying and persistent volatility exhibited by financial returns data, in addition to fat-tailed behavior. Many alternative procedures have been proposed to estimate this model efficiently, and to infer the behavior of (scaled) volatility (e.g., see Jacquier, Polson and Rossi, 1994; Ghysels et al., 1996; Pitt and Shephard, 1999; Kim, Shephard and Chib, 1998; and Liesenfeld and Richard, 2003). Thus it provides a natural testing ground for us as well.

Filtered values for volatility are obtained by replacing $h(s_t)$ by $\exp(s_t/2)$ in (7). These are obtained using Gaussian EIS samplers. Due to ease of implementation in the present context, we construct separate samplers for the numerator and denominator of (7).

Consider the approximation of the denominator: the period-t likelihood $f(y_t|Y_{t-1})$. Let $g_t(s_t; a_{d,t})$ denote the EIS sampling density used to approximate the integrand $\varphi_t(s_t) = f(y_t|s_t, Y_{t-1})f(s_t|Y_{t-1})$. The associated log kernel is parameterized as

$$-2\ln k(s_t; a_{d,t}) = \alpha_{d,t}s_t^2 - 2\beta_{d,t}s_t, \tag{57}$$

where $a_{d,t} = \big(\alpha_{d,t}, \beta_{d,t}\big)$. The corresponding Gaussian sampler has mean $\mu_{d,t} = \beta_{d,t}/\alpha_{d,t}$ and variance $\sigma_{d,t}^2 = \alpha_{d,t}^{-1}$. Thus the auxiliary regression for the computation of $\hat{a}_{d,t}$ consists of a bivariate OLS regression of simulated values of $\ln\varphi_t(s_t)$ on simulated values of s_t^2, s_t and a constant.

We use a constant-weight approximation to approximate the prediction density $f(s_t|Y_{t-1})$; the approximation obtains as

$$f(s_t|Y_{t-1}) \approx \int f(s_t|s_{t-1}, Y_{t-1})g(s_{t-1}; \hat{a}_{d,t-1})ds_{t-1}. \tag{58}$$

Since $f(s_t|s_{t-1}, Y_{t-1})$ is a conditional normal density for $s_t|s_{t-1}, Y_{t-1}$, and $g(s_{t-1}; \hat{a}_{d,t-1})$ is a normal density for s_{t-1}, the integral in (58) has an analytical solution given by a Gaussian density for s_t with mean $\phi\hat{\mu}_{d,t-1}$ and variance $\phi^2\hat{\sigma}_{d,t-1}^2 + \sigma_v^2$. Assuming $f(s_1|Y_0) = f(s_1)$, the initial values are $\hat{\mu}_{d,0} = 0$ and $\hat{\sigma}_{d,0}^2 = \sigma_v^2/(1-\phi^2)$.

Given an EIS sampler for the denominator, the Gaussian EIS kernel for the numerator in (7), denoted by $k(s_t; \hat{a}_{n,t})$ and designed to approximate $h(s_t)\cdot\varphi_t(s_t)$, is obtained analytically, since in the present context $\ln h(s_t)$ is a linear function in s_t. In particular, optimal values for the mean and variance of the sampler for the numerator $(\hat{\mu}_{n,t}, \hat{\sigma}_{n,t}^2)$ are given by $\hat{\sigma}_{n,t}^2 = \hat{\sigma}_{d,t}^2$ and $\hat{\mu}_{n,t} = \hat{\sigma}_{n,t}^2(\hat{\mu}_{d,t}/\hat{\sigma}_{d,t}^2 + 1/2)$. Hence the construction of the optimal sampler for the numerator is obtained without incurring additional computing costs.

Based on these EIS samplers, a consistent MC estimate of the filtered values obtains as

$$\widehat{E}\left(\exp(s_t/2)|Y_t\right) = \frac{\sum_{i=1}^{N} w_{n,t}\left(\widetilde{s}_{n,t}^{0,i}\right)}{\sum_{i=1}^{N} w_{d,t}\left(\widetilde{s}_{d,t}^{0,i}\right)}, \tag{59}$$

where $w_{n,t}(s_t) = \exp(s_t/2) \cdot \varphi_t(s_t)/g(s_t; \hat{a}_{n,t})$ and $w_{d,t}(s_t) = \varphi_t(s_t)/g(s_t; \hat{a}_{d,t})$. The sets $\{\widetilde{s}_{n,t}^{0,i}\}$ and $\{\widetilde{s}_{d,t}^{0,i}\}$ denote swarms of iid draws from $g(s_t; \hat{a}_{n,t})$ and $g(s_t; \hat{a}_{d,t})$, generated under a single set of CRNs.

We demonstrate the performance of the EIS filter for the SV model in an application to sets of artificial data simulated from the model. Performance is characterized relative to that of the particle filter, using the exact experimental design used by Pitt and Shephard (1999) to characterize the performance of their auxiliary and adapted filters.

The model is parameterized as $\phi = 0.9702$, $\sigma_\upsilon = 0.178$, and $\beta = 0.5992$; the sample size is $T = 50$. We draw $R = 40$ different data sets $\{Y_T^i\}_{i=1}^{40}$, all based on one simulated trajectory of the latent variable $\{s_t\}_{t=1}^{50}$. In the measurement-error series, we artificially insert a single outlier: $u_{21} = 2.5$. For each data set we produce 100 i.i.d. MC estimates of the filtered values for volatility using EIS-1K and PF-20K, where the numbers following the acronyms indicate the number of particles (computing times are similar given these settings). For each procedure, the 100 MC estimates are obtained using 100 different CRNs. Comparing MC estimates generated by these procedures with "true" values of the filtered means yields Mean Squared Error (MSE) comparisons identical to those used by Pitt and Shephard (1999).

Let $\widetilde{\ell}_t^i$, $(i : 1 \rightarrow 40,\ t : 1 \rightarrow 50)$ denote "true" filtered means for volatility. These must be computed with high numerical accuracy in order to validate the MSE comparisons that follow. Exploiting the relatively high numerical accuracy of EIS (highlighted below), we estimate "true" filtered means as the arithmetic means of 100 i.i.d. EIS-10K estimates. Corresponding standard deviations are several orders of magnitude lower than those of the estimates we propose to compare. In order to reach similar precision using the particle

filter, we must use the arithmetic means of 100 i.i.d. PF-4 million estimates. We ran this experiment to verify that "true" values produced by both EIS and PF estimators are numerically identical. The latter number far exceeds the PF-120K value employed by Pitt and Shephard (1999), but turns out to be needed in order to eliminate significant and persistent biases characterizing PF estimates of filtered means (illustrated below).

MSE comparisons are constructed as follows. Let $\bar{\ell}_{t,k}^{i,j}$ denote the MC estimate of the filtered mean, for data set i, for replication j, at time t, for procedure $k = \{\text{EIS-1K}, \text{PF-20K}\}$. The log mean squared error (LMSE) for procedure k, at time t is obtained as

$$\text{LMSE}_{t,k} = \ln \left\{ \frac{1}{40} \sum_{i=1}^{40} \left[\frac{1}{100} \sum_{j=1}^{100} \left(\bar{\ell}_{t,k}^{i,j} - \tilde{\ell}_t^i \right)^2 \right] \right\}. \tag{60}$$

Figure 2 (bottom panel) depicts LMSEs for the five procedures against time. As expected, the move from estimates obtained using the particle filter to those obtained using the EIS filter leads to a large reduction in LMSEs: the average difference between PF-20K and EIS-1K is 1.9 on log scale. These differences are far larger than those reported by Pitt and Shephard (1999, Figure 4) in their comparison of the particle filter with their adapted particle filter, both implemented using the same number of particles. In particular, their adapted filter yields a maximal reduction of 1.0 (0.8) on log scale relative to the particle filter using 2K (4K) particles. Note also that the EIS filter is considerably less susceptible to the injected outlier in the measurement error process than is the particle filter.

To identify the source of the large differences in LMSEs, we computed separately MC variances and squared biases for EIS-1K and PF-20K. Logged variances and logged MSE/variance ratios are plotted for both procedures in Figure 2 (top panels). The logged MSE/variance ratio can be interpreted as a "bias multiplier", indicating the extent to which biases amplify differences in logged variances in yielding corresponding LMSEs. Figure 2 indicates that in nearly all periods the logged MC variance for the EIS filter is substantially smaller than for the particle filter. Further, EIS-1K exhibits logged MSE/variance ratios close to zero for all

time periods, indicating near-complete absence of bias. In contrast, for PF-20K this ratio is significantly larger than zero in approximately half of the time periods. Note in particular the comparably large value of the ratio for PF-20K in the time period infected by the outlier ($t = 21$). These results indicate that, in addition to MC variance, bias represents a significant component of the large differences in LMSEs generated by the adoption of EIS.

5.4 Example 4: Bearings-Only Tracking

The bearings-only tracking problem has received much attention in the literature on particle filters, and raises challenging numerical issues. References include Gordon et al. (1993), Carpenter et al. (1999), and Pitt and Shephard (1999); we consider here the scenario described by Gordon et al. (1993).

A ship moves in the (x, z) plane with speed following a bivariate random walk process. Let $\lambda_t = \left(x_t, \quad z_t, \quad \dot{x}_t, \quad \dot{z}_t\right)'$ denote the quadrivariate latent state variable (shortly we shall re-parameterize, and revert to the use of s_t to denote the state). The discrete version of the model we consider first is characterized by the transition

$$\lambda_{t+1} = \begin{pmatrix} I_2 & I_2 \\ 0 & I_2 \end{pmatrix} \lambda_t + \sigma \begin{pmatrix} \frac{1}{2} I_2 \\ I_2 \end{pmatrix} u_t, \tag{61}$$

with $u_t \sim iidN\left(0, I_2\right)$. The initial state vector is distributed as

$$\lambda_1 \sim N\left(\mu_1, \Delta_1\right), \tag{62}$$

with (μ_1, Δ_1) known and Δ_1 diagonal.

An observer located at the origin of the (x, z) plane measures with error the angle $\theta_t = \arctan\left(z_t / x_t\right)$. The measured angle y_t is assumed to be wrapped Cauchy with density

$$f\left(y_t | \theta_t\right) = \frac{1}{2\pi} \frac{1 - r^2}{1 + r^2 - 2r\cos(y_t - \theta_t)}, \tag{63}$$

with $0 \leq (y_t, \theta_t) \leq 2\pi$ and $0 \leq r \leq 1$. Accordingly, we shall introduce a (partial) reparameterization in polar coordinates. Let

$$\begin{aligned} \lambda_t &= (\alpha_t, \quad \beta_t)' \qquad \alpha_t = (x_t, \quad z_t)' \qquad \beta_t = \left(\dot{x}_t, \quad \dot{z}_t\right)', \\ \alpha_t &= \rho_t e\left(\theta_t\right), \qquad \theta_t \in [0, 2\pi], \end{aligned}$$

with $e\left(\theta_t\right) = (\cos\theta_t, \quad \sin\theta_t)'$ and $\rho_t = \left(x_t^2 + z_t^2\right)^{1/2} \geq 0$. The following notation will be used for the transformed state vector:

$$s_t = h(\lambda_t) = (\theta_t, \quad \rho_t, \quad \beta_t') = (\theta_t, \quad \delta_t'). \tag{64}$$

Note that (61) is based on a discretization over a time interval that coincides with the interval between successive measurements. It implies that the transition from λ_t to λ_{t+1} is degenerate. We reinterpret this transition as the combination of a proper bivariate transition

$$\alpha_{t+1}|\lambda_t \sim N\left(A\lambda_t, \Omega\right) \tag{65}$$

and a Dirac transition

$$\beta_{t+1} \equiv \phi\left(\alpha_{t+1}, \lambda_t\right) = 2\left(\alpha_{t+1} - \alpha_t\right) - \beta_t, \tag{66}$$

with $A = (I_2, \quad I_2)$ and $\Omega = \frac{1}{4}\sigma^2 I_2$. Below we shall consider an alternative version of the model discretized on a finer grid than that defined by observation times. This produces a non-degenerate transition, and allows for observations that are spaced unequally over time.

The degenerate version just described is numerically challenging on three counts. First, measurement is non-informative on three out of the four state components. Second, under parameter values typically used in the literature, the density of $\theta_t|Y_t$ is much tighter (though with fat tails) than that of $\theta_t|Y_{t-1}$. This situation yields "sample impoverishment", and thus

(very) high numerical inefficiency for the particle filter. Finally, the degenerate transition creates additional numerical problems since it implies a zero-measure support in $\mathbb{R}^4$ for the density $f(\lambda_{t+1}|\lambda_t)$.

Despite these challenges, we can implement an EIS version of the particle filter that can accommodate these pathologies. While conceptually simple, the algebra of our implementation is somewhat tedious. The text presents the broad lines of our implementation; full technical details are regrouped in the Appendix.

EIS computation of $f(y_t|Y_{t-1})$

We momentarily take as given that $f(s_t|Y_{t-1})$ can be computed for any s_t (as described below). The period-t likelihood function is then given by

$$\ell_t \equiv f(y_t|Y_{t-1}) = \int f(y_t|\theta_t) f(s_t|Y_{t-1}) ds_t. \tag{67}$$

Note that while $f(s_t|s_{t-1})$ is degenerate, $f(s_t|Y_{t-1})$ is not. In the absence of observations, $f(s_t)$ would be quadrivariate Normal. The observation y_t only measures θ_t, thus we shall implement a (sequential) EIS sampler $g(s_t; a_t)$ as the product of a trivariate Gaussian density for $\delta_t|\theta_t$ and a univariate piecewise loglinear density for θ_t. For ease of notation, the auxiliary EIS parameter a_t is deleted from all subsequent equations.

The conditional EIS sampler $g(\delta_t; \theta_t)$ is constructed as follows (accounting for the transformation from λ_t to s_t): (i) We draw a swarm $\left\{\widetilde{\lambda}_t^{1,i}\right\}_{i=1}^N$. Specifically, the period-$(t-1)$ EIS swarm $\left\{\widetilde{s}_{t-1}^{0,i}\right\}_{i=1}^N$ is transformed into a swarm $\left\{\widetilde{\lambda}_{t-1}^{0,i}\right\}_{i=1}^N$ by means of the inverse transformation $\lambda_t = h^{-1}(s_t)$. Then $\widetilde{\lambda}_t^{1,i}$ is drawn from the (degenerate) transition density $f\left(\lambda_t|\widetilde{\lambda}_{t-1}^{0,i}\right)$ associated with (61).

(ii) We construct an auxiliary quadrivariate EIS Gaussian kernel $k_{\lambda,t}(\lambda_t)$ approximating $f(\lambda_t|Y_{t-1})$. To do so, we use the swarm $\left\{\widetilde{\lambda}_t^{1,i}\right\}_{i=1}^N$ to construct an auxiliary OLS regression of $\left\{\ln f\left(\widetilde{\lambda}_t^{1,i}|Y_{t-1}\right)\right\}_{i=1}^N$ on $\left\{\widetilde{\lambda}_t^{1,i}\right\}_{i=1}^N$ and the lower triangle of $\left\{\left(\widetilde{\lambda}_t^{1,i}\right)\left(\widetilde{\lambda}_t^{1,i}\right)'\right\}_{i=1}^N$, for a total of 14 regressors plus one intercept. Let μ_t denote the unconditional mean of this quadrivariate

kernel and P_t its precision matrix. The kernel is then written as

$$k_{\lambda,t}(\lambda_t) = \exp\left\{-\frac{1}{2}\left(\lambda_t' P_t \lambda_t - 2\lambda_t' q_t\right)\right\}, \tag{68}$$

with $q_t = P_t^{-1}\mu_t$.

(iii) We introduce the transformation from λ_t to $s_t = h(\lambda_t)$, with Jacobian $\rho_t > 0$. Let

$$k_{s,t}(s_t) = \rho_t k_{\lambda,t}\left(h_t^{-1}(s_t)\right). \tag{69}$$

The conditional EIS sampler for $\delta_t|\theta_t$ is then given by

$$g_t(\delta_t|\theta_t) = \frac{k_{s,t}(s_t)}{\chi_t(\theta_t)}, \tag{70}$$

with

$$\chi_t(\theta_t) = \int_{\Delta} k_{s,t}(s_t)\, d\beta_t d\rho_t, \tag{71}$$

where $\Delta = \mathbb{R}^2 \times \mathbb{R}_+$.

(iv) The likelihood integral in (67) is rewritten as

$$\ell_t = \int \left[f(y_t|\theta_t)\chi_t(\theta_t)\right] \frac{f(s_t|Y_{t-1})}{k_{s,t}(s_t)} g_t(\delta_t|\theta_t)\, d\delta_t d\theta_t. \tag{72}$$

The next EIS step consists of approximating the product $f(y_t|\theta_t)\chi_t(\theta_t)$ on $[0, 2\pi]$ by a piecewise loglinear EIS sampler $g_t(\theta_t)$. Equation (72) is rewritten as

$$\ell_t = \int w_{s,t}(s_t)\, g_t(\delta_t|\theta_t)\, g_t(\theta_t)\, ds_t, \tag{73}$$

with

$$w_{s,t}(s_t) = \left[\frac{f(y_t|\theta_t)\chi_t(\theta_t)}{g_t(\theta_t)}\right] \frac{f(s_t|Y_{t-1})}{k_{s,t}(s_t)}. \tag{74}$$

Its EIS-MC estimate obtains as

$$\widehat{\ell_t} = \frac{1}{N}\sum_{i=1}^{N} w_{s,t}\left(\widetilde{s}_t^{0,i}\right), \tag{75}$$

where $\left\{\widetilde{s}_t^{0,i}\right\}_{i=1}^{N}$ denotes a swarm of i.i.d.N. draws (under CRNs) from the EIS sampler $g\left(\delta_t|\theta_t\right) g_t\left(\theta_t\right)$.

In view of the structure of the problem (non-observability of 3 out of 4 Gaussian state variables, and flexibility of the piecewise loglinear sampler along the fourth), we anticipate close fit between the numerator and denominator of $w_{s,t}\left(s_t\right)$ as given in (74). Relatedly, we anticipate dramatic reduction in the MC sampling variance of filtered values relative to that of estimates obtained under the particle filter and commonly used extensions.

EIS computation of $f\left(\lambda_{t+1}|Y_t\right)$

Having just discussed EIS for period t, it is notationally more convenient to discuss the computation of $f\left(\lambda_{t+1}|Y_t\right)$ rather than that of $f\left(\lambda_t|Y_{t-1}\right)$. The reason for initially discussing $f\left(\lambda_{t+1}|Y_t\right)$ rather than $f\left(s_{t+1}|Y_t\right)$ is simply that Gaussian algebraic manipulations are more transparent under the λ parametrization. Moreover, $f\left(s_{t+1}|Y_t\right)$ obtains directly from $f\left(\lambda_{t+1}|Y_t\right)$ via the transformation $s_{t+1} = h(\lambda_{t+1})$ with Jacobian $\rho_t > 0$. Relatedly, the weights $w_{s,t}\left(s_t\right)$ in (74) can trivially be transformed into weights for λ_t. Let

$$w_{\lambda,t}\left(\lambda_t\right) = w_{s,t}\left(h^{-1}\left(s_t\right)\right) = \frac{f\left(\lambda_t|Y_{t-1}\right)}{k_{\lambda,t}\left(\lambda_t\right)}\left[\frac{f\left(y_t|\theta_t\right)\chi_t\left(\theta_t\right)}{g_t\left(\theta_t\right)}\right]_{\theta_t=\theta_t(\alpha_t)}, \tag{76}$$

with $\theta_t\left(\alpha_t\right) = \arctan\left(z_t/x_t\right)$. Whence the density $f\left(\lambda_t|Y_t\right)$, given by

$$f\left(\lambda_t|Y_t\right) = \frac{f\left(\lambda_t|Y_{t-1}\right) f\left(y_t|\theta_t\left(\alpha_t\right)\right)}{\ell_t}, \tag{77}$$

can be rewritten as

$$f\left(\lambda_t|Y_t\right) = \frac{w_{\lambda,t}\left(\lambda_t\right)}{\ell_t} k_{\lambda,t}\left(\lambda_t\right) \frac{g_t\left(\theta_t\right)}{\chi_t\left(\theta_t\right)}|_{\theta_t=\theta_t(\alpha_t)}. \tag{78}$$

Under a non degenerate transition from λ_t to λ_{t+1}, $f(\lambda_{t+1}|Y_t)$ obtains as

$$f(\lambda_{t+1}|Y_t) = \int_{\mathbb{R}^4} f(\lambda_t|Y_t) f(\lambda_{t+1}|\lambda_t) d\lambda_t. \tag{79}$$

In the present case, however, we have to properly account for the fact that the transition from λ_t to λ_{t+1} is degenerate. As discussed above, degeneracy is addressed by replacing β_t in (78) by the inverse of the Dirac transition in (66):

$$\phi^{-1}(\lambda_{t+1}, \alpha_t) = \beta_t = 2(\alpha_{t+1} - \alpha_t) - \beta_{t+1}, \tag{80}$$

and integrating only with respect to α_t. Furthermore, since $w_{\lambda,t}(\lambda_t)$ is expected to be near constant over the support of λ_t, we can safely rely upon a "constant weight" approximation whereby the ratio $w_{\lambda,t}(\lambda_t)/\ell_t$ in (78) is set equal to 1. Whence $f(\lambda_{t+1}|Y_t)$ can be accurately evaluated by the following bivariate integral:

$$f(\lambda_{t+1}|Y_t) = \int \frac{g_t(\theta_t)}{\chi_t(\theta_t)} k_{\lambda,t}(\lambda_t) f(\alpha_{t+1}|\lambda_t)|_{\theta_t=\theta_t(\alpha_t), \beta_t=\phi^{-1}(\lambda_{t+1},\alpha_t)} d\alpha_t. \tag{81}$$

Numerically efficient evaluation of this integral requires the following additional steps:

(i) Combine analytically $k_{\lambda,t}(\lambda_t)$ and $f(\alpha_{t+1}|\lambda_t)$ into a Gaussian kernel in $(\alpha_{t+1}, \lambda_t)$;

(ii) Introduce the transformation from α_t into (ρ_t, θ_t) with Jacobian $\rho_t > 0$;

(iii) Given $(\lambda_{t+1}, \theta_t)$, integrate analytically in $\rho_t > 0$;

(iv) Given λ_{t+1}, use $g_t(\theta_t)$ as a natural sampler and compute the integral using the draws of θ_t obtained in the previous round.

Note that the sequence of operations just described must be repeated for any value of λ_{t+1} for which $f(\lambda_{t+1}|Y_t)$ is to be evaluated for period-$(t+1)$ EIS evaluation of ℓ_{t+1}. However, as illustrated below, the numerical efficiency of the EIS procedures we have just described results in dramatic reductions in the number of MC draws required to reach a preassigned level of numerical accuracy, and thus in significant reductions in overall computing time

relative to the particle filter.

Filtered Values

Filtered values for $\{\lambda_t\}_{i=1}^N$ are defined as

$$E\left(\lambda_t|Y_t\right) = \int \lambda_t f\left(\lambda_t|Y_t\right) d\lambda_t. \tag{82}$$

Substituting (78) for $f\left(\lambda_t|Y_t\right)$ and introducing the transformation from λ_t to s_t produces the following operational expression for the filtered values of λ_t:

$$E\left(\lambda_t|Y_t\right) = \frac{1}{\ell_t} \int h(s_t) w_{s,t}\left(s_t\right) g_t(\delta_t|\theta_t) g_t(\theta_t) ds_t. \tag{83}$$

EIS estimates of these filtered values obtain as

$$\widehat{E\left(\lambda_t|Y_t\right)} = \frac{\sum_{i=1}^N h(\widetilde{s}_t^{0,i}) w_{s,t}\left(\widetilde{s}_t^{0,i}\right)}{\sum_{i=1}^N w_{s,t}\left(\widetilde{s}_t^{0,i}\right)}. \tag{84}$$

Filtered values of s_t are obtained by replacing $h(s_t)$ by s_t in (83) and (84). Note that, in contrast with the evaluation of $f\left(\lambda_{t+1}|Y_t\right)$, we do not implement here the "constant weight" approximation under which $w_{\lambda,t}\left(\lambda_t\right)/\ell_t$ is be set to 1 in (83) and filtered values simplify into the arithmetic mean of $\left\{h\left(\widetilde{s}_t^{0,i}\right)\right\}_{i=1}^N$. As discussed, e.g., by Geweke (1989), the reason for preferring the ratio form in (84) is that the use of CRNs in the evaluation of the numerator and denominator typically induces positive correlation between their respective MC estimates, thereby reducing further the MC variance of the ratio.

A non-degenerate version of the problem

The singularity of the transition in (61) is a (spurious) consequence of a model specification that assumes measurements at each division point of the grid used for discretization of the random walk for speed. We now consider the case in which a finer grid for discretization

is used relative to that used for measurement, while also allowing for measurements made at varying time intervals.

For ease of notation, we focus on two successive measurements separated by D discretization intervals. Equation (61) then must be transformed into a transition density for $\lambda_{t+D}|\lambda_t$ by implicit marginalization with respect to the state sequence $\{\lambda_{t+j}\}_{j=1}^{D-1}$. The random walk process for speed is given by

$$\beta_{t+1} = \beta_t + \varepsilon_{t+1}, \qquad \varepsilon_t \sim N\left(0, \sigma^2 I_2\right), \tag{85}$$

and position is discretized as

$$\alpha_{t+1} = \alpha_t + \frac{1}{2}\left(\beta_t + \beta_{t+1}\right). \tag{86}$$

It follows that

$$\beta_{t+D} = \beta_t + u_{t+D}, \tag{87}$$

$$\alpha_{t+D} = \alpha_t + D\beta_t + v_{t+D}, \tag{88}$$

with

$$u_{t+D} = \sum_{j=1}^{D} \varepsilon_{t+j}, \qquad v_{t+D} = \frac{1}{2}\sum_{j=1}^{D} \left[2\left(D-j\right)+1\right]\varepsilon_{t+j}. \tag{89}$$

The covariance matrix of (u_{t+D}, v_{t+D}) obtains by application of standard formulae for the sums and sums of squares of natural numbers - see e.g. Gradshteyn and Ryzhik (1979, 0.122, 1 and 2). It follows that the transition density from λ_t to λ_{t+D} is given by

$$\lambda_{t+D}|\lambda_t \sim N\left(A_D\lambda_t, \sigma^2 V_D\right), \tag{90}$$

with

$$A_D = \begin{pmatrix} I_2 & DI_2 \\ 0 & I_2 \end{pmatrix}, \qquad V_D = D \cdot \begin{pmatrix} \frac{4D^2-1}{12} I_2 & \frac{D}{2} I_2 \\ \frac{D}{2} I_2 & I_2 \end{pmatrix}. \tag{91}$$

The case $D = 1$ obviously coincides with the degenerate transition in (61). The generalization from (61) to (90) does not affect EIS evaluation of the likelihood function. However, the evaluation of $f(\lambda_{t+D}|Y_t)$ now requires four-dimensional integration, and is given by

$$f(\lambda_{t+D}|Y_t) = \int \frac{g_t(\theta_t)}{\chi_t(\theta_t)} k_{\lambda,t}(\lambda_t) f(\lambda_{t+D}|\lambda_t) d\lambda_t. \tag{92}$$

The numerical evaluation of equation (92) parallels that of equation (81) with the additional (analytical) integration with respect to β_t.

Application

We demonstrate our methodology in an application designed essentially along the lines of that constructed by Gordon et al. (1993), and modified by Pitt and Shephard (1999). For the singular and non-singular cases, σ in (61) and (90) is set to 0.001; and r in (63) is set to $1 - (0.005)^2$. The initial latent vector λ_1 is normally distributed with mean vector $(-0.05, 0.2, 0.001, -0.055)$ and diagonal covariance matrix with standard deviations

$(0.05, 0.03, 0.0005, 0.001)$. In the non-singular case, the number D_t of discretization intervals between measurements t and $t+D$ is drawn from a multinomial $\{2, 3, ..., 11\}$ with equal probabilities $p_i = 0.1$, $i = 2, ..., 11$. (Actually, the non-singular case need not be restricted to integer values of D_t, and we have verified that solutions for the non-singular case converge to those for the singular case as D_t tends towards 1.)

We set $T = 10$, and draw two sets of latent vectors $\{\lambda_t^s\}_{t=1}^{10}$, one for the singular case ($s = 1$) and one for the non-singular case ($s = 2$). Both sets are linear transformations of a single set of $N(0, 1)$ draws.

As in Pitt and Shephard (1999), we draw $R = 40$ different data sets $\left\{Y_T^{s,i}\right\}_{i=1}^{40}$ based on the latent vectors $\{\lambda_t^s\}_{t=1}^{10}$ for $s = 1, 2$. For each data set, we produce 100 i.i.d. estimates of

the filtered means (differing by the seeds initializing the MC draws) using EIS-1K and PF-40K (computing times associated with these procedures are similar). As in the SV example, comparing estimates generated by these procedures with "true" filtered means for the latent variables yields LMSE comparisons analogous to those employed by Pitt and Shephard (1999) to demonstrate the gains in precision and efficiency yielded by their extensions of the particle filter. (Details regarding the construction of LMSEs in this case correspond precisely with those described for the SV application.)

Graphs of MC variances and squared biases are presented in Figure 3, and LMSEs in Figure 4, both for the singular case (similar results were obtained for the non-singular case, and thus are not reported). Note the large reductions in LMSEs yielded by the move from PF to EIS estimates: differences average between 4 and 6 on the log scale. These differences are once again much larger than those reported by Pitt and Shephard (1999): their auxiliary particle filter yielded reductions averaging between 0.5 and 1 relative to the particle filter.

Regarding the source of the large differences in LMSEs we obtain, differences in logged variances are typically of the order of 2 to 2.5 in favor of EIS (corresponding roughly to a 10-fold reduction in variance), except for $t = 1$. Logged bias ratios are virtually all close to zero for EIS filter, while they typically lie between 1 and 4 (and as high as 10 for $t = 1$) for the particle filter. Thus biases remain significant for the particle filter even using 40K draws, and are the dominant component of the large differences in LMSEs generated by the adoption of EIS. This is a manifestation of the "sample impoverishment" problem that results from the very tight distribution of $\lambda_t|Y_t$ relative to that of $\lambda_t|Y_{t-1}$ along the θ_t dimension.

6 Conclusion

We have proposed an efficient means of facilitating likelihood evaluation and filtering 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.

7 Appendix

7.1 Bearings-Only Tracking, Singular Case

Derivation of $\chi_t(\theta_t)$

For ease of notation we suppress t subscripts. The kernel $g_s(s)$ defined in (69) depends upon the quadratic form

$$\gamma(s) = \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix}' P \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix} - 2 \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix} q. \tag{93}$$

We partition P and q conformably with $(\rho e_\theta' \quad \beta')$ into

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \qquad q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}. \tag{94}$$

Standard Gaussian algebra operations (square completion in β and ρ successively) produce the following expressions for $\gamma(s)$:

$$\gamma(s) = (\beta - b_\theta)' P_{22} (\beta - b_\theta) + a_\theta (\rho - r_\theta)^2 - s_\theta^2, \tag{95}$$

$$b_\theta = P_{22}^{-1}(q_2 - \rho P_{21} e_\theta), \qquad a_\theta = e_\theta' P_{11.2} e_\theta, \tag{96}$$

$$P_{11.2} = P_{11} - P_{12} P_{22}^{-1} P_{21}, \tag{97}$$

$$r_\theta = \frac{1}{a_\theta}\left(q_1 - P_{12} P_{22}^{-1} q_2\right)' e_\theta, \qquad s_\theta^2 = a_\theta r_\theta^2 + q_2' P_{22}^{-1} q_2. \tag{98}$$

It follows that $\chi(\theta)$, as defined in (71), is given by

$$\chi(\theta) = 2\pi|P_{22}|^{-1} d_\theta \exp\left(\frac{1}{2}s_\theta^2\right), \tag{99}$$

$$d_\theta = \int_0^\infty \rho \exp\left(-\frac{1}{2}a_\theta(\rho - r_\theta)^2\right) d\rho. \tag{100}$$

Introducing the transformation of variables

$$\phi = \sqrt{a_\theta}(\rho - r_\theta), \tag{101}$$

d_θ can be written as

$$d_\theta = \frac{1}{a_\theta}\int_{-c_\theta}^\infty (\phi + c_\theta)\exp\left(-\frac{1}{2}\phi^2\right) d\phi \tag{102}$$

$$= \frac{1}{a_\theta}\left[\exp\left(-\frac{1}{2}c_\theta^2\right) + c_\theta\sqrt{\frac{\pi}{2}}\left(1 + \mathrm{erf}(\frac{c_\theta}{\sqrt{2}})\right)\right], \tag{103}$$

with $c_\theta = r_\theta\sqrt{a_\theta} > 0$, and erf() denoting the error function

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-\phi^2) d\phi. \tag{104}$$

(The properties of erf() are discussed, e.g., in Abramowitz and Segun, 1968, Ch. 7.) In deriving (103), we have exploited the fact that $r_\theta > 0$.

CRN-EIS draws of (β, ρ, θ)

An EIS draw of (β, ρ, θ) obtains from a CRN draw (u_1, u_2, u_3, u_4), where (u_1, u_2) denotes two $U(0,1)$ draws and (u_3, u_4) two i.i.d.N(0,1) draws, through the following sequence of transformations: (i) θ obtains from u_1 by inversion of the cdf associated with the piecewise loglinear EIS sampler $m(\theta)$. (ii) $\rho|\theta$ obtains from u_2 by inversion of the cdf associated with

$$m(\rho|\theta) = \frac{1}{d_\theta}\rho\exp\left(-\frac{1}{2}a_\theta(\rho - r_\theta)^2\right), \qquad \rho > 0. \tag{105}$$

Details of this transformation are provided below. (iii) $\beta|\rho,\theta$ obtains from the transformation

$$\beta = b_\theta + L \begin{pmatrix} u_3 \\ \\ u_4 \end{pmatrix}, \tag{106}$$

where L denotes the Cholesky decomposition of P_{22}^{-1}.

Regarding step (ii), $\rho|\theta$ obtains from the transformation of (101), rewritten as

$$\rho = \frac{1}{\sqrt{a_\theta}} (\phi + c_\theta), \tag{107}$$

where the density of $\phi|\theta$ is given by

$$f_\phi (\phi|\theta) = \frac{1}{d_\theta a_\theta} (\phi + c_\theta) \exp\left(-\frac{1}{2}\phi^2\right), \qquad \phi > -c_\theta, \tag{108}$$

with cdf

$$\begin{array}{c} F_\phi (\phi|\theta) = \frac{1}{d_\theta a_\theta}\{\left[\exp\left(-\frac{1}{2}c_\theta^2\right) - \exp\left(-\frac{1}{2}\phi^2\right)\right] \\ +c_\theta\sqrt{\frac{\pi}{2}}\left[\operatorname{erf}\left(\frac{\phi}{\sqrt{2}}\right) + \operatorname{erf}\left(\frac{c_\theta}{\sqrt{2}}\right)\right]\} \end{array}, \tag{109}$$

accounting for the fact that $\operatorname{erf}(-z) = -\operatorname{erf}(z)$. For the application described in Section 5.4, c_θ turns out to be significantly larger than zero, so that ϕ is nearly $N(0,1)$. Thus for the inversion of the CRN $u_2 \sim U(0,1)$, we take as a starting value the corresponding (inverse) Gaussian draw $\phi^{(0)} \sim N(0,1)$ and iterate once or twice by Newton

$$\phi^{(k+1)} = \phi^{(k)} - \frac{F\left(\phi^{(k)}|\theta\right) - u_2}{F\left(\phi^{(k)}|\theta\right)}. \tag{110}$$

Derivation of $f(\lambda_{t+1}|Y_t)$

We again suppress t subscripts for ease of notation; accordingly, the index $t+1$ is replaced

by +1. The product $g_\lambda(\lambda) f(\alpha_{+1}|\lambda)$ in (81) depends on the quadratic form

$$\delta(\alpha_{+1}, \lambda) = (\lambda' P \lambda - 2\lambda' q) + (\alpha_{+1} - A\lambda)' \Omega^{-1} (\alpha_{+1} - A\lambda). \tag{111}$$

It is transformed into a quadratic form in (α, λ_{+1}) via the inverse Dirac transformation (80)

$$\delta_1(\alpha, \lambda_{+1}) = \delta(\alpha_{+1}, \lambda)|_{\beta=\psi(\lambda_{+1},\alpha)}. \tag{112}$$

This implies the following two transformations:

$$\lambda|_{\beta=\psi(\lambda_{+1},\alpha)} = C \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}, \qquad (\alpha_{+1} - A\lambda)|_{\beta=\psi(\lambda_{+1},\alpha)} = D \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}, \tag{113}$$

$$\tag{114}$$

with C and D respectively being 4×6 and 2×6 matrices partitioned in 2×2 blocks:

$$C = \begin{pmatrix} I_2 & 0 & 0 \\ -2I_2 & 2I_2 & -I_2 \end{pmatrix}, \qquad D = (I_2 \quad -I_2 \quad I_2).$$

Thus

$$\delta_1(\alpha, \lambda_{+1}) = \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}' M \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix} - 2 \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}' m, \tag{115}$$

$$M = C' P C + D' \Omega^{-1} D, \quad m = C' q. \tag{116}$$

Note that $\delta_1(\alpha, \lambda_{+1})$ is functionally similar to $\gamma(s)$ in (93), with β replaced by λ_{+1}. Therefore, the subsequent transformations of $\delta_1(\alpha, \lambda_{+1})$ are similar to those of $\gamma(s)$ outlined above, except that integration in (ρ, θ) is conditional on λ_{+1}, since it is $f(\lambda_{+1}|Y)$ that is

now being evaluated. M and m are partitioned conformably with $\begin{pmatrix}\alpha' & \lambda'_{+1}\end{pmatrix}$ as

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \qquad m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}.$$

After transformation from α to (ρ, θ), $\delta_1(\alpha, \lambda_{+1})$ becomes

$$\delta_*(\rho, \theta_{+1}, \lambda_{+1}) = \lambda'_{+1} M_{22} \lambda_{+1} - 2\lambda'_{+1} m_2 + a^*_\theta (\rho - r^*_{\theta\lambda})^2 - c^*_{\theta\lambda}, \tag{117}$$

with

$$a^*_\theta = e'_\theta M_{11} e_\theta, \qquad r^*_{\theta\lambda} = \frac{1}{a^*_\theta}(m_1 - M_{12}\lambda_{+1})' e_\theta, \qquad c^*_{\theta\lambda} = r^*_{\theta\lambda}\sqrt{a^*_\theta}.$$

Regrouping (114) and integrating with respect to ρ, we obtain

$$\begin{aligned} f(\lambda_{+1}|Y) &= \frac{|\Omega|^{-\frac{1}{2}}}{2\pi} \exp\left\{-\frac{1}{2}\left(\lambda'_{+1} M_{22} \lambda_{+1} - 2\lambda'_{+1} m_2\right)\right\} \\ &\quad \times \int \left[\frac{1}{\chi(\theta)} d^*_{\theta\lambda} \exp\left(\frac{1}{2} c^{*2}_{\theta\lambda}\right)\right] m(\theta)\, d\theta, \end{aligned} \tag{118}$$

where $d^*_{\theta\lambda}$ obtains from (103) by substituting $(a^*_\theta, c^*_{\theta\lambda})$ for (a_θ, c_θ). Since $m(\theta)$ is typically a tight density in our application, the variance of the terms between brackets under the integral sign is expected to be minimal, and the integral in (118) can be estimated accurately by MC using the same EIS draws from $m(\theta)$ used for the evaluation of ℓ_t.

7.2 Bearings-Only Tracking, Non-Singular Case

The computation of $\chi(\theta)$ and CRN-EIS draws of (β, ρ, θ) are the same as for the singular case. The derivation of $f(\lambda_{t+D}|Y_t)$ under the non-singular transition defined in (90) is straightforward. As above, we suppress the index t, and replace $t + D$ by $+D$. The product $g_\lambda(\lambda) f(\lambda_{+D}|\lambda)$ in (92) depends on the quadratic form

$$\delta(\lambda_{+D}, \lambda) = (\lambda' P \lambda - 2\lambda' q) + (\lambda_{+D} - A_D \lambda) V_D^{-1} (\lambda_{+D} - A_D \lambda), \tag{119}$$

which is rewritten as

$$\delta\left(\lambda_{+D},\lambda\right)=\left(\lambda'P_0\lambda-2\lambda'q_0\right)+\lambda'_{+D}V_D^{-1}\lambda_{+D}, \tag{120}$$

with

$$P_0=P+A'_DV_D^{-1}A_D,\qquad q_0=q+A'_DV_D^{-1}\lambda_{+D}.$$

The integration with respect to λ in (118) proceeds exactly as described in the singular case, except that (P,q) are replaced by (P_0,q_0). Thus $f\left(\lambda_{+D}|Y\right)$ is given by

$$\begin{aligned} f\left(\lambda_{+D}|Y\right) &= \frac{\left|V_D\right|^{-\frac{1}{2}}}{2\pi}\exp\left\{-\frac{1}{2}\left(\lambda'_{+D}V_D^{-1}\lambda_{+D}\right)\right\} \\ &\quad\times\int\left[\frac{1}{\chi\left(\theta\right)}d_{\theta\lambda}^0\exp\left(\frac{1}{2}\left(s_{\theta\lambda}^0\right)^2\right)\right]m\left(\theta\right)d\theta, \end{aligned} \tag{121}$$

where $s_{\theta\lambda}^0$ and $d_{\theta\lambda}^0$ are defined by (98) and (100), with (P,q) replaced by (P_0,q_0). The EIS evaluation of (121) parallels that of $f\left(\lambda_+|Y\right)$ in (118).

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., 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, 2004, "Sequential Monte Carlo Filtering: An Example", University of Pennsylvania Working Paper.

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

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

[10] 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.

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

[12] Ghysels, E., A.C. Harvey, and E. Renault, 1996, "Stochastic Volatility", in Maddala, G., and C.R. Rao, Eds. *Handbook of Statistics,* Vol. 14. Amsterdam: Elsevier.

[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] 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.

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

[16] Jacquier, E., N.G. Polson, and P.E. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics.* 12, 371-417.

[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, 1032-1041.

[19] Liesenfeld, R. and J.F. Richard, 2006, "Improving MCMC Using Efficient Importance Sampling", University of Pittsburgh Working Paper.

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

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

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

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

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

[25] 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.

[26] Taylor, S. J., 1982, "Financial Returns Modeled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices", in Anderson, O.D., Ed. *Time Series Analysis : Theory and Practice* Vol. 1. Amsterdam: North Holland.

[27] Taylor, S.J., 1986, *Modeling Financial Time Series.* Chichester: John Wiley and Sons.

8 Tables and Figures

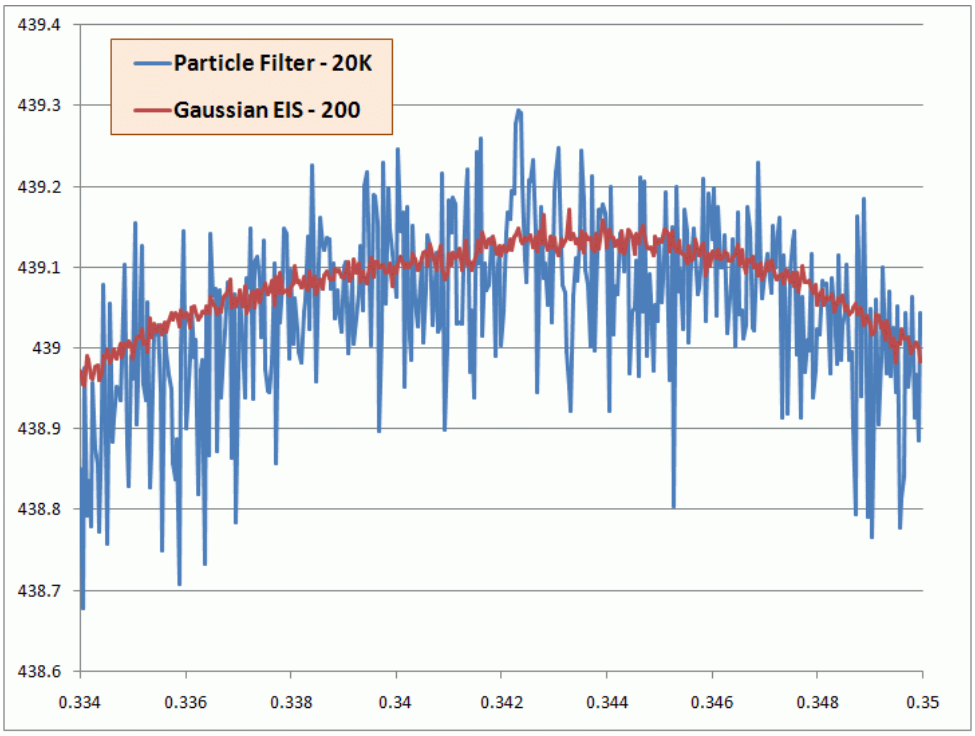

Figure 1. Conditional Log Likelihood Function for α.

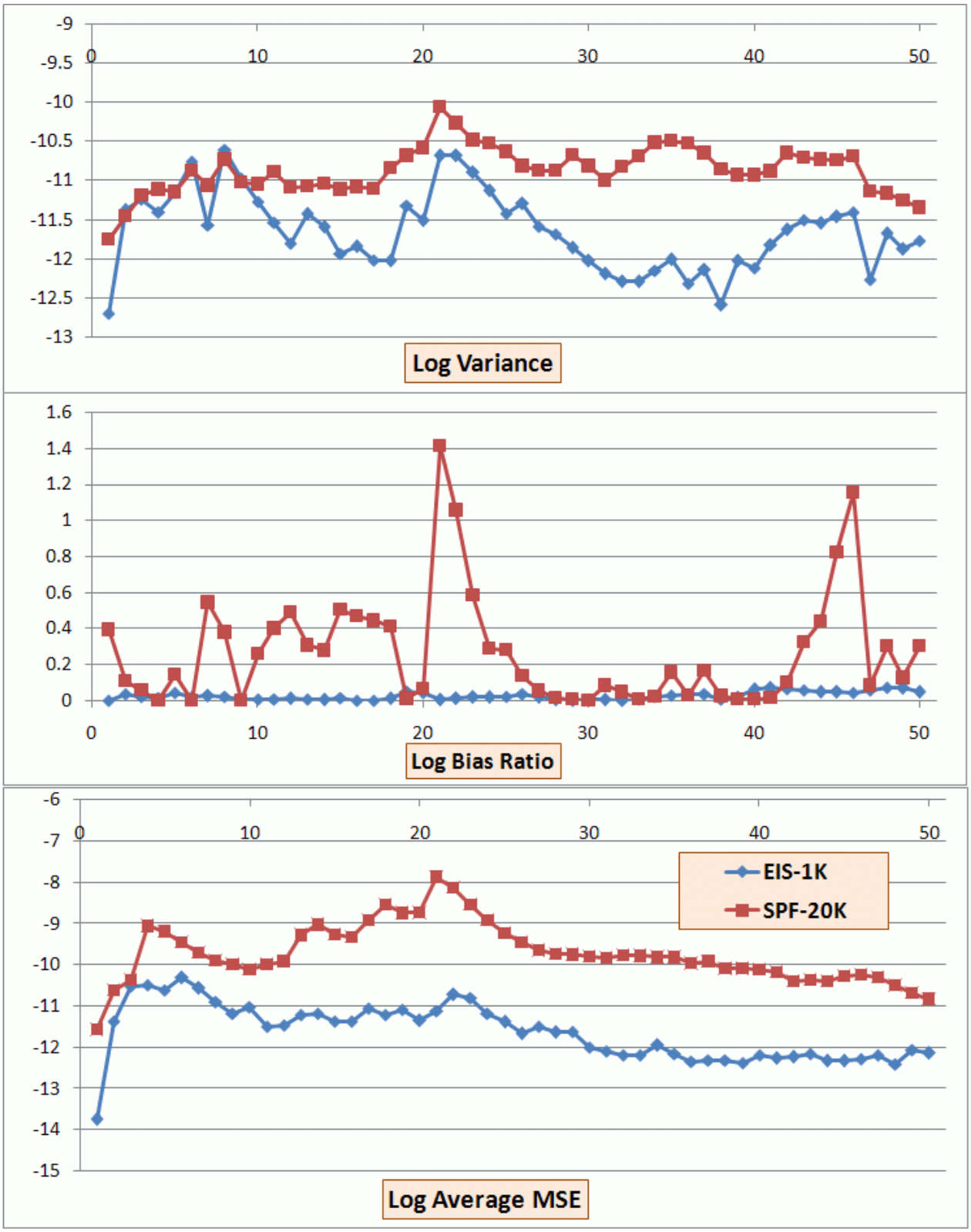

Figure 2. MSE decompositions, Stochastic Volatility Model.

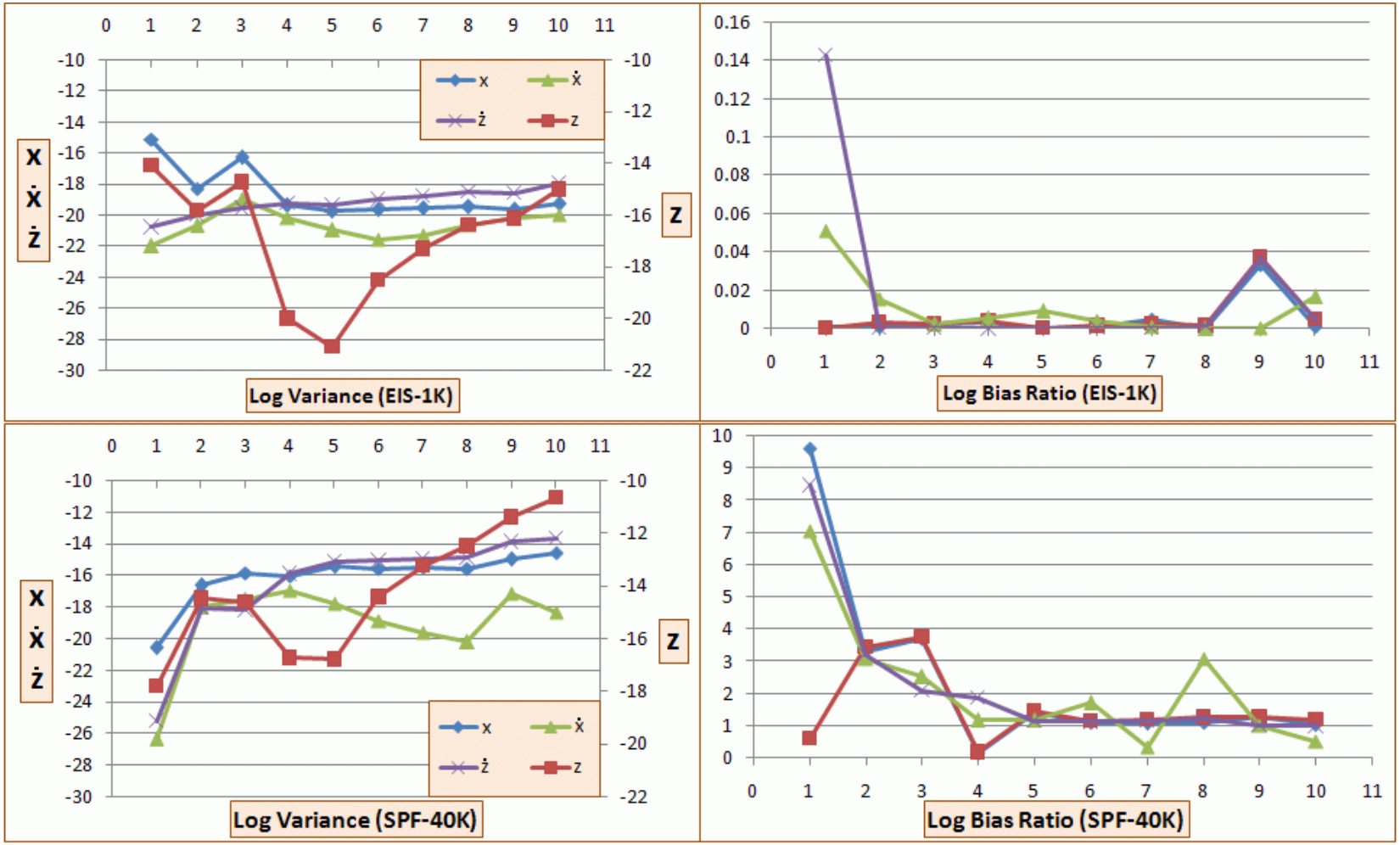

Figure 3. Log Variance and Log Bias Ratio, Singular Case.

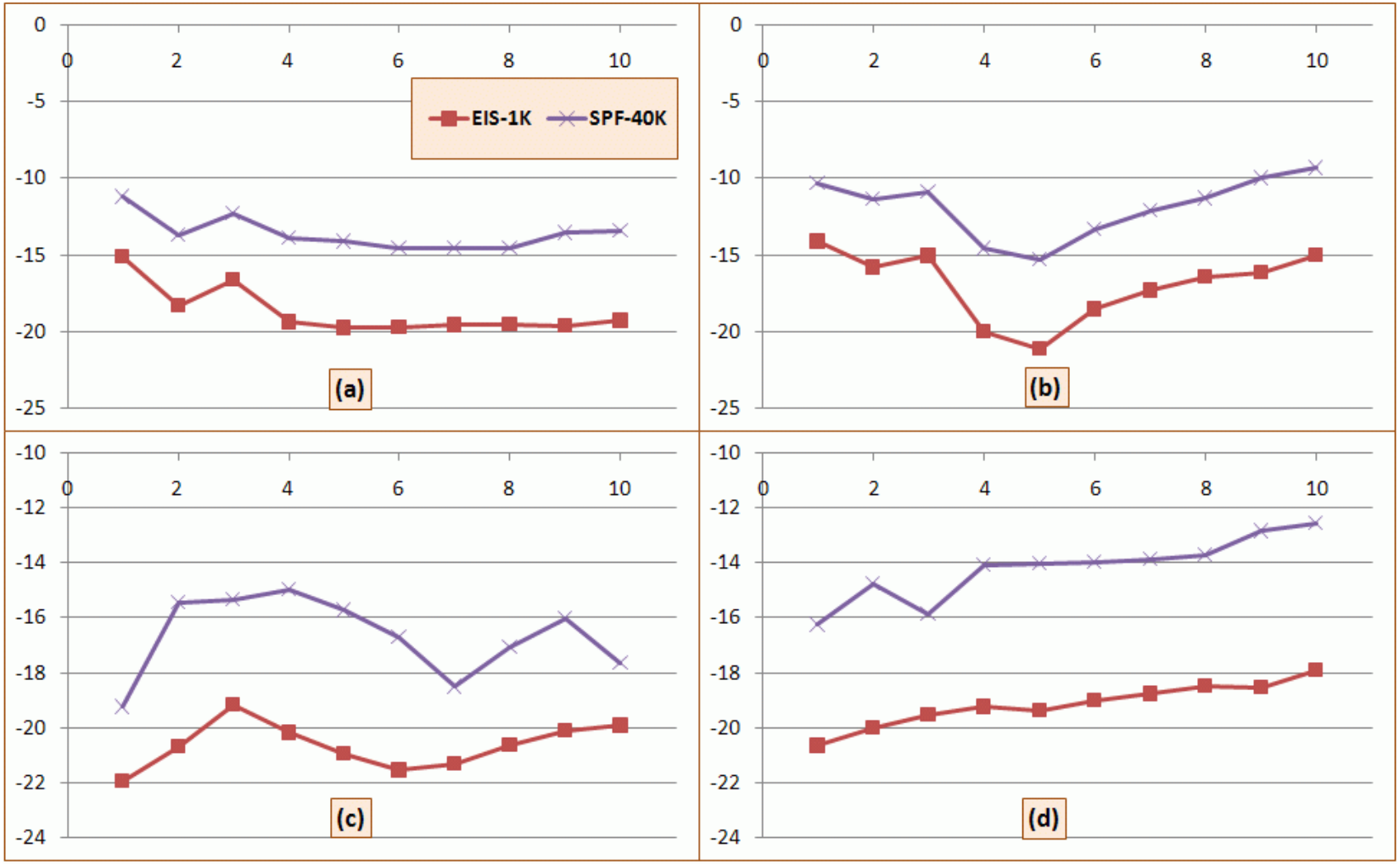

Figure 4. Log-Avg.MSE Comparisons, Singular Case. Panel (a) $\rightarrow x$, (b)$\rightarrow z$, (b)$\rightarrow \dot{x}$, (b)$\rightarrow \dot{z}$.

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	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time
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
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												
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
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
Relative Time Efficiency			0.748			0.699			0.775			0.713
ADAPTED PARTICLE FILTER												
20000	-195.839	0.0059	2.710	-213.600	0.0306	2.730	-263.635	0.3902	2.740			
200000	-195.840	0.0017	29.945	-213.604	0.0113	29.935	-263.543	0.1312	29.778			
Relative Time Efficiency			8.182			1.497			0.0543			
PIECEWISE-EIS PARTICLE FILTER (R=100, S=100)												
100	-195.845	0.0409	0.552	-213.595	0.0788	0.5392	-263.530	0.0344	0.549	-368.516	0.0107	0.550
1000	-195.845	0.0139	2.262	-213.602	0.0261	2.180	-263.526	0.0108	2.228	-368.521	0.0049	2.241
Relative Time Efficiency			1.644			3.846			107.32			2001.93
GAUSSIAN-EIS PARTICLE FILTER (R=100, S=100)												
1000	-195.825	0.043	0.717	-213.815	0.145	0.860	-266.016	0.459	1.699	-372.937	0.5117	2.278
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
Relative 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	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time
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
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												
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
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
Relative Time Efficiency			1.067			0.929			0.707			0.697
ADAPTED PARTICLE FILTER												
20000	-141.168	0.0063	2.629	-166.144	0.0425	2.643						
200000	-141.170	0.0020	29.892	-166.145	0.0159	30.015						
Relative Time Efficiency			11.590			1.157						
PIECEWISE-EIS PARTICLE FILTER (R=100, S=100)												
100	-141.167	0.0485	0.5156	-166.147	0.0727	0.568	-241.046	0.0325	0.513	-356.803	0.0173	0.587
1000	-141.164	0.0167	2.0878	-166.128	0.0248	2.324	-241.038	0.0109	2.099	-356.796	0.0061	2.388
Relative Time Efficiency			2.330			6.075			109.97			1401.81
GAUSSIAN-EIS PARTICLE FILTER (R=100, S=100)												
1000	-141.171	0.051	0.833	-166.160	0.087	0.791	-241.088	0.034	0.865	-356.868	0.0215	0.896
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
Relative Time Efficiency			0.782			1.976			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.

Table 3. DSGE Model

STANDARD PARTICLE FILTER				
N	Mean	Stdev	Time	Rel. Time Efficiency
20000	438.561	0.206	5.7651	1.000
100000	438.545	0.0774	29.036	1.417
GAUSSIAN-EIS PARTICLE FILTER ($R = 100$)				
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.

Table 4. MLE Comparisons

.								
STANDARD PARTICLE FILTER (N=20000)								
	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 (N=200)								
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.