

Efficient Filtering in State-Space Representations

David N. DeJong *
University of Pittsburgh

Hariharan Dharmarajan
Bates and White, Wash. D.C.

Roman Liesenfeld
Universität Kiel

Jean-François Richard
University of Pittsburgh

November 2008; Revised August 2009

Abstract

We develop a numerical procedure that facilitates efficient filtering in applications involving non-linear and non-Gaussian state-space representations. 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 grants SES-0516642 and SES-004514360; likewise DeJong under SES-004514360. For helpful comments, we thank four anonymous referees, Jesus Fernandez-Villaverde, Hiroyuki Kasahara, Juan Rubio-Ramirez, and seminar and conference participants at the Universitites of Comenius, Di Tella, Texas (Dallas), Pennsylvania, the 2008 Econometric Society Summer Meetings (CMU), the 2008 Conference on Computations in Economics (Paris), and the 2008 Vienna Macroeconomics Workshop. For assistance with manuscript preparation, we thank Angelika Reinmüller and Georgia Spears.

1 Introduction

In applications involving non-linear and non-Gaussian state-space representations, the objective of filtering requires numerical integration over unobservable state variables. Dating at least to the sequential importance sampling (SIS) algorithm developed by Handschin and Dayne (1969) and Handschin (1970), sequential Monte Carlo (SMC) methods have served as a workhorse for achieving this objective. These methods entail the construction of operational approximations to filtering densities in the form of mixtures of Dirac measures. Such approximations are trivially amenable to Monte Carlo simulation, and when applied to state-transition densities, enable the straightforward sequential estimation of filtered trajectories of state variables.

The simplicity and flexibility associated with approximations constructed using mixtures of Dirac measures make them an attractive option for achieving filtering. Moreover, extensive efforts have gone towards refining the baseline SIS algorithm, with the goal of enhancing accuracy and numerical efficiency in the face of challenging scenarios that give rise to problems known as degeneracy and sample impoverishment. A sketch of these problems, and methods for overcoming them, are provided below (for extensive surveys, see Ristic, Arulampalan and Gordon, 2004; and Cappé, Godsill and Moulines, 2007). In the terminology of Pitt and Shephard (1999), the goal of these efforts is to achieve adaption.

While the quest for adaption has produced effective algorithms in a wide range of applications, adherence to the use of mixtures of Dirac measures for approximating targeted filtering densities entails a significant limitation: the support of the period- $(t - 1)$ mixtures of Dirac measures carried over for the (sequential) construction of period- t approximations can no longer be changed in period t . Only the weights attached to the period- $(t - 1)$ discrete elements of the state space (known as particles) can be adjusted in period t . To be sure, there exist a wide range of algorithms designed to optimize (adapt) these weights in light of new information available at time t . Nevertheless, optimality is restricted as being conditional on the individual particles established in period $(t - 1)$. That is to say, mixture-of-Dirac

approximations cannot achieve unconditionally optimality, or in other words, full adaption.

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

The paper is organized as follows. The generic problem of filtering is outlined in the next section. We then provide a brief overview of SMC filtering, including unadapted filters, the auxiliary particle filter, and the optimal particle filter (where optimality is conditional on the fixed- and prespecified-support constraint). While the latter is rarely operational, it provides a useful conceptual framework for highlighting the difference between conditional optimality and the unconditional optimality the EIS filter targets. The EIS methodology is then introduced generically, and described specifically in the context of filtering. Example applications follow, along with concluding remarks.

2 Filtering

The objective of filtering is to track the dynamic evolution of unobservable state variables using (typically) noisy measurements of observables. This requires the computation of integrals over the unobserved states, which in general must be approximated numerically. Let s_t be a m -dimensional vector of state variables, and denote $\{s_j\}_{j=1}^t$ as S_t . Likewise, let y_t be an n -dimensional vector of measurements, and denote $\{y_j\}_{j=1}^t$ as Y_t . The state space is characterized by a state transition density $f(s_t|s_{t-1}, Y_{t-1})$, initialized by a known marginal density $f(s_0)$. The model is completed by a measurement density $f(y_t|s_t, Y_{t-1})$.

Let $h(s_t)$ denote a function of interest for tracking. Our objective is that of producing numerically efficient estimates of the filtered values

$$\bar{h}_t = E[h(s_t) | Y_t] = \int h(s_t) f(s_t | Y_t) ds_t, \quad (1)$$

where $f(s_t | Y_t)$ denotes the filtering density. Filtering densities $\{f(s_t | Y_t)\}_{t=1}^T$ are constructed sequentially; computations in period t take the period- $(t-1)$ filtering density as input and proceed as follows.

- Predictive density:

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

- Likelihood integral:

$$f(y_t | Y_{t-1}) = \int f(y_t | s_t, Y_{t-1}) f(s_t | Y_{t-1}) ds_t. \quad (3)$$

- Filtering density:

$$f(s_t | Y_t) = \frac{f(y_t | s_t, Y_{t-1}) f(s_t | Y_{t-1})}{f(y_t | Y_{t-1})}. \quad (4)$$

- Given $f(s_t|Y_t)$, filtering obtains as in (1).

Note that the inclusion of Y_{t-1} as a conditioning argument in the state-transition and measurement densities is critical for verifying internal consistency between the left- and right-hand sides of (2) - (4). This being said, it is often the case that the model specification imposes additional conditional independence assumptions such as $f(y_t|s_t, Y_{t-1}) = f(y_t|s_t)$ and $f(s_t|s_{t-1}, Y_{t-1}) = f(s_t|s_{t-1})$. Such assumptions are not needed to validate the following analysis, since all that is required is that state transitions be (sequentially) Markovian, conditionally on the observed measurements. Note also that the inclusion of Y_{t-1} as a conditioning argument in state transitions allows for control applications wherein measurements trigger actions that feed back on states.

When the model is linear and Gaussian this sequence of operations can be executed analytically via the Kalman filter (Kalman, 1960). Otherwise approximations must be implemented. One approach to approximation is based upon (local) linearization and/or deterministic representation of filtering densities. Examples include the extended Kalman filter (Jazwinski, 1970); the unscented Kalman filter (Van der Merwe, Doucet, De Freitas and Wan, 2000); and the Gaussian quadrature Kalman filter (Ito and Xiong, 2000). Alternative approaches, like ours, are based upon SMC methods. Following is a brief overview of these methods.

3 Overview of SMC Methods

As noted in the introduction, following the development of the SIS algorithm pioneered by Handschin and Dayne (1969) and Handschin (1970), SMC methods have employed mixtures of Dirac measures for approximating filtering densities. Efforts to build upon the baseline SIS algorithm have sought to overcome two challenges. The first is known as degeneracy: as the sequential approximation of filtering densities advances through time, the variance of importance sampling weights can only increase (Bobrovsky and Zakai, 1975).

Thus: “In practical terms this means that after a certain number of recursive steps, all but one particle will have negligible normalized weights.” [Ristic, Arulampalam and Gordon, 2004, p.40].

A major advance in overcoming degeneracy was achieved by Gordon, Salmond and Smith (1993), through the development of the bootstrap particle filter, which incorporates the resampling technique of Rubin (1987) in the sequential filtering process. The bootstrap particle filter has since become a benchmark sampling importance resampling (SIR) algorithm; for important variations, e.g., see Kitagawa (1996); Del Moral (1996); Blake and Isard (1998); and Liu and Chen (1995).

While SIR algorithms are effective in alleviating degeneracy, they are prone to a second challenge known as sample impoverishment: particles with relatively high associated weights end up being selected many times in the resampling step, leading to a loss of diversity among the resultant sample. The well-known source of this challenge is that the period- $(t - 1)$ swarm of particles $\{s_{t-1}^i\}_{i=1}^N$ used for the construction of approximations to period- t predictive densities is established absent information conveyed by the period- t observables y_t . Pitt and Shephard (1999) refer to the establishment of these particles as ‘blind’.

Efforts to overcome the challenge of sample impoverishment, known as the achievement of adaption, entail the use of guidance from y_t in constructing period- t sampling densities. Optimal adaption, conditional upon $\{s_{t-1}^i\}_{i=1}^N$, obtains through the construction of a sampler designed to eliminate the conditional MC variance of period- t filtering approximations (e.g., see Zaritskii, Svetnik and Shimelevich, 1975; and Akasaki and Kumamoto, 1977). However, conditional optimality relies upon an auxiliary density $f(s_t|s_{t-1}, Y_t)$ which in general is intractable analytically. Partial conditional adaption obtains via the implementation of approximations to this factorization (e.g., see Pitt and Shephard, 1999; the collection of papers in Doucet, de Freitas and Gordon, 2001; and Santos and Smith, 2006).

Following is a brief overview of filtering from a period- t algorithmic prospective, which describes these challenges in further detail and provides context for the EIS-PF algorithm we

develop in Section 4. For more detailed presentations, along with comprehensive theoretical validation of the algorithms, see, e.g., Pitt and Shephard (1999), Ristic, Arulampalan and Gordon (2004), or Cappé, Godsill and Moulines (2007).

3.1 General Principle

The filtering algorithms described here share the critical characteristic that period- t computations take as input an approximation of the period- $(t-1)$ filtering density. The approximation is in the form of a mixture of Dirac measures associated with the swarm $\{s_{t-1}^i\}_{i=1}^N$, which is fixed in period t :

$$\hat{f}(s_{t-1}|Y_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \delta_{s_{t-1}^i}(s_{t-1}), \quad (5)$$

where $\delta_{s_{t-1}^i}(s)$ denotes the Dirac measure at point s_{t-1}^i and ω_{t-1}^i the weight associated with particle s_{t-1}^i ($\sum_{i=1}^N \omega_{t-1}^i = 1$). If a potential resampling step is included in the algorithm, and took place in period- $(t-1)$, then $\omega_{t-1}^i = \frac{1}{N}$.

For the purpose of comparing the EIS filter with existing SMC methods, as is our aim here, the incorporation of a potential resampling step is unimportant, thus for ease of presentation we do not explicitly describe this potential step in the following presentation. What is important is the pursuit of conditional adaption. Thus we first describe an unadapated filtering algorithm, then characterize the pursuit of conditional adaption.

3.2 Unadapated Filtering

- Propagation: Inherit $\{\omega_{t-1}^i, s_{t-1}^i\}_{i=1}^N$ from period $(t-1)$. Then for each particle s_{t-1}^i , draw a particle s_t^i from the transition density $f(s_t|s_{t-1}^i, Y_{t-1})$. The corresponding approximation of the predictive density is given by the mixture of Dirac measures

$$\hat{f}(s_t|Y_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \cdot \delta_{s_t^i}(s_t). \quad (6)$$

- Likelihood integral: Substitution of $\widehat{f}(s_t|Y_{t-1})$ for $f(s_t|Y_{t-1})$ in (3), followed by analytical integration in s_t , yields the following approximation for the likelihood integral:

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

- Filtering: Equation (4) amounts to a direct application of Bayes' Theorem to the particles $\{s_{t-1}^i\}_{i=1}^N$, resulting in the following discrete approximation for the period- t filtering density:

$$\widehat{f}(s_t|Y_t) = \sum_{i=1}^N \omega_t^i \delta_{s_t^i}(s_t), \quad (8)$$

with

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

The filtered value in (1) is approximated by

$$\widehat{h}_t = \sum_{i=1}^N \omega_t^i h(s_t^i). \quad (10)$$

- Pass the weights and mass points $\{\omega_t^i, s_t^i\}_{i=1}^N$ to the period- $(t+1)$ propagation step and proceed through period T .

Note: When the algorithm is extended to incorporate an automatic resampling step, the bootstrap particle filter of Gordon, Salmond and Smith (1993) is obtained. Alternative SIR algorithms allow for a conditional resampling step, depending on the realization of a period-by-period degeneracy diagnostic.

3.2.1 The Conditionally Optimal Particle Filter

Notice that the measurement density incorporates the critical conditional independence assumption that y_t is independent of s_{t-1} given (s_t, Y_{t-1}) :

$$f(y_t|s_t, Y_{t-1}) = f(y_t|s_t, s_{t-1}, Y_{t-1}). \quad (11)$$

It follows that

$$f(s_t|s_{t-1}, Y_{t-1}) \cdot f(y_t|s_t, Y_{t-1}) = f(s_t|s_{t-1}, Y_t) \cdot f(y_t|s_{t-1}, Y_{t-1}). \quad (12)$$

For cases in which this factorization is tractable analytically, it is possible to achieve conditionally optimal adaption. To see how, substitute the predictive density (2) into the likelihood integral (3) to obtain

$$f(y_t|Y_{t-1}) = \int \int [f(y_t|s_t, Y_{t-1}) f(s_t|s_{t-1}, Y_{t-1})] f(s_{t-1}|Y_{t-1}) ds_t ds_{t-1}. \quad (13)$$

Next, substituting for the bracketed term in (13) using the factorization (12), we obtain

$$f(y_t|Y_{t-1}) = \int \int f(y_t|s_{t-1}, Y_{t-1}) \cdot [f(s_t|s_{t-1}, Y_t) f(s_{t-1}|Y_{t-1})] ds_t ds_{t-1}. \quad (14)$$

Noting that $f(s_t|s_{t-1}, Y_t)$ integrates to 1 in s_t , we obtain

$$f(y_t|Y_{t-1}) = \int f(y_t|s_{t-1}, Y_{t-1}) f(s_{t-1}|Y_{t-1}) \cdot ds_{t-1}, \quad (15)$$

whose estimate is given by

$$\hat{f}(y_t|Y_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \cdot f(y_t|s_{t-1}^i, Y_{t-1}). \quad (16)$$

The propagation stage for the conditionally optimal filter is defined by the bracketed term in (14), and proceeds as follows: for each particle s_{t-1}^i , draw a particle s_t^i from $f(s_t|s_{t-1}^i, Y_t)$. The implied weights are given by

$$\omega_t^i = \frac{\omega_{t-1}^i \cdot f(y_t|s_{t-1}^i, Y_{t-1})}{\sum_{j=1}^N \omega_{t-1}^j \cdot f(y_t|s_{t-1}^j, Y_{t-1})}, \quad (17)$$

and are used to estimate the period- t filtering density and filtered values as given by (8) and (10).

Note that since ω_t^i does not depend on s_t^i , but only on s_{t-1}^i , its conditional variance is zero given $\{s_{t-1}^i\}_{i=1}^N$. This is referenced as the optimal sampler, following Zaritskii, Svetnik and Shimelevich (1975) and Akaski and Kumamoto (1977). However, since the factorization in (12) is tractable analytically only in special cases, this sampler represents a theoretical rather than an operational benchmark (for approaches to achieving approximate conditional optimality, see Doucet, 1998; and Vaswani, 2008). But more importantly, this sampler is only optimal conditionally: a different weight is associated with each particle s_{t-1}^i , implying that while the conditional MC variance of ω_t^i is zero, its unconditional MC variance is not. The limitation to conditional optimality is inherently linked to the use of mixtures of Dirac measures in representing filtering densities. As noted, the EIS implementation we propose below targets unconditional optimality.

3.2.2 Conditionally Suboptimal Particle Filters

Since the optimal kernel is generally intractable, much effort has been devoted to its approximation. The class of approximations we now discuss still relies upon mixture-of-Dirac approximations for filtering densities, but relaxes the propagation steps leading to (6).

The predictive density is now approximated by the continuous mixture

$$\hat{f}(s_t|Y_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \cdot f(s_t|s_{t-1}^i, Y_{t-1}). \quad (18)$$

In turn, the likelihood integral in (3) is approximated by

$$\widehat{f}(y_t|Y_t) = \sum_{i=1}^N \int \omega_{t-1}^i \cdot f(y_t|s_t, Y_{t-1}) \cdot f(s_t|s_{t-1}^i, Y_{t-1}) ds_t. \quad (19)$$

Next, one reinterprets this integrand as the kernel of a mixed density for the pair (s_t, k_t) , where k_t denotes the multinomial $MN\left(N; \{\omega_{t-1}^i\}_{i=1}^N\right)$. The likelihood integral is evaluated via importance sampling using an auxiliary mixed kernel of the form

$$\gamma(s, k; Y_t) = \omega_{t-1}^k \cdot p(s, k; Y_t) \cdot f(s|s_{t-1}^k, Y_{t-1}), \quad (20)$$

where $p(s, k; Y_t)$ is introduced with the objective of minimizing the MC variance of the likelihood estimate conditionally on k . The kernel $\gamma(s, k; Y_t)$ must be normalized to produce the corresponding importance sampling density $g(s, k|Y_t)$. The simplest case obtains when $p(\cdot)$ does not depend on s :

$$p(s, k; Y_t) \equiv p_t^k. \quad (21)$$

In this case,

$$g(s, k|Y_t) = \pi_t^k \cdot f(s|s_{t-1}^k, Y_{t-1}), \quad (22)$$

$$\pi_t^k = D_t^{-1} \cdot \omega_{t-1}^k \cdot p_t^k, \quad D_t = \sum_{j=1}^N \omega_{t-1}^j \cdot p_t^j. \quad (23)$$

Let $\widetilde{\omega}(s, k; Y_t)$ denote the ratio between the integrand in (19) and the IS density

$$\widetilde{\omega}(s, k; Y_t) = D_t \cdot \frac{f(y_t|s, Y_{t-1})}{p_t^k}; \quad (24)$$

also, let $\{s_t^i, k_t^i\}$ denote N i.i.d draws from $g(s, k|Y_t)$. Specifically k_i is drawn from the multinomial $MN(N; \{\pi_t^i\}_{i=1}^N)$ and s_t^i from $f(s_t|s_{t-1}^{k_i}, Y_{t-1})$. The corresponding IS estimates

of the likelihood function and filtered values are given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N \widetilde{\omega}_t^i, \quad (25)$$

$$\widehat{h}_N = \sum_{i=1}^N \omega_t^i h(s_t^i), \quad (26)$$

$$\widetilde{\omega}_t^i = \widetilde{\omega}(s_t^i, k_t^i, Y_t), \quad \omega_t^i = \frac{\widetilde{\omega}_t^i}{\sum_{j=1}^N \widetilde{\omega}_t^j}. \quad (27)$$

Finally, the period- t filtering density is given by the mixture of Dirac measures

$$\widehat{f}_N(s_t|Y_t) = \sum_{i=1}^N \omega_t^i \delta_{s_t^i}(s_t). \quad (28)$$

As a special case, the auxiliary particle filter of Pitt and Shephard (1999) uses the following specification for p_t^k in (21):

$$p_t^k = f(y_t|\mu_t^k, Y_{t-1}), \quad (29)$$

$$\mu_t^k = E(s_t|s_{t-1}^k, Y_{t-1}). \quad (30)$$

Following (21)-(24) the IS weights are then given by

$$\widetilde{\omega}_t^i = D_t \cdot \frac{f(y_t|s_{t-1}^k, Y_{t-1})}{f(y_t|\mu_t^k, Y_{t-1})} \Big|_{k=k_t^i}, \quad (31)$$

$$D_t = \sum_{j=1}^N \omega_{t-1}^j f(y_t|\mu_t^j, Y_{t-1}). \quad (32)$$

Under special circumstances further adaption can be achieved. If the transition density $f(s_t|s_{t-1}, Y_{t-1})$ belongs to a family of densities closed under multiplication, then selecting $p(s, k; Y_t)$ in (20) from that same family can produce an operational and improved sampler. Its integrating constant obtains by first integrating the product $p(s, k; Y_t) \cdot f(s|s_{t-1}^k, Y_{t-1})$ with respect to s , then summing the remainder over k . Examples for the case in which

$f(s_t|s_{t-1}, Y_{t-1})$ is Gaussian in s_t are discussed in Pitt and Shephard (1999) and Smith and Santos (2006). Using for $p(s, k; Y_t)$ a first-order Taylor series expansion of $\ln f(y_t|s_t, Y_{t-1})$ in s_t around μ_t^j yields Pitt and Shephard’s adapted particle filter; likewise, Smith and Santos demonstrate implementation of a second-order expansion.

3.3 Sample Impoverishment and Bias

As mentioned above, sample impoverishment is a direct consequence of resampling. It occurs when the number of *distinct* particles following resampling is far smaller than N . This problem arises when the measurement density $f(y_t|s_t, Y_{t-1})$ is tightly distributed relative to the state-transition density $f(s_t|s_{t-1}, Y_{t-1})$. With references to (2) and (3), this scenario implies that the predictive density $f(s_t|Y_{t-1})$ exhibits significantly greater dispersion than $f(y_t|s_t, Y_{t-1})$, treated as a function of s_t given Y_t . By implication, $f(s_t|Y_{t-1})$ or a particle-based representation thereof will serve as inefficient IS samplers both for computing (3) and for approximating the filtering density $f(s_t|Y_t)$.

As noted, algorithms designed to achieve conditional adaptation seek to alleviate this problem by accounting for y_t in establishing period- t sampling densities. However, even in the best-case scenario the gains from adaption are restricted due to the reliance on Dirac representations of filtering densities. Specifically, while conditional adaptation enables adaptation of the discrete mixture probabilities or even the conditional IS samplers for s_t given (s_{t-1}^k, Y_t) , the conditioning swarm $\{s_{t-1}^k\}$ remains fixed in period t . This is especially critical when s_t turns out to be an outlier, which is not uncommon given relatively diffuse state-transition densities. Given an outlier, $f(y_t|s_t, Y_{t-1})$ takes the form of a needle located in the far tails of particle representations. As the examples below indicate, under such circumstances filtered estimates will not only have large MC variances, but will also exhibit even larger squared biases, whose elimination can require the implementation of prohibitively large particle swarms.

A final point bears noting. Insofar as measurement densities are more tightly distributed

than transition densities, MC imprecision and inaccuracy are far less of an issue for the approximation of predictive densities. With reference to (2), the density $f(s_{t-1}|Y_{t-1})$ or discrete representations thereof are expected to be tight relative to the transition density $f(s_t|s_{t-1}, Y_{t-1})$, viewed as a function of s_{t-1} given (s_t, Y_{t-1}) . This explains why efforts to refine SMC algorithms have largely emphasized (conditional) efficiency in the construction of importance samplers for likelihood integrals, leaving the representation of predictive densities (6) unaffected.

4 Efficient Importance Sampling and Filtering

As noted, the EIS filter we propose departs from traditional SMC algorithms in that it employs tailored continuous approximations of targeted filtering densities in place of mixtures of Dirac measures, and thus targets the construction of unconditionally rather than conditionally optimal importance samplers. Here we briefly describe the generic EIS algorithm, and then characterize its extension to the specific problem of filtering.

4.1 Efficient Importance Sampling

Consider an analytically intractable integral of the form

$$I(\theta) = \int \phi(s; \theta) ds, \quad (33)$$

where θ regroups all variables (parameters and observations) upon which the integral is conditioned. The goal of EIS is to construct a parametric continuous IS density $g(s|\hat{a}(\theta))$ designed to minimize the MC sampling variance of the ratio

$$\omega(s; \theta, a) = \frac{\phi(s; \theta)}{g(s|a)} \quad (34)$$

under IS draws obtained from $g(s|a)$. Construction is achieved as follows. One initially selects a parametric class of auxiliary density kernels $\mathcal{K} = \{k(s; a); a \in A\}$ amenable to MC simulation and with known integrating constant $\chi(a)$. The relationship between kernel and density is given by

$$\begin{aligned} g(s|a) &= \frac{k(s; a)}{\chi(a)}, \\ \chi(a) &= \int k(s; a) ds. \end{aligned} \tag{35}$$

Following Richard and Zhang (2007), an (approximate) optimal value $\hat{a}(\theta)$ obtains as the solution of the following auxiliary Least Squares (LS) problem

$$(\hat{a}(\theta), \hat{c}(\theta)) = \underset{a \in A, c \in \mathbb{R}}{\text{ArgMin}} \sum_{i=1}^R [\ln \phi(s_i; \theta) - c - \ln k(s_i; a)]^2. \tag{36}$$

Here, $\{s_i\}_{i=1}^R$ denotes i.i.d. draws from $g(s|\hat{a}(\theta))$, with R typically set at 3 to 5 times the dimension of a (and thus typically smaller than the number N of draws subsequently used for IS estimation). The fact that these draws are conditional on $\hat{a}(\theta)$ implies that the latter obtains as the fixed point solution of a sequence $\{\hat{a}_v(\theta)\}_{v=1,2,\dots}$, where $\hat{a}_v(\theta)$ solves (36) under draws from $g(s|\hat{a}_{v-1}(\theta))$. In turn, c is a normalizing constant that accounts for factors in $\phi()$ and $k()$ that do not depend upon s . The sequence can be initialized, e.g., using a Taylor series expansion of $\ln \phi(s; \theta)$ around its mode (possibly adjusting the Hessian with a tuning constant specified to ensure sufficient coverage of the sample space).

In order to achieve smooth convergence, all draws under $\hat{a}_v(\theta)$ are obtained by transformation of a single set $\{u_i\}_{i=1}^R$ of Common Random Numbers (CRNs), where u_i denotes a draw from a canonical distribution - i.e. one that does not depend on a (typically $U(0, 1)$ or $N(0, 1)$) - to be transformed into draws from $g(s|a)$ using techniques described, e.g., in Devroye (1986). Under these conditions convergence turns out to be fast for reasonably well-behaved problems, with typically less than 5 iterations. Most importantly, these itera-

tions turn out to be the main driver for full adaption (within the class of densities \mathcal{K}), as it progressively repositions $k(s; a)$ on the region of importance of $\phi(s; \theta)$.

Convergence can be assessed by monitoring $\hat{a}_v(\theta)$ across successive iterations v , and a stopping rule can be established using a relative-change threshold. Although there is no guarantee nor formal proof that the sequence $\{\hat{a}_v(\theta)\}$ converges for every possible pair $(\phi(\cdot), k(\cdot))$, we have found repeatedly that it never fails to occur short of a complete mismatch between the target ϕ and the kernel k (e.g., ϕ is bimodal and h is unimodal), in which case failure to converge serves as a signal that the class \mathcal{K} needs to be adjusted. Moreover, what matters is that the resulting sampler deliver accurate and numerically efficient approximations of the targeted integrand. This can be assessed using various diagnostic measures, including the R^2 statistic associated with the auxiliary regression (36), and summary statistics regarding the dispersion of the importance sampling weights $\omega(s_i; \theta, \hat{a}(\theta))$ (see Richard and Zhang, 2007, for further details).

An algorithm for implementing EIS is as follows.

Algorithm 1: Efficient Importance Sampling

A. Exponential family of distributions (Lehman, 1986): $\ln k(s; a) = b(s) + a' \cdot T(s)$, where $T(s)$ denotes a fixed-dimensional vector of sufficient statistics.

- Initialize: $\hat{a}_0(\theta)$, $v = 1$.
- Step v :
 - Draw: $\{s_i\}_{i=1}^R$ from $g(s|\hat{a}_{v-1}(\theta))$.
 - Regress: $\{\ln \phi(s_i; \theta)\}_{i=1}^R$ on $\{1, T(s_i)\}_{i=1}^R$, obtain $(\hat{c}_v(\theta), \hat{a}_v(\theta))$.
 - Check Convergence: $\|\hat{a}_v(\theta) - \hat{a}_{v-1}(\theta)\| < \varepsilon$ (or similar).
 - * If No: $v = v + 1$; repeat step v .
 - * If Yes: end iteration, set $\hat{a}(\theta) = \hat{a}_v(\theta)$.

- Estimation: Draw $\{s_i\}_{i=1}^N$ from $g(s|\hat{a}(\theta))$, compute the corresponding weights $\{\tilde{\omega}_i\}_{i=1}^N$, where

$$\tilde{\omega}_i = \frac{\phi(s_i; \theta)}{g(s_i|\hat{a}(\theta))},$$

and obtain

$$\hat{I}_N(\theta) = \frac{1}{N} \sum_{i=1}^N \tilde{\omega}_i.$$

B. Other distributions: LS regression replaced by one-step Gauss-Newton iteration:

$$\{\ln \phi(s_i; \theta) - \hat{c}_{v-1}(\theta) - \ln k(s_i; \hat{a}_{v-1}(\theta))\}_{i=1}^R \text{ on } \left\{ 1, \frac{\partial \ln k(s_i; a)}{\partial a} \Big|_{a=\hat{a}_{v-1}(\theta)} \right\}_{i=1}^R.$$

4.2 Classes of EIS Kernels

Currently EIS is fully operational only for auxiliary kernels from the exponential family of distributions, since the auxiliary LS problem (36) is then linear in a , where a denotes a natural parametrization in the sense of Lehmann (1986, Section 2.7). A flexible (low-dimensional) extension outside of the exponential family of distributions is described below; others are under development using Gauss-Newton iterations. A particularly promising extension entails the specification of $g(s|a)$ as a mixture of (multivariate) Gaussian densities, in which case a denotes the collection of mixing and individual component parameters. As is well known, mixtures provide a straightforward yet highly flexible extension of linear Gaussian models (e.g., see Frühwirth-Schnatter, 2006). Initial pilot applications, implemented using one-step Gauss-Newton EIS iterations, have yielded promising results.

Here we adopt an extension that has been successfully applied to particle filters in cases wherein a subset of state variables are linear Gaussian, facilitating the development of mixture Kalman filters (e.g., see Doucet et al., 2000; Chen and Liu, 2000; and Andrieu and Coucet, 2002). In the context of EIS, this amounts to partitioning s into $s = (s_1, s_2)$, and

factorizing $g(s|a)$ conformably into

$$g(s|a) = g(s_2|s_1, a_2) \cdot g(s_1|a_1), \quad (37)$$

where $g(s_2|s_1, a_2)$ is Gaussian in s_2 given s_1 . Then $\hat{a}_2(\theta)$ obtains either analytically by direct factorization of the target ϕ (as in the mixture Kalman filters), or by Gaussian EIS. Below we present a bearings-only tracking application wherein s is four-dimensional, and extension beyond the exponential family is required along a single dimension. The extension we implement is a piecewise log-linear approximation to the targeted integrand.

Specifically, the sampler is characterized by a kernel $k(s; a)$ with parameter $a' = (a_0, \dots, a_R)$ consisting of a grid $a_0 < a_1 < \dots < a_R$, where the interval covers the support of the relevant margin. Extensions to infinite support require either truncation or tinkering with tail approximations, as is commonly done for random variate generation – e.g., see Devroye (1986). Deleting the argument θ for the ease of notation, let $\phi(s)$ denote the target kernel. We first describe the kernel $k(s; a)$ for a preassigned grid a . It is defined as follows:

$$\ln k_j(s; a) = \alpha_j + \beta_j s, \quad \forall s \in [a_{j-1}, a_j] \quad (38)$$

$$\beta_j = \frac{\ln \phi(a_j) - \ln \phi(a_{j-1})}{a_j - a_{j-1}} \quad (39)$$

$$\alpha_j = \ln \phi(a_j) - \beta_j a_j.$$

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

$$F_j(s; a) = \frac{\chi_j(s; a)}{\chi_n(a)}, \quad \forall s \in [a_{j-1}, a_j] \quad (40)$$

$$\chi_j(s; a) = \chi_{j-1}(a) + \frac{1}{\beta_j} [k_j(s; a) - k_j(a_{j-1}; a)] \quad (41)$$

$$\chi_0(a) = 0, \quad \chi_j(a) = \chi_j(a_j, a). \quad (42)$$

Its inverse c.d.f. is given by

$$s = \frac{1}{\beta_j} \{ \ln[k_j(a_{j-1}; a) + \beta_j(u \cdot \chi_R(a) - \chi_{j-1}(a))] - \alpha_j \} \quad (43)$$

$$u \in]0, 1[, \quad \chi_{j-1}(a) < u \cdot \chi_R(a) < \chi_j(a). \quad (44)$$

Next we discuss the iterative selection of the grid parameter a . Note that $k(s; a)$ is nonlinear in a and that a close fit might require a sufficiently large number of intervals (in the application described below, $R = 100$ suffices to produce very accurate results). This precludes implementing EIS iterations as described above. Instead we pursue the recursive construction of an equal-probability-division kernel $k(s; \hat{a})$, which mimics the standard EIS objective function. This approach entails the non-random equal division of $[\varepsilon, 1 - \varepsilon]$ with $u_i = \varepsilon + (2 - \varepsilon)i/R$ for $i = 1, \dots, R - 1$, with ε sufficiently small (typically $\varepsilon = 10^{-4}$) to avoid tail intervals of excessive length. It proceeds as follows:

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

$$\hat{a}_i^{v+1} = F_v^{-1}(u_i), \quad i = 1 \dots, R - 1. \quad (45)$$

The algorithm iterates until (approximate) convergence.

4.3 EIS Filtering

Referring back to (2) and (3), computations in period t require the evaluation of two nested integrals: the likelihood integral and the predictive integral (evaluation of filtered values rely upon the same IS sampler as that of the likelihood values). In both cases, targeted integrands are not tractable analytically, and thus entail the following adjustments to achieve implementation.

4.3.1 Likelihood Integral (Outer EIS)

Application of EIS to the period- t likelihood integral (3) requires that the predictive density $f(s_t|Y_{t-1})$ be evaluated at any particle s_t^i produced by EIS iterations; thus we assume momentarily that an operational continuous approximation $\hat{f}(s_t|Y_{t-1})$ is available. For ease of notation, we use a subscript t to represent all of the information upon which period- t calculations are conditioned. Application of the EIS algorithm described above produces at convergence an EIS sampler $g(s_t|\hat{a}_t)$, with weights

$$\tilde{\omega}_t(s_t; \hat{a}_t) = \frac{f(y_t|s_t, Y_{t-1}) \cdot \hat{f}(s_t|Y_{t-1})}{g(s_t|\hat{a}_t)}. \quad (46)$$

The EIS draws $\{s_t^i\}_{i=1}^N$ and associated weights $\{\tilde{\omega}_t^i\}_{i=1}^N$ are then used to produce EIS estimates of likelihood and filtered values according to (25)-(27).

4.3.2 Predictive Integral (inner EIS)

The initial step in evaluating the predictive integral consists of replacing the lagged filtering density in (2) by its expression as given by (3) and (4):

$$f(s_t|Y_{t-1}) = \frac{\int f(s_t|s_{t-1}, Y_{t-1}) \cdot [f(y_{t-1}|s_{t-1}, Y_{t-2}) \cdot f(s_{t-1}|Y_{t-2})] ds_{t-1}}{\int f(y_{t-1}|s_{t-1}, Y_{t-2}) \cdot f(s_{t-1}|Y_{t-2}) \cdot ds_{t-1}}. \quad (47)$$

Note that this can be reinterpreted as a filtering problem from the perspective of information available in period $(t-1)$:

$$f(s_t|Y_{t-1}) = E_{s_{t-1}|Y_{t-1}}[f(s_t|s_{t-1}, Y_{t-1})], \quad (48)$$

and can be computed for arbitrary s_t using period- $(t-1)$ draws $\{s_{t-1}^i\}_{i=1}^S$ and normalized weights $\{\omega_{t-1}^i\}_{i=1}^S$ defined in (27), where it is expected that $S < N$ following the discussion

in Section 3.2. This produces an easy-to-implement functional approximation of the form

$$\widehat{f}_N(s_t|Y_{t-1}) = \sum_{i=1}^S \omega_{t-1}^i \cdot f(s_t|s_{t-1}^i, Y_{t-1}). \quad (49)$$

While this estimate bears resemblance to the particle-based estimate in (6), it is distinct in a critical respect: it relies upon a continuous density approximation $g(s_{t-1}|\widehat{a}_{t-1})$ to the period- $(t-1)$ filtering density, rather than a fixed-support mixture of Dirac measures. This property can be exploited further in order to produce more efficient estimates of the predictive density. Foremost, if the weights $\{\omega_{t-1}^i\}_{i=1}^N$ are near-constant over the region of importance, which typically obtains if $g(s_{t-1}|\widehat{a}_{t-1})$ is an efficient sampler for the period- $(t-1)$ likelihood integral, then we can replace the weights in the IS version of (47) by a constant, yielding the functional approximation

$$\widehat{f}(s_t|Y_{t-1}) = \int f(s_t|s_{t-1}, Y_{t-1}) \cdot g(s_{t-1}|\widehat{a}_{t-1}) ds_{t-1}. \quad (50)$$

This expression offers a critical advantage: if $g(s_{t-1}|\widehat{a}_{t-1})$ and $f(s_t|s_{t-1}, Y_{t-1})$ - treated as a kernel in s_{t-1} given (s_t, Y_{t-1}) - belong in whole or in part to a common class of density kernels closed under multiplication, then (partial) integration with respect to s_{t-1} is feasible analytically, thus reducing the dimensions for required numerical integration.

When these approximations prove insufficient, we can resort to ‘inner EIS’ to approximate (47), or instead, ‘joint EIS’ in (s_{t-1}, s_t) applied directly to (13).

4.3.3 Singular Transitions

State transitions often include identities that effectively reduce the dimension of integration in (47) and (50), since for any given s_t the vector s_{t-1} is then restricted to a strict subset of \mathcal{R}^m . Let s_t partition into $s_t = (p_t, q_t)$ in such a way that the transition identities can be written as

$$q_t = \varphi(p_t, s_{t-1}). \quad (51)$$

We reinterpret these identities as the limit of a uniform density for $q_t|p_t, s_{t-1}$ on the interval $[\varphi(p_t, s_{t-1}) - \epsilon, \varphi(p_t, s_{t-1}) + \epsilon]$. Under the most operational scenario, $\varphi(p_t, s_{t-1})$ is differentiable and strictly monotone in q_{t-1} , with inverse

$$q_{t-1} = \psi(s_t, p_{t-1}), \quad (52)$$

in which case we can take the limit of the integral in (50) as ϵ tends to zero, producing

$$\begin{aligned} \widehat{f}(s_t|Y_{t-1}) &= \int J(s_t, p_{t-1}) \cdot f(p_t|s_{t-1}, Y_{t-1}) \\ &\quad \cdot g(s_{t-1}; \widehat{a}_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}, \end{aligned} \quad (53)$$

where

$$J(s_t, p_{t-1}) = \left\| \frac{\partial}{\partial q_t} \psi(s_t, p_{t-1}) \right\|. \quad (54)$$

Since $g(p_{t-1}, \psi(s_t, p_{t-1}); \widehat{a}_{t-1})$ is not a sampler for $p_{t-1}|s_t$, we must evaluate (53) by a mix of analytical, quadrature and/or EIS methods. A more general weighted version of (53) would include the weight function $\widetilde{\omega}_{t-1}(s_{t-1}; \widehat{a}_{t-1})$, which would have to be interpolated from the swarm $\{s_{t-1}^i\}_{i=1}^N$ and associated weights $\{\widetilde{\omega}_{t-1}^i\}_{i=1}^N$.

While the construction of a continuous approximation for $f(s_t|Y_{t-1})$ under singular transitions may appear complicated, note that singular transitions render the particle filter particularly prone to sample impoverishment: for any value of s_t outside of the particle filter swarm $\{s_t^i\}$, the support of the s_t integral in (47) and (50) has zero measure relative to that of the filtering density $f(s_{t-1}|Y_{t-1})$.

Following is a general nested algorithm for implementing the EIS filter.

Algorithm 2: EIS Filtering, Period t

2.1. Likelihood and filtered values (outer EIS)

Requirement: Function subprogram (from Algorithm 2.2 below) for functional approximation $\widehat{f}(s_t|Y_{t-1})$.

- Target: $\phi_t(s_t) = f(y_t|s_t, Y_{t-1}) \cdot \widehat{f}(s_t|Y_{t-1})$.
- If $s_t = (s_{t1}, s_{t2})$, such that $\phi_t(s_t)$ takes the form of an analytical IS kernel for s_{t2} given s_{t1} , then

$$g(s_t|a_t) = g_2(s_{t2}|s_{t1}, a_{t2}) \cdot g_1(s_{t1}|a_{t1}), \quad (55)$$

with

$$\begin{aligned} g_2(s_{t2}|s_{t1}, a_{t2}) &= \frac{\phi_t(s_t)}{\phi_t^1(s_{t1})}, \\ \phi_t^1(s_{t1}) &= \int \phi_t(s_t) ds_{t2}, \end{aligned}$$

where a_{t2} denotes parameters induced by the transformation of the kernel $\phi_t(s_t)$ into a density for s_{t2} given s_{t1} .

- Else $s_{t1} = s_t$, $g_1(\cdot) = g(\cdot)$, $\phi_t^1(\cdot) = \phi_t(\cdot)$.
- EIS target: $\phi_t^1(s_{t1})$

– Sampler: $g_1(s_{t1}|\widehat{a}_{t1})$ (from Algorithm 1)

– Weight:

$$\widetilde{\omega}_t(s_{t1}; \widehat{a}_t) = \frac{\phi_t^1(s_{t1})}{g_1(s_{t1}|\widehat{a}_{t1})} = \frac{\phi_t(s_t)}{g(s_t|\widehat{a}_t)}.$$

- Obtain draws and weights $\left\{s_t^i, \widetilde{\omega}_t^i\right\}_{i=1}^N$, and perform EIS evaluation following (25).

2.2 Function subprogram $\widehat{f}(s_t|Y_{t-1})$ (inner EIS)

Inputs: EIS continuous sampler $g(s_{t-1}|\widehat{a}_{t-1})$, with EIS swarm $\left\{s_t^i, \widetilde{\omega}_t^i\right\}_{i=1}^N$.

- Select from the following options:
 - Normalize weights and obtain filtered value estimate using (49).
 - Obtain constant-weight estimate using (50) (possibly partial analytical integration).

- Obtain estimate using inner EIS (for any s_t required for outer EIS).
- If Inner EIS:
 - (i) Target: $\phi_t^p(s_t) \equiv f(s_t|s_{t-1}, Y_{t-1}) \cdot g(s_{t-1}|\widehat{a}_{t-1}) \cdot \widehat{\omega}_{t-1}(s_{t-1})$, with $\widehat{\omega}_{t-1}$ either constant-weight or interpolated from $\{s_t^i, \widetilde{\omega}_t^i\}_{i=1}^N$.
 - (ii) Construct EIS sampler $g_p(s_{t-1}|\widehat{b}_t)$ (from Algorithm 1).
 - (iii) Obtain S inner EIS draws and weights, and estimate $\widehat{f}(s_t|Y_{t-1})$.

Note: Under singular transitions, as in (53), the only available option is inner EIS (or when feasible, a combination of partial analytical integration and/or low-dimensional quadrature).

5 Examples

Here we present two examples that illustrate the relative performance of the bootstrap particle of Gordon, Salmond and Smith (1993); the auxiliary and adapted particle filters of Pitt and Shephard (1999); and the EIS filter. The two examples mimic examples used by Pitt and Shephard (1999) to demonstrate the value added of their auxiliary and adapted filters relative to the bootstrap filter. In both examples we replicate their experimental design to demonstrate the value added of the EIS filter relative to the bootstrap filter. In so doing, we obtain a triangular comparison of the value added of the EIS filter relative to the auxiliary filter.

We begin with two lessons gleaned through these examples regarding the selection of the two auxiliary sample sizes employed under the outer EIS filter: N , the number of draws used for outer EIS; and R , the number of draws used in the EIS optimization step. 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 several orders of magnitude. This typically 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).

5.1 Example 1: Bearings-Only Tracking

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

A ship moves in the (x, z) plane with speed following a bivariate random walk process. Let $s_t = (x_t, z_t, \dot{x}_t, \dot{z}_t)'$ denote the quadrivariate latent state variable. The discrete version of the model we consider first is characterized by the singular transition

$$s_t = \begin{pmatrix} I_2 & I_2 \\ 0 & I_2 \end{pmatrix} s_{t-1} + \sigma \begin{pmatrix} \frac{1}{2}I_2 \\ I_2 \end{pmatrix} u_t, \quad (56)$$

with $u_t \sim \text{i.i.d.} N(0, I_2)$. The initial state vector is distributed as

$$s_0 \sim N(\mu_0, \Delta_0), \quad (57)$$

with (μ_0, Δ_0) known.

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

$$f(y_t | s_t, Y_{t-1}) = f(y_t | \theta_t) = \frac{1}{2\pi} \frac{1 - r^2}{1 + r^2 - 2r \cos(y_t - \theta_t)}, \quad (58)$$

with $0 \leq (y_t, \theta_t) \leq 2\pi$ and $0 \leq r \leq 1$. As discussed in Section 4.3.3 the computation of a continuous approximation for the predictive density requires partitioning s_t conformably with (51) and (52). As we shall subsequently introduce polar coordinates, it proves more operational to elicit the Dirac identity (51) in terms of the ship's coordinates. Therefore, let

$$s_t = (p_t, q_t), \quad p'_t = (x_t, z_t), \quad q'_t = (\dot{x}_t, \dot{z}_t).$$

The transition (56) is then reinterpreted as the combination of a proper bivariate transition

$$p_t | s_{t-1} \sim N(A s_{t-1}, \Omega), \tag{59}$$

$$A = (I_2, I_2), \quad \Omega = \frac{1}{4} \sigma^2 I_2, \tag{60}$$

and a Dirac transition

$$q_t \equiv \varphi(p_t, s_{t-1}) = 2(p_t - p_{t-1}) - q_{t-1}. \tag{61}$$

Below we shall also consider a non-singular version of the model.

This model is numerically challenging on two counts. First, under parameter values typically used in the literature, the measurement density $f(y_t | \theta_t)$ - viewed as a function of θ_t given y_t - is much tighter (though with fat tails) than that of $\theta_t | Y_{t-1}$. As discussed in Section 3.3, this is a situation conducive to sample impoverishment. As we shall illustrate below, this induces substantial numerical inefficiency for the bootstrap particle filter that, while partially alleviated via implementation of the auxiliary particle filter, nevertheless remains substantial. Second, the degenerate transition creates additional numerical problems since it implies a zero-measure support in \mathcal{R}^4 for the density $f(s_t | s_{t-1}, Y_{t-1})$.

Despite these challenges, we can implement a version of the EIS filter that overcomes these pathologies. While conceptually simple, the algebra of our implementation is somewhat tedious. Next we present the broad lines of our implementation; full technical details are regrouped in the Appendix.

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

We momentarily take as given that a continuous approximation $\widehat{f}(s_t|Y_{t-1})$ can be computed for any s_t (as described below). The period- t likelihood function is then given by (3). In order to account for the fact that y_t only measures $\theta_t = \arctan(z_t/x_t)$, we reparameterize the ship location in terms of polar coordinates. Let

$$p_t = \nu(\rho_t, \theta_t) = \rho_t e(\theta_t), \quad \theta_t \in [0, 2\pi], \quad \rho_t \geq 0, \quad (62)$$

with $e(\theta_t) = (\cos \theta_t, \sin \theta_t)'$ and $\rho_t = (x_t^2 + z_t^2)^{1/2}$. To avoid confusion, the transformed state vector is denoted

$$\lambda_t = \ell(s_t) = (\theta_t, \rho_t, q_t') = (\theta_t, \delta_t'). \quad (63)$$

In line with (37), the EIS sampler $g(\lambda_t|a_t)$ we implement is the product of a trivariate Gaussian sampler for $\delta_t|\theta_t$ - subject to the positivity constraint $\rho_t \geq 0$ - and a univariate piecewise loglinear density for θ_t .

Algorithm 3: Construction of $g(\lambda_t|a_t)$

- Propagation: The period- $(t-1)$ EIS swarm $\{\lambda_{t-1}^i\}_{i=1}^N$ is transformed into a swarm $\{s_{t-1}^i\}_{i=1}^N$ by means of the inverse transformation $s = \ell^{-1}(\lambda)$. An initial period- t swarm $\{s_t^i\}_{i=1}^N$ obtains by drawing s_t^i from the singular transition density $f(s_t|s_{t-1}^i)$ associated with (56).
- Initial EIS: Since the measurement density only depends on θ_t , the EIS kernel for $\delta_t|\theta_t$ needs only to approximate the filtering density. Whence our first step is that of constructing an EIS Gaussian kernel in s_t from which we shall extract the EIS kernel for $\delta_t|\theta_t$ through reparametrization and normalization. We construct an auxiliary quadri-variate Gaussian kernel $k_s(s_t; a_t)$ by regressing $\{\ln f(s_t^i|Y_{t-1})\}_{i=1}^N$ on $\{s_t^i\}_{i=1}^N$ and the lower triangle of $\{s_t^i \cdot s_t^{i'}\}_{i=1}^N$, for a total of 14 regressors plus one intercept.
- Polar coordinates: We introduce the transformation from s_t to $\lambda_t = \ell(s_t)$, with Jaco-

bian $\rho_t > 0$. Let

$$k_\lambda(\lambda_t; a_t^s) = \rho_t \cdot k_s(\ell^{-1}(\lambda_t); a_t^s). \quad (64)$$

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

$$g_\delta(\delta_t|\theta_t, a_t^s) = \frac{k_\lambda(\lambda_t; a_t^s)}{\chi_\lambda(\theta_t; a_t^s)}, \quad (65)$$

with

$$\chi_\lambda(\theta_t; a_t^s) = \int_{\Delta} k_\lambda(\lambda_t; a_t^s) d\delta_t, \quad (66)$$

where $\Delta = \mathcal{R}^2 \times \mathcal{R}_+$. The likelihood integral in (3) is rewritten as

$$f(y_t|Y_{t-1}) = \int [f(y_t|\theta_t) \cdot \chi_\lambda(\theta_t; a_t^s)] \cdot \frac{f(\ell^{-1}(\lambda_t)|Y_{t-1})}{k_\lambda(\lambda_t; a_t^s)} \cdot g_\delta(\delta_t|\theta_t; a_t^s) d\delta_t d\theta_t. \quad (67)$$

- Piecewise loglinear EIS for θ_t : We apply the algorithm introduced in Section 4.2 in order to construct a piecewise loglinear approximation $g(\theta_t|a_t^\theta)$ for the product $f(y_t|\theta_t) \cdot \chi_\lambda(\theta_t; a_t^s)$. The likelihood integral is then rewritten as

$$f(y_t|Y_{t-1}) = \int \omega_t(\lambda_t; a_t) \cdot g_\delta(\delta_t|\theta_t; a_t^s) \cdot g_\theta(\theta_t|a_t^\theta) d\lambda_t, \quad (68)$$

with

$$\omega_t(\lambda_t; a_t) = \left[\frac{f(y_t|\theta_t) \cdot \chi_\lambda(\theta_t; a_t^s)}{g_\theta(\theta_t|a_t^\theta)} \right] \cdot \frac{f(h^{-1}(\lambda_t)|Y_{t-1})}{k_{t\lambda}(\lambda_t; a_t^s)}, \quad (69)$$

and $a_t = (a_t^s, a_t^\theta)$.

- EIS filtered values and period-t swarm: Let $\{\lambda_t^i\}_{i=1}^N$ denote a swarm of N i.i.d. draws

from the EIS sampler $g_\delta(\delta_t|\theta_t; a_t^s) \cdot g_\theta(\theta_t|a_t^\theta)$. Filtered values obtain as:

$$\bar{h}_N = \frac{\sum_{i=1}^N h(\ell^{-1}(\lambda_t^i)) \cdot \omega_t^i}{\sum_{i=1}^N \omega_t^i}, \quad (70)$$

where $\omega_t^i = \omega_t(\lambda_t^i; a_t)$. The swarm $\{\lambda_t^i\}_{i=1}^N$ is then passed to period- $(t+1)$.

Detailed derivations for this algorithm are given in the Appendix.

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 $\omega_t(\lambda_t; a_t)$ as given in (69). 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.

5.1.2 Computation of $\hat{f}(s_t|Y_{t-1})$

Since we can expect an excellent fit between period- $(t-1)$ filtering density $f(s_{t-1}|Y_{t-1})$ and its EIS approximation $g(s_{t-1}|a_{t-1})$, we can approximate $f(s_t|Y_{t-1})$ according to (53):

$$\hat{f}(s_t|Y_{t-1}) = \int f(p_t|s_{t-1}, Y_{t-1}) \cdot g_s(s_{t-1}|a_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}. \quad (71)$$

Here, $q_{t-1} = \psi(s_t, p_{t-1})$ denotes the inverse of Dirac transition (61) with unit Jacobian, and $g_s(s_{t-1}|a_{t-1})$ obtains from the period- $(t-1)$ EIS sampler by the transformation $s_{t-1} = \ell^{-1}(\lambda_{t-1})$ with Jacobian ρ_t^{-1} , together with (64) and (66):

$$g_s(s_{t-1}|a_{t-1}) = k_s(s_{t-1}|a_{t-1}^s) \cdot \frac{g_\theta(\theta_{t-1}|a_{t-1}^\theta)}{\chi_\lambda(\theta_{t-1}; a_{t-1}^s)}|_{\theta_{t-1}=\arctan(z_{t-1}/x_{t-1})}. \quad (72)$$

The evaluation of the predictive integral is now conceptually straightforward (details are regrouped in the Appendix).

Algorithm 4: Evaluation of $\hat{f}(s_t|Y_{t-1})$

- Since $f(p_t|s_{t-1}, Y_{t-1})$ and $k_s(s_{t-1}; a_{t-1}^s)$ are Gaussian kernels in (p_t, s_{t-1}) , and the transformation $q_{t-1} = \psi(s_t, p_{t-1})$ is linear, compute the following Gaussian kernel in (s_t, p_{t-1}) :

$$\kappa_{t-1}(s_t, p_{t-1}; a_{t-1}^s) = f(p_t|s_{t-1}, Y_{t-1}) \cdot k_s(s_{t-1}; a_{t-1}^s) |_{q_{t-1}=\psi(s_t, p_{t-1})}. \quad (73)$$

- Introduce the transformation $p_{t-1} = \nu(\rho_{t-1}, \theta_{t-1})$ in (62), with Jacobian $\rho_{t-1} > 0$. The predictive density is then approximated by

$$\hat{f}(s_t|Y_{t-1}) = \int \rho_{t-1} \cdot \kappa_{t-1}^*(s_t, \rho_{t-1}, \theta_{t-1}; a_{t-1}^s) \cdot \frac{g_\theta(\theta_{t-1}|a_{t-1}^\theta)}{\chi_\lambda(\theta_{t-1}; a_{t-1}^s)} d\rho_{t-1} d\theta_{t-1},$$

where

$$\kappa_{t-1}^*(s_t, \rho_{t-1}, \theta_{t-1}; a_{t-1}^s) = \kappa_{t-1}(s_t, p_{t-1}; a_{t-1}^s) |_{p_{t-1}=\nu(\rho_{t-1}, \theta_{t-1})}. \quad (74)$$

- Given (s_t, θ_{t-1}) , integrate analytically in $\rho_{t-1} > 0$ to obtain

$$\hat{f}(s_t|Y_{t-1}) = \int_0^{2\pi} d(s_t, \theta_{t-1}; a_{t-1}^s) \cdot \frac{g_\theta(\theta_{t-1}|a_{t-1}^\theta)}{\chi_\lambda(\theta_{t-1}; a_{t-1}^s)} d\theta_{t-1},$$

where

$$d(s_t, \theta_{t-1}; a_{t-1}^s) = \int_0^\infty \rho_{t-1} \cdot \kappa_{t-1}^*(s_t, \rho_{t-1}, \theta_{t-1}; a_{t-1}^s) d\rho_{t-1}. \quad (75)$$

- Given s_t , use $g_\theta(\theta_{t-1}|a_{t-1}^\theta)$ as the auxiliary IS sampler. The corresponding estimate of $\hat{f}(s_t|Y_{t-1})$ is given by

$$\hat{f}_N(s_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N \frac{d(s_t, \theta_{t-1}^i; a_{t-1}^s)}{\chi_\lambda(\theta_{t-1}^i; a_{t-1}^s)}, \quad (76)$$

where $\{\theta_{t-1}^i\}_{i=1}^N$ denotes period- $(t-1)$ EIS draws of θ_{t-1} .

5.1.3 Non-Singular Version

The singularity of the transition in (56) is a 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 (56) then must be transformed into a transition density for $s_t|s_{t-D}$ by implicit marginalization with respect to the state sequence $\{s_{t-j}\}_{j=1}^{D-1}$. The random walk process for speed is given by

$$q_t = q_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2 I_2), \quad (77)$$

and position is discretized as

$$p_t = p_{t-1} + \frac{1}{2}(q_t + q_{t-1}). \quad (78)$$

It follows that

$$q_t = q_{t-D} + u_{t,D}, \quad (79)$$

$$p_t = p_{t-D} + Dq_{t-D} + v_{t,D}, \quad (80)$$

with

$$u_{t,D} = \sum_{j=0}^{D-1} \varepsilon_{t-j}, \quad v_{t,D} = \frac{1}{2} \sum_{j=0}^{D-1} (2j+1) \varepsilon_{t-j}. \quad (81)$$

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. It follows that the transition density from

s_{t-D} to s_t is given by

$$s_t | s_{t-D} \sim N(A_D s_{t-D}, \sigma^2 V_D), \quad (82)$$

with

$$A_D = \begin{pmatrix} I_2 & DI_2 \\ 0 & I_2 \end{pmatrix}, \quad V_D = D \begin{pmatrix} \frac{4D^2-1}{12} I_2 & \frac{D}{2} I_2 \\ \frac{D}{2} I_2 & I_2 \end{pmatrix}. \quad (83)$$

The case $D = 1$ obviously coincides with the degenerate transition in (56). The generalization from (56) to (82) does not affect EIS evaluation of the likelihood function, except that $t - 1$ is replaced by $t - D$. The evaluation of $\hat{f}(s_t | Y_{t-D})$ proceeds as described in Algorithm 4 except that: (i) the inverse transformation in step 1 is omitted; and (ii) the (analytical) integral in step 3 is trivariate in (q_{t-D}, ρ_{t-D}) . Notice that we need not restrict D to integer values. We can apply (77)-(83) for real values of $D > 1$. This offers the advantage that the singular case can be reinterpreted as the limit for $D \rightarrow 1_+$ of the non-singular case (allowing for numerical verification of the validity of our Dirac treatment for the former).

5.1.4 Application

As noted, we demonstrate our methodology in an application designed along the lines of that Pitt and Shephard (1999). For the singular and non-singular cases, σ in (56) and (82) is set to 0.001; and r in (58) is set to $1 - (0.005)^2$. The initial latent vector s_0 is distributed as Gaussian 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 $\{2, 3, \dots, 11\}$ with equal probabilities $p_i = 0.1$, $i = 2, \dots, 11$. (As noted, 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 $\{s_t^c\}_{t=1}^{10}$, one for the singular case ($c = 1$) and one for the non-singular case ($c = 2$). Both sets are linear transformations of a single set of $N(0, 1)$ draws. We draw $I = 40$ different data sets $\{Y_T^{c,i}\}_{i=1}^{40}$ based on the

latent vectors $\{s_t^c\}_{t=1}^{10}$ for $c = 1, 2$. 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, where the acronyms indicate respectively the EIS and bootstrap particle filters, and the numbers following acronyms indicate the number of draws N used to implement the filters (K indicating thousands); computing times associated with these procedures are similar.

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) to demonstrate the performance of the auxiliary particle filter. Let \bar{s}_t^i , ($i : 1 \rightarrow 40$, $t : 1 \rightarrow 10$) denote ‘true’ filtered means for the states. 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. The latter number turns out to be needed in order to eliminate significant biases characterizing PF estimates of filtered means (illustrated below). We ran this experiment to verify that ‘true’ values produced by both EIS and PF estimators are numerically identical.

MSE comparisons are constructed as follows. Let $\tilde{s}_{t,k}^{i,j}$ denote the MC estimate of the filtered mean, for data set i , for replication j , in period t , for procedure $k = \{\text{EIS-1K, PF-40K}\}$. The log mean squared error (LMSE) for procedure k , in period t is obtained as

$$\text{LMSE}_{t,k} = \ln \left\{ \frac{1}{40} \sum_{i=1}^{40} \left[\frac{1}{100} \sum_{j=1}^{100} (\tilde{s}_{t,k}^{i,j} - \bar{s}_t^i)^2 \right] \right\}. \quad (84)$$

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.

Figure 1 depicts LMSEs for the two procedures applied to the singular case against time. As expected, the move from estimates obtained using the bootstrap particle filter to those obtained using the EIS filter leads to a large reduction in LMSEs: differences average between 4 and 6 on the log scale. These differences are far larger than those reported by Pitt and Shephard (1999): their auxiliary particle filter yielded reductions averaging between 0.5 and 1 relative to the bootstrap 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-40K. Logged variances and logged MSE/variance ratios are plotted for both procedures in Figure 2. 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 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 bootstrap 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.

The results obtained for the non-singular case of the bearings-only tracking are, as expected, very similar to those obtained for the singular case. Figure 3 plots the LMSEs and Figure 4 the logged MC variances and the logged MSE/variance ratio for the non-singular case.

We conclude this application with a discussion of diagnostics. Since the EIS sampler in this case is not obtained by application of the least-squares algorithm described in Section 4.1, we do not have goodness-of-fit measures such as R^2 statistics to report. Instead we explore

whether filtered estimates obtained using the bootstrap particle filter converge to the ‘true’ filtered values \bar{s}_t^i constructed using the EIS filter as N increases under the bootstrap particle filter. Persistent differences would raise doubts regarding the validity of the EIS filter in this application, since convergence under the bootstrap particle filter is well-documented in the literature. The comparison was conducted for a small set of artificial data sets generated as described above. For all time periods under each data set, convergence was found to obtain, but slowly: much slower than would be expected under \sqrt{N} convergence. The finding of convergence confirms the validity of the estimates generated by the EIS filter; the slow speed of convergence underscores the severity of the sample impoverishment problem suffered by the bootstrap particle filter in this application.

5.2 Stochastic Volatility

The baseline version of the Stochastic Volatility (SV) model is characterized by the distributions

$$y_t | s_t, Y_{t-1} \sim N_1(0, \beta^2 \exp s_t) \quad (85)$$

$$s_t | s_{t-1}, Y_{t-1} \sim N_1(r s_{t-1}, \sigma^2), \quad (86)$$

where β , r and σ^2 represent median volatility, volatility persistence and variance of volatility shocks, respectively. This model was introduced by Taylor (1982, 1986) to account for the time-varying and persistent volatility exhibited by high-frequency 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; 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. Persistence is typically very high - $r \in [0.95, 1.00)$ - implying that observed returns are highly informative on volatilities, again a situation prone to sample impoverishment. Nevertheless, the SV

model turns out to be exceptionally well-behaved for EIS implementation and computation of the filtered values $E[\exp s_t | Y_t]$. The EIS auxiliary regression under a Gaussian EIS kernel obtains from (36) together with

$$\phi(s_t, Y_t) = f(y_t | s_t, Y_{t-1}) \cdot f(s_t | s_{t-1}, Y_{t-1}) \quad (87)$$

$$-2 \ln k(s_t; a_t) = \alpha_t s_t^2 - 2\beta_t s_t, \quad (88)$$

where $a'_t = (\alpha_t, \beta_t)$. The corresponding EIS Gaussian sampler has mean $\mu_t = \beta_t / \alpha_t$ and variance $\sigma_t^2 = \alpha_t^{-1}$. The filtered volatility values are then given by (26) and (27), with

$$\tilde{\omega}_t^i = \frac{\phi(s_t^i; Y_t)}{g(s_t^i | \hat{a}_t)}. \quad (89)$$

This EIS Gaussian kernel turns out to provide a close (global) fit to $\phi(s_t; Y_t)$, thus we can employ the approximation (50) to compute the predictive density $\hat{f}(s_t | Y_{t-1})$. This offers the key advantage that the integrand becomes a product of two Gaussian kernels in s_{t-1} , and can be integrated analytically.

Notice that $h(s_t) = \exp s_t > 0$. This enables us to implement an additional refinement to EIS filtering. Specifically, note that if $g(s_t | \hat{a}_t)$ is an EIS approximation for $\phi(s_t; Y_t)$, then $g(s_t | \hat{a}_t^n)$, with $\hat{a}_t^n = (\hat{\alpha}_t, \hat{\beta}_t + 1)$, is an EIS approximation for the product $\phi(s_t; Y_t) \cdot \exp s_t$. Whence a numerically more efficient estimate of filtered values obtains as

$$\hat{h}_N = \frac{\sum_{i=1}^N \tilde{\omega}_t^i}{\sum_{i=1}^N \hat{\omega}_t^i}, \quad \hat{\omega}_t^i = \frac{\phi(\hat{s}_t^i; Y_t) \cdot \exp \hat{s}_t^i}{g(\hat{s}_t^i | \hat{a}_t^n)}, \quad (90)$$

where $\{\hat{s}_t^i\}_{i=1}^N$ denotes N i.i.d. draws from $g(s_t | \hat{a}_t^n)$.

5.2.1 Application

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

The model is parameterized as $\phi = 0.9702$, $\sigma_v = 0.178$, 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 (computing times are similar given these settings). For each procedure, the 100 MC estimates are obtained using 100 different CRNs. As in the tracking 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 bearings-only tracking application.)

Figure 5 (bottom panel) depicts LMSEs for the 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 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 again 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 5 (top panels). Figure 5 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.

As with the bearings-only application, we conclude with a diagnostic exercise. Since in this case the EIS sampler was obtained via application of the least-squares algorithm outlined in Section 4.1, R^2 measures are available for assessing goodness-of-fit. Using selected artificial samples generated as described above, and multiple sets of filtered estimates obtained using alternative sets of random numbers, period-by-period R^2 measures were found to exceed 0.999 in all cases, indicating excellent global fit. Moreover, differences in filtered estimates obtained across the EIS and bootstrap particle filters were found once again to disappear as N increases under the bootstrap particle filter, this time more rapidly than in the bearings-only tracking application.

6 Conclusion

We have proposed an efficient means of facilitating filtering in applications involving non-linear and/or non-Gaussian state-space representations: the EIS filter. The filter is adapted unconditionally using an optimization procedure designed to minimize numerical standard errors associated with targeted integrals. Implementation of the filter is conceptually straightforward, and the payoff of can be substantial. To date, the selection of auxiliary

EIS samplers is restricted primarily to the exponential family of distributions. Extensions to more flexible classes of samplers (e.g. mixtures of Gaussian samplers) relying upon Gauss-Newton iterations are currently under development.

7 Appendix

Herein we provide algebraic details regarding implementation of the EIS filter in the bearings-only tracking problem.

7.1 Singular Case

A. Derivation of $\chi_\lambda(\theta_t; a_t^s)$ in (66). For ease of notation we delete t subscripts. The kernel $k^\lambda(\lambda; a^s)$ defined in (64) is given by

$$\gamma(\lambda; a^s) = -2 \ln \left[\frac{1}{\rho} k^\lambda(\lambda; a^s) \right] = \begin{pmatrix} \rho e(\theta) \\ q \end{pmatrix}' H \begin{pmatrix} \rho e(\theta) \\ q \end{pmatrix} - 2 \begin{pmatrix} \rho e(\theta) \\ q \end{pmatrix}' m, \quad (91)$$

with a^s consisting of m and the Cholesky decomposition of H . We partition H and m conformably with $(\rho e'(\theta) \quad q')$ into

$$P = (P_{ij}), \quad m = (m_i), \quad i, j = 1, 2. \quad (92)$$

Successive square completions in q and ρ produce the following expression for $\gamma(\lambda; a^s)$:

$$\gamma(\lambda; a^s) = (q - q_\theta)' H_{22} (q - q_\theta) + a_\theta (\rho - r_\theta)^2 - s_\theta^2 \quad (93)$$

$$q_\theta = H_{22}^{-1} [m_2 - \rho H_{21} e(\theta)], \quad a_\theta = e'(\theta) H_{11.2} e(\theta) \quad (94)$$

$$r_\theta = \frac{1}{a_\theta} (m_1 - \Delta_{12} m_2)' e(\theta), \quad s_\theta^2 = a_\theta r_\theta^2 + m_2' H_{22}^{-1} m_2 \quad (95)$$

$$H_{11.2} = H_{11} - H_{12} H_{22}^{-1} H_{21}, \quad \Delta_{12} = H_{12} H_{22}^{-1}. \quad (96)$$

It follows that $\chi_\lambda(\theta; a^s)$, as defined in (66), is given by

$$\chi_\lambda(\theta; a^s) = 2\pi |H_{22}|^{-\frac{1}{2}} d_\theta \exp\left(\frac{1}{2} s_\theta^2\right), \quad (97)$$

with

$$d_\theta = \int_0^\infty \rho \exp\left(-\frac{1}{2} a_\theta (\rho - r_\theta)^2\right) d\rho. \quad (98)$$

We introduce the transformation of variables

$$z = \sqrt{a_\theta} \cdot (\rho - r_\theta), \quad (99)$$

and assuming that $r_\theta > 0$ (which is the case for the application considered here), rewrite d_θ as

$$\begin{aligned} d_\theta &= \frac{1}{a_\theta} \cdot \int_{-c_\theta}^\infty (z + c_\theta) \exp\left(-\frac{1}{2} z^2\right) dz \\ &= \frac{1}{a_\theta} \left[\exp\left(-\frac{1}{2} c_\theta^2\right) + c_\theta \sqrt{\frac{\pi}{2}} \left(1 + \operatorname{erf}\left(\frac{c_\theta}{\sqrt{2}}\right)\right) \right], \end{aligned} \quad (100)$$

with $c_\theta = r_\theta \sqrt{a_\theta}$, and

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du. \quad (101)$$

The properties of $\operatorname{erf}(x)$ are discussed, e.g. in Abramowitz and Segun (1968, ch.7).

B. CRN-EIS Draws of (θ, ρ, q)

Let (u_1, u_2) denote two $U(0, 1)$ draws, and (u_3, u_4) two $N(0, 1)$ draws. The corresponding EIS draw of (θ, ρ, q) obtains through the following sequence of transformations:

- θ obtains from u_1 by inversion of the c.d.f. associated with the piecewise loglinear sampler for θ , as given in (40) to (42).

- $\rho|\theta$ obtains from u_2 by inversion of the c.d.f. associated with the density

$$f(\rho|\theta) = \frac{1}{d_\theta} \rho \exp\left(-\frac{1}{2}a_\theta(\rho - r_\theta)^2\right) \quad \rho > 0. \quad (102)$$

Details of this inversion are provided below.

- $q|\rho, \theta$ obtains from the transformation

$$q = q_\theta + L \begin{pmatrix} u_3 \\ u_4 \end{pmatrix}, \quad (103)$$

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

C. Inversion of the c.d.f. of $\rho|\theta$

$\rho|\theta$ obtains by inversion of the transformation (99)

$$\rho = r_\theta + \frac{z}{\sqrt{a_\theta}}, \quad (104)$$

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

$$f(z|\theta) = \frac{1}{d_\theta a_\theta} (z + c_\theta) \exp\left(-\frac{1}{2}z^2\right), \quad z > -c_\theta, \quad (105)$$

with c.d.f.

$$F(z|\theta) = \frac{1}{d_\theta a_\theta} \left\{ \left[\exp\left(-\frac{1}{2}c_\theta^2\right) - \exp\left(-\frac{1}{2}z^2\right) \right] + c_\theta \cdot \sqrt{\frac{\pi}{2}} \left[\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) + \operatorname{erf}\left(\frac{c_\theta}{\sqrt{2}}\right) \right] \right\}, \quad (106)$$

accounting for the fact that $\operatorname{erf}(-x) = -\operatorname{erf}(x)$. For the application described in Section 5.1.4, c_θ turns out to be significantly larger than zero, so that z is nearly $N(0, 1)$. Thus for inversion of the CRN $u_2 \sim U(0, 1)$, we take as starting value the corresponding inverse

Gaussian draw, say $z^{(0)} \sim N(0, 1)$ and iterate once or twice by Newton:

$$z^{(k+1)} = z^{(k)} - \frac{F(z^{(k)}|\theta) - u_2}{F'(z^{(k)}|\theta)}. \quad (107)$$

D. Derivation of $\hat{f}(s_t|Y_{t-1})$ in (71)

We again suppress t subscripts for ease of notation; accordingly the index $t-1$ is replaced by -1 . We also delete the argument Y_{t-1} and follow closely the notation and successive steps outlined in Algorithm 4. First,

$$-2 \ln \left[\frac{2\pi}{|\Omega|^{\frac{1}{2}}} \cdot f(p|s_{-1}) \cdot k_s(s_{-1}; a_{-1}^s) \right] = (p - As_{-1})' \Omega^{-1} (p - As_{-1}) + s_{-1}' H_{-1} s_{-1} - 2s_{-1}' m_{-1}, \quad (108)$$

with

$$A = \begin{pmatrix} I_2 & I_2 \\ 0 & I_2 \end{pmatrix}, \quad \Omega = \sigma^2 \begin{pmatrix} \frac{1}{4}I_2 & \frac{1}{2}I_2 \\ \frac{1}{2}I_2 & I_2 \end{pmatrix}. \quad (109)$$

Next, we introduce the inverse transformation

$$q_{-1} = \psi(s, p_{-1}) = 2(p - p_{-1}) - q, \quad (110)$$

whereby

$$s|_{q_{-1}=\psi(s, p_{-1})} = C \begin{pmatrix} p_{-1} \\ s \end{pmatrix}, \quad p - As_{-1}|_{q_{-1}=\psi(s, p_{-1})} = D \begin{pmatrix} p_{-1} \\ s \end{pmatrix}, \quad (111)$$

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

$$C = \begin{pmatrix} I_2 & 0 & 0 \\ -2I_2 & 2I_2 & -I_2 \end{pmatrix}, \quad D = \begin{pmatrix} I_2 & -I_2 & I_2 \end{pmatrix}. \quad (112)$$

Thus, the kernel κ_{-1} introduced in (73) is given by

$$-2 \ln \left[\frac{2\pi}{|\Omega|^{\frac{1}{2}}} \kappa_{-1}(s, p_{-1}; a_{-1}^s) \right] = \begin{pmatrix} p_{-1} \\ s \end{pmatrix}' H_* \begin{pmatrix} p_{-1} \\ s \end{pmatrix} - 2 \begin{pmatrix} p_{-1} \\ s \end{pmatrix}' m_*, \quad (113)$$

with

$$H_* = C' H_{-1} C + D' \Omega^{-1} D, \quad m_* = C' m_{-1}.$$

Next, we introduce the transformation from p_{-1} to (ρ_{-1}, θ_{-1}) . In view of the close functional similarity between (91) and (113), the subsequent derivations parallel those from equations (91) to (101). We partition H_* and m_* conformably with (p_{-1}, s) :

$$H_* = (H_{ij}^*), \quad m_* = (m_i^*), \quad i, j = 1, 2 \quad (114)$$

The kernel κ_{-1}^* introduced in (74) is then given by

$$-2 \ln \left[\frac{2\pi}{|\Omega|^{\frac{1}{2}}} \kappa_{-1}^*(s, \rho_{-1}, \theta_{-1}; a_{-1}^s) \right] = s' H_{22}^* s - 2s' m_2^* + a_\theta^* (\rho_{-1} - r_{\theta s}^*)^2 - c_{\theta s}^{*2}, \quad (115)$$

with

$$a_\theta^* = e'(\theta_{-1}) \cdot H_{11}^* \cdot e(\theta_{-1}), \quad r_{\theta s}^* = \frac{1}{a_\theta^*} (m_1^* - H_{12}^* s)' e(\theta_{-1}), \quad c_{\theta s}^* = r_{\theta s}^* \cdot \sqrt{a_\theta^*}. \quad (116)$$

Integration of κ_{-1}^* with regard to ρ_{-1} produces the following expression for $d(\cdot)$ in (75):

$$d(s, \theta_{-1}; a_{-1}^s) = \frac{|\Omega|^{-\frac{1}{2}}}{2\pi} \cdot \exp \left[-\frac{1}{2} (s' H_{22}^* s - 2s' m_2^*) \right] \cdot \left[d_{\theta s}^* \exp \left(\frac{1}{2} c_{\theta s}^{*2} \right) \right], \quad (117)$$

where $d_{\theta s}^*$ obtains from (100) by substituting $(a_\theta^*, c_{\theta s}^*)$ for (a_θ, c_θ) . The approximation $\widehat{f}(s_t | Y_{t-1})$ is then given by (76).

7.2 Non-Singular Case

The computation of $\chi_\lambda(\theta; a^s)$ and the CRN draws of (θ, ρ, q) are the same as for the singular case. The derivation of $\hat{f}(s|Y_{t-D})$ under the non-singular transition defined by (82) is straightforward, since it no longer requires an inverse Dirac transformation. The quadratic form in (115) is replaced by

$$(s - A_D s_{-D})' V_D^{-1} (s - A_D s_{-D}) + s'_{-D} H_{-D} s_{-D} - 2s'_{-D} m_{-D} = s'_{-D} H_D^* s_{-D} - 2s'_{-D} m_D^* + s' V_D^{-1} s, \quad (118)$$

with

$$H_D^* = H_{-D} + A'_D V_D^{-1} A_D, \quad m_D^* = m_{-D} + A'_D V_D^{-1} s. \quad (119)$$

This expression is functionally similar to the quadratic form in (91). Integration with regard to s_{-D} proceeds as described by equations (93)-(101), with q replaced by s_{-D} . The functional $d(s, \theta_{-D}; a^s_{-D})$ defined as in (75) for D_{-1} is then given by

$$d(s, \theta_{-D}; a^s_{-D}) = \frac{|V_D|^{-\frac{1}{2}}}{2\pi} \exp \left[-\frac{1}{2} s' V_D^{-1} s \right] \cdot d_{\theta_s}^D \exp \frac{1}{2} c_{\theta_s}^D,$$

where $(d_{\theta_s}^D, c_{\theta_s}^D)$ are defined as (d_θ, s_θ) in (95) and (100), with (H, m) replaced by (H_D^*, m_D^*) . $\hat{f}(s|Y_{t-D})$ is then given by (76) with the lag adjustment described above.

References

- [1] Abramowitz, M. and I.A. Stegun, 1968, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, New York: Dover Publications.
- [2] Akashi, H., and H. Kumamoto, 1977, *Random sampling approach to state estimation in switching environment*, Automatica, 13, 429-434.

- [3] Andrieu, C. and A. Doucet, 2002, *Particle filtering for partially observed Gaussian state space models*, Journal of the Royal Statistical Society (Series B),64, 827-836.
- [4] Blake, A. and M. Isard, 1998, *Active Contours*, Springer.
- [5] Bobrovsky, B.Z. and M. Zakai, 1975, *A Lower Bound on the Estimation Error for Markov Processes*, IEEE Transactions on Automatic Control, 20, 785-788.
- [6] Cappé, O., S.J. Godsill and E. Moulines, 2007, *An overview of existing methods and recent advances in sequential Monte Carlo*, Proceedings of the IEEE, 95, 899-924.
- [7] 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.
- [8] Chen, R. and J. S. Liu, 2000, *Mixture Kalman filter*, Journal of the Royal Statistical Society (Series B), 62, 493-508.
- [9] Del Moral, P., 1996, *Nonlinear filtering: interacting particle solution*, Markov Processes and Related Fields, 2, 555-579.
- [10] Devroye, L., 1986, *Non-Uniform Random Variate Generation*, New York: Springer.
- [11] Doucet, A., 1998, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Cambridge University, Dep. of Engineering, Cambridge, U.K.
- [12] Doucet, A., N. de Freitas and N. Gordon, 2001, *Sequential Monte Carlo Methods in Practice*, New York: Springer.
- [13] Doucet, A., S. Godsill, and C. Andrieu, 2000, *On sequential Monte-Carlo sampling methods for Bayesian filtering*, Statistics and Computing, 10, 197-208.
- [14] Frühwirth-Schnatter, S., 2006, *Finite Mixture and Markov Switching Models*, New York: Springer.

- [15] Gordon, N.J., D.J. Salmond and A.F.M. Smith, 1993, *A novel approach to non-linear and non-Gaussian Bayesian state estimation*, IEEE Proceedings F. 140, 107-113.
- [16] Handschin, J. and D. Mayne, 1969, *Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering*, International Journal of Control, 9, 547-559.
- [17] Handschin, J., 1970, *Monte Carlo techniques for prediction and filtering of non-linear stochastic processes*, Automatica, 6, 555-563.
- [18] Ito, K. and K. Xiong, 2000, *Gaussian filters for nonlinear filtering problems*, IEEE Transactions on Automatic Control, 45, 910-927.
- [19] Jacquier, E., N.G. Polson and P.E. Rossi, 1994, *Bayesian analysis of stochastic volatility models (with discussion)*, Journal of Business and Economic Statistics, 12, 371-389.
- [20] Jazwinski, A., 1970, *Stochastic Processes and Filtering Theory*, New York: Academic Press.
- [21] Kalman, R.E., 1960, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering, 82, 35-45.
- [22] Kim, S., N. Shephard and S. Chib, 1998, *Stochastic volatility: Likelihood inference and comparison with ARCH models*, Review of Economic Studies, 65, 361-393.
- [23] Kitagawa, G., 1996, *Monte Carlo Filter and Smoother for Non-Gaussian Non-Linear State-Space Models*, Journal of Computational and Graphical Statistics, 5, 1-25.
- [24] Lehmann, E.L., 1986, *Testing Statistical Hypotheses*, John Wiley & Sons.
- [25] Liesenfeld, R. and J.F. Richard, 2003, *Univariate and multivariate stochastic volatility models: estimation and diagnostics*, Journal of Empirical Finance, 10, 505-531.

- [26] Liu, J. and R. Chen, 1995, *Blind deconvolution via sequential imputations*, Journal of the Royal Statistical Society (Series B), 430, 567-576.
- [27] Pitt, M.K. and N. Shephard, 1999, *Filtering via simulation: auxiliary particle filters*, Journal of the American Statistical Association, 94, 590-599.
- [28] Richard, J.-F. and W. Zhang, 2007, *Efficient high-dimensional Monte Carlo importance sampling*, Journal of Econometrics, 141, 1385-1411.
- [29] Ristic, B., S. Arulampalam and N. Gordon, 2004, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Boston: Artech House Publishers.
- [30] Rubin, D.B., 1987, *A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm (discussion of Tanner and Wong)*, Journal of the American Statistical Association, 82, 543-546.
- [31] Smith, J.Q. and A.A.F. Santos, 2006, *Second-Order Filter Distribution Approximations for Financial Time Series With Extreme Outliers*, Journal of Business and Economic Statistics, 24, 329-337.
- [32] Taylor, S.J., 1982, *Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices*, in Time Series Analysis: Theory and Practice 1, O.D. Anderson, Eds., Amsterdam: North-Holland.
- [33] Taylor, S.J., 1986, *Modelling Financial Time Series*, Chichester: John Wiley and Sons.
- [34] Van der Merwe, R., A. Doucet, N. De Freitas and E. Wan, 2000, *The unscented particle filter*, in Advances in Neural Information Processing Systems, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. MIT Press.

- [35] Vaswani, N., 2008, *Particle filtering for large-dimensional state spaces with multimodal observation likelihoods*, IEEE Transactions on Signal Processing, 56, 4583-4597.
- [36] Zaritskii, V., V. Svetnik and L. Shimelevich, 1975, *Monte-Carlo techniques in problems of optimal data processing*, Automation Remote Control, 12, 2015-2022.

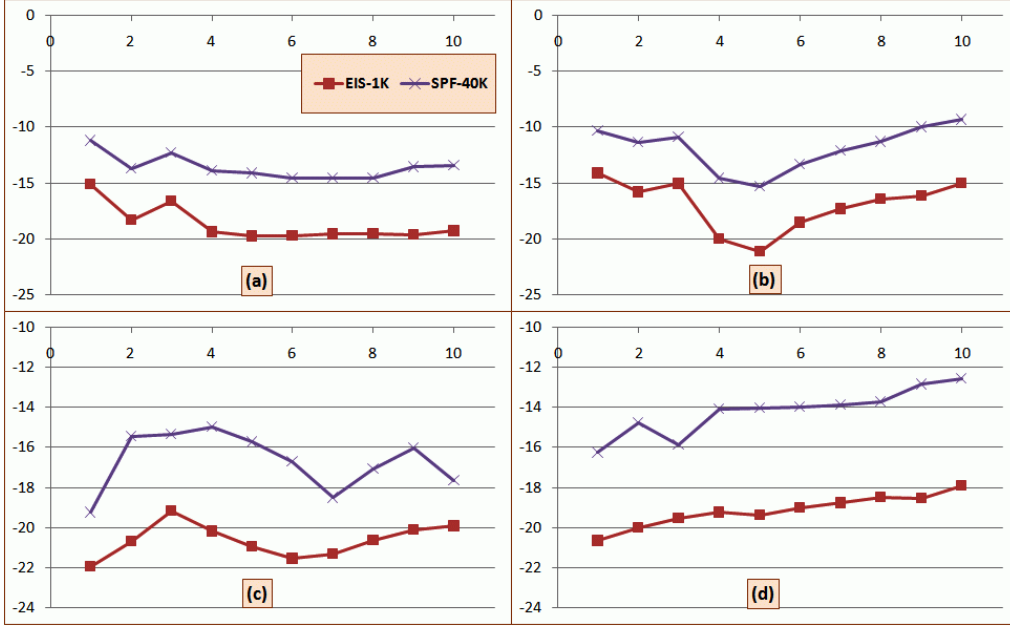


Figure 1: Log-Avg. MSE Comparisons, Singular Case. Panel Panel (a) $\rightarrow x$, (b) $\rightarrow z$, (b) $\rightarrow \dot{x}$, (b) $\rightarrow \dot{z}$

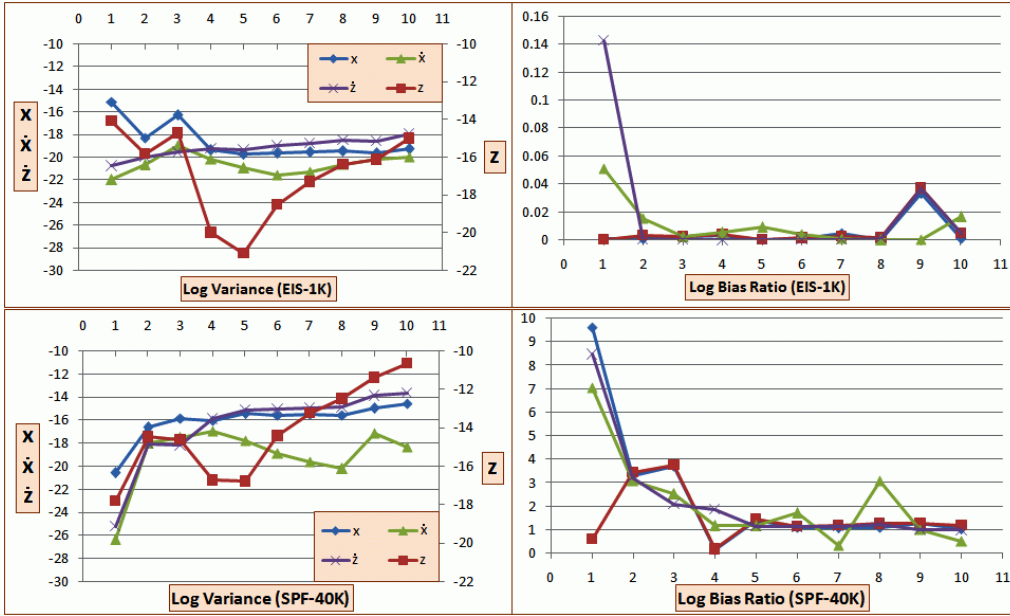


Figure 2: Log Variance and Log Bias Ratio, Singular Case.

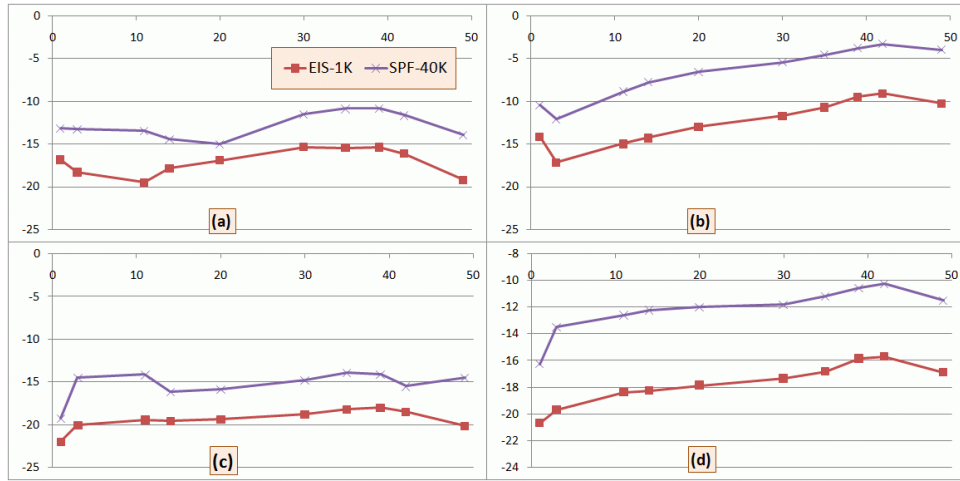


Figure 3: Log-Avg. MSE Comparisons, Non-Singular Case. Panel (a) $\rightarrow x$, (b) $\rightarrow z$, (c) $\rightarrow \dot{x}$, (d) $\rightarrow \dot{z}$

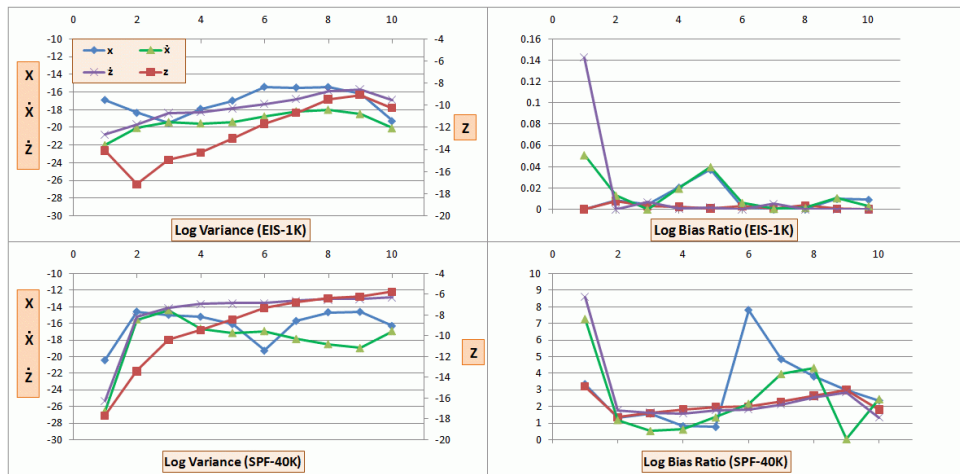


Figure 4: Log Variance and Log Bias Ratio, Non-Singular Case.

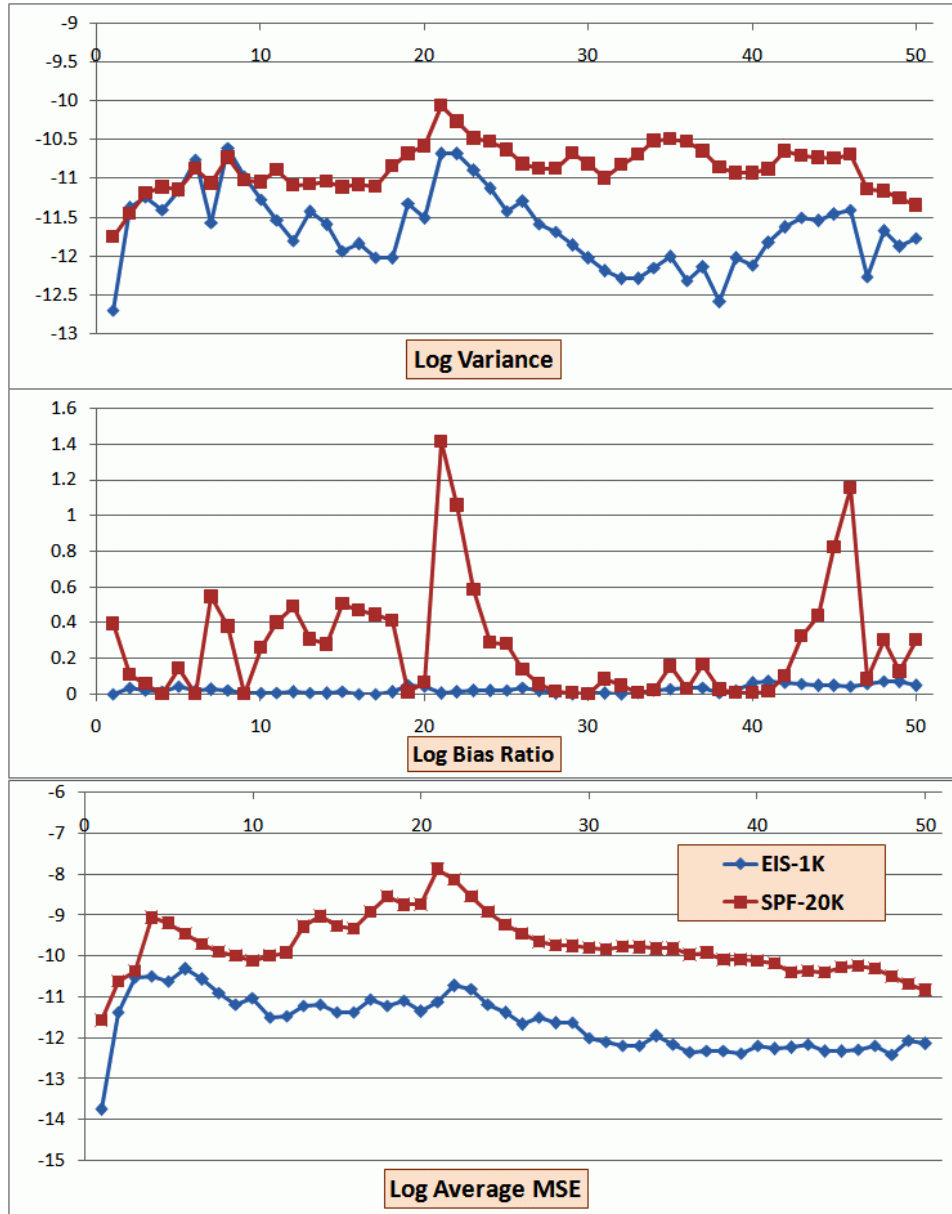


Figure 5: Log Variance, Log Bias Ratio and Log Avg. MSE Comparison, Stochastic Volatility Model