Efficient Filtering in 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

November 2008

Abstract

We develop a numerical procedure that facilitates efficient 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. We illustrate our procedure for the bearings-only tracking problem.

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. For helpful comments, we thank 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 Macroecnomics Workshop. All reported computational times are based on runs executed on a 1.86 GHz Opteron 165 processor, using Compaq Visual Fortran 6.6. Documentation, pseudocode and Fortran code used to execute the EIS filter are available at `www.pitt.edu/~dejong/wp.htm`

1 Introduction

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 state. When models are linear and gaussian, these integrals obtain analytically from the Kalman filter; departures entail integrals that must be approximated numerically. A broad range of numerical techniques have been proposed to accomplish filtering given departures from linearity/gaussianity. Here we propose an enhancement of the particle filter (PF) developed by Gordon et al. (1993) and Kitagawa (1996).

The PF achives filtering via the construction of discrete approximations to the densities that appear in the predictive and updating stages of the filtering process. The discrete supports of these approximations are known as swarms of particles. While the PF is conceptually simple and easy to program, it suffers from a major and well-documented shortcoming: the discrete supports of period-t approximations are determined using period-$(t-1)$ information, and thus ignore crucial information provided by period-t measurements. That is to say, the supports are not adapted. This shortcoming can produce large biases in the filtered estimates of the state; the elimination of these biases can require prohibitively large numbers of particles.

A number of innovative techniques have been proposed to achieve adaption. For examples, see Pitt and Shephard (1999); the collection of papers in Doucet et al. (2001); Pitt (2002); Ristic et al. (2004); and the collection housed at `http://www.sigproc.eng.cam.ac.uk/smc/papers.html`. These techniques typically rely upon local (Taylor Series) approximations and/or perturbations (jittering particles) to produce partial adaption, but they lack the flexibility needed for achieving full adaption.

Here we propose a generalization of the particle filter designed to achieved full adaption in filtering applications. In a companion paper, DeJong et al. (2008), we discuss the achievement of adaption in the context of likelihood evaluation. The generalization features two key extensions: the density approximations it constructs are continuous; and integrals are calcu-

lated using the Efficient Importance Sampling (EIS) methodology developed by Richard and Zhang (2007). The latter entails the iterative construction of global approximations of targeted integrands, which then serve as importance sampling densities that are fully adapted within the selected class of auxiliary samplers.

Full adaption comes at a cost: EIS entails conceptually simple but relatively time-consuming auxiliary calculations, in the form of least squares regressions. Moreover, the class of fully operational EIS approximations is limited largely to members of the exponential family of distributions (though flexible extensions beyond that family are currently being developed). Nevertheless, as we shall illustrate in the context of a benchmark bearings-only tracking problem, the EIS filter can produce unbaised estimates of state trajectories at high degrees of numerical precision. In particular, the EIS filter can produce dramatic reductions in Mean Squared Errors (MSEs) – by several orders of magnitude – relative to existing techniques. Due to these reductions, its computational complexity can be more than offset by its accuracy: e.g., to achieve reductions in MSEs on the scale delived by the EIS filter, the PF routinely requires the use of enormous swarms entailing prohibitive computational costs.

The rest of the paper in organized as follows. Section 2 provides a brief overview of the filtering problem using continuous densities. The PF and EIS filters are introduced in Section 3. Section 4 discusses the critical continuous predictive stage. Application of the EIS filter to the standard singular version of a bearings-only tracking problem is presented in Section 5, and a non-singular version of the problem is analyzed in Section 6. Section 7 concludes.

2 Filtering

It proves convenient for our purposes to present filtering in terms of continuous densities. Let s_t be a m-dimensional vector of unobserved state (target) 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. We use the generic notation $f(\cdot)$ and $f(\cdot|\cdot)$ to denote marginal and conditional densities, respectively.

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 $f(s_0)$. It is often the case that s_t is independent of Y_{t-1} conditionally on s_{t-1}. Nevertheless, inclusion of Y_{t-1} as an explicit conditioning variable is required for internal consistency of the filter. It also allows for control applications wherein measurements initiate actions that feed back on the state transitions. 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 accurate sequential estimates of the filtered values $\mathrm{E}[h(s_t)|Y_t]$.

Updating the filter in period t requires the following operations:

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

$$f\left(s_t|Y_t\right) = \frac{f\left(y_t|s_t, Y_{t-1}\right) f\left(s_t|Y_{t-1}\right)}{f\left(y_t|Y_{t-1}\right)}, \tag{2}$$

where

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

Equations (1) and (2) characterize the filter's predictive and updating stages, respectively. Filtered values of interest are then computed as

$$\mathrm{E}\left[h\left(s_t\right)|Y_t\right] = \int h\left(s_t\right) f\left(s_t|Y_t\right) ds_t, \tag{4}$$

which amounts to a ratio of integrals, in view of (1) and (3).

The particle filter approximates the densities $f(s_t|Y_{t-1})$ and $f(s_t|Y_t)$ by discrete swarms of equiprobable particles, denoted by $\{s_t^{1,i}\}_{i=1}^N$ and $\{s_t^{0,i}\}_{i=1}^N$, respectively (the first superscript denotes the time lag between observation and prediction). Most importantly, $\{s_t^{0,i}\}_{i=1}^N$ obtains by drawing with replacement from $\{s_t^{1,i}\}_{i=1}^N$ using the re-sampling probabilities $\{\pi_t^{0,i}\}_{i=1}^N$

obtained from the discretized version of (2); that is,

$$\pi_t^{0,i} = \frac{f\left(y_t|s_t^{1,i}, Y_{t-1}\right)}{\sum_{j=1}^{N} f\left(y_t|s_t^{1,j}, Y_{t-1}\right)}. \tag{5}$$

Note that by construction the support of $\{s_t^{0,i}\}_{i=1}^N$ is a subset of that of $\{s_t^{1,i}\}_{i=1}^N$, and can become very small under either of two conditions. First, due to the realization of an outlier y_t, $f(y_t|s_t, Y_{t-1})$ assigns the bulk of its weight in the far tails or even outside of the range covered by $\{s_t^{1,i}\}_{i=1}^N$. Second, $f(y_t|s_t, Y_{t-1})$ can be tightly distributed relative to $f(s_t|Y_{t-1})$, in which case very few elements within $\{s_t^{1,i}\}_{i=1}^N$ receive appreciable weight. In either case, this eventuality is known as sample impoverishment, and results from the absence of adaption intrinsic to the particle filter.

As noted, many procedures have been developed in efforts to achieve partial adaption. We now characterize the EIS filter, which is designed to achieve full adaption within a preselected class of continuous auxiliary samplers.

3 EIS-PF

3.1 Baseline algorithm

The EIS filter aims at constructing a fully adapted parametric importance sampling density $g(s_t|\hat{a}_t)$ for the product of densities in (3). Filtered values are then rewritten as

$$\mathrm{E}\left[h\left(s_t\right)|Y_t\right] = \frac{\int h(s_t)\omega\left(s_t;\hat{a}_t\right) g\left(s_t|\hat{a}_t\right) ds_t}{\int \omega\left(s_t;\hat{a}_t\right) g\left(s_t|\hat{a}_t\right) ds_t}, \tag{6}$$

where

$$\omega\left(s_t;a_t\right) = \frac{f\left(y_t|s_t, Y_{t-1}\right) f\left(s_t|Y_{t-1}\right)}{g\left(s_t|a_t\right)}. \tag{7}$$

The corresponding filtered estimate is then given by

$$\hat{\mathrm{E}}_N\left[h(s_t)|Y_t\right] = \frac{\sum_{i=1}^N h\left(\tilde{s}_t^{0,i}\right)\omega\left(\tilde{s}_t^{0,i};\hat{a}_t\right)}{\sum_{i=1}^N \omega\left(\tilde{s}_t^{0,i};\hat{a}_t\right)}, \tag{8}$$

where $\{\tilde{s}_t^{0,i}\}_{i=1}^N$ now denotes i.i.d. draws from the auxiliary importance sampling density $g(s_t|a_t)$.

The construction of $\hat{a}_t$ is achieved via minimization of the MC sampling variance of the ratio $\omega(s_t;a_t)$ given draws from $g(s_t|a_t)$. Construction is achieved as follows. One initially selects a parametric class of auxiliary density kernels $\mathcal{K} = \{k(s;a); a \in A\}$ with analytical integrating constant $\chi(a)$. The relationship between kernel and density is given by

$$g\left(s|a\right) = \frac{k(s;a)}{\chi(a)}, \quad \text{with} \quad \chi(a) = \int k(s;a)ds. \tag{9}$$

The selection of $\mathcal{K}$ is inherently problem-specific since these kernels are meant to provide operational (global) approximations to the product of densities $f(y_t|s_t,Y_{t-1})f(s_t|Y_{t-1})$.

Following Richard and Zhang (2007), an (approximate) optimal value for a_t obtains as the solution of the following auxiliary least squares problem:

$$(\hat{a}_t,\hat{c}_t) = \arg\min_{a\in A, c\in\mathbb{R}} \sum_{i=1}^R \left\{\ln\left[f\left(y_t|\tilde{s}_t^{0,i},Y_{t-1}\right)f(\tilde{s}_t^{0,i}|Y_{t-1})\right] - c - \ln k\left(\tilde{s}_t^{0,i};a\right)\right\}^2, \tag{10}$$

where, as above $\{\tilde{s}_t^{0,i}\}_{i=1}^R$ denotes i.i.d. draws from $g(s|\hat{a}_t)$ with R typically smaller than N in (8). The fact that these draws are conditional on $\hat{a}_t$ implies that the latter obtains as the fixed point solution of a sequence $\{\hat{a}_t^{(v)}\}_{v=1}^V$, where $\hat{a}_t^{(v)}$ solves (10) given draws from $g(s|\hat{a}_t^{(v-1)})$. This sequence can be initialized e.g. by (local) Taylor Series expansion of the product $f(y_t|s_t,Y_{t-1})f(s_t|Y_{t-1})$. The iterative search for $\hat{a}_t$ could be viewed as drawback of the EIS filter. However, for reasonably well-behaved problems convergence is fast, with V typically less than 5. Moreover, iteration turns out to be the main driver for full adaption (within the class $\mathcal{K}$), as it allows $k(s;a)$ to progressively reposition itself on the region of

importance of the product being approximated.

Note from (10) that adaption requires that the density $f(s_t|Y_{t-1})$ be computed for all draws from $g(s_t|a_t)$. The computational expense inherent in this requirement is unavoidable if adaption is to be achieved: the fact that $f(s_t|Y_{t-1})$ is represented by a predetermined swarm $\{\tilde{s}_t^{0,i}\}_{i=1}^N$ is the root source of the numerical inefficiencies suffered by the particle filter. In the next section we outline operational procedures for evaluating $f(s_t|Y_{t-1})$.

We conclude this brief presentation of the EIS filter with two important remarks.

(i) In order to achieve smooth EIS convergence all draws under $\hat{a}_t^{(v)}$ must be obtained via transformation of a single set of $\{\tilde{u}_t^i\}_{i=1}^R$ Common Random Numbers (CRNs), where the $\tilde{u}_t^i$s denote draws from a canonical distribution – i.e. one that does not depend on a (typically $U(0,1)$ or $N(0,1)$) – to be transformed as needed into draws from $g(s_t|a_t)$ using techniques described, e.g., in Devroye (1986).

(ii) As discussed, e.g., in Geweke (1989), the use of a common set of CRNs for the numerator and denominator of the filtered estimates in (8) typically induces positive correlation between the two, reducing further the MC variance of the ratio.

3.2 Auxiliary EIS samplers

In its current form EIS is fully operational only for kernels from the exponential family of distributions, in which case the auxiliary regressions are linear in a, where a denotes a natural parametrization in the sense of Lehmann (1986, Section 2.7). Extensions outside the exponential family of distributions (multivariate-t, mixtures, semiparametric, etc.) are currently being developed. Essentially, they amount to embedding a one-step Gauss-Newton iteration within each EIS-iteration step, thereby preserving the linearity of the auxiliary regressions. Preliminary results in this direction are promising.

In the bearings-only tracking application described below, we specify a four-dimensional EIS sampler as the product of a trivariate gaussian sampler for three 'well-behaved' state variables, conditionally on a fourth one for which we introduce a highly flexible piecewise

log-linear kernel. Additional flexibility in the selection of margins could be gained with the introduction of copulas, in which case EIS iterations would alternate between adjusting the copula and fitting the margins (for progress along this dimension, see Dharmarajan, 2008).

Finally, we are currently exploring the implementation of nonparametric importance sampling densities. This is an idea that has already been partially exploited to construct 'regularized' particle filters (jittering particles) – e.g., see Ristic et al. (2004, Section 3.5.3). We propose to push this further by implementing EIS iterations relying on nonparametric samplers to achieve full adaption, exploiting the fact that it is trivial to combine a nonparametric density with a piecewise log-linear Taylor Series expansion of the target densities. See DeJong et al. (2008) for further discussion of some of these proposed extensions. The point is that, as illustrated below, (full) adaption can produce such large reductions in MSEs for filtered values that it fully justifies investing in the development of increasingly flexible auxiliary samplers for the EIS filter.

3.3 Piecewise log-linear approximations

In line with the preceeding discussion we introduce here a flexible one-dimensional importance sampler that can be used to approximate ill-behaved margins as part of a higher-dimensional problem. This sampler consists of a piecewise-continuous log-linear approximation to the margin under consideration. It is characterized by a kernel $k(s;a)$ whose parameter $a' = (a_0, \ldots, a_R)$ consists of a grid $a_0 < a_1 < \ldots < a_R$, where the interval covers the support of the relevant margin[1]. We first describe the kernel $k(s;a)$ for a preassigned grid a. It is defined as follows:

$$\ln k_j(s;a) = \alpha_j + \beta_j s \quad \forall s \in [a_{j-1}, a_j] \tag{11}$$

$$\beta_j = \frac{\ln \varphi(a_j) - \ln \varphi(a_{j-1})}{a_j - a_{j-1}}, \quad \alpha_j = \ln \varphi(a_j) - \beta_j a_j. \tag{12}$$

[1]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).

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

$$K_j(s;a) = \frac{\chi_j(s;a)}{\chi_n(a)}, \quad \forall s \in [a_{j-1}, a_j], \tag{13}$$

$$\chi_j(s;a) = \chi_{j-1}(a) + \frac{1}{\beta_j}\left[k_j(s;a) - k_j(a_{j-1};a)\right], \tag{14}$$

$$\chi_0(a) = 0, \quad \chi_j(a) = \chi_j(a_j, a). \tag{15}$$

Its inverse c.d.f. is given by

$$s = \frac{1}{\beta_j}\left\{\ln\left[k_j(a_{j-1};a) + \beta_j\left(u \cdot \chi_R(a) - \chi_{j-1}(a)\right)\right] - \alpha_j\right\}, \tag{16}$$

$$u \in]0,1[\quad \text{and} \quad \chi_{j-1}(a) < u \cdot \chi_R(a) < \chi_j(a). \tag{17}$$

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 in Section 3.1 above. Instead we rely upon the following iterative procedure designed to achieve the same objective.

The recursive construction of an equal-probability-division kernel $k(s;\hat{a})$ is based upon the non-random equal division of $[\varepsilon, 1-\varepsilon]$ with $u_i = \varepsilon + (2-\varepsilon)i/R$ for $i = 1, \ldots, R-1$, 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 $\hat{a}^l$, construct the density kernel k and its c.d.f. K as described above. The step-$l+1$ grid is then computed as

$$\hat{a}_i^{l+1} = K^{-1}(u_i), \quad i = 1 \ldots, R-1. \tag{18}$$

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 paramet-

ric classes of samplers. A two-dimensional version of this algorithm has been successfully developed. Beyond two dimensions we propose to investigate the use of copulas to combine together univariate continuous-piecewise approximations as well as kernels from the exponential family of distributions. See Dharmarajan (2008) for a preliminary bivariate application of copulas in this context.

4 Constant weight approximation of $f(s_t|Y_{t-1})$

4.1 Principle

In the companion paper by DeJong et al. (2008), we discuss several alternatives to evaluate $f(s_t|Y_{t-1})$ at any value of s_t needed for EIS iterations. In the application below we rely upon a constant-weight approximation that we briefly present here. Combining (1) to (3) we can rewrite $f(s_t|Y_{t-1})$ as a ratio of integrals:

$$f(s_t|Y_{t-1}) = \frac{\int f(s_t|s_{t-1}, Y_{t-1}) f(y_{t-1}|s_{t-1}, Y_{t-2}) f(s_{t-1}|Y_{t-2}) ds_{t-1}}{\int f(y_{t-1}|s_{t-1}, Y_{t-2}) f(s_{t-1}|Y_{t-2}) ds_{t-1}}. \tag{19}$$

Next note that if EIS delivers a good approximation $g(s_{t-1}; \hat{a}_{t-1})$ to the product of densities in the denominator, then the weight $\omega(s_{t-1}; \hat{a}_{t-1})$, as defined in (7), will be near-constant over the the relevant range of integration. In such a case, we can replace the product $f(y_{t-1}|s_{t-1}, Y_{t-2}) f(s_{t-1}|Y_{t-2})$ by $\bar{\omega}_{t-1} \cdot g(s_{t-1}; \hat{a}_{t-1})$, where $\bar{\omega}_{t-1}$ denotes the average EIS weight for period $t-1$. It follows that $f(s_t|Y_{t-1})$ is approximated by

$$f(s_t|Y_{t-1}) \simeq \int f(s_t|s_{t-1}, Y_{t-1}) g(s_{t-1}; \hat{a}_{t-1}) ds_{t-1}, \tag{20}$$

whose MC approximation is given by

$$\hat{f}_S(s_t|Y_{t-1}) = \frac{1}{S} \sum_{i=1}^{S} f(s_t|s_{t-1}^{0,i}, Y_{t-1}). \tag{21}$$

As illustrated below it is often possible to exploit an operational expression for the EIS sampler $g(s_{t-1}; \hat{a}_{t-1})$ to evaluate the integral in (20) by combining partial analytical integration with MC integration of the remainder.

4.2 Singular transitions

State transitions often include identities that effectively reduce the dimension of integration in (20), since for any given s_t, the vector s_{t-1} is then restricted to a strict subset of $\mathbb{R}^m$. With only minor loss of generality, let partition s_t into $s_t = (p_t, q_t)$ in such a way that the transition identities can be written as

$$q_t = \phi(p_t, s_{t-1}). \tag{22}$$

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

$$q_{t-1} = \psi(s_t, p_{t-1}), \tag{23}$$

we can take the limit of the integral in (20) as ϵ tends to zero, producing

$$f(s_t|Y_{t-1}) \simeq \int J(s_t, p_{t-1}) f(p_t|s_{t-1}, Y_{t-1}) g(s_{t-1}; \hat{a}_{t-1})|_{q_{t-1}=\psi(s_t,p_{t-1})} dp_{t-1}, \tag{24}$$

where

$$J(s_t, p_{t-1}) = \frac{\partial}{\partial q_t}\psi(s_t, p_{t-1}). \tag{25}$$

Since $g(p_{t-1}, \psi(s_t, p_{t-1}); \hat{a}_{t-1})$ is not a sampler for $p_{t-1}|s_t$, we must evaluate (24) by a mix of analytical, quadrature and/or own EIS methods.

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 precisely

those most prone to significant sample impoverishment problems (since given (p_t, s_{t-1}), there is no jittering in q_t.)

5 Application to 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

$$\lambda_{t+1} = \begin{pmatrix} I_2 & I_2 \\ 0 & I_2 \end{pmatrix} \lambda_t + \sigma \begin{pmatrix} \frac{1}{2} I_2 \\ I_2 \end{pmatrix} u_t, \tag{26}$$

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

$$\lambda_1 \sim N(\mu_1, \Delta_1), \tag{27}$$

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 $\theta_t = \arctan(z_t/x_t)$. The measured angle y_t is assumed to be wrapped Cauchy with density

$$f(y_t|\theta_t) = \frac{1}{2\pi} \frac{1 - r^2}{1 + r^2 - 2r\cos(y_t - \theta_t)}, \tag{28}$$

with $0 \leq (y_t, \theta_t) \leq 2\pi$ and $0 \leq r \leq 1$. Accordingly, we shall introduce a (partial) re-

parametrization in polar coordinates. Let

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

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

$$s_t = h(\lambda_t) = (\theta_t, \quad \rho_t, \quad \beta_t') = (\theta_t, \quad \delta_t'). \tag{29}$$

Note that (26) 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

$$\alpha_{t+1}|\lambda_t \sim N(A\lambda_t, \Omega), \tag{30}$$

and a Dirac transition

$$\beta_{t+1} \equiv \phi(\alpha_{t+1}, \lambda_t) = 2(\alpha_{t+1} - \alpha_t) - \beta_t, \tag{31}$$

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 next 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

$$\ell_t \equiv f(y_t|Y_{t-1}) = \int f(y_t|\theta_t)\, f(s_t|Y_{t-1})\, ds_t. \tag{32}$$

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 $\{\widetilde{\lambda}_t^{1,i}\}_{i=1}^N$. Specifically, the period-$(t-1)$ EIS swarm $\{\widetilde{s}_{t-1}^{0,i}\}_{i=1}^N$ is transformed into a swarm $\{\widetilde{\lambda}_{t-1}^{0,i}\}_{i=1}^N$ 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 $f(\lambda_t|\widetilde{\lambda}_{t-1}^{0,i})$ associated with (26).

(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 $\{\widetilde{\lambda}_t^{1,i}\}_{i=1}^N$ to construct an auxiliary OLS regression of $\{\ln f(\widetilde{\lambda}_t^{1,i}|Y_{t-1})\}_{i=1}^N$ on $\{\widetilde{\lambda}_t^{1,i}\}_{i=1}^N$ and the lower triangle of $\{(\widetilde{\lambda}_t^{1,i})(\widetilde{\lambda}_t^{1,i})'\}_{i=1}^N$, 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

$$k_{\lambda,t}(\lambda_t) = \exp\{-\frac{1}{2}(\lambda_t' P_t \lambda_t - 2\lambda_t' q_t)\}, \tag{33}$$

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

$$k_{s,t}(s_t) = \rho_t k_{\lambda,t}\left(h_t^{-1}(s_t)\right). \tag{34}$$

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

$$g_t(\delta_t|\theta_t) = \frac{k_{s,t}(s_t)}{\chi_t(\theta_t)}, \tag{35}$$

with

$$\chi_t(\theta_t) = \int_{\Delta} k_{s,t}(s_t)\, d\beta_t d\rho_t, \tag{36}$$

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

(iv) The likelihood integral in (32) is rewritten as

$$\ell_t = \int \left[f(y_t|\theta_t)\chi_t(\theta_t)\right] \frac{f(s_t|Y_{t-1})}{k_{s,t}(s_t)} g_t(\delta_t|\theta_t)\, d\delta_t d\theta_t. \tag{37}$$

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 (37) is rewritten as

$$\ell_t = \int w_{s,t}(s_t)\, g_t(\delta_t|\theta_t)\, g_t(\theta_t)\, ds_t, \tag{38}$$

with

$$w_{s,t}(s_t) = \left[\frac{f(y_t|\theta_t)\chi_t(\theta_t)}{g_t(\theta_t)}\right] \frac{f(s_t|Y_{t-1})}{k_{s,t}(s_t)}. \tag{39}$$

Its EIS-MC estimate obtains as

$$\widehat{\ell_t} = \frac{1}{N}\sum_{i=1}^{N} w_{s,t}\left(\widetilde{s}_t^{0,i}\right), \tag{40}$$

where $\{\widetilde{s}_t^{0,i}\}_{i=1}^N$ 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 (39). 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 (39) can trivially be transformed into weights for λ_t. Let

$$w_{\lambda,t}\left(\lambda_t\right) = w_{s,t}\left(h^{-1}\left(s_t\right)\right) = \frac{f\left(\lambda_t|Y_{t-1}\right)}{k_{\lambda,t}\left(\lambda_t\right)}\left[\frac{f\left(y_t|\theta_t\right)\chi_t\left(\theta_t\right)}{g_t\left(\theta_t\right)}\right]_{\theta_t=\theta_t(\alpha_t)}, \tag{41}$$

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

$$f\left(\lambda_t|Y_t\right) = \frac{f\left(\lambda_t|Y_{t-1}\right) f\left(y_t|\theta_t\left(\alpha_t\right)\right)}{\ell_t}, \tag{42}$$

can be rewritten as

$$f\left(\lambda_t|Y_t\right) = \frac{w_{\lambda,t}\left(\lambda_t\right)}{\ell_t} k_{\lambda,t}\left(\lambda_t\right) \frac{g_t\left(\theta_t\right)}{\chi_t\left(\theta_t\right)}|_{\theta_t=\theta_t(\alpha_t)}. \tag{43}$$

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

$$f(\lambda_{t+1}|Y_t) = \int_{\mathbb{R}^4} f(\lambda_t|Y_t)\, f(\lambda_{t+1}|\lambda_t)\, d\lambda_t. \tag{44}$$

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 (43) by the inverse of the Dirac transition in (31):

$$\phi^{-1}(\lambda_{t+1}, \alpha_t) = \beta_t = 2(\alpha_{t+1} - \alpha_t) - \beta_{t+1}, \tag{45}$$

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 (43) is set equal to 1. Whence $f(\lambda_{t+1}|Y_t)$ can be accurately evaluated by the following bivariate integral:

$$f(\lambda_{t+1}|Y_t) = \int \frac{g_t(\theta_t)}{\chi_t(\theta_t)} k_{\lambda,t}(\lambda_t)\, f(\alpha_{t+1}|\lambda_t)\,|_{\theta_t=\theta_t(\alpha_t),\beta_t=\phi^{-1}(\lambda_{t+1},\alpha_t)} d\alpha_t. \tag{46}$$

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

$$E\left(\lambda_t|Y_t\right) = \int \lambda_t f\left(\lambda_t|Y_t\right) d\lambda_t. \tag{47}$$

Substituting (43) 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:

$$E\left(\lambda_t|Y_t\right) = \frac{1}{\ell_t}\int h(s_t) w_{s,t}\left(s_t\right) g_t(\delta_t|\theta_t) g_t(\theta_t) ds_t. \tag{48}$$

EIS estimates of these filtered values obtain as

$$\widehat{E\left(\lambda_t|Y_t\right)} = \frac{\sum_{i=1}^N h(\tilde{s}_t^{0,i}) w_{s,t}\left(\tilde{s}_t^{0,i}\right)}{\sum_{i=1}^N w_{s,t}\left(\tilde{s}_t^{0,i}\right)}. \tag{49}$$

Filtered values of s_t are obtained by replacing $h(s_t)$ by s_t in (48) and (49). 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 (48) and filtered values simplify into the arithmetic mean of $\{h\left(\tilde{s}_t^{0,i}\right)\}_{i=1}^N$. As discussed, e.g., by Geweke (1989), the reason for preferring the ratio form in (49) 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 (26) is a (spurious) consequence of a model specifica-

tion 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 (26) 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

$$\beta_{t+1} = \beta_t + \varepsilon_{t+1}, \qquad \varepsilon_t \sim N\left(0, \sigma^2 I_2\right), \tag{50}$$

and position is discretized as

$$\alpha_{t+1} = \alpha_t + \frac{1}{2}\left(\beta_t + \beta_{t+1}\right). \tag{51}$$

It follows that

$$\beta_{t+D} = \beta_t + u_{t+D}, \tag{52}$$

$$\alpha_{t+D} = \alpha_t + D\beta_t + v_{t+D}, \tag{53}$$

with

$$u_{t+D} = \sum_{j=1}^{D} \varepsilon_{t+j}, \qquad v_{t+D} = \frac{1}{2}\sum_{j=1}^{D} \left[2\left(D-j\right)+1\right]\varepsilon_{t+j}. \tag{54}$$

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

$$\lambda_{t+D}|\lambda_t \sim N\left(A_D\lambda_t, \sigma^2 V_D\right), \tag{55}$$

with

$$A_D = \begin{pmatrix} I_2 & DI_2 \\ 0 & I_2 \end{pmatrix}, \qquad V_D = D \cdot \begin{pmatrix} \frac{4D^2-1}{12} I_2 & \frac{D}{2} I_2 \\ \frac{D}{2} I_2 & I_2 \end{pmatrix}. \tag{56}$$

The case $D = 1$ obviously coincides with the degenerate transition in (26). The generalization from (26) to (55) 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

$$f(\lambda_{t+D}|Y_t) = \int \frac{g_t(\theta_t)}{\chi_t(\theta_t)} k_{\lambda,t}(\lambda_t) f(\lambda_{t+D}|\lambda_t) d\lambda_t. \tag{57}$$

The numerical evaluation of equation (57) parallels that of equation (46) 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 (26) and (55) is set to 0.001; and r in (28) 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 $I = 40$ 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, where the numbers following acronyms indicate the number of particles (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). Let $\bar{l}_t^i$, $(i : 1 \rightarrow 40,\ t : 1 \rightarrow 10)$ 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. We ran this experiment to verify that 'true' values produced by both EIS and PF estimators are numerically identical. The latter number turns out to be needed in order to eliminate significant biases characterizing PF estimates of filtered means (illustrated below).

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

$$\text{LMSE}_{t,k} = \ln \left\{ \frac{1}{40} \sum_{i=1}^{40} \left[\frac{1}{100} \sum_{j=1}^{100} \left(\tilde{l}_{t,k}^{i,j} - \bar{l}_t^i \right)^2 \right] \right\}. \tag{58}$$

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 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 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 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.

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 using an optimization procedure designed to minimize numerical standard errors associated

with targeted integrals. Implementation of the filter is straightforward, and the payoff of adoption can be substantial as illustrated by means of the bearings-only tracking problem.

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 (34) depends upon the quadratic form

$$\gamma(s) = \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix}' P \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix} - 2 \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix} q. \tag{59}$$

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

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \qquad q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}. \tag{60}$$

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

$$\begin{aligned}
\gamma(s) &= (\beta - b_\theta)' P_{22} (\beta - b_\theta) + a_\theta (\rho - r_\theta)^2 - s_\theta^2, && (61)\\
b_\theta &= P_{22}^{-1} (q_2 - \rho P_{21} e_\theta), \qquad a_\theta = e_\theta' P_{11.2} e_\theta, && (62)\\
P_{11.2} &= P_{11} - P_{12} P_{22}^{-1} P_{21}, && (63)\\
r_\theta &= \frac{1}{a_\theta} \left(q_1 - P_{12} P_{22}^{-1} q_2\right)' e_\theta, \qquad s_\theta^2 = a_\theta r_\theta^2 + q_2' P_{22}^{-1} q_2. && (64)
\end{aligned}$$

It follows that $\chi(\theta)$, as defined in (36), is given by

$$\chi(\theta) = 2\pi|P_{22}|^{-1} d_\theta \exp\{\frac{1}{2}s_\theta^2\}, \tag{65}$$

$$d_\theta = \int_0^\infty \rho \exp\{-\frac{1}{2}a_\theta(\rho - r_\theta)^2\}d\rho. \tag{66}$$

Introducing the transformation of variables

$$\phi = \sqrt{a_\theta}(\rho - r_\theta), \tag{67}$$

d_θ can be written as

$$d_\theta = \frac{1}{a_\theta}\int_{-c_\theta}^\infty (\phi + c_\theta)\exp\{-\frac{1}{2}\phi^2\}d\phi \tag{68}$$

$$= \frac{1}{a_\theta}\left[\exp(-\frac{1}{2}c_\theta^2) + c_\theta\sqrt{\frac{\pi}{2}}\{1 + \mathrm{erf}(\frac{c_\theta}{\sqrt{2}})\}\right], \tag{69}$$

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

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-\phi^2)d\phi. \tag{70}$$

(The properties of erf() are discussed, e.g., in Abramowitz and Segun, 1968, Ch. 7.) In deriving (69), 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

$$m(\rho|\theta) = \frac{1}{d_\theta}\rho\exp\{-\frac{1}{2}a_\theta(\rho - r_\theta)^2\}, \qquad \rho > 0. \tag{71}$$

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

$$\beta = b_\theta + L \begin{pmatrix} u_3 \\ \\ u_4 \end{pmatrix}, \tag{72}$$

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

Regarding step (ii), $\rho|\theta$ obtains from the transformation of (67), rewritten as

$$\rho = \frac{1}{\sqrt{a_\theta}} (\phi + c_\theta), \tag{73}$$

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

$$f_\phi (\phi|\theta) = \frac{1}{d_\theta a_\theta} (\phi + c_\theta) \exp\{-\frac{1}{2}\phi^2\}, \qquad \phi > -c_\theta, \tag{74}$$

with cdf

$$\begin{aligned} F_\phi (\phi|\theta) = \frac{1}{d_\theta a_\theta} \Big\{ & [\exp(-\frac{1}{2}c_\theta^2) - \exp(-\frac{1}{2}\phi^2)] \\ & + c_\theta \sqrt{\frac{\pi}{2}} [\mathrm{erf}(\frac{\phi}{\sqrt{2}}) + \mathrm{erf}(\frac{c_\theta}{\sqrt{2}})] \Big\}, \end{aligned} \tag{75}$$

accounting for the fact that $\mathrm{erf}(-z) = -\mathrm{erf}(z)$. For the application described in Section 6, 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

$$\phi^{(k+1)} = \phi^{(k)} - \frac{F(\phi^{(k)}|\theta) - u_2}{F(\phi^{(k)}|\theta)}. \tag{76}$$

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 (46) depends on the quadratic form

$$\delta(\alpha_{+1}, \lambda) = (\lambda' P \lambda - 2\lambda' q) + (\alpha_{+1} - A\lambda)' \Omega^{-1} (\alpha_{+1} - A\lambda). \tag{77}$$

It is transformed into a quadratic form in (α, λ_{+1}) via the inverse Dirac transformation (45)

$$\delta_1(\alpha, \lambda_{+1}) = \delta(\alpha_{+1}, \lambda)|_{\beta=\psi(\lambda_{+1},\alpha)}. \tag{78}$$

This implies the following two transformations:

$$\lambda|_{\beta=\psi(\lambda_{+1},\alpha)} = C\begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}, \qquad (\alpha_{+1} - A\lambda)|_{\beta=\psi(\lambda_{+1},\alpha)} = D\begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}, \tag{79}$$

$$\tag{80}$$

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

$$\delta_1(\alpha, \lambda_{+1}) = \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}' M \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix} - 2\begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}' m, \tag{81}$$

$$M = C'PC + D'\Omega^{-1}D, \quad m = C'q. \tag{82}$$

Note that $\delta_1(\alpha, \lambda_{+1})$ is functionally similar to $\gamma(s)$ in (59), 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

$$\delta_*(\rho, \theta_{+1}, \lambda_{+1}) = \lambda'_{+1} M_{22} \lambda_{+1} - 2\lambda'_{+1} m_2 + a^*_\theta (\rho - r^*_{\theta\lambda})^2 - c^*_{\theta\lambda}, \tag{83}$$

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 (80) and integrating with respect to ρ, we obtain

$$\begin{aligned} f(\lambda_{+1}|Y) &= \frac{|\Omega|^{-\frac{1}{2}}}{2\pi} \exp\{-\frac{1}{2}\left(\lambda'_{+1} M_{22}\lambda_{+1} - 2\lambda'_{+1} m_2\right)\} \\ &\quad \times \int \Big[\frac{1}{\chi(\theta)} d^*_{\theta\lambda} \exp(\frac{1}{2} c^{*2}_{\theta\lambda})\Big] m(\theta)\, d\theta, \end{aligned} \tag{84}$$

where $d^*_{\theta\lambda}$ obtains from (69) 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 (84) 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 (55) 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 (57) depends on the quadratic form

$$\delta(\lambda_{+D}, \lambda) = (\lambda' P \lambda - 2\lambda' q) + (\lambda_{+D} - A_D\lambda) V_D^{-1} (\lambda_{+D} - A_D\lambda), \tag{85}$$

which is rewritten as

$$\delta(\lambda_{+D}, \lambda) = (\lambda' P_0 \lambda - 2\lambda' q_0) + \lambda'_{+D} V_D^{-1} \lambda_{+D}, \tag{86}$$

with

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

The integration with respect to λ in (84) proceeds exactly as described in the singular case, except that (P, q) are replaced by (P_0, q_0). Thus $f(\lambda_{+D}|Y)$ is given by

$$\begin{aligned} f(\lambda_{+D}|Y) &= \frac{|V_D|^{-\frac{1}{2}}}{2\pi} \exp\{-\frac{1}{2}\left(\lambda_{+D}' V_D^{-1} \lambda_{+D}\right)\} \\ &\quad \times \int \Big[\frac{1}{\chi(\theta)} d_{\theta\lambda}^0 \exp\{\frac{1}{2}(s_{\theta\lambda}^0)^2\}\Big] m(\theta)\, d\theta, \end{aligned} \tag{87}$$

where $s_{\theta\lambda}^0$ and $d_{\theta\lambda}^0$ are defined by (64) and (66), with (P, q) replaced by (P_0, q_0). The EIS evaluation of (87) parallels that of $f(\lambda_+|Y)$ in (84).

References

[1] 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.

[2] DeJong, D.N., H. Dharmarajan, R. Liesenfeld, J.-F. Richard, G. Valle-Moura, 2008, "Efficient Likelihood Evaluation of State-Space Representations", University of Pittsburgh Working Paper.

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

[4] Dharmarajan , H., 2008, *Essays on Estimation of Non-Linear State-Space Models*. PhD Dissertation, University of Pittsburgh.

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

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

[7] 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.

[8] Gradshteyn, I.S. and I.M. Ryzhik, 1979, *Table of integrals, series and products*. Academic Press, San Diego.

[9] 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.

[10] Lehmann, E.L., 1986, *Testing Statistical Hypotheses*. John Wiley & Sons.

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

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

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

[14] Ristic, B., S. Arulampalam, and N. Gordon, 1984, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston: Artech House Publishers.

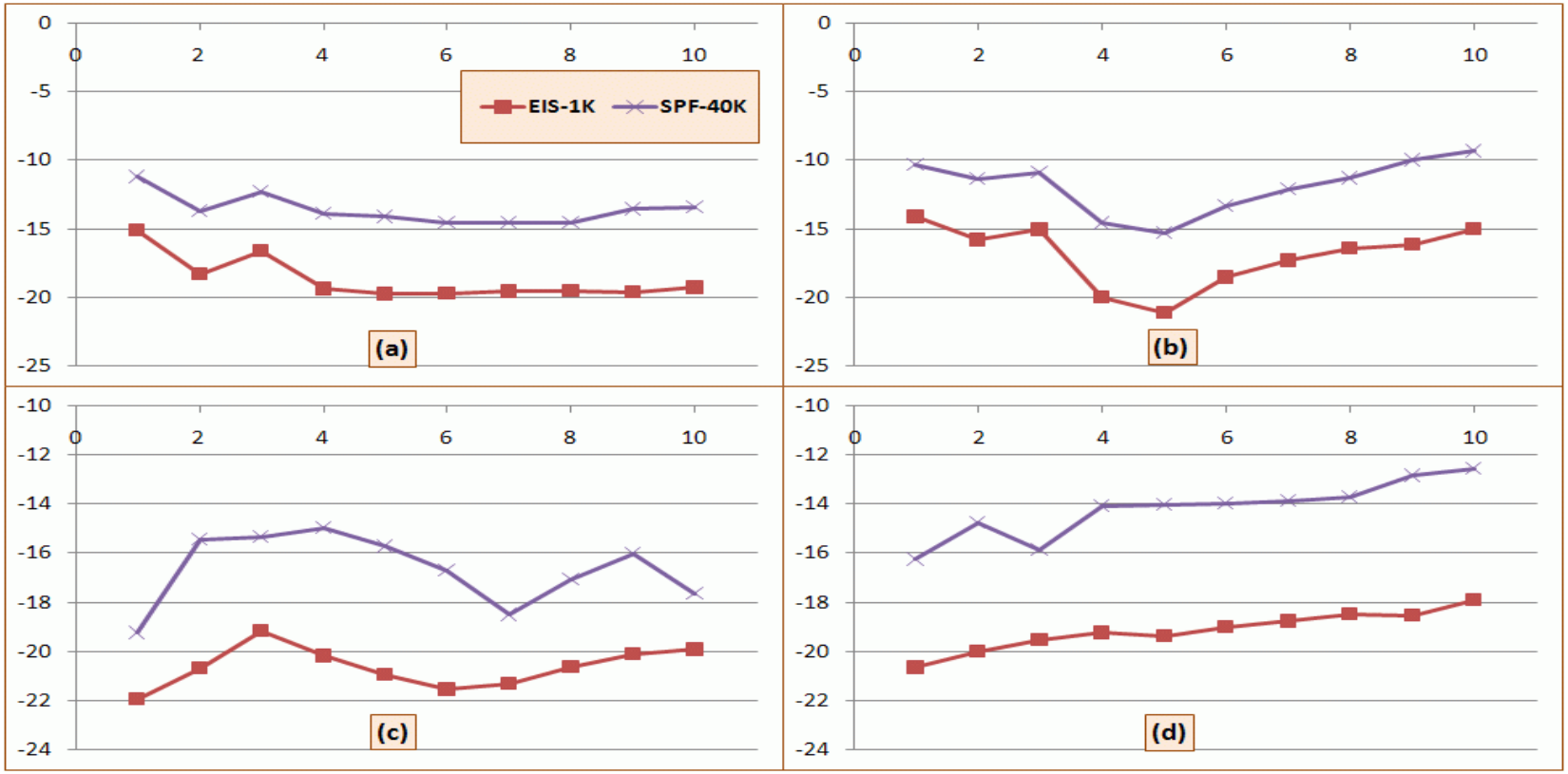

Figure 1. Log-Avg.MSE Comparisons, Singular Case. Panel (a) $\rightarrow x$, (b)$\rightarrow z$, (b)$\rightarrow \dot{x}$, (b)$\rightarrow \dot{z}$.

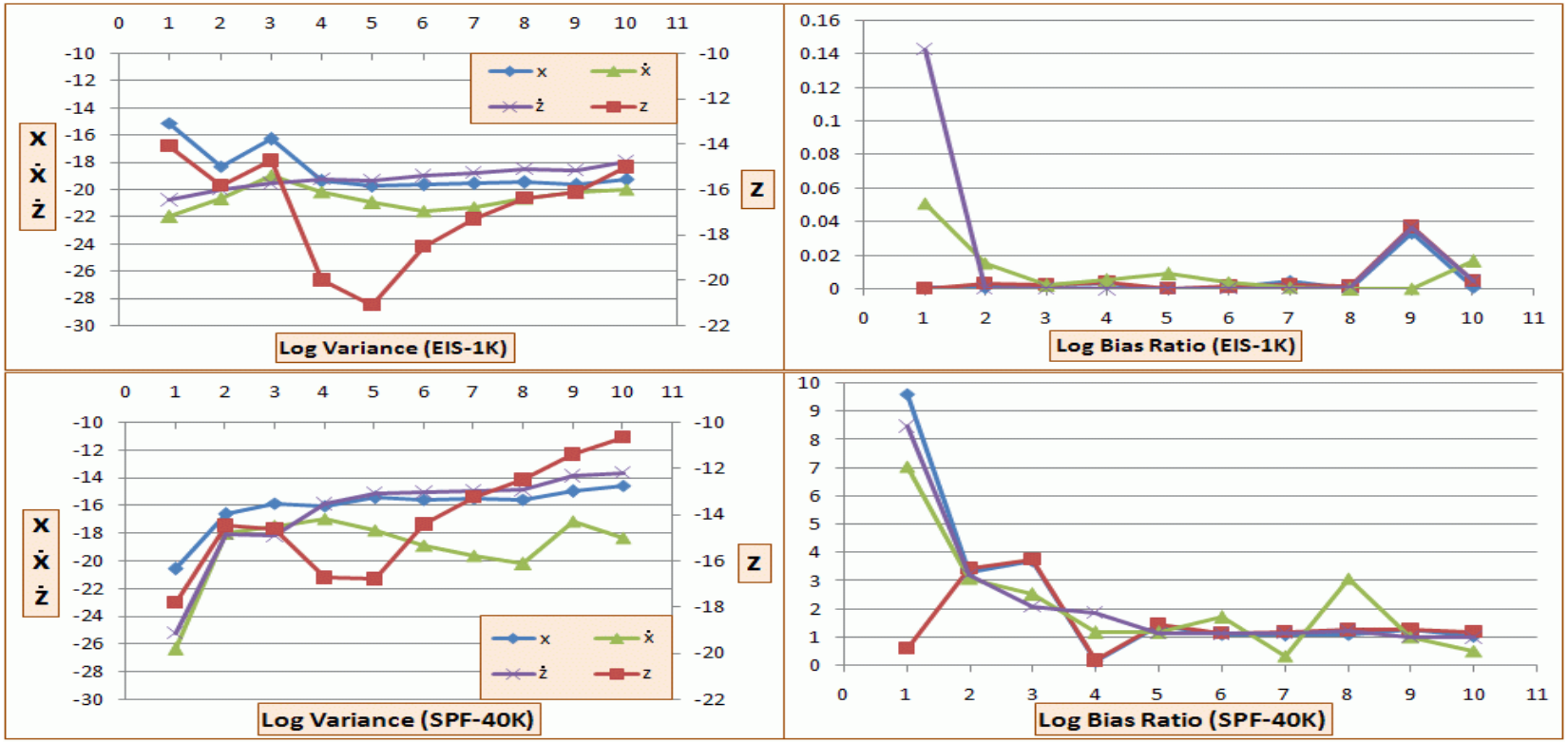

Figure 2. Log Variance and Log Bias Ratio, Singular Case.

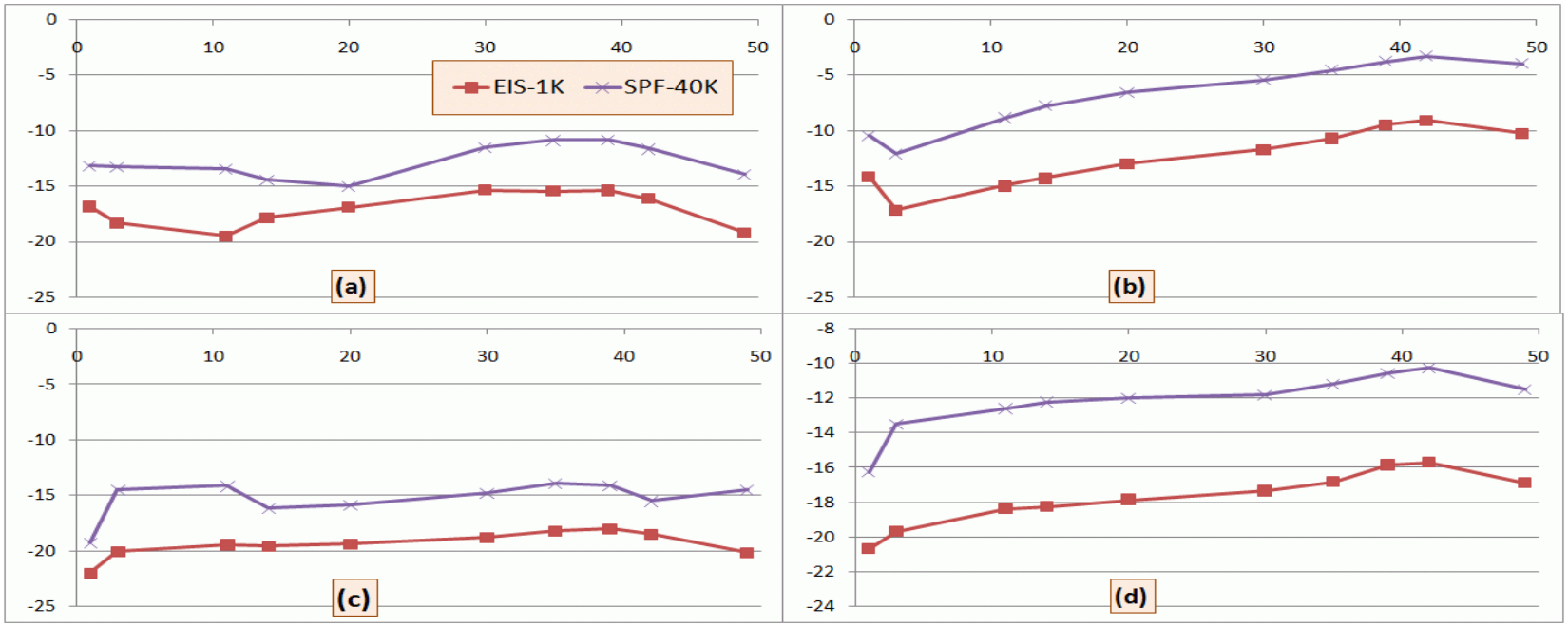

Figure 3. Log-Avg.MSE Comparisons, Non-Singular Case. Panel (a) $\rightarrow x$, (b)$\rightarrow z$, (b)$\rightarrow \dot{x}$, (b)$\rightarrow \dot{z}$.

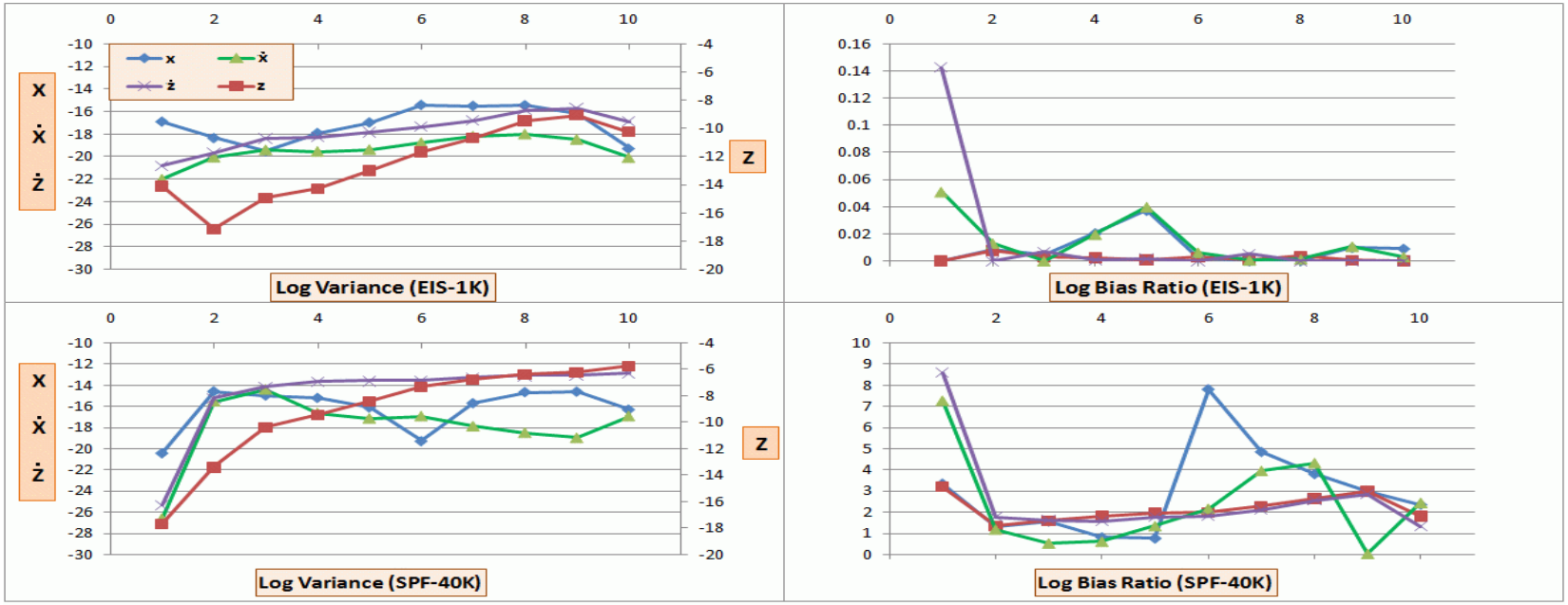

Figure 4. Log Variance and Log Bias Ratio, Non-Singular Case.