Particle Efficient Importance Sampling*

Marcel Scharth

The University of Sydney Business School

Robert Kohn

Department of Economics, University of New South Wales

Abstract

The efficient importance sampling (EIS) method is a general principle for the numerical evaluation of high-dimensional integrals that uses the sequential structure of target integrands to build variance minimising importance samplers. Despite a number of successful applications in high dimensions, it is well known that importance sampling strategies are subject to an exponential growth in variance as the dimension of the integration increases. We solve this problem by recognising that the EIS framework has an offline sequential Monte Carlo interpretation. The particle EIS method is based on non-standard resampling weights that take into account the construction of the importance sampler as a sequential approximation to the state smoothing density. We apply the method for a range of univariate and bivariate stochastic volatility specifications. We also develop a new application of the EIS approach to state space models with Student's t state innovations. Our results show that the particle EIS method strongly outperforms both the standard EIS method and particle filters for likelihood evaluation in high dimensions. We illustrate the efficiency of the method for Bayesian inference using the particle marginal Metropolis-Hastings and importance sampling squared algorithms.

KEYWORDS: Bayesian inference, particle filters, particle marginal Metropolis-Hastings, sequential Monte Carlo, stochastic volatility.

*We would like to thank two anonymous referees for helping to improve the presentation and technical content of the paper. Both authors were partially supported by ARC discovery grant DP120104014. Sample code for replication of the bivariate stochastic volatility example and performing Bayesian estimation using the particle EIS method can be downloaded on http://www.marcelscharth.com.

1 Introduction

This paper introduces the particle efficient importance sampling (P-EIS) method as a tool for likelihood evaluation and state inference in nonlinear non-Gaussian state space model applications. The approach is based on the EIS algorithm of Richard and Zhang (2007), which is an importance sampling method for the estimation of high-dimensional integrals that have a sequential structure. The EIS method constructs a global approximation to the target integrand by iterating a sequence of least-squares regressions, which are linear and computationally efficient for a wide range of models. The essential idea of the particle EIS method is that the high-dimensional EIS approach has a sequential Monte Carlo (SMC) interpretation amenable to resampling steps. We show that it is crucial to use non-standard resampling weights that take into account the construction of the importance sampler as a sequential approximation to the smoothing density of the states.

The use of importance sampling to evaluate the likelihood of nonlinear non-Gaussian state space models for long time series dates back to the method of Shephard and Pitt (1997) and Durbin and Koopman (1997), which relies on a local approximation method. The use of the global approximation technique in the EIS method has expanded the scope of high-dimensional importance sampling and a range of applications are now available in the literature. Some examples include stochastic volatility models in Liesenfeld and Richard (2003), the stochastic conditional duration model in Bauwens and Galli (2009), probit models with correlated errors in Liesenfeld and Richard (2010), DSGE models in DeJong, Liesenfeld, Moura, Richard, and Dharmarajan (2012), state space models with mixture measurement densities in Kleppe and Liesenfeld (2013) and discrete dependent variable models with spatial correlation in Liesenfeld, Richard, and Vogler (2013).

Despite these successful applications, the use of importance sampling has so far been limited by the exponential increase in the variance of the likelihood estimate as the dimension of the integration problem increases. See for example Chopin (2004), Section 3.3. The SMC approach of the particle EIS method solves this problem by introducing resampling when generating draws from a high-dimensional importance density. Like the standard EIS method, the particle EIS algorithm is an offline approach that approximates the smoothing density of the states and explicitly minimises the

variance of the likelihood estimate given the sequential structure of the sampler and all the information up to end of the available series. In contrast, particle filters are designed for online sequential processing of the data by considering low dimensional proposals at every period given the current particle system and the series up to that period. As an extension, these proposals may include some look-ahead information to increase efficiency (see Lin, Chen, and Liu 2013). We maximise efficiency for likelihood evaluation by using both an offline global approximation method and resampling steps that previously were only available for particle filters.

We base the particle EIS method on the auxiliary particle filter algorithm (APF) introduced by Pitt and Shephard (1999). Even though the APF was originally designed for filtering, we adopt it as general sequential Monte Carlo method to implement our offline approach. The APF algorithm provides a correct procedure for likelihood estimation in the presence of the non-standard resampling weights required by the particle EIS method. The particle EIS method provides an unbiased estimate of the likelihood following the general result for auxiliary particle filters in Del Moral (2004). The unbiasedness property is essential for applications to Bayesian inference using the particle marginal Metropolis-Hastings (PMMH) method of Andrieu, Doucet, and Holenstein (2010) and the importance sampling squared (IS^2) method of Tran, Scharth, Pitt, and Kohn (2014). See also the discussion in Flury and Shephard (2011). Given the offline nature of the PMMH and IS^2 methods, the particle EIS method has a natural advantage over particle filters for likelihood evaluation in those contexts.

We present a detailed study of the numerical efficiency of the particle EIS method compared to the EIS algorithm and standard particle filters. We base our analysis on a simulation study for a range of univariate stochastic volatility (SV) models and a bivariate SV specification, for which we also present an empirical application. Our general univariate specification allows for a fat-tailed measurement density, a two-factor log-volatility process, leverage effects (which imply a nonlinear state transition) and additive Student's t state innovations, highlighting the flexibility of the EIS framework. The application of EIS for models with additive Student's t state disturbances is new to the literature. We develop the EIS algorithm for this model using a data augmentation idea initially proposed by Kleppe and Liesenfeld (2013).

The simulation study leads to three main conclusions. First, the particle EIS method brings large reductions in variance over the standard EIS method. For a time

series of $10,000$ observations, the decrease in variance ranges from 80% for the univariate SV model with Student's t state disturbances to 95% for the bivariate specification. These gains come with a negligible increase in computational time. Second, the EIS and P-EIS methods strongly outperform standard particle filters for these models. Our result show that the P-EIS method outperforms the best particle filter in our analysis by factors of approximately 100 to 6,000 depending on the specification under consideration. Third, the particle EIS method maintains a constant performance relative to the particle filters for all time series dimensions.

The empirical application for the bivariate SV model uses $5,797$ daily observations of the IBM and General Electric stock returns. We focus on posterior inference using the PMMH and IS^2 methods. Using the theory on the optimal implementation of these two methods developed by Pitt, Silva, Giordani, and Kohn (2012) and Tran, Scharth, Pitt, and Kohn (2014) respectively, we find that for this example the particle EIS method needs only 10 particles to achieve the same numerical performance for posterior inference as a bootstrap filter with 15,000 particles. This result shows that particle EIS can make Bayesian estimation of complex state space models feasible in situations in which simple particle methods require excessive computing times for accuracy. As in the simulation study, our empirical analysis shows substantial gains from using the P-EIS method in comparison with the standard EIS algorithm.

We organise the paper as follows. Section 2 presents the notation and estimation objective and reviews the EIS method. Section 3 introduces and motivates the particle EIS method. Section 4 studies the relative performance of the new method for likelihood evaluation for univariate and bivariate stochastic volatility models in a simulated setting. Section 5 presents an empirical application to posterior inference via IS^2 and PMMH.

2 Importance sampling

2.1 State Space Model

Consider a discrete-time Markov process $\{X_t\}_{t\geq 1}$ such that

$$X_1 \sim p(x_1), \qquad X_t|(X_{t-1} = x_{t-1}) \sim p(x_t|x_{t-1}).$$

We assume that n observations are generated by the measurement density

$$Y_t|(X_t = x_t) \sim p(y_t|x_t).$$

The state and measurement densities implicitly depend on a parameter vector $\theta \in \Theta \subseteq \mathbb{R}^d$, which we omit from the notation whenever possible for conciseness. Define $x_{1:t} = (x_1', \ldots, x_t')'$ and $y_{1:t} = (y_1', \ldots, y_t')'$. The likelihood for the state space model is given by the integral

$$\begin{aligned} L(y_{1:n}) &= \int p(y_{1:n}, x_{1:n}) \, \mathrm{d}x_{1:n} = \int p(y_{1:n}|x_{1:n}) p(x_{1:n}) \, \mathrm{d}x_{1:n} \\ &= \int p(y_1|x_1) p(x_1) \prod_{t=2}^{n} p(y_t|x_t) p(x_t|x_{t-1}) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n, \end{aligned} \tag{1}$$

which is typically analytically intractable. Our objective in this paper to obtain an accurate and unbiased Monte Carlo estimate $\widehat{L}(y)$ of this integral for a wide class of models.

2.2 Efficient High Dimensional Importance Sampling

To evaluate the likelihood function by importance sampling, we consider a high-dimensional importance distribution $q(x_{1:n}|y_{1:n})$ and rewrite the likelihood function as

$$\begin{aligned} L(y_{1:n}) &= \int \frac{p(y_{1:n}|x_{1:n}) p(x_{1:n})}{q(x_{1:n}|y_{1:n})} q(x_{1:n}|y_{1:n}) \, \mathrm{d}x_{1:n} \\ &= \int \omega(x_{1:n}, y_{1:n}) q(x_{1:n}|y_{1:n}) \, \mathrm{d}x_{1:n}, \end{aligned} \tag{2}$$

where the importance weight function is given by

$$\omega(x_{1:n}, y_{1:n}) = \frac{p(y_{1:n}|x_{1:n}) p(x_{1:n})}{q(x_{1:n}|y_{1:n})}. \tag{3}$$

We estimate the likelihood function (1) by generating N independent trajectories $x_{1:n}^{(1)}, \ldots, x_{1:n}^{(N)}$ from the importance density $q(x_{1:n}|y_{1:n})$ and computing

$$\widehat{L}(y_{1:n}) = \bar{\omega}, \qquad \bar{\omega} = \frac{1}{N} \sum_{i=1}^{N} \omega_i, \qquad \omega_i = \omega(x_{1:n}^{(i)}, y_{1:n}),$$

where ω_i is the realised importance weight function in (3) for $x_{1:n} = x_{1:n}^{(i)}$. Geweke (1989) showed that a central limit theorem applies to the importance sampling estimate provided that

$$\int \omega(x_{1:n}, y_{1:n})^2 q(x_{1:n}|y_{1:n}) \, \mathrm{d}x < \infty,$$

in which case the estimate is asymptotically normal and converges at the regular parametric rate to the true likelihood. A sufficient condition for the integral above to be finite is that the importance weight function is bounded from above. Koopman, Shephard, and Creal (2009) used extreme value theory to develop diagnostic tests to validate the existence of the variance of the importance weights.

The high-dimensional efficient importance sampling method of Richard and Zhang (2007) considers an importance sampler with the following form

$$q(x_{1:n}|y_{1:n}) = q(x_1|y_{1:n}) \prod_{t=2}^{n} q(x_t|x_{t-1}, y_{1:n}). \tag{4}$$

It follows that the we can factorise the importance weight as

$$\omega(x_{1:n}, y_{1:n}) = \frac{p(y_1|x_1)p(x_1)}{q(x_1|y_{1:n})} \prod_{t=2}^{n} \frac{p(y_t|x_t)p(x_t|x_{t-1})}{q(x_t|x_{t-1}, y_{1:n})}. \tag{5}$$

Richard and Zhang (2007) write the conditional densities $q(x_t|x_{t-1}, y_{1:n})$ in terms of a kernel in x_t and an integration constant

$$q(x_t|x_{t-1}, y_{1:n}) = \frac{k(x_t, x_{t-1}; a_t)}{\chi(x_{t-1}; a_t)}, \tag{6}$$

where

$$\chi(x_{t-1}; a_t) = \int k(x_t, x_{t-1}; a_t) \, \mathrm{d}x_t \tag{7}$$

and a_t is a vector of importance parameters which depends on $y_{1:n}$. At the initial period, we have the density

$$q(x_1|y_{1:n}) = \frac{k(x_1; a_1)}{\chi(a_1)}, \qquad \chi(a_1) = \int k(x_1; a_1) \, \mathrm{d}x_1.$$

Using (5) and (6), we express the importance sampling identity (2) as

$$\begin{aligned} & \int \frac{p(y_1|x_1)p(x_1)}{q(x_1|y_{1:n})} \prod_{t=2}^{n} \frac{p(y_t|x_t)p(x_t|x_{t-1})}{q(x_t|x_{t-1}, y_{1:n})} q(x_{1:n}|y_{1:n}) \, \mathrm{d}x_{1:n} \qquad (8) \\ = \quad & \chi(a_1) \int \frac{p(y_1|x_1)p(x_1)\chi(x_1; a_2)}{k(x_1; a_1)} \prod_{t=2}^{n} \frac{p(y_t|x_t)p(x_t|x_{t-1})\chi(x_t; a_{t+1})}{k(x_t, x_{t-1}; a_t)} q(x_{1:n}|y_{1:n}) \, \mathrm{d}x_{1:n}, \end{aligned}$$

with the convention that $\chi(x_n; a_{n+1}) \equiv 1$.

The EIS method seeks to find importance parameters a_t which minimise the variance of the ratio

$$\frac{p(y_t|x_t)p(x_t|x_{t-1})\chi(x_t; a_{t+1})}{k_t(x_t, x_{t-1}; a_t)}. \qquad (9)$$

The backward shifting of the period $t+1$ integration constant $\chi(x_t; a_{t+1})$ is essential for obtaining a numerically efficient estimate of (1) because the integration constant depends on the lagged state and therefore affects the variance of the global importance weight in (8). Koopman, Lucas, and Scharth (2014) note that when both the measurement and transition densities are linear and Gaussian, letting $k_t(x_t, x_{t-1}; a_t) \propto p(y_t|x_t)p(x_t|x_{t-1})\chi(x_t; a_{t+1})$ leads to an analytical backward-forward smoother and an efficient simulation smoother for this class of models. More generally, the EIS method aims to approximate the smoothing density $p(x_{1:n}|y_{1:n})$ as well as possible given the use of an importance density with the sequential form (4), which is required by the new SMC method of Section 3.

Richard and Zhang (2007) propose Algorithm 1 for selecting the importance parameters $a_{1:n}$. We highlight some critical aspects of it. The use of common random numbers (CRN) ensure the smoothness of the criterion function across successive iterations, facilitating the convergence of the algorithm. In some cases we can only implement CRNs via the inverse cumulative distribution method, which is computationally demanding. In this situation we can instead fix the number iterations beforehand; the convergence of the algorithm is not crucial, as typically only the initial iterations generate substantial reductions in the variance of the likelihood estimate (DeJong, Liesenfeld, Moura, Richard, and Dharmarajan 2012). For this reason, we recommend a non-strict convergence criterion in Algorithm 1.

Algorithm 1 can be subject to numerical instability leading to the divergence of

$a_{1:n}$, especially when the state vector x_t is multivariate and when using the natural sampler $p(x_{1:n})$ to draw the initial set of state trajectories. Koopman, Lucas, and Scharth (2013) argue that we can typically eliminate this problem by reducing the step size at the initial iterations of the algorithm. We can achieve this by replacing the measurement density $p(y_t|x_t^{(s)})$ in (11) by $p(y_t|x_t^{(s)})^{\zeta_k}$, where $\zeta_k \in (0, 1]$ gradually increases with k. Numerical errors may also indicate the use of an excessively low number of samples S to compute the regressions.

Even though we have not made any additional assumptions regarding the state space model, the practical applicability of the EIS method relies on the availability of a kernel $k(x_t, x_{t-1}; a_t)$ that is able to accurately approximate the numerator in (9), leads to a tractable least squares regression within Algorithm 1, and ensures finite variance in the estimation of the likelihood. The EIS method becomes less interesting when the minimisation problem is nonlinear, in which case the procedure becomes computationally expensive. That suggests that the EIS method is potentially applicable when the approximating kernel belongs to the exponential family. Existing applications focus on kernels which are conjugate with $p(x_t|x_{t-1})$ or $p(x_t|x_{t-1})\chi(x_t; a_{t+1})$.[1] In Section 4, we build on the ideas in Kleppe and Liesenfeld (2013) to consider a new case in which the state transition has an additive error that follows the Student's t distribution, leading to a conditionally Gaussian setting that is amenable to the use of exponential family kernels.

The simplest way to guarantee the boundedness of the importance weight 5 is the defensive mixture approach of Hesterberg (1995). In this method, we replace the the sequential proposal $q(x_t|x_{t-1}, y_{1:n})$ by the two-component mixture

$$q^{\triangle}(x_t|x_{t-1}, y_{1:n}) = \pi q(x_t|x_{t-1}, y_{1:n}) + (1 - \pi)p(x_t|x_{t-1}),$$

with $0 < \pi < 1$. The importance weight then becomes bounded since

$$\omega_t = \frac{p(y_t|x_t)p(x_t|x_{t-1})}{\pi q(x_t|x_{t-1}, y_{1:n}) + (1 - \pi)p(x_t|x_{t-1})} < \frac{p(y_t|x_t)}{1 - \pi}.$$

The defensive mixture approach has a cost in terms of efficiency since it relies in part on

[1] Liesenfeld and Richard (2003) and Richard and Zhang (2007) originally considered linear Gaussian and inverse Gamma transitions respectively. Nonlinear transitions with additive Gaussian innovations follow easily from the linear case. Liesenfeld and Richard (2010) consider truncated normal states.

Algorithm 1 Efficient importance parameters

▷ Initialise the iteration index $k \leftarrow 0$.

▷ Set the initial values for the importance parameters $a_{1:n}^{[0]}$ and denote the associated importance density as $q^{[0]}(x_{1:n}|y_{1:n})$. A generic and easy to implement choice to initialise the algorithm is the natural sampler, i.e. $q^{[0]}(x_{1:n}|y_{1:n}) = p(x_{1:n})$.

▷ Draw a set of common random numbers (CRN) $u_{1:S}$.

while convergence criterion is not met **do**

▷ $k \leftarrow k + 1$

▷ Obtain S trajectories $x_{1:n}^{(s)} \sim q^{[k-1]}(x_{1:n}|y_{1:n})$ using the CRNs $u_{1:S}$.

for t=n:-1:1 **do**

▷ Solve the least squares problem

$$a_t^{[k]}, \gamma_t^{[k]} = \underset{a_t, \gamma_t}{\operatorname{argmin}} \sum_{s=1}^{S} \lambda(y_t, x_t^{(s)}, x_{t-1}^{(s)}, a_t, a_{t+1}^{[k]}, \gamma_t)^2 \tag{10}$$

where

$$\lambda(y_t, x_t^{(s)}, x_{t-1}^{(s)}, a_t, a_{t+1}^{[k]}, \gamma_t) = \log \left(\frac{p(y_t|x_t^{(s)}) p(x_t^{(s)}|x_{t-1}^{(s)}) \chi(x_t^{(s)}; a_{t+1}^{[k]})}{\gamma_t k(x_t^{(s)}, x_{t-1}^{(s)}; a_t)} \right), \tag{11}$$

with $\chi(x_n^{(s)}; a_{n+1}^{[k]}) \equiv 1$. The normalising constant γ_t plays no further role in the method.

end for

end while

▷ Set the efficient importance density as $q(x_{1:n}|y_{1:n}) = q^{[k]}(x_{1:n}|y_{1:n})$.

blind draws from $p(x_t|x_{t-1})$. This disadvantage is mitigated by particle EIS framework through resampling. When π is zero we obtain the bootstrap filter, which is commonly used in the particle filter due to its simplicity.

Finally, we note that more efficient procedures are available when the state transition equation is linear and Gaussian. In this situation the marginal importance density $q(x_t|y_{1:n})$ is available analytically for a Gaussian sampler, enabling numerical and computational gains over the standard algorithm using the results in Koopman, Lucas, and Scharth (2014).

3 Particle efficient importance sampling

The particle efficient importance sampling method in this section consists of embedding the period t proposal $q(x_t|x_{t-1}, y_{1:n})$ obtained by the efficient importance sampling method of Richard and Zhang (2007) into a sequential Monte Carlo algorithm that combines the numerical efficiency of these sequential densities as approximations to $p(y_t|x_t)p(x_t|x_{t-1})\chi(x_t; a_{t+1})$ with resampling steps that ensure that the variance of the target estimate does not grow exponentially with the time series dimension. We base the computations on the auxiliary particle filter (APF) algorithm, which allows for the use of nonstandard resampling weights as required by our method. The APF leads to an appropriate procedure for unbiased likelihood evaluation even though our offline proposal is by construction not geared to filtering. Sections 3.1 and 3.2 motivate and describe the new method. Algorithm 2 provides pseudo code for implementation.

3.1 Particle methods

Particle filtering methods recursively obtain a sequence of particles $\{x_{1:t}^i\}_{i=1}^N$ and associated weights $\{W_t^i\}_{i=1}^N$ that approximate the filtering distribution $p(x_{1:t}|y_{1:t})$ at each time period as

$$\widehat{p}(x_{1:t}|y_{1:t}) = \sum_{i=1}^{N} W_t^i \delta_{x_{1:t}^i}(x_{1:t}),$$

where $\delta_{x_{1:t}^i}(x_{1:t})$ denotes the Dirac delta mass located at $x_{1:t}^i$.

The basic particle filter method is based on the sequential importance sampling (SIS) algorithm. Suppose that at the end of period $t-1$ we have a particle system

$\{x^i_{1:t-1}, W^i_{t-1}\}_{i=1}^N$ which approximates the filtering density $p(x_{1:t-1}|y_{1:t-1})$. Upon the arrival of a new observation y_t, SIS updates the particle system by propagating the particles $x^i_{1:t-1}$ using the importance distribution

$$q(x^i_t|x^i_{t-1}, y_{1:n}) \propto k(x^i_t, x^i_{t-1}; a^i_t)$$

and reweighing each particle trajectory $x^i_{1:t}$ according to

$$w^i_t = W^i_{t-1} \frac{p(y_t|x^i_t)p(x^i_t|x^i_{t-1})}{q(x^i_t|x^i_{t-1}, y_{1:n})}, \tag{12}$$

with corresponding normalised weights calculated as

$$W^i_t = w^i_t / \sum_{i=1}^N w^i_t.$$

At each period, we can also estimate the likelihood contribution $p(y_t|y_{1:t-1})$ as

$$\widehat{p}(y_t|y_{1:t-1}) = \sum_{i=1}^N w^i_t.$$

It is straightforward to recognise that the efficient high-dimensional importance sampling method of Section 2.2 is a SIS method in which the sequential proposal density $q(x^i_t|x^i_{t-1}, y_{1:n})$ has the kernel $k(x^i_t, x^i_{t-1}; a_t)$ which we construct according to Algorithm 1. In the EIS method, which approximates the smoothing density, we obtain all the importance parameters $a_1, \ldots, a_n$ jointly using the full sample information $y_{1:n}$. However, these auxiliary parameters do not depend on the particle trajectory i. That contrasts with the use of SIS in the particle filter literature, in which $q(x^i_t|x^i_{t-1}, y_{1:n}) = q(x^i_t|x^i_{t-1}, y_t)$. We refer to this case as online sequential importance sampling. In the online SIS method, we can tailor the importance parameters in the proposal kernel $k(x^i_t, x^i_{t-1}; a^i_t)$ to each inherited particle (indexed by i), but do not use the future observations $y_{t+1:n}$ when selecting a^i_t.

The second fundamental ingredient of particle methods is resampling, which reduces the impact of the weight degeneracy problem on the performance of the filter in subsequent periods. It can be shown that as the number of iterations of the SIS method increases, the normalised weights of the particle system become concentrated

on fewer particles. Eventually, the weight of a single particle converges to one; see for example Chopin (2004). It follows that the variance of estimates obtained using the SIS method grows exponentially in time. Resampling solves this problem by randomly replicating particles from the current population according to their weights, therefore discarding particles with low probability mass.

The standard sequential importance sampling with resampling (SISR) method resamples N particles $\{x_t^n\}_{i=1}^N$ with probabilities $\{W_t^i\}_{i=1}^N$ and assigns equal weights $W_t^i = 1/N$ to all particles at the end of each time period. Several unbiased resampling schemes that improve upon multinomial resampling are proposed in the literature; some examples are systematic resampling (Kitagawa 1996) and residual resampling (Liu and Chen 1998). The effective sample size defined as $ESS = 1/\sum_{i=1}^N (W_t^i)^2$ is a standard tool for monitoring the degeneracy of particle systems. Since resampling introduces its own source of error by reducing the number of distinct particles at the current period, a straightforward improvement to the basic algorithm is to perform resampling only when he particle weights reach a certain degeneracy threshold.

3.2 Particle EIS

Since the EIS algorithm is a sequential importance sampler, a SISR version of the method based on the global importance density $q(x_{1:n}|y_{1:n})$ which uses (12) as resampling weights follows immediately by using the procedure described in Section 3.1. Even though this approach leads to a valid algorithm, we argue that the standard SISR resampling weights are inefficient in this case because the EIS kernel $k(x_t^i, x_{t-1}^i; a_t^i)$ targets $p(y_t^i|x_t^i)p(x_t^i|x_{t-1}^i)\chi(x_t^i; a_{t+1})$, which contrasts to $p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)$ for an online SIS kernel.

The critical step in the particle efficient importance sampling method is the introduction of the forward weights

$$w_t^{+i} = W_t^i \ \chi(x_t^i; a_{t+1}), \tag{13}$$

leading to the normalised resampling weights

$$W_t^{+i} = w_t^{+i} / \sum_{i=1}^N w_t^{+i}.$$

We now track the degeneracy of the particle system using the forward effective sample size $ESS^+ = 1/\sum_{i=1}^N (W_t^{+i})^2$.

The justification for the forward weights follows immediately from the construction of the efficient importance sampler. Since

$$w_t^{+i} \propto W_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)\chi(x_t^i; a_{t+1})}{q(x_t^i|x_{t-1}^i, y_{1:n})},$$

the introduction of the integration constant for the next period $\chi(x_t^i; a_{t+1})$ matches the importance density $q(x_t^i|x_{t-1}^i, y_{1:n})$ to its target in the minimisation problem (10), appropriately balancing the resampling weights.

The use of alternative resampling weights leads us to implement the particle EIS method using the auxiliary particle filter (APF) algorithm introduced by Pitt and Shephard (1999). The auxiliary particle filter algorithm was designed to improve the efficiency of online particle filters by incorporating period t information when resampling the particles after period $t-1$, anticipating which particles will be in regions of high probability mass after propagation. Here, we use the APF framework simply to obtain correct importance weights and likelihood increment estimates when using the forward weights for resampling; the particle EIS obtains the sequential proposals offline and therefore is not a filter. Conceptually, considering a proposal $q(x_t^i|x_{t-1}^i, y_{1:n})$ instead of $q(x_t^i|x_{t-1}^i, y_{1:t})$ is a straightforward extension of the APF algorithm.

When the forward effective sample size falls below a threshold after period $t-1$, we store the forward weights $\{w_{t-1}^{+i}\}_{i=1}^N$ and resample N particles $\{x_{t-1}^i\}_{i=1}^N$ with probabilities $\{W_{t-1}^{+i}\}_{i=1}^N$ and set $W_{t-1}^i = 1/N$ for all the particles. From the APF algorithm, the importance weights after resampling at the end of period $t-1$ and propagating the particles using the importance density $q(x_t^i|x_{t-1}^i, y_{1:n})$ are

$$w_t^i = W_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{\chi(x_{t-1}^i; a_t)q(x_t^i|x_{t-1}^i, y_{1:n})} = W_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{k(x_t^i, x_{t-1}^i, a_t)},$$

where x_{t-1}^i are the particles after resampling. When we perform no resampling at the end of the previous iteration, the calculation of the weights and the estimation follows exactly as in the sequential importance sampling algorithm in Section 3.1

Pitt, Silva, Giordani, and Kohn (2012) gives an estimator of the likelihood contri-

bution $p(y_t|y_{t-1})$ based on the auxiliary particle filtering algorithm.[2] After resampling with weights (13), the estimate is

$$\widehat{p}(y_t|y_{1:t-1}) = \left(\sum_{i=1}^{N} w_{t-1}^{+i}\right)\left(\sum_{i=1}^{N} w_t^i\right),$$

where $\{w_{t-1}^{+i}\}_{i=1}^N$ are the forward weights prior to resampling. Proposition 7.4.1 of Del Moral (2004) establishes the unbiasedness of the general auxiliary particle filter estimator, while Pitt, Silva, Giordani, and Kohn (2012) provide an alternative proof of the same result. The use of a proposal $q(x_t^i|x_{t-1}^i, y_{1:n})$ does not affect the assumptions behind these theorems since the data $y_{1:n}$ is fixed and the unbiasedness is with respect to the random variables generated by the particle filter. Unbiasedness with respect to the random EIS auxiliary parameters obtained from Algorithm 1 follows in a straightforward way from the law of iterated expectations. Algorithm 2 provides the pseudo code for the particle EIS method.

3.3 Particle EIS, particle filters and look-ahead approaches

We intuitively expect importance sampling methods that construct a global approximation to the smoothing density $p(x_{1:n}|y_{1:n})$ to perform better for likelihood estimation than online sequential importance sampling methods, which construct a proposal at each time t using y_t. To better understand the numerical performance of efficient importance sampling, we start by considering the optimal (but often infeasible) online sequential importance sampler. The conditionally optimal importance distribution for online SIS, in the sense of minimising the variance of the importance weights at each period, is

$$q^*(x_t|x_{t-1}, y_t) = p(x_t|x_{t-1}, y_t) = \frac{p(y_t|x_t)p(x_t|x_{t-1})}{p(y_t|x_{t-1})}, \tag{14}$$

in which case the importance weight is

$$w_t^{i*} = p(y_t|x_{t-1}) = \chi^*(x_{t-1}^i; a_t^i). \tag{15}$$

It is well known that the conditionally optimal importance density does not guar-

[2]This estimator was previously introduced in the working paper by Pitt (2002)

Algorithm 2 Particle Efficient Importance Sampling

▷ Obtain the efficient importance density $q(x_{1:n}|y_{1:n})$ using Algorithm 1 or one of its variations.

At time $t = 1$:

for i=1:N **do**

▷ Draw $x_1^i \sim q(x_1|y_{1:n})$.

▷ Compute the importance weight:

$$w_1^i = \frac{p(y_t|x_1^i)p(x_1^i)}{q(x_1^i|y_{1:n})}.$$

end for

▷ Calculate the estimate of the likelihood contribution as $\widehat{p}(y_1) = \sum_{i=1}^N w_1^i/N$.

▷ Compute the normalised weights $W_1^i = w_1(x_1^i)/\sum_{i=1}^N w_1(x_1^i)$, $i = 1, \ldots, N$.

▷ Compute the forward weights $w_1^{+i} = W_1^i \cdot \chi(x_1; a_2)$, $i = 1, \ldots, N$.

▷ Compute the normalised forward weights $W_1^{+i} = w_1^{+i}/\sum_{i=1}^N w_1^{+i}$, $i = 1, \ldots, N$.

▷ Compute the effective sample size $ESS = 1/\sum_{i=1}^N (W_1^{+i})^2$.

At time $t \geq 2$:

▷ If the effective sample size is below a certain threshold, resample N particles $\{x_{t-1}^i\}_{i=1}^N$ with probabilities $\{W_{t-1}^{+i}\}_{i=1}^N$ and set $W_{t-1}^i = 1/N$, for $i = 1, \ldots, N$. Store $\{w_{t-1}^{+i}\}_{i=1}^N$.

for n=1:N **do**

▷ Draw $x_t^i \sim q(x_t^i|y_t, x_{t-1}^i)$.

if resampling **then**

▷ Compute the importance weight

$$w_t^i = W_{t-1}^i \times \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{k(x_t^i, x_{t-1}^i, a_t)}.$$

else

▷ Compute the importance weight

$$w_t^i = W_{t-1}^i \times \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i, y_{1:n})}.$$

end if

end for

(continued on the next page)

Algorithm 2 (continued)

▷ Calculate the estimate of the likelihood contribution as $\widehat{p}(y_t|y_{1:t-1}) = \sum_{i=1}^{N} w_t^i$ if there is no resampling and as $\widehat{p}(y_t|y_{1:t-1}) = (\sum_{i=1}^{N} w_{t-1}^{+i})(\sum_{i=1}^{N} w_t^i)$ if there is resampling.
▷ Compute the normalised weights $W_1^i = w_t^i / \sum_{i=1}^{N} w_t^i$, $i = 1, \ldots, N$.
▷ Compute the forward weights $w_t^{+i} = W_t^i \cdot \chi(x_t; a_{t+1})$, $i = 1, \ldots, N$.
▷ Compute the normalised forward weights $W_t^{+i} = w_t^{+i} / \sum_{i=1}^{N} w_t^{+i}$, $i = 1, \ldots, N$.
▷ Compute the effective sample size $ESS^+ = 1/\sum_{i=1}^{N} (W_t^{+i})^2$.

antee good performance even when used within an SISR algorithm. The reason is transparent from the efficient importance sampling framework of Section 2.2 and equation (15): the online sequential importance sampling proposal ignores the integration constants $\chi(x_t^i; a_{t+1}^i)$, which may have high variance. Hence, the conditionally optimal importance density can still result in inaccurate estimation and frequent resampling if the variance of $p(y_t|x_{t-1})$ is high.

The EIS method of Richard and Zhang (2007) directly increases efficiency by incorporating the integration constant $\chi(x_t; a_{t+1})$ into the variance minimisation problem (10) for each t. These least squares problems are all interlinked by the shifted integration constants, which is how the method obtains a global approximation. It straightforward to see that the optimal kernel $k^*(x_t, x_{t-1}^i, a_t) \propto p(y_t|x_t)p(x_t|x_{t-1})\chi(x_t^i; a_{t+1})$ which the EIS method approximates leads to importance weights with zero variance. The particle efficient importance sampling method therefore fully combines the numerical efficiency of global importance densities targeting the smoothing distribution with the benefits of resampling.

We make three qualifications. First, we focus on likelihood estimation, for which importance sampling and particle filters are competing approaches. If the goal is to sequentially process the data, then particle filters are the appropriate tool (possibly by incorporating look-ahead information). Second, global importance sampling comes at the cost of greater difficulty in designing a high-dimensional proposal $q(x_{1:n}|y_{1:n})$ in comparison with devising the low-dimensional sequential proposal densities $q(x_t|x_{t-1}, y_t)$ used in online sequential importance sampling. This task will inevitably be highly model specific. Third, we recall that the importance parameters in the online sequen-

tial importance kernel $k(x_t^i, x_{t-1}^i, a_t^i)$ can depend on the particle index, which is not the case with efficient importance sampling. Therefore, online SIS can lead to a more accurate approximation to $p(y_t|x_t)p(x_t|x_{t-1})$ in particular for any given kernel $k(x_t^i, x_{t-1}^i, a_t^i)$ when compared to existing global importance sampling methods. However, we argue that this extra flexibility in the online SIS method is typically of limited practical value, as it is computationally costly to obtain efficient importance parameters for every particle.

Our approach is distinct to looka-head SMC methods proposed in literature, such as the block-sampling method of Doucet, Briers, and Sénécal (2006). See Lin, Chen, and Liu (2013) for a review. These algorithms have the goal of sequentially approximating $p(x_t|y_{1:t+\Delta})$ at every period t, for a small number of steps Δ, leading to more accurate sequential inference on the states as Δ increases. These methods are based on constructing at each period t block proposals or other schemes that explore future information. In the particle EIS method, we obtain an approximation to the smoothing density $p(x_{1:n}|y_{1:n})$ using a high-dimensional importance density $q(x_{1:n}|y_{1:n})$ which has the sequential form (4). In contrast to the lookahead algorithms, we obtain all the sequential proposals $q(x_t|x_{t-1}, y_{1:n})$ jointly using the EIS algorithm.

3.4 Antithetic variables

Antithetic sampling is a variance reduction method based on generating negatively correlated draws from a sampling density. The technique is often instrumental for the success of importance sampling strategies, see for example Durbin and Koopman (2000). In this section we propose a modification of the particle EIS method in order to incorporate the use of antithetic variables. We focus on a particular setting that encompasses our illustrations in Sections 4 and 5.

Suppose that we can formulate the dynamics of x_t under the sequential importance density $q(x_t|x_{t-1}, y_{1:n})$ using the equation

$$x_t = H_t(x_{t-1}, \xi_t; a_t), \tag{16}$$

where $H_t(.)$ is a nonlinear function and ξ_t is a random variable following a symmetric distribution, which we assume to have mean zero without loss of generality. Both H_t and the distribution of ξ_t depend on the state transition equation and the importance

parameters.

Starting from a particle system $\{x_{t-1}^i, W_{t-1}^i\}_{i=1}^N$, we implement antithetic variables at period t by drawing $N/2$ innovations $\xi_t^1, \ldots, \xi_t^{N/2}$ and propagating the corresponding first half of particles by using equation (16) to calculate $x_t^i = H_t(x_{t-1}^i, \xi_t^i; a_t)$ for $i \leq N/2$. We then compute the antithetic draws as $x_t^i = H_t(x_{t-1}^i, -\xi_t^{i-N/2}; a_t)$ for $N/2 < i \leq N$. When the forward effective sample size reaches the defined degeneracy level at the end of period t, we resample only $N/2$ particles $\{x_t^j\}_{j=1}^{N/2}$ with probabilities $\{W_t^{+i}\}_{i=1}^N$ and duplicate each of then so that $x_t^{N/2+j} = x_t^j$ for $j \leq N/2$ after resampling. As before, normalised weights after resampling are $W_t^i = 1/N$ for all N particles. We follow this procedure at every period, where at time $t = 1$ we replace the sampling equation in (16) by $x_1 = H_1(\xi_1; a_1)$.

In the context of particle EIS, antithetic variables have the side effect of amplifying the loss of information when performing resampling because we reduce particle diversity through duplication. Nevertheless, we have found experimentally that this version of the algorithm strongly outperforms the standard version without variance reduction for the models we consider in Sections 4 and 5. The reason for the efficiency gain is that resampling takes place infrequently within the particle EIS method. We therefore adopt antithetic sampling throughout the rest of the paper.

3.5 Particle smoothing

When a full proposal $q(x_{1:n}|y_{1:n})$ is available, it is straightforward to apply importance sampling to estimate functionals based on the full smoothing density $p(x_{1:n}|y_{1:n})$ in $O(N)$ operations after we calculate the importance weights (3). See for example Durbin and Koopman (2001). It is therefore natural to also consider the use of particle EIS for smoothing. Estimating smoothing densities becomes computationally more challenging for particle methods due to resampling, which progressively reduces the number of distinct particles in earlier parts of the sample. Particle methods provide an accurate approximation $p(x_{j:n}|y_{1:n})$ only for j relatively close to n. While we can expect the particle EIS method to degenerate slower due to infrequent resampling, the standard algorithm will still suffer from this problem for $j \ll n$.

Alternative smoothing schemes that avoid this problem based on forward filtering-backward smoothing recursions and the generalised two-filter formula have been de-

veloped in the literature, e.g. Godsill, Doucet, and West (2004). These algorithms often have a computational cost which is proportional to $N^2(R+1)$, where R is the number of resampling steps. More sophisticated algorithms with computing time proportional to $N(R+1)$ are now also available, see for example Fearnhead, Wyncoll, and Tawn (2010). While an investigation of particle smoothing is out of the scope of this paper, we note that when we are able to successfully implement the particle EIS method and directly target the smoothing distribution $p(x_{1:n}|y_{1:n})$ by an importance sampling approximation, we can expect both the number of resampling steps and the number of particles required to reach a certain level of numerical accuracy to be lower than what is the case for standard algorithms, so that the new method can be a useful tool for particle smoothing and the particle Gibbs algorithm of Andrieu, Doucet, and Holenstein (2010).

4 Simulation study

This section investigates how particle EIS compares to the EIS method and standard particle filters for likelihood estimation. The motivation for contrasting these online and offline approaches is that they are unbiased likelihood estimation methods that we can apply to Bayesian inference using algorithms such as the particle marginal Metropolis-Hastings (PMMH) method of Andrieu, Doucet, and Holenstein (2010). We consider these applications directly in Section 5. Section 4.1 describes the models in the simulation study, Section 4.2 discusses the alternative methods and implementation details and Section 4.4 provides the comparison methodology and presents the results.

4.1 Univariate and bivariate stochastic volatility models

We consider two stochastic volatility (SV) models in our simulation study: a univariate two-factor stochastic volatility model with leverage effects and a simple bivariate specification. We assume the following measurement and transition equations for the univariate specification

$$
\begin{aligned}
y_t &= \exp([c + x_{1,t} + x_{2,t}]/2)\varepsilon_t, \qquad t = 1, \ldots, n \\
x_{i,t+1} &= \phi_1 x_{i,t} + \rho_1 \sigma_{i,\eta} \varepsilon_t + \sqrt{1-\rho_i^2}\sigma_{i,\eta}\eta_{i,t}, \qquad i = 1, 2
\end{aligned}
$$

where the return innovations are i.i.d. and have the standardised Student's t distribution with ν degrees of freedom and $1 > \phi_1 > \phi_2 > -1$ for stationarity and identification. Likewise, we consider two settings for the state disturbances: in the first, they follow the $\mathrm{N}(0,1)$ distribution, while in the second they follow the standardised Student's t distribution with 10 degrees of freedom. The parameters for the simulation exercise reflect typical values found by empirical studies: $\phi_1 = 0.995$, $\sigma^2_{1,\eta} = 0.005$, $\rho_1 = -0.2$, $\phi_2 = 0.9$, $\sigma^2_{2,\eta} = 0.03$, $\rho_2 = -0.5$ and $\nu = 10$.

The bivariate stochastic volatility model follows the specification originally suggested by Harvey, Ruiz, and Shephard (1994). The model is

$$y_t \sim \mathcal{MVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2_{1,t} & \rho_t\sigma_{1,t}\sigma_{2,t} \\ \rho_t\sigma_{1,t}\sigma_{2,t} & \sigma^2_{2,t} \end{bmatrix}\right), \qquad t = 1, \ldots, n,$$

$$\sigma^2_{1,t} = \exp(c_1 + x_{1,t}), \qquad \sigma^2_{2,t} = \exp(c_2 + x_{2,t}), \qquad \rho_t = \frac{1 - \exp(-c_3 - x_{3,t})}{1 + \exp(-c_3 - x_{3,t})},$$

where each state follows an AR(1) process,

$$x_{i,t+1} = \phi_i x_{i,t} + \eta_{i,t}, \qquad \eta_{i,t} \sim \mathrm{N}(0, \sigma^2_{i,\eta}), \qquad i = 1, 2, 3.$$

The parameters for the simulation study DGP are $c_1 = c_2 = 0$, $\phi_1 = \phi_2 = 0.98$, $\sigma_{1,\eta} = \sigma_{2,\eta} = 0.15$, $c_3 = 1$, $\phi_3 = 0.99$, $\sigma_{3,\eta} = 0.05$.

4.2 Alternative methods and implementation details

We implement four alternative likelihood estimation methods: the standard EIS method described in Section 2.2 and three particle filter algorithms. The first particle filter algorithm is the bootstrap filter (BF), which corresponds to the sequential importance sampling resampling (SISR) method outlined in Section 3.1 with the state transition density as a proposal distribution, that is $q(x_t|x_{t-1}, y_t) = p(x_t|x_{t-1})$. The BF has the advantage of being very simple to implement. The second particle filter method is SISR using a Gaussian proposal which we construct via a second order Taylor expansion of $p(y_t|x^i_t)p(x^i_t|x^i_{t-1})$ around the mode (conditional on x^i_{t-1}). We only consider this method for the univariate SV model with Gaussian state innovations. We label it SISR(2) in

the tables. The final particle filter method is a zero order auxiliary particle filter as in Pitt and Shephard (1999). As in the BF, the proposal is $q(x_t|x_{t-1}, y_t) = p(x_t|x_{t-1})$, but the resampling weights become $W_{t-1}^i\, p(y_t|\mu_t(x_{t-1}^i))$, where $\mu_t(x_{t-1}^i)$ is the mean of x_t given x_{t-1}^i according to the state transition. We denote this method by APF(0) in the tables.

Appendix A provides the implementation for details for a general EIS algorithm that allows for nonlinear state transitions and additive Student's t state innovations. The algorithm follows Kleppe and Liesenfeld (2013) and uses a data augmentation scheme that treats the state disturbances as normal-inverse gamma mixtures. We consider two versions of the method. The first only approximates the Gaussian part of the state transition, while the second does importance sampling for both the Gaussian and inverse gamma components. We refer to the two algorithms partial and full EIS respectively. The implementation for the models with Gaussian state disturbances follows simply by omitting the inverse gamma component and the associated steps. For the bivariate SV model the state transition equation is also linear, so that we adopt the results in Koopman, Lucas, and Scharth (2014) to improve computational efficiency. The number of samples for Algorithm 1 is $S = 50$. We find that it is important to use the step size reduction modification to Algorithm 1 mentioned in Section 2.2 to ensure that all EIS implementations are free of occasional numerical instability. We also recommend setting the leverage effect coefficients to zero at the initial iterations of the algorithm for the univariate SV model. We carefully assessed the behaviour of the importance weights in all the examples below and found no indication of possible infinite variance problems.

We adopt systematic resampling in all the particle methods. When running the particle filters, we resample when the effective sample size divided by the number of particles falls below 0.5. In the particle efficient importance sampling method, we resample if the forward effective sample size divided by the number of particles is under 0.9, a choice based on experimentation. We use antithetic variables for variance reduction in the EIS and particle EIS methods. We have implemented all methods efficiently using MATLAB mex files coded in C. All the reported computing times are based on a computer equipped with an Intel Xeon 3.40 GHz processor with four cores. They do not involve any parallel processing, except in Table 6 of Section 5.

4.3 Illustration

To illustrate how the particle EIS method compares to the standard EIS algorithm as the dimension of integration increases, we focus on a single simulated trajectory for the bivariate SV model with $n = 10,000$; the same results hold generally. At every period t, we compute 5,000 independent log-likelihood estimates $\widehat{L}(y_{1:t})$ using the EIS and particle EIS methods with $N = 50$. To put the magnitudes in context, we note that recent theoretical results discussed in Section 5.1 indicate that the optimal variance of the log-likelihood estimate for Bayesian inference is typically between 0.5 and 1 under certain assumptions.

Figure 1 displays the results. The first panel shows the effective sample size for the EIS method as we increase the size of the time series. We base this on a single run with $N = 10,000$ importance samples. The plot suggests that the EIS method provides an almost perfect approximation to the smoothing density in very short samples, obtaining an ESS of nearly one. However, as predicted by the theory in Chopin (2004) the ESS decays exponentially as t increases. By the end of the sample, less than 5% of the importance samples have relevant weight.

The second panel compares the EIS and particle EIS mean squared errors (MSE) for estimating the log-likelihood (where we approximate the true likelihood up to a negligible error by using all the particles generated by our replications). It shows that even though the MSEs are similar in short samples, the EIS method is subject to the expected exponential increase in MSE, while the particle EIS errors remain low for a large t . The following panel zooms in on the MSE for the particle EIS method, showing that it grows linearly with the time series dimension. We note that even though we observe an exponential deterioration in the performance of the EIS method, the method is still remarkably numerically efficient: the MSE is negligible for a small t. Even at the end of the sample, the MSE for EIS is around 2, suggesting that only 100-250 importance samples would be needed to obtain a theoretically optimal variance for Bayesian inference when $n = 10,000$. This result illustrates why the method has been successfully applied to a variety of models in large time series, despite its adverse properties as n gets larger.

The last two panels plot the histograms of the EIS and particle EIS log-likelihood estimates for the entire simulated time series ($n = 10,000$). While the particle EIS es-

timates are approximately Gaussian (we cannot reject the normality of the distribution using the Jarque-Bera test), the EIS estimates are right skewed and subject to some outliers. In addition to robustness considerations, the Gaussianity of the log-likelihood estimates is important because results on the optimal implementation Bayesian estimation methods rely on this assumption, see Section 5.1. The non-Gaussianity of the EIS estimates is a natural consequence of the exponential decay in ESS seen in the first panel. Because we take the log of the likelihood estimate, this non-Gaussianity is associated with an adverse effect: the variance and MSE of the EIS method estimate does not scale at rate $1/N$ at this range. For this example, our results indicate that doubling the number of importance samples from $N = 50$ to $N = 100$ only reduces the MSE by 30%. This implies that larger values of N are favourable to the particle EIS method when comparing the efficiency of the two approaches.

4.4 Likelihood estimation analysis

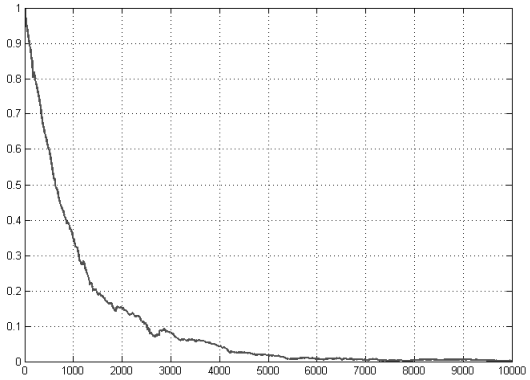
We implement a simulation study as follows. We draw 500 trajectories of time series dimensions $n = 2,500$, $5,000$ and $10,000$ using the three univariate SV and the bivariate SV data generating processes described in Section 4.1. For each realisation, we perform 100 independent log-likelihood evaluations at the DGP parameters using particle efficient importance sampling and the alternative methods listed in Section 4.2. The number of particles is $N = 50$ for all methods.

We compare the expected MSE and variance of the log-likelihood estimates with respect to DGP. We compute the estimates

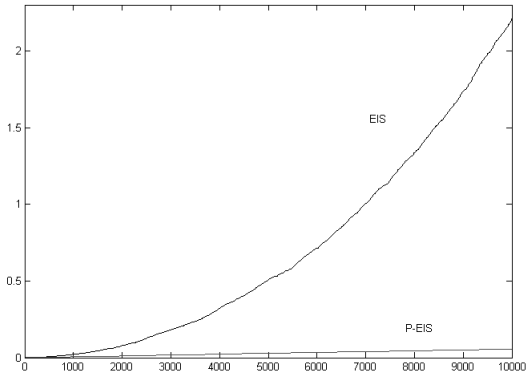
$$\widehat{E}\left[\mathrm{MSE}(\log \widehat{L})\right] = \sum_{i=1}^{500}\left(\sum_{j=1}^{20}\frac{(\widehat{\log L_{i,j}} - \log \widetilde{L_i})^2}{100}\right)\Big/ 500$$

where i indexes the DGP realisations, j the independent likelihood evaluations, $\widehat{\log L_{i,j}}$ are the corresponding likelihood estimates, and $\widetilde{\log L_i}$ is a highly accurate approximation to the true log-likelihood for trajectory i obtained by combining all the 20 replications of the EIS and particle EIS methods.

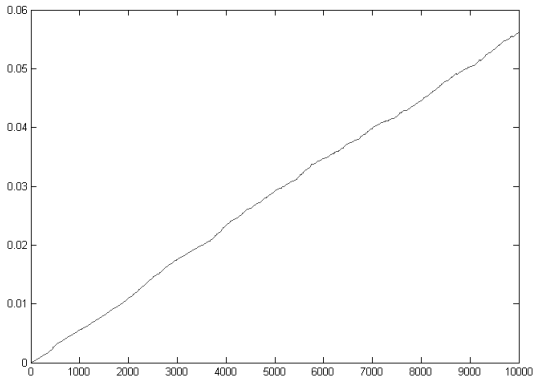
It is also important to take the computing times into account when comparing the likelihood estimation methods. Let h index the method. The computing time is



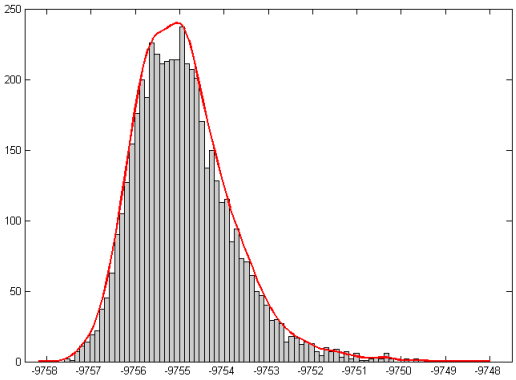
(a) EIS effective sample size as a function of t



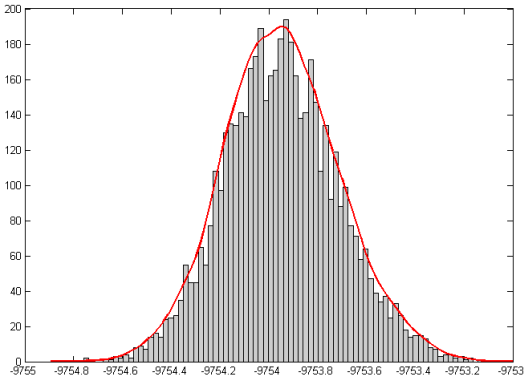
(b) Log-likelihood MSE as a function of t



(c) Log-likelihood MSE as a function of t (P-EIS only)



(d) EIS log-likelihood estimates ($n = 10,000$) and kernel density plot



(e) Particle EIS log-likelihood estimates ($n = 10,000$) and kernel density plot

Figure 1: Illustrating the performance of particle EIS

(approximately) an affine function of the number of particles

$$\text{Computing time}^{h,N} = \tau_1^h + N\tau_2^h, \tag{17}$$

where τ_1^h is the overhead cost per likelihood evaluation, which mainly corresponds to the time to run Algorithm 1, and τ_2^h is the additional computing time per particle. We have that $\tau_1^h = 0$ for the particle filters. In the tables we label τ_1^h and $N\tau_2^h$ as EIS density time and likelihood time respectively. We calculate these quantities by running the methods with two different values of N. We define the efficiency relative to the standard EIS method benchmark as

$$\text{Efficiency}^{h,N} = \frac{\widehat{\text{MSE}}(\log \widehat{L}^{h,N})}{\widehat{\text{MSE}}(\log \widehat{L}^{b,N})} \left(1 + \frac{\tau_1^b + N\tau_2^b - \tau_1^h - N\tau_2^h}{N\tau_2^h}\right)^{-1}, \tag{18}$$

where h indexes the method, b indexes the benchmark and $\widehat{\text{Var}}(\log \widehat{L}^{h,N})$ denotes the estimated variance of method h with N particles. Assuming that the MSE of the log of the likelihood estimate scales at rate $1/N$, the efficiency measure estimates the variance associated with algorithm h for a number of particles N' such that $\tau_1^h + N'\tau_2^h = \tau_1^b + N\tau_2^b$. It therefore estimates the variance of method h when we give it the same total computing time as the benchmark.

Tables 1-3 present the results. Three main findings appear in all cases. First, the particle EIS method brings large reductions in MSE over the standard EIS method. When $n = 10,000$ the decrease in variance ranges from 91% for the univariate SV model with Student's t state disturbances to 97% for the bivariate SV specification. These gains come with almost no increase in computational time since the new method resamples infrequently. Second, the use of a global approximation in the EIS and particle EIS methods leads to substantial gains in numerical efficiency over the particle filters. The simulations reveal that even after taking the larger computing times into account, the particle EIS method is more than 100 times more efficient than the bootstrap filter in terms of MSE for the model in Table 2, going up to 16000 times more efficient for the bivariate SV model. In contrast, the use of a better importance density for particle filtering in the SISR(2) method is counterproductive when taking into account the computational burden of constructing the proposals and computing the importance weights for each particle separately. Following our illustration in Section

Table 1: Two-factor stochastic volatility with leverage effects and Student's t return innovations ($\nu = 10$): likelihood evaluation.

The table compares the efficiency of different likelihood estimation methods. The methods are the bootstrap filter (BF), sequential importance sampling with resampling based on a Laplace approximation (SISR (2)), a zero order auxiliary particle filter (APF), efficient importance sampling (EIS) and particle EIS (P-EIS).

			$n = 2,500$		
	BF	SISR (2)	APF (0)	EIS	P-EIS
MSE ratio	13036	9761	12530	1.000	0.446
EIS density time	0.000	0.000	0.000	0.391	0.391
Likelihood time	0.030	0.761	0.026	0.048	0.049
Efficiency ($N = 50$)	904	16952	742	1.000	0.461
Efficiency ($N \to \infty$)	8330	156173	6836	1.000	0.461
			$n = 5,000$		
	BF	SISR (2)	APF (0)	EIS	P-EIS
MSE ratio	11040	8195	10747	1.000	0.246
EIS density time	0.000	0.000	0.000	0.759	0.759
Likelihood time	0.059	1.496	0.052	0.090	0.097
Efficiency ($N = 50$)	769	14445	654	1.000	0.267
Efficiency ($N \to \infty$)	7288	136917	6199	1.000	0.267
			$n = 10,000$		
	BF	SISR (2)	APF (0)	EIS	P-EIS
MSE ratio	10317	7486	9837	1.000	0.088
EIS density time	0.000	0.000	0.000	1.384	1.384
Likelihood time	0.109	2.917	0.100	0.155	0.170
Efficiency ($N = 50$)	731	14185	642	1.000	0.096
Efficiency ($N \to \infty$)	7251	140695	6365	1.000	0.096

4.3 the relative performance of the EIS method deteriorates quickly with the time series dimension, despite its numerical efficiency. The particle EIS method solves this problem, approximately maintaining a constant relative performance compared to the particle filters for all time series dimensions.

Table 4 shows the average effective sample sizes for the EIS method. The ESS indicates how challenging each model is for the importance sampling approach. As expected, we find the largest efficiency gains for the particle EIS method over EIS when the effective sample size decreases fast in n. Figure 2 plots the distribution

Table 2: Two-factor stochastic volatility with leverage effects and Student's t return and state innovations: likelihood evaluation.

The table compares the efficiency of different likelihood estimation methods for the two-factor stochastic volatility model with Student's t state disturbances. The methods are the bootstrap filter (BF), a zero order auxiliary particle filter (APF), efficient importance sampling (EIS) and particle EIS (P-EIS). The EIS methods are based on a data augmentation scheme for the transition density. The full EIS method performs importance sampling in both the Gaussian and the inverse-gamma state components, whereas the partial EIS method performs importance sampling only on the Gaussian component(see Appendix A for the details).

			$n = 2,500$			
			EIS		P-EIS	
	BF	APF (0)	partial	full	partial	full
MSE ratio	343	332	1.000	0.417	0.529	0.207
EIS density time	0.000	0.000	1.245	1.258	1.245	1.258
Likelihood time	0.031	0.031	0.183	0.181	0.180	0.184
Efficiency ($N = 50$)	7.350	7.304	1.000	0.444	0.522	0.224
Efficiency ($N \to \infty$)	57	57	1.000	0.413	0.522	0.209
			$n = 5,000$			
			EIS		P-EIS	
	BF	APF (0)	partial	full	partial	full
MSE ratio	364	348	1.000	0.454	0.328	0.125
EIS density time	0.000	0.000	2.184	2.249	2.184	2.249
Likelihood time	0.056	0.057	0.303	0.319	0.331	0.338
Efficiency ($N = 50$)	8.147	7.991	1.000	0.609	0.359	0.178
Efficiency ($N \to \infty$)	67	66	1.000	0.478	0.359	0.140
			$n = 10,000$			
			EIS		P-EIS	
	BF	APF (0)	partial	full	partial	full
MSE ratio	364	347	1.000	0.578	0.180	0.070
EIS density time	0.000	0.000	3.556	3.605	3.556	3.605
Likelihood time	0.097	0.102	0.511	0.529	0.534	0.556
Efficiency ($N = 50$)	8.684	8.734	1.000	0.661	0.188	0.085
Efficiency ($N \to \infty$)	69	70	1.000	0.598	0.188	0.077

Table 3: Bivariate stochastic volatility: likelihood evaluation.

The table compares the efficiency of different likelihood estimation methods for the bivariate stochastic volatility model. The methods are the bootstrap filter (BF), a zero order auxiliary particle filter (APF), efficient importance sampling (EIS) and particle EIS (P-EIS).

	$n = 2,500$			
	BF	APF (0)	EIS	P-EIS
MSE Ratio	5292	4778	1.000	0.128
EIS density time	0.000	0.000	0.146	0.146
Likelihood time	0.027	0.028	0.032	0.032
Efficiency ($N = 50$)	804	750	1.000	0.131
Efficiency ($N \to \infty$)	4518	4217	1.000	0.131
	$n = 5,000$			
	BF	APF (0)	EIS	P-EIS
MSE Ratio	5077	4547	1.000	0.064
EIS density time	0.000	0.000	0.385	0.385
Likelihood time	0.056	0.058	0.073	0.072
Efficiency ($N = 50$)	622	575	1.000	0.063
Efficiency ($N \to \infty$)	3892	3600	1.000	0.063
	$n = 10,000$			
	BF	APF (0)	EIS	P-EIS
MSE Ratio	4338	3890	1.000	0.030
EIS density time	0.000	0.000	0.833	0.833
Likelihood time	0.108	0.111	0.152	0.149
Efficiency ($N = 50$)	477	440	1.000	0.029
Efficiency ($N \to \infty$)	3093	2856	1.000	0.029

Table 4: Effective sample size for the EIS method (N=50).

	n=2,500	n=5,000	n=10,000
SV-t with Gaussian states	0.758	0.598	0.410
SV-t with t states (partial)	0.326	0.189	0.106
SV-t with t states (full)	0.435	0.268	0.156
Bivariate SV	0.261	0.145	0.084

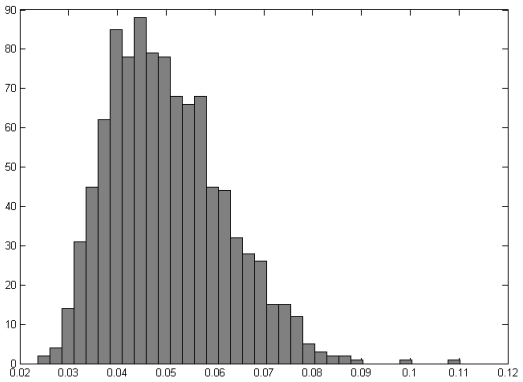


Figure 2: Histogram of the P-EIS/EIS variance ratio across trajectories for the bivariate SV model with $n = 10,000$.

of the particle EIS/EIS relative estimated variances for the bivariate SV model with $n = 10,000$, showing that the improvement in efficiency is uniform across trajectories.

5 Empirical application

This section studies the performance of the particle EIS method as a tool for Bayesian inference. We consider an empirical application of the bivariate stochastic volatility model of Section 4.1 using daily holding period returns for IBM and General Electric stocks between 1990 and 2012. The total number of bivariate time series observations is 5,797. The source of the series is the Center for Research in Security Prices (CRSP)

database. We adopt the following independent priors for each parameter

$$c_i \sim \mathrm{N}(0,1), \qquad \phi_i \sim \mathrm{Unif}(0,1), \quad i=1,2,3,$$
$$\sigma^2_{i,\eta} \sim \mathrm{IG}(2.5, 0.035), \quad i=1,2, \qquad \sigma^2_{3,\eta} \sim \mathrm{IG}(2.5, 0.0075),$$

where $\mathrm{IG}(a,b)$ denotes the inverse Gamma distribution with shape a and scale b.

We investigate two approaches for posterior inference: particle marginal Metropolis-Hastings (PMMH, Andrieu, Doucet, and Holenstein 2010) and importance sampling squared (IS^2, Tran, Scharth, Pitt, and Kohn 2014). The key idea of both PMMH and IS^2 is that replacing the unknown true likelihood by an unbiased estimator in standard Metropolis-Hastings and IS algorithms still leads to valid procedures that target the correct posterior distribution of the parameters. Let $p(\theta)$ be the prior distribution, $p(y_{1:n}|\theta)$ the likelihood (1) and $\pi(\theta) \propto p(y_{1:n}|\theta)p(\theta)$ the posterior distribution of the parameters defined on Θ. Suppose we want to calculate the integral

$$\pi(\varphi) = \int_\Theta \varphi(\theta)\pi(\theta)\,\mathrm{d}\theta.$$

The IS^2 method requires the following steps

1. Draw M parameter samples $\theta_1, \ldots, \theta_M$ from an importance density $q(\theta|y_{1:n})$.
2. Compute an unbiased estimate $\widehat{p}(y_{1:n}|\theta_i)$ of the likelihood function for $i = 1, \ldots, M$.
3. Compute the importance weights for $i = 1, \ldots, M$

$$\omega(\theta_i, y) = \frac{\widehat{p}(y_{1:n}|\theta_i)p(\theta_i)}{q(\theta_i|y_{1:n})}$$

4. Compute the importance sampling estimator

$$\widehat{\pi}(\varphi) = \frac{\sum_{i=1}^M \varphi(\theta_i)\omega(\theta_i, y_{1:n})}{\sum_{i=1}^M \omega(\theta_i, y_{1:n})}.$$

5. We can also estimate the marginal likelihood $p(y_{1:n}) = \int_\Theta p(y_{1:n}|\theta)p(\theta)\,\mathrm{d}\theta$ as $\widehat{p}(y_{1:n}) = \sum_{i=1}^M \omega(\theta_i, y_{1:n})/M$.

To obtain the parameter proposals $q(\theta|y_{1:n})$ for the IS^2 and particle independent Metropolis-Hastings (PIMH) methods, we consider the mixture of t by importance sampling weighted expectation maximisation (MitISEM) method of Hoogerheide, Opschoor, and van Dijk (2012). The MitISEM method implements a recursive sequence of importance weighted expectation maximisation that minimises the Kullback-ŰLeibler divergence between the posterior distribution and a mixture of Student's t densities proposal.

We implement the basic proposal training algorithm in that paper, but replace the true likelihood used in the original method by estimates obtained by the EIS and particle EIS methods with $N = 50$ particles. We label these two cases MitISEM (EIS) and MitISEM (P-EIS) respectively. We use 250 points from a Halton sequence with 9 dimensions and 250 antithetic draws to generate samples from the candidate densities within the training phase of the algorithm. We found that a multivariate Student's t density provides a good approximation to the posterior for the current problem. In our illustrations the likelihood estimation algorithm which we use when running the IS^2 and PIMH algorithms does not necessarily correspond to the one we adopt for training the MitISEM proposal. Our objective in doing so is to study the performance of different unbiased likelihood estimation methods when the proposal is fixed.

5.1 Choosing the number of particles

Pitt, Silva, Giordani, and Kohn (2012) and Tran, Scharth, Pitt, and Kohn (2014) study efficient implementations of Markov chain Monte Carlo and importance sampling when using unbiased likelihood estimators and general parameter proposals. The idea behind these papers is that the choice of the number of particles for likelihood estimation is a trade-off between variance reduction and computing time, which we may best allocate running more iterations of the Markov chain or generating additional importance samples for the parameters.

Assume that the log of the likelihood estimator is normal and that its variance is constant across different values of θ. The main finding in these papers is that the optimal number of particles to minimise the computing time for any given target Monte Carlo variance is such that the variance of the log-likelihood estimator is approximately equal to one when using particle filters. The EIS and particle EIS methods involve the

additional complication of the overhead associated with Algorithm 1, which does not depend on N. Let the variance of the log-likelihood estimator be $\text{Var}(\log \widehat{L}^h)/N$. Tran, Scharth, Pitt, and Kohn (2014) obtain the optimal number of particles

$$N_h^{\text{opt}} = \frac{\text{Var}(\log \widehat{L}^h)\left(1 + \sqrt{1 + 4\text{Var}(\log \widehat{L}^h)^{-1}(\tau_1^h/\tau_2^h)}\right)}{2},$$

where τ_1^h and τ_2^h are defined in (17). Note that $N_h^{\text{opt}} > \text{Var}(\log \widehat{L}^h)$ when $\tau_1^h > 0$. By dividing the variance of the log-likelihood by N_h^{opt}, we can see that the optimal variance of the log-likelihood estimate is lower than one when there is an overhead cost for estimating the likelihood.

Table 5 summarises a limited simulation study of how the variance of the log of the estimated likelihood depends on the method. The motivation for the study is to determine the number of particles for the empirical example. We carry out the simulation study as follows. First, we obtain a proposal density that approximates the posterior distribution of the parameters using the MitISEM (EIS) method. We then generate $M = 100$ draws from this proposal. For each sampled parameter vector, we perform 20 independent log-likelihood evaluations using the bootstrap filter, the EIS and the particle EIS algorithms. We use $S = 32$ simulations to obtain the importance parameters in the EIS method. We report the average of the sample variances across the 100 parameters draws, the corresponding variance ratios (with the EIS method as the benchmark), the computing time for obtaining the efficient importance density (τ_1^h), the computing time for the likelihood estimation step ($N\tau_2^h$), and the relative efficiency as defined in (18).

Consistent with Table 3, we find a 96.5% reduction in average variance for particle EIS in comparison with the EIS method. The results imply that the optimal number of particles is approximately $14,800$ for the bootstrap filter, 310 for EIS, and 42 for particle EIS. That leads us to use $N = 150$ and $N = 300$ samples for the EIS method and $N = 10$ and $N = 50$ particles for P-EIS, with the lower number of particles indicating the case for which the variance of the log-likelihood estimate is approximately one on average. For the bootstrap filter, we set the number of particles sub-optimally to $N = 5,000$ due to the excessively high computational cost of an ideal implementation for this problem.

Table 5: BIVARIATE SV - LIKELIHOOD EVALUATION FOR THE PARAMETERS SAMPLED FROM THE MULTIVARIATE t PROPOSAL.

	BF (N=100)	EIS (N=100)	P-EIS (N=100)
Variance	14.786	1.582	0.055
Variance ratio	9.349	1.000	0.035
EIS density time (τ_1^h)	-	0.567	0.567
Likelihood time ($N\tau_2^h$)	1.039	0.188	0.205
Efficiency	12.858	1.000	0.038

The theoretical results on the optimal implementation of PMMH and IS^2, in conjunction with Tables 1-2, highlight that the EIS method is remarkably efficient for Bayesian inference in the univariate SV model with Student's t return innovations. Based on the variance estimates for the EIS method in those tables, the standard algorithm with no resampling requires only 2 to 16 samples (including antithetic draws) to achieve a log-likelihood variance of approximately one for n as large as $10,000$. For particle EIS, only two particles are typically sufficient in this scenario. For this reason, we focus on the more challenging bivariate specification in this section.

5.2 Posterior analysis

Table 6 presents estimates of selected posterior distribution statistics estimated by the IS^2 method. We estimate the likelihood for a given set of parameters using the particle EIS method with $N = 50$ particles. We estimated the posterior distribution using $M = 10,000$ importance samples for the parameters, which required a total computing time of 21 minutes (parallelising the computations over 4 cores). We also estimate the Monte Carlo standard errors by bootstrapping the importance samples. The low MC standard errors confirm the efficiency of IS^2 approach using particle EIS. Figure 3 estimates the kernel smoothing density estimates of the marginal posteriors.

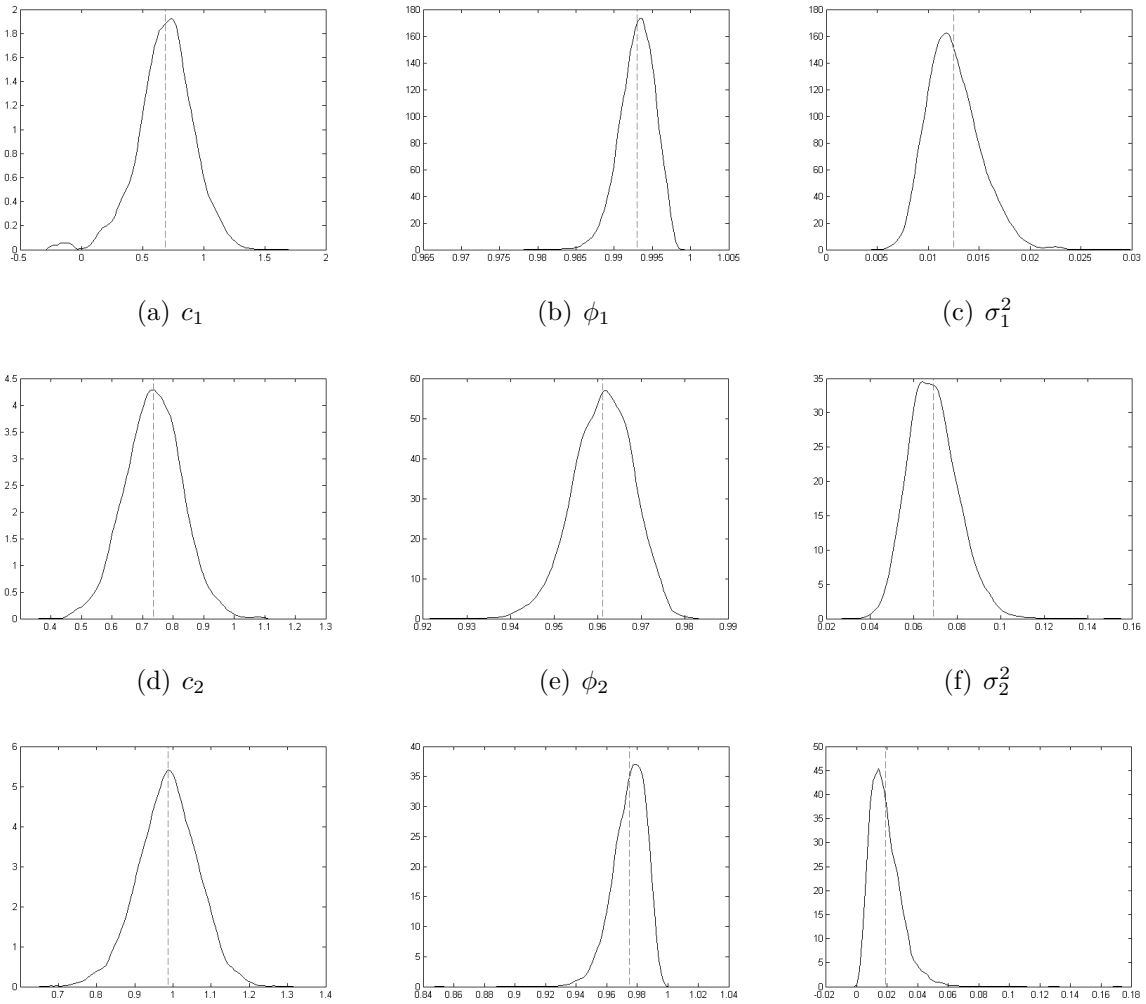
5.3 Importance sampling squared

This section compares the use of the bootstrap filter, EIS and P-EIS algorithms for the IS^2 method. We consider the efficiency of each method for estimating the posterior mean of all the parameters and the marginal likelihood. We estimate the Monte Carlo

Table 6: Bivariate SV - posterior statistics estimated by importance sampling squared.

The table presents estimates of selected posterior distribution statistics for the bivariate stochastic volatility application. The Monte Carlo standard errors are in brackets.

	Mean	Std. Dev.	Skew.	Kurt.	90% Credible Interval	
c_1	0.687 [0.014]	0.229 [0.008]	−0.469 [0.051]	4.188 [0.131]	0.299 [0.044]	1.048 [0.007]
ϕ_1	0.993 [<0.001]	0.002 [<0.001]	−0.465 [0.039]	3.237 [0.070]	0.989 [<0.001]	0.997 [<0.001]
σ_1^2	0.013 [<0.001]	0.003 [<0.001]	0.647 [0.046]	3.753 [0.155]	0.009 [<0.001]	0.017 [<0.001]
c_2	0.736 [0.002]	0.093 [0.001]	0.008 [0.059]	3.152 [0.087]	0.585 [0.003]	0.886 [0.005]
ϕ_2	0.961 [<0.001]	0.007 [<0.001]	−0.288 [0.047]	3.079 [0.049]	0.950 [<0.001]	0.972 [<0.001]
σ_2^2	0.069 [<0.001]	0.011 [<0.001]	0.415 [0.034]	3.246 [0.079]	0.052 [<0.001]	0.089 [<0.001]
c_3	0.987 [0.001]	0.078 [0.001]	−0.124 [0.060]	3.270 [0.159]	0.855 [0.004]	1.110 [0.002]
ϕ_3	0.975 [<0.001]	0.011 [<0.001]	−0.764 [0.031]	3.866 [0.084]	0.955 [<0.001]	0.990 [0.001]
σ_3^2	0.019 [<0.001]	0.010 [<0.001]	1.145 [0.036]	5.277 [0.137]	0.006 [<0.001]	0.037 [<0.001]



(a) c_1 (b) ϕ_1 (c) σ_1^2

(d) c_2 (e) ϕ_2 (f) σ_2^2

(g) c_3 (h) ϕ_3 (i) σ_3^2

Figure 3: Kernel smoothing density estimates of the marginal posterior distributions of the bivariate SV model parameters (estimated by importance sampling squared). The dashed lines indicate the posterior means.

variances associated with each method as in Section 4.4 by running 250 independent replications of the importance sampling algorithm using $M = 500$ importance samples for the parameters. We consider two versions of the MitISEM method: one using the EIS method for estimating the likelihood the training step of the method the importance density and another using the P-EIS method for the same purpose. Our efficiency measure is the time normalised variance of estimates, which we define as the product of Monte Carlo variance and the computational time. We report all the time normalised variances as relative to the EIS method with $N = 150$ samples. The time normalised variance determines the total computing time required for obtaining any given MC variance for the posterior mean and marginal likelihood using each method.

Table 7 summarises the results. Focusing on the MitISEM (EIS) proposal, the table shows reductions in time normalised variance which range from 79% to 97% for the P-EIS method relative to the EIS method. When considering the P-EIS method for constructing the proposal, we find further reductions in time normalised variance of as much as 60%. Compared to the bootstrap filter, the reductions in time normalised variance range from 99.09% to 99.81%. The table also shows that the EIS method with $N = 300$ samples, which should be approximately the optimal number of samples according to Section 5.1, has an inferior performance to the implementation with $N = 150$. We conjecture that this is because the EIS log-likelihood estimates are skewed for $N = 150$ and $N = 300$, whereas the theoretical result for the optimal number of particles is based on a normality assumption. In this setting, increasing the number of samples from 150 to 300 reduces the EIS log-likelihood variance by less than 50%, so that the benefit of increasing N is lower than assumed by the result on the optimal number of samples.

5.4 Particle Marginal Metropolis-Hastings

We now consider the use of the bootstrap filter, EIS and particle EIS algorithms for the PMMH estimation of the posterior distribution of the bivariate stochastic volatility model. We implement two Metropolis-Hastings (M-H) algorithms: the adaptive random walk method of Roberts and Rosenthal (2009) and the independent M-H method using the MitISEM proposal. We run 50,000 iterations of the two algorithms and discard a burn-in sample of 5,000 iterations.

Table 7: Bivariate SV - relative time normalised variances for posterior inference using IS^2.

The table shows the performance of different methods for estimating the posterior distribution of the bivariate stochastic volatility model by IS^2. We indicate the number of particles for each method by N. We also report the average computing time in seconds.

		MitISEM (EIS)				MitISEM (PEIS)	
	BF	EIS		P-EIS		P-EIS	
	N=5,000	N=150	N=300	N=10	N=50	N=10	N=50
c_1	14.951	1.000	1.824	0.359	0.210	0.122	0.136
ϕ_1	18.504	1.000	1.381	0.228	0.202	0.170	0.131
σ_1^2	16.773	1.000	1.528	0.131	0.098	0.051	0.039
c_2	17.712	1.000	1.686	0.145	0.083	0.082	0.062
ϕ_2	14.109	1.000	1.442	0.088	0.076	0.068	0.051
σ_2^2	12.829	1.000	1.616	0.082	0.056	0.051	0.046
c_3	15.212	1.000	1.517	0.182	0.087	0.091	0.087
ϕ_3	25.558	1.000	1.111	0.125	0.106	0.077	0.062
σ_3^2	26.736	1.000	0.875	0.116	0.107	0.066	0.050
Marg. Lik.	6.720	1.000	1.075	0.042	0.032	0.033	0.018
Time (s)	5603	285	430	154	194	167	204

Table 8: Bivariate SV - PMMH inefficiencies under different likelihood estimation methods.

The table examines the performance of different particle marginal Metropolis-Hastings (PMMH) methods for estimating the posterior distribution of the bivariate stochastic volatility model. The table reports the acceptance rates, the inefficiency factors for each parameter, and the total computing time in hours.

	Adaptive Random Walk			MitISEM (EIS)			MitISEM (P-EIS)
	BF	EIS	P-EIS	BF	EIS	P-EIS	P-EIS
	N=5,000	N=300	N=50	N=5,000	N=300	N=50	N=50
Acc. rate	0.082	0.116	0.241	0.119	0.199	0.423	0.588
c_1	107.4	94.3	37.1	44.2	60.4	16.4	11.1
ϕ_1	96.4	89.4	32.9	62.8	62.6	9.5	6.0
σ_1^2	82.6	83.8	31.4	54.5	68.8	7.0	4.7
c_2	91.9	97.3	31.9	43.1	47.7	13.3	7.1
ϕ_2	85.9	78.8	32.3	53.0	87.3	7.5	4.2
σ_2^2	85.6	85.4	32.7	56.7	69.1	8.4	5.7
c_3	99.0	89.3	32.9	52.5	63.7	7.3	4.7
ϕ_3	85.8	99.7	33.1	43.4	75.6	10.0	5.4
σ_3^2	77.6	100.2	32.9	48.5	74.3	9.8	5.0
Time (h)	79.6	8.5	5.1	79.5	8.2	4.8	4.4

Table 8 reports the acceptance rates, the inefficiency factors (calculated using the overlapping batch means method), and the total computing times in hours. The results show that only the independent Metropolis-Hastings using the MitISEM proposal in combination with the particle EIS method for estimating the likelihood performs satisfactorily. The particle EIS method achieves an acceptance rate of 0.423 and inefficiency factors between 7.3 and 16.4 when using the MitISEM (EIS) proposal, in a total computing time of 4.8 hours. That compares to acceptance rates of 0.119 and 0.199 and computing times of 79.5 and 8.2 hours for the BF and EIS methods respectively, with inefficiency factors higher than 40 for all the parameters. We also find that the MitISEM (P-EIS) proposal leads to an increase in the acceptance rate to 0.588 and substantial improvements in the inefficiency factors.

References

Andrieu, C., A. Doucet, and R. Holenstein (2010): "Particle Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society - Series B: Statistical Methodology*, 72, 269–342.

Bauwens, L. and F. Galli (2009): "Efficient Importance Sampling for ML Estimation of SCD Models," *Computational Statistics and Data Analysis*, 53, 1974–1992.

Chopin, N. (2004): "Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference," *The Annals of Statistics*, 32, 2385–2411.

DeJong, D. N., R. Liesenfeld, G. V. Moura, J.-F. Richard, and H. Dharmarajan (2012): "Efficient likelihood evaluation of state-space representations," *The Review of Economic Studies*, 80, 538–567.

Del Moral, P. (2004): *Feynman-Kac formulae: genealogical and interacting particle systems with applications*, New York: Springer.

Doucet, A., M. Briers, and S. Sénécal (2006): "Efficient Block Sampling Strategies for Sequential Monte Carlo Methods," *Journal of Computational and Graphical Statistics*, 15, 693–711.

Durbin, J. and S. J. Koopman (1997): "Monte Carlo Maximum Likelihood Estimation for non-Gaussian State Space Models," *Biometrika*, 669–684.

——— (2000): "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives," *Journal of the Royal Statistical Society, Series B*, 3–56.

——— (2001): *Time Series Analysis by State Space Methods*, Oxford University Press.

Fearnhead, P., D. Wyncoll, and J. Tawn (2010): "A sequential smoothing algorithm with linear computational cost," *Biometrika*, 97, 447–464.

Flury, T. and N. Shephard (2011): "Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models," *Econometric Theory*, 27, 933–956.

Geweke, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1739.

Godsill, S. J., A. Doucet, and M. West (2004): "Monte Carlo smoothing for nonlinear time series," *Journal of the American Statistical Association*, 99, 156–168.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Hesterberg, T. (1995): "Weighted Average Importance Sampling and Defensive Mixture Distributions," *Technometrics*, 37, 185–194.

Hoogerheide, L., A. Opschoor, and H. K. van Dijk (2012): "A class of adaptive importance sampling weighted EM algorithms for efficient and robust posterior and predictive simulation," *Journal of Econometrics*, 171, 101–Ű120.

Kitagawa, G. (1996): "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, 5, 1–25.

Kleppe, T. S. and R. Liesenfeld (2013): "Efficient importance sampling in mixture frameworks," *Computational Statistics & Data Analysis*, forthcoming.

Koopman, S. J., A. Lucas, and M. Scharth (2013): "Static and dynamic multivariate Gaussian efficient importance sampling," Working paper.

——— (2014): "Numerically accelerated importance sampling for nonlinear non-Gaussian state space models," *Journal of Business and Economic Statistics*, forthcoming.

Koopman, S. J., N. Shephard, and D. Creal (2009): "Testing the assumptions behind importance sampling," *Journal of Econometrics*, 149, 2–11.

Liesenfeld, R. and J.-F. Richard (2003): "Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics," *Journal of Empirical Finance*, 10, 505–531.

——— (2010): "Efficient estimation of probit models with correlated errors," *Journal of Econometrics*, 156, 367–376.

Liesenfeld, R., J.-F. Richard, and J. Vogler (2013): "Analysis of discrete dependent variable models with spatial correlation," Working paper.

Lin, M., R. Chen, and J. S. Liu (2013): "Lookahead Strategies for Sequential Monte Carlo," *Statistical Science*, 28, 69–94.

Liu, J. S. and R. Chen (1998): "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, 93, 1032–1044.

Pitt, M. K. (2002): "Smooth particle filters for likelihood evaluation and maximisation," Working paper.

Pitt, M. K. and N. Shephard (1999): "Filtering via simulation: auxiliary particle filters," *Journal of the American Statistical Association*, 94, 590–599.

Pitt, M. K., R. d. S. Silva, P. Giordani, and R. Kohn (2012): "On some properties of Markov chain Monte Carlo simulation methods based on the particle filter," *Journal of Econometrics*, 171, 134–151.

Richard, J.-F. and W. Zhang (2007): "Efficient High-Dimensional Importance Sampling," *Journal of Econometrics*, 141, 1385–1411.

Roberts, G. O. and J. S. Rosenthal (2009): "Examples of adaptive MCMC," *Journal of Computational and Graphical Statistics*, 18, 349–367.

Shephard, N. and M. Pitt (1997): "Likelihood analysis of non-Gaussian measurement time series," *Biometrika*, 84, 653–667.

Tran, M.-N., M. Scharth, M. K. Pitt, and R. Kohn (2014): "Importance Sampling Squared for Bayesian Inference in Latent Variable Models," `http://arxiv.org/abs/1309.3339`.

Appendix

A EIS for state space models with nonlinear transition and additive Student's t state disturbances

This appendix develops an efficient importance sampling method for a state space model with nonlinear transition and additive Student's t state disturbances which includes the univariate SV models of Section 4.1 as special cases. The method follows from Kleppe and Liesenfeld (2013), which consider the case in which the measurement density $p(y_t|x_t)$ is a continuous or discrete mixture. They propose a data augmentation scheme in which they explicitly include the mixture components in the integrand of (1). This allows them to approximate the different components of the measurement density separately using the EIS method.

Applying this principle to our current setting, we consider the modified transition density $p(x_t|x_{t-1}, \lambda_t)p(\lambda_t)$, where $p(x_t|x_{t-1}, \lambda_t)$ is a Gaussian density and λ_t is a vector of inverse gamma random variables. The state space model is

$$y_t|x_t \sim p(y_t|Zx_t), \qquad x_t = F(x_{t-1}) + \Lambda_t\eta_t, \qquad x_1 \sim \mathrm{N}(a_1, P_1), \qquad \eta_t \sim \mathrm{N}(0, Q),$$

where y_t is the observation vector, x_t is the $m \times 1$ state vector, Z is a $p \times m$ (with $p \leq m$) and $F(.)$ is a $\mathbb{R}^m \to \mathbb{R}^m$ nonlinear function. The scaling matrix Λ_t is diagonal with entries $\sqrt{\lambda_{1,t}}, \ldots, \sqrt{\lambda_{m,t}}$, where $\lambda_{j,t} \sim \mathrm{IG}(\nu_{\eta,j}/2, \nu_{\eta,j}/2)$. All the random variables $\lambda_{j,t}$ are mutually independent. We have that $\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{m,t})'$. We write the measurement density in terms of the signal vector Zx_t instead of the state vector x_t in order to reduce the computational cost of running the EIS algorithm when $p < m$, see for example Koopman, Lucas, and Scharth (2014).

After data augmentation, the likelihood function (1) becomes

$$= \int p(y_t|x_1)p(x_1) \prod_{t=2}^{n} p(y_t|x_t)p(x_t|x_{t-1}, \lambda_t)p(\lambda_t)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n\, \mathrm{d}\lambda_2 \ldots \mathrm{d}\lambda_n.$$

We consider the sequential importance densities

$$q(x_t, \lambda_t|x_{t-1}, y_{1:n}) = q(x_t|x_{t-1}, \lambda_t, y_{1:n})q(\lambda_t|y_{1:n}),$$

where

$$q(x_t|x_{t-1},\lambda_t,y_{1:n}) = \delta_t(x_{t-1},\lambda_t)\exp\left(b_t' Z x_t - \frac{1}{2}x_t' Z' C_t Z x_t\right) p(x_t|x_{t-1},\lambda_t) \tag{19}$$

and

$$q(\lambda_t|y_{1:n}) = \left(\prod_{j=1}^{m} \varphi_{j,t}\lambda_{j,t}^{\alpha_{j,t}} \exp(\beta_{j,t}/\lambda_{j,t})\right) p(\lambda_t). \tag{20}$$

The importance parameters are b_t, C_t, $\alpha_{1,t},\ldots,\alpha_{m,t}$ and $\beta_{1,t},\ldots,\beta_{m,t}$. The terms $\delta_t(x_{t-1},\lambda_t)$ and $\varphi_{j,t}$ are constants that ensure that $q(x_t,\lambda_t|x_{t-1},y_{1:n})$ integrates to one.

The importance densities in (19) and (20) offset the model transition densities and use conjugate terms to approximate the measurement densities and integration constants. With some algebra, we can show that $q(x_t|x_{t-1},\lambda_t,y_{1:n})$ is a Gaussian density with covariance matrix

$$V_t = [(\Lambda_t Q \Lambda_t)^{-1} + C_t]^{-1}$$

and mean vector

$$\mu_t = V_t[Z' b_t + (\Lambda_t Q \Lambda_t)^{-1} F(x_{t-1})],$$

while the importance density $q(\lambda_t|y_{1:n})$ is such that

$$\lambda_{j,t} \sim \text{IG}(\nu_{\eta,j}/2 - \alpha_{j,t}, \nu_{\eta,j}/2 - \beta_{j,t})$$

for $j = 1,\ldots,m$. The constants are

$$\log \delta_t(x_{t-1},\lambda_t) = \frac{1}{2}\log(|\Lambda_t Q \Lambda_t|/|V_t|) + \frac{1}{2}F(x_{t-1})'(\Lambda_t Q \Lambda_t)^{-1}F(x_{t-1}) - \frac{1}{2}\mu_t' V_t^{-1}\mu_t$$

and

$$\varphi_{j,t} = \frac{\Gamma(\nu_{\eta,j}/2)}{(\nu_{\eta,j}/2)^{\nu_{\eta,j}/2}} \frac{(\nu_{\eta,j}/2 - \beta_{j,t})^{\nu_{\eta,j}/2-\alpha_{j,t}}}{\Gamma(\nu_{\eta,j}/2 - \beta_{j,t})}.$$

To implement Algorithm 1, suppose we generate draws $x^{(1)},\ldots,x^{(S)},\lambda^{(1)},\ldots,\lambda^{(S)}$ from the current candidate density $q^{[k]}(x_{1:n},\lambda_{2:n}|y_{1:n})$. Following an appropriate modification of (10) and (11) for the data augmentation setting, we update the importance parameters by running backwards recursively for every period t ordinary least squared

regressions with dependent variable

$$\log p(y_t|x_t^{(s)}) - \log \delta_{t+1}(x_t^{(s)}, \lambda_{t+1}^{(s)})$$

and regressors

$$\begin{aligned} &Zx_t^{(s)},\ -(1/2)\text{vech}[(Zx_t^{(s)})(Zx_t^{(s)})'], \\ &\log(\lambda_{1,t+1}^{(s)}), \ldots \log(\lambda_{m,t+1}^{(s)}),\ 1/\lambda_{1,t+1}^{(s)}, \ldots, 1/\lambda_{m,t+1}^{(s)} \end{aligned}$$

plus a constant. We need to multiply the coefficients associated with the off-diagonal elements of $-(1/2)\text{vech}[(Zx_t^{(s)})(Zx_t^{(s)})']$ by two because these terms appear twice in the quadratic form in (19). The resulting coefficients after these steps give us b_t, C_t, $\alpha_{1,t+1}, \ldots, \alpha_{m,t+1}$ and $\beta_{1,t+1}, \ldots, \beta_{m,t+1}$ respectively. Because the inverse gamma variables appear directly in $\log \delta_t(x_{t-1}^{(s)}, \lambda_t^{(s)})$ and not $p(y_t|x_t^{(s)})$, it is necessary to estimate the coefficients of $q(\lambda_t|y_{1:n})$ jointly with $q(x_{t-1}|x_{t-2}, \lambda_{t-1}, y_{1:n})$. Though this may initially seem counterintuitive, the need for this design highlights the importance of the integration constants in the EIS method.

We emphasised that Algorithm 1 is based on common random numbers (CRNs). The use of CRNs for the current problem requires computationally expensive inversions of gamma cumulative density functions. We circumvent this issue by first fixing $q(\lambda_t|y_{1:n}) = p(\lambda_t)$ and letting the importance parameters b_t and C_t converge for $t = 1, \ldots, n$. That provides the partial EIS density which we use in Section 4.4. We then run only one iteration of the full EIS regressions described above using the partial EIS parameters as starting values. We have found that additional iterations generate modest gains in efficiency that do not compensate for the additional computational cost.