Filtering via Simulation: Auxiliary Particle Filters

Michael K. PITT and Neil SHEPHARD

This article analyses the recently suggested particle approach to filtering time series. We suggest that the algorithm is not robust to outliers for two reasons: the design of the simulators and the use of the discrete support to represent the sequentially updating prior distribution. Here we tackle the first of these problems.

KEY WORDS: Filtering; Markov chain Monte Carlo; Particle filter; Sampling/importance resampling; Simulation; State space.

1. INTRODUCTION

In this article we model a time series $y_t, t = 1, \ldots, n$, as being conditionally independent given an unobserved sufficient state α_t, which is itself assumed to be Markovian. The task is to use simulation to carry out on-line filtering—that is, to learn about the state given contemporaneously available information. We do this by estimating the difficult-to-compute density (or probability distribution function) $f(\alpha_t|y_1, \ldots, y_t) = f(\alpha_t|Y_t), t = 1, \ldots, n$. We assume parametric forms for both the "measurement" density $f(y_t|\alpha_t)$ and the "transition" density of the state $f(\alpha_{t+1}|\alpha_t)$. The state evolution is initialized by some density $f(\alpha_0)$.

Filtering can be thought of as the repeated application of a two-stage procedure. First, the current density must be propagated into the future via the transition density $f(\alpha_{t+1}|\alpha_t)$ to produce the prediction density

$$f(\alpha_{t+1}|Y_t) = \int f(\alpha_{t+1}|\alpha_t)\, dF(\alpha_t|Y_t). \tag{1}$$

Second, one moves to the filtering density via Bayes theorem,

$$f(\alpha_{t+1}|Y_{t+1}) = \frac{f(y_{t+1}|\alpha_{t+1}) f(\alpha_{t+1}|Y_t)}{f(y_{t+1}|Y_t)},$$

$$f(y_{t+1}|Y_t) = \int f(y_{t+1}|\alpha_{t+1})\, dF(\alpha_{t+1}|Y_t). \tag{2}$$

This implies that the data can be processed in a single sweep, updating our knowledge about the states as we receive more information. This is straightforward if $\alpha_t|\alpha_{t-1}$ has a finite set of known discrete points of support, as the previous calculations can be computed exactly. When the support is continuous and the integrals cannot be analytically solved, then numerical methods must be used.

Numerous attempts have been made to provide algorithms that approximate the filtering densities. Important recent work includes that of Gerlach, Carter, and Kohn (1996), Kitagawa (1987), West (1992), and those papers reviewed by West and Harrison (1997, chaps. 13 and 15).

In this article we use simulation to perform filtering following an extensive recent literature. Our approach is to extend the particle filter that has recently been suggested independently by various authors. In particular, it was used by Gordon, Salmond, and Smith (1993) on non-Gaussian state–space models. The same algorithm, with extensions to the smoothing problem, has been independently proposed by Kitagawa (1996) (and generalized in Hürzeler and Künsch 1995) for use in time series problems. It reappeared and was then discarded by Berzuini, Best, Gilks, and Larizza (1997) in the context of a real-time application of the sequential analysis of medical patients. It was again proposed by Isard and Blake (1996) in the context of robustly tracking motion in visual clutter, under the term the "condensation" algorithm. Some statistical refinements of this general class of algorithm, generically called particle filters, have been given by Carpenter, Clifford, and Fearnhead (1998), Doucet (1998), and Liu and Chen (1998) (which were written independently of this article). The idea of calling this class of algorithm "particle filters" is from Carpenter et al. (1998), although Kitagawa (1996) used the term "particles." Similar ideas (but using stronger assumptions) are used on the blind deconvolution problem of Liu and Chen (1995) and in the sequential importance sampling algorithms of Hendry and Richard (1991) and Kong, Liu, and Wong (1994).

Here we discuss the particle filtering literature and extend it in a number of directions so that it can be used in a much broader context. The article is organized as follows. In Section 2 we analyze the statistical basis of particle filters and focus on its weaknesses. In Section 3 we introduce our main contribution, an auxiliary particle filter method. We give some numerical examples in Section 4 and state some conclusions in Section 5.

2. PARTICLE FILTERS

2.1 Definition of Particle Filters

Particle filters are the class of simulation filters that recursively approximate the filtering random variable $\alpha_t|Y_t = (y_1, \ldots, y_t)'$ by "particles" $\alpha_t^1, \ldots, \alpha_t^M$, with discrete probability mass of $\pi_t^1, \ldots, \pi_t^M$. Hence a continuous variable is

Michael K. Pitt is a Research Associate at Department of Mathematics, Imperial College, Queen's Gate, London SW7 2BZ, U.K. (E-mail: *m.k.pitt@ic.ac.uk*). Neil Shephard is an Official Fellow at Nuffield College, Oxford OX1 1NF, U.K. (E-mail: *neil.shephard@nuf.ox.ac.uk*). The authors thank the Economic and Social Research Council (ESRC) for their financial support through the projects "Estimation via Simulation in Econometrics and Modelling and Analysis of Econometric and Financial Time series;" the European Union (EU) through their grant "Econometric Inference Using Simulation Techniques;" and Robert Kohn, Peter Clifford, and David Firth for various helpful conversations. The comments of the editor, the associate editor, and the two referees on the initial submission were also useful. All of the computations reported in this article were carried out using the Ox language of Doornik (1996).

Journal of the American Statistical Association
June 1999, Vol. 94, No. 446, Theory and Methods

approximated by a discrete one with random support. These discrete points are viewed as samples from $f(\alpha_t|Y_t)$. In the literature all of the π_t^j are assumed to equal $1/M$. Throughout, M is taken to be very large. Then we require that as $M \to \infty$, the particles can be used to increasingly well approximate the density of $\alpha_t|Y_t$.

Particle filters treat the discrete support generated by the particles as the true filtering density. This allows us to produce an approximation to the prediction density, (1), simply by using the discrete support of the particles. We call

$$\hat{f}(\alpha_{t+1}|Y_t) = \sum_{j=1}^{M} f(\alpha_{t+1}|\alpha_t^j)\pi_t^j \tag{3}$$

the "empirical prediction density." This is a mixture of densities and so echoes the earlier filtering work of, for example, Sorensen and Alspach (1971). This can be combined with the measurement density to produce, up to proportionality,

$$\hat{f}(\alpha_{t+1}|Y_{t+1}) \propto f(y_{t+1}|\alpha_{t+1}) \sum_{j=1}^{M} f(\alpha_{t+1}|\alpha_t^j)\pi_t^j, \tag{4}$$

the "empirical filtering density" as an approximation to the true filtering density (2). Generically, particle filters then sample from this density to produce new particles $\alpha_{t+1}^1, \ldots, \alpha_{t+1}^M$ with weights $\pi_{t+1}^1, \ldots, \pi_{t+1}^M$. This procedure can then be iterated through the data. We call a particle filter "normal" if it produces independent and identically distributed samples from the empirical filtering density. There may be advantages in deliberately inducing (negative) correlations among the particles. This approach was first explicitly pointed out by Carpenter et al. (1998), and has been further explored in our earlier work (Pitt and Shephard 1998). We do not discuss this here.

If the particle filter can be made to work, it could be used in a number of different contexts. These could include online tracking problems; estimating the one-step-ahead density $f(y_{t+1}|Y_t)$ and so, via the prediction decomposition, the joint density of the observations; and estimating the corresponding distribution function $F(y_{t+1}|Y_t)$, which can be a useful diagnostic measure of fit for non-Gaussian models (see, e.g., Gerlach et al. 1996, Shephard 1994; Smith 1985).

2.2 Sampling the Empirical Prediction Density

One way of sampling from the empirical prediction density is to think of $\sum_{j=1}^{M} f(\alpha_{t+1}|\alpha_t^j)\pi_t^j$ as a "prior" density $\hat{f}(\alpha_{t+1}|Y_t)$ that is combined with the "likelihood" $f(y_{t+1}|\alpha_{t+1})$ to produce a posterior. We can sample from $\hat{f}(\alpha_{t+1}|Y_t)$ by choosing α_t^j with probability π_t^j and then drawing from $f(\alpha_{t+1}|\alpha_t^j)$. If we can also evaluate $f(y_{t+1}|\alpha_{t+1})$ up to proportionality, then this leaves us with three sampling methods to draw from $f(\alpha_{t+1}|Y_{t+1})$: sampling/importance resampling, acceptance sampling, and Markov chain Monte Carlo (MCMC). In the rest of this section we write the prior as $f(\alpha)$ and the likelihood as $f(y|\alpha)$, abstracting from subscripts and conditioning arguments, to briefly describe these methods in this context.

2.2.1 Sampling/Importance Resampling. The sampling/importance resampling (SIR) method Rubin (1987) draws $\alpha^1, \ldots, \alpha^R$ from $f(\alpha)$ and then associates with each of these draws the weights π_j, where

$$w_j = f(y|\alpha^j), \qquad \pi_j = \frac{w_j}{\sum_{i=1}^{R} w_i}, \qquad j = 1, \ldots, R.$$

Then the weighted sample will converge, as $R \to \infty$, to a nonrandom sample from the desired posterior $f(\alpha|y)$ as $R^{-1}\sum_{i=1}^{R} w_i \xrightarrow{p} f(y)$. The nonrandom sample can be converted into a random sample of size M by resampling the $\alpha^1, \ldots, \alpha^R$ using weights $\pi_1, \ldots, \pi_R$. This requires $R \to \infty$ and $R >> M$. The use of this method has been suggested in the particle filter framework by Berzuini et al. (1997), Gordon et al. (1993), Isard and Blake (1996), and Kitagawa (1996).

To understand the efficiency of the SIR method, it is useful to think of SIR as an approximation to the importance sampler of the moment

$$E_{f\pi}\{h(\alpha)\} = \int h(\alpha)\pi(\alpha)\, dF(\alpha) \quad \text{by} \quad \frac{1}{R}\sum_{j=1}^{R} h(\alpha^j)\pi(\alpha^j),$$

where $\alpha \sim f(\alpha)$ and $\pi(\alpha) = f(y|\alpha)/f(y)$. Liu (1996) suggested that the variance of this estimator is approximately (for slowly varying $h(\alpha)$) proportional to $E_f\{\pi(\alpha)^2\}/R$. Hence the SIR method will become very imprecise when the π_j become very variable. This will happen if the likelihood is highly peaked compared to the prior.

2.2.2 Adaptation. The foregoing SIR algorithm samples from $f(\alpha|y)$ by making *blind* proposals $\alpha^1, \ldots, \alpha^R$ from the prior, ignoring the fact that we know the value of y. This is the main feature of existing particle filters. We say that a particle filter is adapted if we make proposals that take into account the value of y.

An adapted SIR-based particle filter has the following general structure:

1. Draw from $\alpha^1, \ldots, \alpha^R \sim g(\alpha|y)$.
2. Evaluate $w_j = f(y|\alpha^j)f(\alpha^j)/g(\alpha^j|y), j = 1, \ldots, R$.
3. Resample among the $\{\alpha^j\}$ using weights proportional to $\{w_j\}$ to produce a sample of size M.

Although this looks attractive, for a particle filter, $f(\alpha) = \sum_{j=1}^{M} f(\alpha_{t+1}|\alpha_t^j)\pi_t^j$, which implies we must at least evaluate $M \times R$ densities to generate M samples from $f(\alpha|y)$. Given that M and R are typically very large, this implies that adaption is not generally feasible for SIR-based particle filters.

2.2.3 Rejection and Markov Chain Monte Carlo Sampling. Exactly the same remarks hold for rejection sampling. A blind rejection sampling-based particle filter will simulate from $f(\alpha)$ and accept with probability $\pi(\alpha) = f(y|\alpha)/f(y|\alpha_{\max})$, where $\alpha_{\max} = \arg\max_\alpha f(y|\alpha)$. This has been proposed by Hürzeler and Künsch (1995). Again, the rejection becomes worse if the $\text{var}_f\{\pi(\alpha)\}$ is high and adaption is difficult, as it will again typically involve evaluating $f(\alpha)$ and so is computationally infeasible.

Another alternative to SIR is the use of a blind MCMC method (see Gilks, Richardson, and Spiegelhalter 1996 for a review). In this context the MCMC accepts a move from a current state α^i to $\alpha^{i+1} \sim f(\alpha)$ with probability $\min\{1, f(y|\alpha^{i+1})/f(y|\alpha^i)\}$; otherwise, it sets $\alpha^{i+1} = \alpha^i$. Again, if the likelihood is highly peaked, there may be a large amount of rejection, which will mean the Markov chain will have a great deal of dependence. This suggests adapting, when this is possible, the MCMC method to draw from $g(\alpha|y)$ and then accept these draws with probability

$$\min\left\{1, \frac{f(y|\alpha^{i+1})f(\alpha^{i+1})}{f(y|\alpha^i)f(\alpha^i)} \frac{g(\alpha^i|y)}{g(\alpha^{i+1}|y)}\right\}.$$

Again, the problem with this is that evaluating $f(\alpha)$ is very expensive.

2.3 Weaknesses of Particle Filter

The particle filter based on SIR has two basic weaknesses. The first is well known, that when there is an outlier, the weights π_j will be very unevenly distributed and so it will require an extremely large value of R for the draws to be close to samples from the empirical filtering density. This is of particular concern if the measurement density $f(y_{t+1}|\alpha_{t+1})$ is highly sensitive to α_{t+1}. Notice this is not a problem of having too small a value of M. That parameter controls the accuracy of (3). Instead, the difficulty is, given that degree of accuracy, how to efficiently sample from (4). We show how to do this in the next section.

The second weakness holds in general for particle filters for which the π^j are equal and where the states are updated one period at a time. As $R \to \infty$, so the weighted samples can be used to arbitrarily well approximate (4). However, the tails of (3) usually only poorly approximate the true tails of $\alpha_{t+1}|Y_t$, due to the use of the mixture approximation. As a result, (4) can only poorly approximate the true $f(\alpha_{t+1}|Y_{t+1})$ when there is an outlier. Hence the second question is how to improve the empirical prediction density's behavior in the tails. In earlier work (Pitt and Shephard 1998) we analyzed this problem using the so-called fixed lagged filter.

3. AUXILIARY VARIABLE

3.1 The Basics

A fundamental problem with existing particle filters is that their mixture structure means that it is difficult to adapt the SIR, rejection, or MCMC sampling methods without greatly slowing the running of the filter. Here we argue that many of these problems are reduced when we perform particle filtering in a higher dimension.

Our task is to sample from the joint density $f(\alpha_{t+1}, k|Y_{t+1})$, where k is an index on the mixture in (3). We define

$$f(\alpha_{t+1}, k|Y_{t+1}) \propto f(y_{t+1}|\alpha_{t+1})f(\alpha_{t+1}|\alpha_t^k)\pi^k, \qquad k = 1, \ldots, M. \quad (5)$$

If we draw from this joint density and then discard the index, then we produce a sample from the empirical filtering density (4) as required. We call k an auxiliary variable, as it is present simply to aid the task of simulation. We call generic particle filters of this type auxiliary particle filters.

We can now sample from $f(\alpha_{t+1}, k|Y_{t+1})$ using SIR, rejection, or MCMC. The SIR idea is to make R proposals $\alpha_{t+1}^j, k^j \sim g(\alpha_{t+1}, k|Y_{t+1})$ and then construct resampling weights

$$w_j = \frac{f(y_{t+1}|\alpha_{t+1}^j)f(\alpha_{t+1}^j|\alpha_t^{k^j})}{g(\alpha_{t+1}^j, k^j|Y_{t+1})}, \quad \pi_j = \frac{w_j}{\sum_{i=1}^R w_i}, \qquad j = 1, \ldots, R.$$

We have complete control over the design of $g(\cdot)$, which can depend on y_{t+1} and α_t^k, to make the weights even. Thus this method is adaptable and extremely flexible. In the next section we give a convenient generic suggestion for the choice of $g(\cdot)$.

Rejection sampling for auxiliary particle filtering could also be used in this context. An example of this appears in Section 3.3.4. We can also make proposals for an MCMC variate of the auxiliary particle filter from $\alpha_{t+1}^{(i+1)}, k^{(i+1)} \sim g(\alpha_{t+1}, k|Y_{t+1})$, where $g(\alpha_{t+1}, k|Y_{t+1})$ is some arbitrary density; then these moves are accepted with probability

$$\min\left\{1, \frac{f(y_{t+1}|\alpha_{t+1}^{(i+1)})f(\alpha_{t+1}^{(i+1)}|\alpha_t^{k^{(i+1)}})}{f(y_{t+1}|\alpha_{t+1}^{(i)})f(\alpha_{t+1}^{(i)}|\alpha_t^{k^{(i)}})} \times \frac{g(\alpha_{t+1}^{(i)}, k^{(i)}|Y_{t+1})}{g(\alpha_{t+1}^{(i+1)}, k^{(i+1)}|Y_{t+1})}\right\}.$$

A special case of this argument has been given by Berzuini et al. (1997), who put $g(\alpha_{t+1}, k|Y_{t+1}) \propto f(\alpha_{t+1}|\alpha_t^k)$, which means that their method is again blind.

3.2 A Generic Sampling/Importance Resampling–Based Auxiliary Proposal

Here we give a generic $g(\cdot)$ that can be broadly applied. We base our discussion on the SIR algorithm, although we could have used an MCMC method. We approximate (5) by

$$g(\alpha_{t+1}, k|Y_{t+1}) \propto f(y_{t+1}|\mu_{t+1}^k)f(\alpha_{t+1}|\alpha_t^k)\pi^k, \qquad k = 1, \ldots, M,$$

where μ_{t+1}^k is the mean, the mode, a draw, or some other likely value associated with the density of $\alpha_{t+1}|\alpha_t^k$. The form of the approximating density is designed so that

$$g(k|Y_{t+1}) \propto \pi^k \int f(y_{t+1}|\mu_{t+1}^k)\, dF(\alpha_{t+1}|\alpha_t^k) = \pi^k f(y_{t+1}|\mu_{t+1}^k).$$

Thus we can sample from $g(\alpha_{t+1}, k|Y_{t+1})$ by simulating the index with probability $\lambda_k \propto g(k|Y_{t+1})$, and then sampling from the transition density given the mixture $f(\alpha_{t+1}|\alpha_t^k)$. We call the λ_k the first-stage weights.

The implication is that we simulate from particles associated with large predictive likelihoods. Having sampled

the joint density of $g(\alpha_{t+1}, k|Y_{t+1})$ R times, we perform a reweighting, putting on the draw (α_{t+1}^j, k^j) the weights proportional to the so-called second-stage weights

$$w_j = \frac{f(y_{t+1}|\alpha_{t+1}^j)}{f(y_{t+1}|\mu_{t+1}^{k^j})}, \qquad \pi_j = \frac{w_j}{\sum_{i=1}^R w_i}, \qquad j = 1, \ldots, R.$$

The hope is that these second-stage weights are much less variable than for the original SIR method. We might resample from this discrete distribution to produce a sample of size M.

By making proposals that have high conditional likelihoods, we reduce the costs of sampling many times from particles that have very low likelihoods and so will not be resampled at the second stage of the process. This improves the statistical efficiency of the sampling procedure and means that we can reduce the value of R substantially.

To measure the statistical efficiency of these procedures, we argued earlier that we could look at minimizing $E\{\pi(\alpha)^2\}$. Here we compare a standard SIR with a SIR based on our auxiliary variable. For simplicity, we set $\pi^k = 1/M$ in both cases. Then, for a standard SIR-based particle filter, for large M,

$$E\{\pi(\alpha)^2\} = \frac{\frac{1}{M}\sum_{k=1}^M \int f(y_{t+1}|\alpha_{t+1})^2\, dF(\alpha_{t+1}|\alpha_t^k)}{\left\{\frac{1}{M}\sum_{k=1}^M \int f(y_{t+1}|\alpha_{t+1})\, dF(\alpha_{t+1}|\alpha_t^k)\right\}^2}$$

$$= \frac{M\sum_{k=1}^M \lambda_k^2 f_k}{(\sum_{k=1}^M \lambda_k f_k^*)^2},$$

where

$$f_k = \int \left\{\frac{f(y_{t+1}|\alpha_{t+1})}{f(y_{t+1}|\mu_{t+1}^k)}\right\}^2 dF(\alpha_{t+1}|\alpha_t^k)$$

and

$$f_k^* = \int \left\{\frac{f(y_{t+1}|\alpha_{t+1})}{f(y_{t+1}|\mu_{t+1}^k)}\right\} dF(\alpha_{t+1}|\alpha_t^k).$$

The same calculation for a SIR-based auxiliary variable particle filter gives

$$E\{\pi_\alpha(\alpha)^2\} = \frac{\sum_{k=1}^M \lambda_k f_k}{(\sum_{k=1}^M \lambda_k f_k^*)^2},$$

which shows an efficiency gain if

$$\sum_{k=1}^M \lambda_k f_k < M\sum_{k=1}^M \lambda_k^2 f_k.$$

If f_k does not vary over k, then the auxiliary variable particle filter will be more efficient as $\sum_{k=1}^M \lambda_k(1/M) = (1/M) \le \sum_{k=1}^M \lambda_k^2$. More likely is that f_k will depend on k, but only mildly, as $f(\alpha_{t+1}|\alpha_t^k)$ will be typically quite tightly peaked [much more tightly peaked than $f(\alpha_{t+1}|Y_t)$] compared to the conditional likelihood.

3.3 Examples of Adaption

3.3.1 Basics. Although the previous generic scheme can usually reduce the variability of the second-stage weights, other adaption schemes use the specific structure of the time series model to allow us to achieve yet more equal weights. If we can achieve exactly equal weights, then we say that we have *fully adapted* the procedure to the model, for now we can produce iid samples from (4). This situation is particularly interesting, as we are then close to the assumptions made by Kong et al. (1994) for their sequential importance sampler. Although full adaption is of some practical importance, we should remind ourselves that even fully adapted particle filters do not produce iid samples from $f(\alpha_{t+1}|Y_{t+1})$, due to their approximation of $f(\alpha_{t+1}|Y_t)$ by a finite mixture distribution. This is inherent in the construction of this class of filter.

3.3.2 Nonlinear Gaussian Measurement Model. In the Gaussian measurement case, the absorption of the measurement density into the transition equation is particularly convenient. Consider a nonlinear transition density with $\alpha_{t+1}|\alpha_t \sim \mathrm{N}\{\mu(\alpha_t), \sigma^2(\alpha_t)\}$ and $y_{t+1}|\alpha_{t+1} \sim \mathrm{N}(\alpha_{t+1}, 1)$. Then

$$f(\alpha_{t+1}, k|Y_{t+1}) \propto f(y_{t+1}|\alpha_{t+1}) f(\alpha_{t+1}|\alpha_t^k)$$
$$= g_k(y_{t+1}) f(\alpha_{t+1}|\alpha_t^k, y_{t+1}),$$

where

$$f(\alpha_{t+1}|\alpha_t^k, y_{t+1}) = \mathrm{N}(\mu_k^*, \sigma_k^{*2}),$$

$$\mu_k^* = \sigma_k^{*2}\left\{\frac{\mu(\alpha_t^k)}{\sigma^2(\alpha_t^k)} + y_{t+1}\right\},$$

and

$$\sigma_k^{-*2} = 1 + \sigma^{-2}(\alpha_t^k).$$

This implies that the first-stage weights are

$$g_k(y_{t+1}) \propto \frac{\sigma_k^*}{\sigma(\alpha_t^k)} \exp\left\{\frac{\mu_k^{*^2}}{2\sigma_k^{*^2}} - \frac{\mu(\alpha_t^k)^2}{2\sigma^2(\alpha_t^k)}\right\}.$$

The Gaussian measurement density implies that the second-stage weights are all equal.

An example of this is a Gaussian autoregressive conditional heteroscedasticity (ARCH) model (see, e.g., Bollerslev, Engle, and Nelson 1994) observed with independent Gaussian error. So we have

$$y_t|\alpha_t \sim \mathrm{N}(\alpha_t, \sigma^2), \qquad \alpha_{t+1}|\alpha_t \sim \mathrm{N}(0, \beta_0 + \beta_1\alpha_t^2).$$

This model is fully adaptable. It has received a great deal of attention in the econometric literature, as it has some attractive multivariate generalizations; see the work by Diebold and Nerlove (1989), Harvey, Ruiz, and Sentana (1992) and King, Sentana, and Wadhwani (1994). As far as we know, no likelihood methods exist in the literature for the analysis of this type of model (and its various generalizations), although a number of very good approximations have been suggested.

3.3.3 Log-Concave Measurement Densities. Suppose again that $f(\alpha_{t+1}|\alpha_t^k)$ is Gaussian, but the measurement density is log-concave as a function of α_{t+1}. Then we might extend the foregoing argument by Taylor expanding $\log f(y_{t+1}|\alpha_{t+1})$ to a second-order term, again around μ_{t+1}^k, to give the approximation

$$\begin{aligned}\log g(y_{t+1}|\alpha_{t+1}, \mu_{t+1}^k) &= \log f(y_{t+1}|\mu_{t+1}^k) + (\alpha_{t+1} - \mu_{t+1}^k)' \\ &\quad \times \frac{\partial \log f(y_{t+1}|\mu_{t+1}^k)}{\partial \alpha_{t+1}} \\ &\quad + \frac{1}{2}(\alpha_{t+1} - \mu_{t+1}^k)' \\ &\quad \times \frac{\partial^2 \log f(y_{t+1}|\mu_{t+1}^k)}{\partial \alpha_{t+1} \partial \alpha_{t+1}'}(\alpha_{t+1} - \mu_{t+1}^k).\end{aligned}$$

Then

$$g(\alpha_{t+1}, k|Y_{t+1}) \propto g(y_{t+1}|\alpha_{t+1}; \mu_{t+1}^k) f(\alpha_{t+1}|\alpha_t^k).$$

Rearranging, we can express this as

$$g(\alpha_{t+1}, k|Y_{t+1}) \propto g(y_{t+1}|\mu_{t+1}^k) g(\alpha_{t+1}|\alpha_t^k, y_{t+1}; \mu_{t+1}^k),$$

which means that we could simulate the index with probability proportional to $g(y_{t+1}|\mu_{t+1}^k)$ and then draw from $g(\alpha_{t+1}|\alpha_t^k, y_{t+1}, \mu_{t+1}^k)$. The resulting reweighted sample's second-stage weights are proportional to the, hopefully fairly even, weights

$$\begin{aligned} w_j &= \frac{f(y_{t+1}|\alpha_{t+1}^j) f(\alpha_{t+1}|\alpha_t^{k_j})}{g(y_{t+1}|\mu_{t+1}^{k_j}) g(\alpha_{t+1}^j|\alpha_t^{k_j}, y_{t+1}, \mu_{t+1}^k)} \\ &= \frac{f(y_{t+1}|\alpha_{t+1}^j)}{g(y_{t+1}|\alpha_{t+1}^j; \mu_{t+1}^{k_j})}, \\ \pi_j &= \frac{w_j}{\sum_{i=1}^R w_i}, \qquad j = 1, \ldots, R. \end{aligned}$$

Thus we can exploit the special structure of the model, if available, to improve upon the auxiliary particle filter.

3.3.4 Stochastic Volatility and Rejection Sampling. The same argument carries over when we use a first-order Taylor expansion to construct $g(y_{t+1}|\alpha_{t+1}, \mu_{t+1}^k)$, but in this case we know that $g(y_{t+1}|\alpha_{t+1}, \mu_{t+1}^k) \geq f(y_{t+1}|\alpha_{t+1})$ for any value of μ_{t+1}^k due to the assumed log-concavity of the measurement density. Thus

$$\begin{aligned} f(\alpha_{t+1}, k|Y_{t+1}) &\propto f(y_{t+1}|\alpha_{t+1}) f(\alpha_{t+1}|\alpha_t^k) \\ &\leq g(y_{t+1}|\alpha_{t+1}; \mu_{t+1}^k) f(\alpha_{t+1}|\alpha_t^k) \\ &= g(y_{t+1}|\mu_{t+1}^k) g(\alpha_{t+1}|\alpha_t^k, y_{t+1}; \mu_{t+1}^k) \\ &\propto g(\alpha_{t+1}, k|Y_{t+1}). \end{aligned}$$

Thus we can perform rejection sampling from $f(\alpha_{t+1}, k|Y_{t+1})$ by simply sampling k with probability proportional to $g(y_{t+1}|\mu_{t+1}^k)$ and then drawing α_{t+1} from $g(\alpha_{t+1}|\alpha_t^k, y_{t+1}; \mu_{t+1}^k)$. This pair is then accepted with probability $f(y_{t+1}|\alpha_{t+1})/g(y_{t+1}|\alpha_{t+1}; \mu_{t+1}^k)$.

This argument applies to the stochastic volatility (SV) model,

$$y_t = \varepsilon_t \beta \exp(\alpha_t/2), \qquad \alpha_{t+1} = \phi \alpha_t + \eta_t, \tag{6}$$

where ε_t and η_t are independent Gaussian processes with variances 1 and σ^2. Here β has the interpretation as the modal volatility, ϕ is the persistence in the volatility shocks, and σ_η^2 is the volatility of the volatility. This model has attracted much recent attention in the econometrics literature as a way of generalizing the Black–Scholes option pricing formula to allow volatility clustering in asset returns (see, e.g., Hull and White 1987). MCMC methods have been used on this model by, for instance, Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998), and Shephard and Pitt (1997).

For this model $\log f(y_{t+1}|\alpha_{t+1})$ is concave in α_{t+1} so that, for $\mu_{t+1}^k = \phi \alpha_t^k$,

$$\begin{aligned}\log g(y_{t+1}|\alpha_{t+1}; \mu_{t+1}^k) &= \text{const} - \frac{1}{2}\alpha_{t+1} \\ &\quad - \frac{y_t^2}{2\beta^2} \exp(-\mu_{t+1}^k)\{1 - (\alpha_{t+1} - \mu_{t+1}^k)\}.\end{aligned}$$

The implication is that

$$\begin{aligned} g(\alpha_{t+1}|\alpha_t^k, y_{t+1}; \mu_{t+1}^k) &= \mathrm{N}\left[\mu_{t+1}^k + \frac{\sigma^2}{2}\left\{\frac{y_t^2}{\beta^2}\exp(-\mu_{t+1}^k) - 1\right\}, \sigma^2\right] \\ &= \mathrm{N}(\mu_{t+1}^{*k}, \sigma^2). \end{aligned}$$

Likewise,

$$\begin{aligned} g(y_{t+1}|\mu_{t+1}^k) &= \exp\left\{\frac{1}{2\sigma^2}(\mu_{t+1}^{*k2} - \mu_{t+1}^{k2})\right\} \\ &\quad \times \exp\left\{-\frac{y_t^2}{2\beta^2}\exp(-\mu_{t+1}^k)(1 + \mu_{t+1}^k)\right\}. \end{aligned}$$

Finally, the log-probability of acceptance is

$$-\frac{y_t^2}{2\beta^2}[\exp(-\alpha_{t+1}) - \exp(-\mu_{t+1}^k)\{1 - (\alpha_{t+1} - \mu_{t+1}^k)\}].$$

Notice that as σ^2 falls to 0, so the acceptance probability goes to 1.

Finally, the same argument holds when we use a SIR algorithm instead of rejection sampling. The proposals are made in exactly the same way, but now instead of computing log-probabilities of accepting, these become log–second-stage weights.

3.3.5 Limited Dependent Processes. A less trivial example of full adaption is a special case of limited dependent processes, where the observations are deterministic functions of the states. A simple example of this is a probit time series where $y_t = I(\alpha_t > 0)$, where α_t is Gaussian and univariate and $I(\cdot)$ denotes an indicator function. Then

if $y_{t+1} = 1$, we have, exactly,

$$\Pr(\alpha_{t+1}, k|Y_{t+1}) \propto w^k \Pr(\alpha_{t+1}|\alpha_t^k, \alpha_{t+1} > 0),$$

$$w^k = \Pr(\alpha_{t+1} > 0|\alpha_t^k).$$

Hence we choose k with probability proportional to w^k and then draw from a truncated distribution conditional on k. If y_{t+1} is negative, then the weights w^k would be $\Pr(\alpha_{t+1} < 0|\alpha_t^k)$, and while the truncated draw would be from $\Pr(\alpha_{t+1}|\alpha_t^k, \alpha_{t+1} < 0)$. This style of argument carries over to ordered probit and censored models where we observe, for example, $\min(0, \alpha_t)$.

Adaption can be very important in these types of models, for naively implemented particle and auxiliary variable filters are generally vulnerable to tightly peaked measurement densities. In the censored model, where $y_{t+1} = \min(0, \alpha_{t+1})$, the measurement density is degenerate when $y_{t+1} > 0$, and so the particle filter will degenerate to give all of its mass on the simulation that is closest (but because they are simulated from $\Pr(\alpha_{t+1}|\alpha_t^k)$ not equal) to y_{t+1}. Adaption overcomes this problem instantly.

Adaption is also essential for the following problem. Suppose that $\alpha_{t+1}|\alpha_t$ is Gaussian, α_{t+1} is bivariate, and we observe $y_{t+1} = \min(\alpha_{t+1})$. Such models are called disequilibrium models in economics. (Recent work in this area includes Laroque and Salanié 1993 and Manrique and Shephard 1998.) Then

$$\Pr(\alpha_{t+1}, k|Y_{t+1}) \propto \Pr(y_{t+1}|\alpha_{t+1}) \Pr(\alpha_{t+1}|\alpha_t^k).$$

Then we have that w^k should be proportional to the probability of $\alpha_{t+1}|\alpha_t^k$ having its minimum exactly at y_{t+1}. This can be shown to be exactly

$$w^k = f_{\alpha^k_{1,t+1|\alpha_t}}(y_{t+1})\{1 - \Pr \alpha_{2,t+1|\alpha_t}(y_{t+1})\} + f_{\alpha_{2,t+1|\alpha_t}}(y_{t+1})\{1 - \Pr \alpha_{1,t+1|\alpha_t}(y_{t+1})\},$$

while, having selected k, we sample from

$$\alpha_{1,t+1} = y_{t+1} \quad \text{with probability}$$

$$\lambda_{t+1} = \frac{f_{\alpha_{1,t+1|\alpha_t^k}}(y_{t+1})\{1 - \Pr \alpha_{2,t+1|\alpha_t^k}(y_{t+1})\}}{w^k},$$

and then from

$$\alpha_{2,t+1}|\alpha_{1,t+1} = y_{t+1}, \alpha_t^k, \alpha_{2,t+1} > y_{t+1}.$$

Likewise, $\alpha_{2,t+1} = y_{t+1}$ with probability $1 - \lambda_{t+1}$.

3.3.6 Mixtures of Normals. Suppose that $f(\alpha_{t+1}|\alpha_t)$ is Gaussian, but the measurement density is a discrete mixture of normals $\sum_{j=1}^{P} \lambda_j f_j(y_{t+1}|\alpha_{t+1})$. Then we can perfectly sample from $f(\alpha_{t+1}, k|Y_{t+1})$ by working with

$$f(\alpha_{t+1}, k, j|Y_{t+1}) \propto \lambda_j f_j(y_{t+1}|\alpha_{t+1}) f(\alpha_{t+1}|\alpha_t^k) = w_{j,k} f_j(\alpha_{t+1}|\alpha_t^k, y_{t+1}).$$

Then we sample from $f(\alpha_{t+1}, k, j|Y_{t+1})$ by selecting the index k, j with probability proportional to $w_{j,k}$ and then drawing from $f_j(\alpha_{t+1}|\alpha_t^k, y_{t+1})$. The disadvantage of this approach is that the complete enumeration and storage of $w_{j,k}$ involves $P \times M$ calculations. This approach can be trivially extended to cover the case where $f(\alpha_{t+1}|\alpha_t)$ is a mixture of normals. MCMC smoothing methods

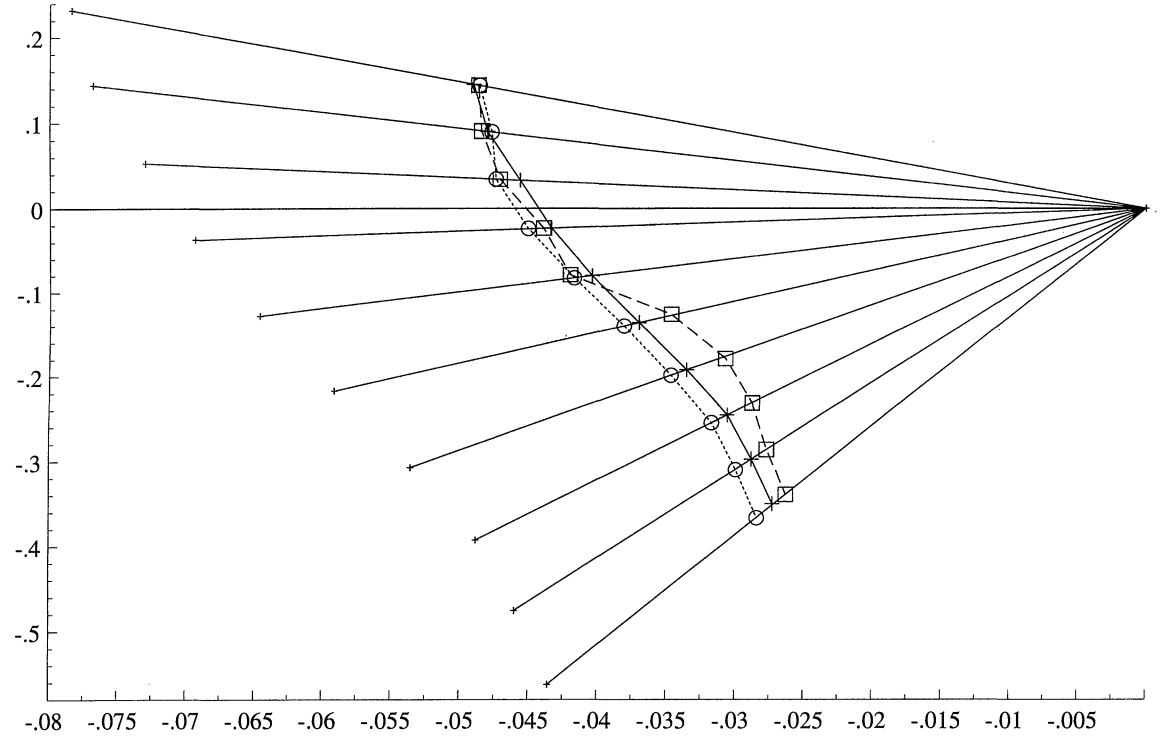

Figure 1. Plot of the Angular Measurements From Origin, the True Trajectory (Solid Line, Crosses), the Particle Filtered Mean Trajectory (Dashed Line, Boxes), and the Auxiliary Particle Mean Trajectory (Dotted Line, Circles). Ship moving southeast. T = 10, M = 300, R = 500.

for state–space models with mixtures have been studied by, for example, Carter and Kohn (1994) and Shephard (1994).

4. NUMERICAL EXAMPLE

4.1 A Time Series of Angles

4.1.1 The Model. In this section we compare the performance of the particle and auxiliary particle filter methods for an angular time series model: the bearings-only model. We consider the simple scenario described by Gordon et al. (1993). The observer is considered stationary at the origin of the $x - z$ plane, and the ship is assumed to gradually accelerate or decelerate randomly over time. We use the following discretisation of this system, where $\alpha_t = (x_t, vx_t, z_t, vz_t)'$:

$$\alpha_{t+1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \alpha_t + \sigma_\eta \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \\ 0 & 1 \end{pmatrix} u_t,$$

$$u_t \sim \text{NID}(0, 1). \quad (7)$$

In obvious notation x_t and z_t represent the ship's horizontal and vertical positions at time t and vx_t and vz_t represent the corresponding velocities. The state evolution is thus a VAR(1) of the form $\alpha_{t+1} = T\alpha_t + Hu_t$. The model indicates that the state evolution error arises because the accelerations are white noise. The initial state describes the ship's starting positions and velocities $\alpha_1 \sim \text{NID}(a_1, P_1)$. This prior, together with the state evolution of (7), describe the overall prior for the states.

Our model is based on a mean direction $\mu_t = \tan^{-1}(z_t/x_t)$. The measured angle is assumed to be wrapped Cauchy with density (see, e.g., Fisher 1993, p. 46)

$$f(y_t|\mu_t) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho\cos(y_t - \mu_t)},$$

$$0 \le y_t < 2\pi, \qquad 0 \le \rho \le 1, \quad (8)$$

where ρ is the mean resultant length.

4.1.2 The Simulated Scenario. To assess the relative efficiency of the particle filter and the basic auxiliary method discussed in Section 3.2, we have closely followed the setup described by Gordon et al. (1993). They considered $\sigma_\eta = .001$ and $\sigma_\varepsilon = .005$, where $z_t|\mu_t \sim \text{NID}(\mu_t, \sigma_\varepsilon^2)$. We choose $\rho = 1 - \sigma_\varepsilon^2$ (yielding the same circular dispersion) for our wrapped Cauchy density. The actual initial starting vector of this is taken to be $\alpha_1 = (-.05, .001, .2, -.055)'$. In contrast to the method of Gordon et al. (1993), however, we wish to have an extremely accurate and tight prior for the initial state. This is because we want the variance of quantities arising from the filtered

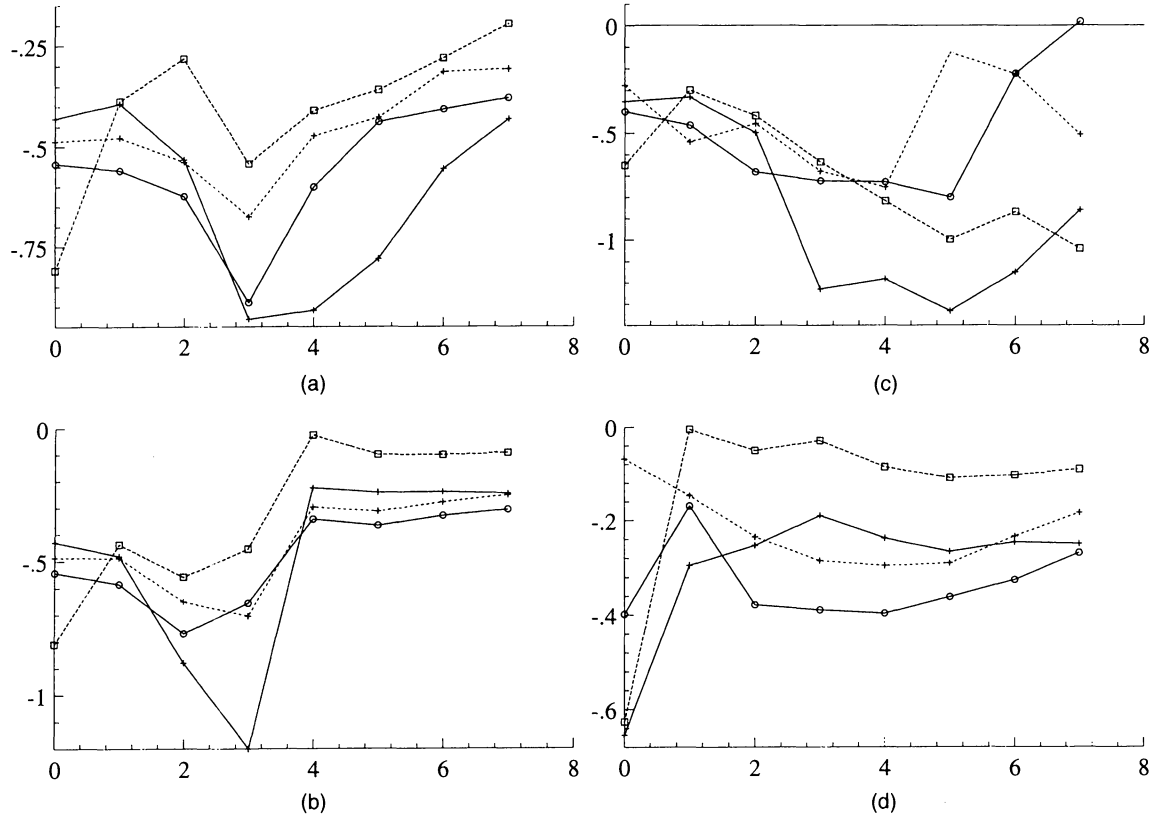

Figure 2. Plot of the Relative MSE Performance (on the Log Scale) of the Particle Filter and the Auxiliary-Based Particle Filter for the Bearings-Only Tracking Problem. Numbers below 0 indicate a superior performance by the auxiliary particle filter. In these graphs M = 4,000 or 8,000 and R = M or R = 2M. Throughout SIR is used as the sampling mechanism. (a) $\alpha_{t1} = x_t$; (b) $\alpha_{t3} = z_t$; (c) $\alpha_{t2} = vx_t$; (d) $\alpha_{t4} = vz_t$. + — — — +, M and R = 4,000; ○ — — — ○, (b) M and R = 8,000; (c) □ — — — □, M = 4,000 and (d) R = 8,000; + - - - +, M = 8,000 and R = 16,000.

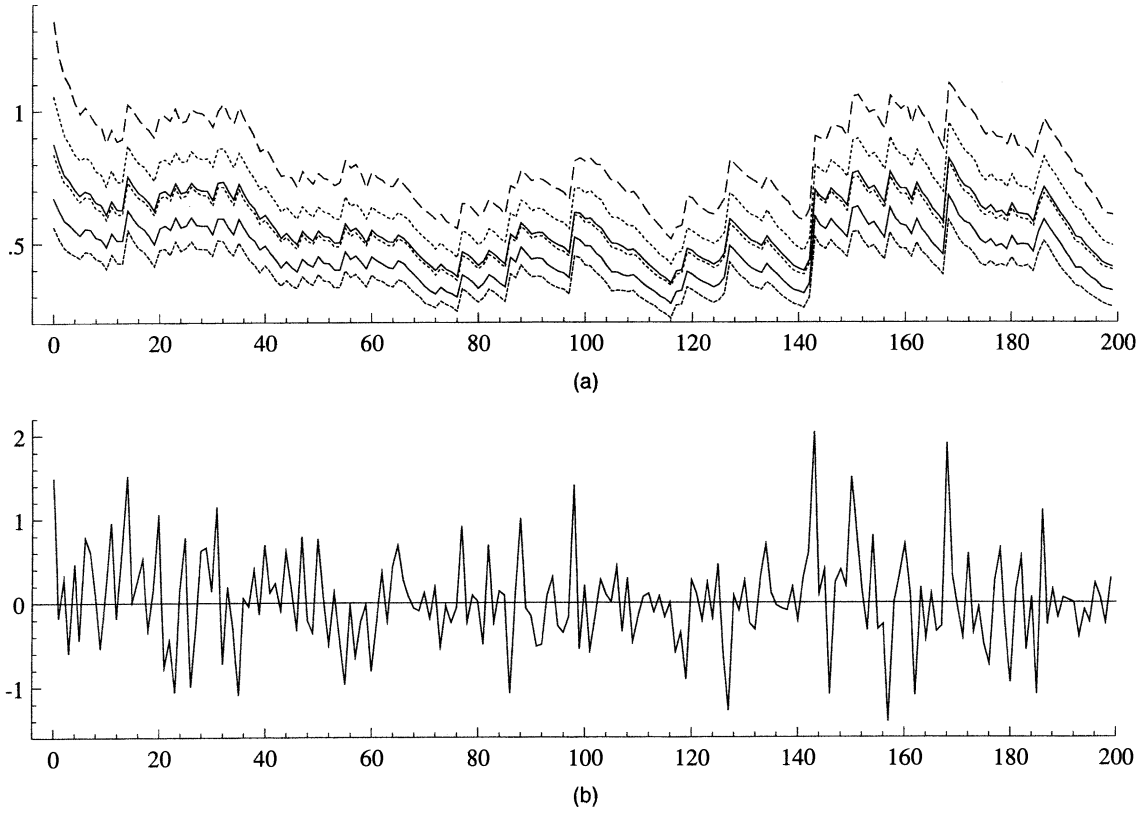

Figure 3. (a) The Posterior Filtered Mean (Heavy Line) of $\beta \exp(\alpha_t/2)|Y_t$, Together with the 5, 20, 50, 80, and 95 Percentage Points of the Distribution and (b) The Daily Returns on the U.S. dollar Against the U.K. pound sterling from the first Day of Trading in 1997 for 200 Trading Days. Notice the median is always below the mean. M = 5,000; R = 6,000.

posterior density to be small, allowing us to formulate reasonably conclusive evidence about the relative efficiency of the auxiliary method to the standard method. Thus we take $a_1 = \alpha_1$ and have a diagonal initial variance P_1 with the elements $.01 \times (.5^2, .005^2, .3^2, .01^2)$ on the diagonal.

Figure 1 illustrates a realization of the model for the foregoing scenario with $T = 10$. The ship is moving in a southeasterly direction over time. The trajectories given by the posterior filtered means from the particle method and the auxiliary method ($M = 300$ and $R = 500$ in both cases) are both fairly close to the true path despite the small amount of simulation used.

4.1.3 Monte Carlo Comparison. We now compare the two methods using a Monte Carlo study of the foregoing scenario with $T = 10$. The "true" filtered mean is calculated for each replication by using the auxiliary method with $M = 100{,}000$ and $R = 120{,}000$. Within each replication, the mean squared error (MSE) for the particle method for each component of the state over time is evaluated by running the method, with a different random number seed, S times and recording the average of the resulting squared difference between the resulting particle filter's estimated mean and the "true" filtered mean. Hence for replication i, state component j, at time t, we calculate

$$\mathrm{MSE}^P_{i,j,t} = \frac{1}{S}\sum_{s=1}^{S} (\bar{\alpha}^i_{t,j,s} - \tilde{\alpha}^i_{t,j})^2,$$

where $\bar{\alpha}^i_{t,j,s}$ is the particle mean for replication i, state component j, at time t, for simulation s and $\tilde{\alpha}^i_{t,j}$ is the "true" filtered mean replication i, state component j, at time t. The log mean squared error (LMSE) for component j at time t is obtained as

$$\mathrm{LMSE}^P_{j,t} = \log \frac{1}{\mathrm{REP}} \sum_{i=1}^{\mathrm{REP}} \mathrm{MSE}^{\,P}_{i,j,t}.$$

The same operation is performed for the auxiliary method to deliver the corresponding quantity $\mathrm{LMSE}^{\mathrm{AM}}_{j,t}$. For this study, we use $\mathrm{REP} = 40$ and $S = 20$. We allow $M = 4{,}000$ or $8{,}000$, and for each of these values we set $R = M$ or $2M$. Figure 2 shows the relative performance of the two methods for each component of the state vector over time. For each component j, the quantity $\mathrm{LMSE}^{\mathrm{AM}}_{j,t} - \mathrm{LMSE}^P_{j,t}$ is plotted against time. Values close to 0 indicate that the two methods are broadly equivalent in performance; negative values indicate that the auxiliary method performs better than the standard particle filter.

The graphs give the expected result, with the auxiliary particle filter typically being more precise, but with the difference between the two methods falling as R increases.

4.2 Stochastic Volatility

The basic SV model was defined in Section 3.3.4. Here we construct 100 times the compound daily returns on the U.S. dollar against the U.K. pound sterling from the first day of trading in 1997 and for the next 200 days of active trading. (This dataset is discussed in more detail in Pitt

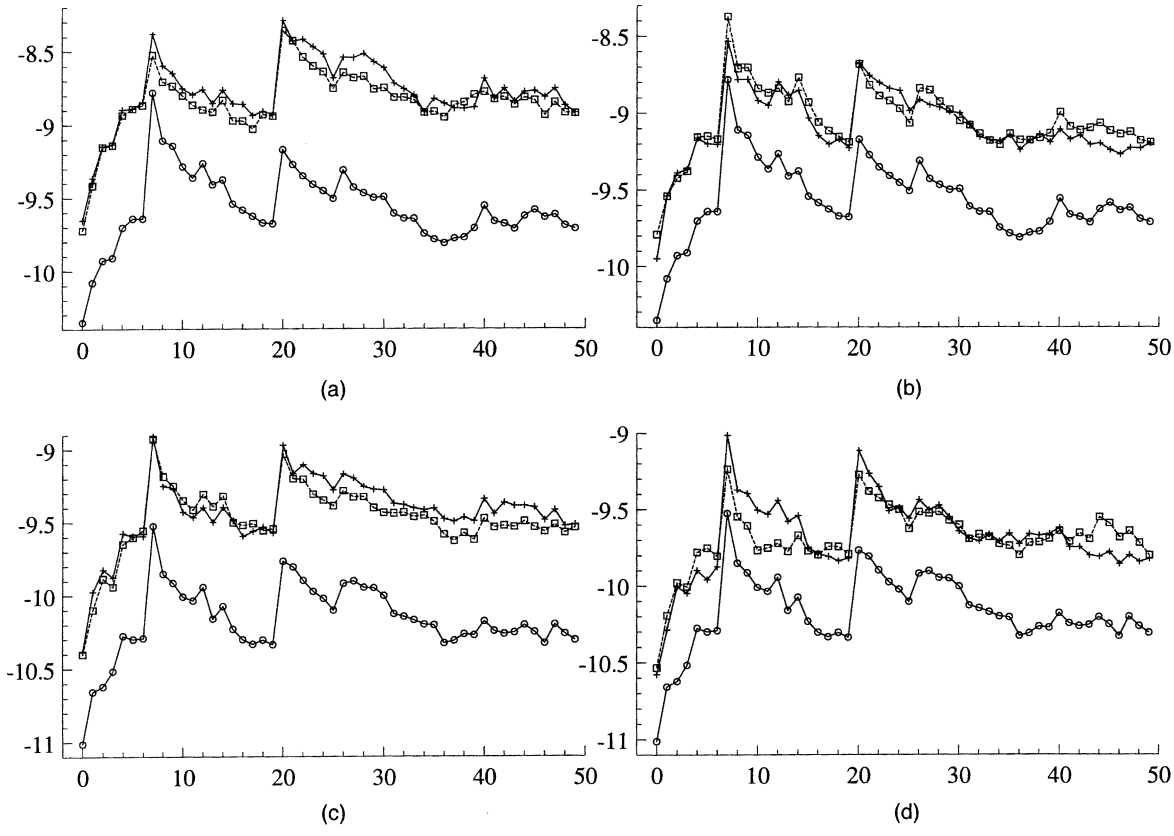

Figure 4. Plot of the MSE Performance (on the Log Scale) of the Particle Filter to the Auxiliary-Based Particle Filter and an Adapted Particle Filter. The lower the number, the more efficient the method. (a) and (b) have M = 2,000; (c) and (d) have M = 4,000. (a) and (c) have R = M; (b) and (d) have R = 2M. (+- - -+ SIR; □ — — — □ auxiliary SIR; ○- - -○ adapted SIR.)

and Shephard 1999, where we estimated the parameters of the model using Bayesian methods.) Throughout we take $\phi = .9702, \sigma_\eta = .178$, and $\beta = .5992$, the posterior means of the model for a long time series of returns up until the end of 1996.

Figure 3 graphs these daily returns against time. The figure also displays the estimated quantiles of the filtering density, $f\{\beta \exp(\alpha_t/2)|Y_t\}$ computed using an auxiliary particle filter. Throughout the series, we set $M = 5{,}000$ and $R = 6{,}000$. We have also displayed the posterior mean of the filtering random variable. This is always very slightly above the posterior median, as $\alpha_t|Y_t$ is very close to being symmetric.

The figure shows that the filtered volatility jumps up more quickly than it tends to go down. This reflects the fact that the volatility is modeled on the log scale.

4.2.1 Simulation Experiment. To compare the efficiency of the simple particle filter, our basic auxiliary particle filter, and the (rejection-based) fully adapted particle filter discussed in Section 3.3.4, we again conducted a simulation experiment measuring MSE for each value of t using the foregoing model and again having $n = 50$. We simulated the data using the model parameters discussed earlier. Figure 4 shows the results (using a log scale). To make the problem slightly more realistic and challenging, we set $\varepsilon_{21} = 2.5$ for each series, so there is a significant outlier at that point. For this study, we set REP $= 40$ and $S = 20$. We allow $M = 2{,}000$ or $4{,}000$, and for each of these values we set $R = M$ or $2M$. For the rejection-based particle filter algorithm, it only makes sense to take $M = R$, and so when $R > M$, we repeat the calculations as if $M = R$. Finally, the rejection-based method takes approximately twice the time of the SIR-based particle filter when $M = R$.

The plot shows that the fully adapted particle filter is considerably more accurate than the other particle filters. It also has the advantage of not depending on R. The auxiliary particle filter is more efficient than the plain particle filter, but the difference is small, reflecting the fact that for the SV model, the conditional likelihood is not very sensitive to the state.

5. CONCLUSION

In this article we have studied the weaknesses of the very attractive particle filtering method proposed by Gordon et al. (1993). The SIR implementation of this method is not robust to outliers for two different reasons: sampling efficiency and the unreliability of the empirical prediction density in the tails of the distribution. We have introduced an auxiliary variable into the particle filter to overcome the first of these problems, providing a powerful framework that is as simple as SIR, but more flexible and reliable.

[Received October 1997. Revised September 1998.]

REFERENCES

Berzuini, C., Best, N. B., Gilks, W. R., and Larizza, C. (1997), "Dynamic Conditional Independence Models and Markov Chain Monte Carlo Methods," *Journal of the American Statistical Association*, 92, 1403–1412.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994), "ARCH Models," in *The Handbook of Econometrics, Vol. 4*, eds. R. F. Engle and D. McFadden, Amsterdam: North-Holland, pp. 2959–3038.

Carpenter, J. R., Clifford, P., and Fearnhead, P. (1998), "An Improved Particle Filter for Non-Linear Problems," working paper, University of Oxford, Dept. of Statistics.

Carter, C. K., and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Diebold, F. X., and Nerlove, M. (1989), "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model," *Journal of Applied Econometrics*, 4, 1–21.

Doornik, J. A. (1996), *Ox: Object Oriented Matrix Programming, 2.0*, London: Timberlake Consultants Press.

Doucet, A. (1998), "On Sequential Simulation-Based Methods for Bayesian Filtering," unpublished manuscript, University of Cambridge, Dept. of Engineering.

Fisher, N. I. (1993), *Statistical Analysis of Circular Data*, Cambridge, U.K.: Cambridge University Press.

Gerlach, R., Carter, C., and Kohn, R. (1996), "Diagnostics for Time Series Analysis," unpublished manuscript, University of New South Wales, Australian Graduate School of Management.

Gilks, W. K., Richardson, S., and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993), "A Novel Approach to Non-Linear and Non-Gaussian Bayesian State Estimation," *IEEE Proceedings F*, 140, 107–113.

Harvey, A. C., Ruiz, E., and Sentana, E. (1992), "Unobserved Component Time Series Models With ARCH Disturbances," *Journal of Econometrics*, 52, 129–158.

Hendry, D. F., and Richard, J. F. (1991), "Likelihood Evaluation for Dynamic Latent Variable Models," in *Computational Economics and Econometrics*, Dordrecht: Kluwer.

Hull, J., and White, A. (1987), "The Pricing of Options on Assets With Stochastic Volatilities," *Journal of Finance*, 42, 281–300.

Hürzeler, M., and Künsch, H. R. (1995), "Monte Carlo Approximations for General State Space Models," unpublished manuscript, Seminar für Statistik, ETH.

Isard, M., and Blake, A. (1996), "Contour Tracking by Stochastic Propagation of Conditional Density," *Proceedings of the European Conference on Computer Vision, Cambridge*, 1, 343–356.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models" (with discussion), *Journal of Business and Economic Statistics*, 12, 371–417.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison With ARCH Models," *The Review of Economic Studies*, 65, 361–393.

King, M., Sentana, E., and Wadhwani, S. (1994), "Volatility and Links Between National Stock Markets," *Econometrica*, 62, 901–933.

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

——— (1996), "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, 5, 1–25.

Kong, A., Liu, J. S., and Wong, W. H. (1994), "Sequential Imputations and Bayesian Missing-Data Problems," *Journal of the American Statistical Association*, 89, 278–288.

Laroque, G., and Salanié, B. (1993), "Simulation-Based Estimation of Models With Lagged Latent Variables," *Journal of Applied Econometrics*, 8, S119–S133.

Liu, J. (1996), "Metropolized Independent Sampling With Comparison to Rejection Sampling and Importance Sampling," *Statistics and Computing*, 6, 113–119.

Liu, J., and Chen, R. (1995), "Blind Deconvolution via Sequential Imputation," *Journal of the American Statistical Association*, 90, 567–576.

——— (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93, 1032–1044.

Manrique, A., and Shephard, N. (1998), "Simulation Based Likelihood Inference for Gaussian Limited Dependent Processes," *Econometrics Journal*, 1, C174–C202.

Pitt, M. K., and Shephard, N. (1998), "A Fixed Lag Auxiliary Particle Filter With Deterministic Sampling Rules," unpublished manuscript, Nuffield College.

——— (1999), "Time Varying Covariances: A Factor Stochastic Volatility Approach" (with discussion), in Bayesian Statistics 6, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press.

Rubin, D. B. (1987), "Comment on 'The Calculation of Posterior Distributions by Data Augmentation' by M. A. Tanner and W. H. Wong," *Journal of the American Statistical Association*, 82, 543–546.

Shephard, N. (1994), "Partial Non-Gaussian State Space," *Biometrika*, 81, 115–131.

Shephard, N., and Pitt, M. K. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84, 653–667.

Smith, J. Q. (1985), "Diagnostic Checks of Non-Standard Time Series Models," *Journal of Forecasting*, 4, 283–291.

Sorensen, H. W., and Alspach, D. L. (1971), "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, 7, 465–479.

West, M. (1992), "Mixture Models, Monte Carlo, Bayesian Updating and Dynamic Models," *Computer Science and Statistics*, 24, 325–333.

West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer-Verlag.