

Auxiliary variable based particle filters

Michael K. Pitt & Neil Shephard

Prepared for a book on “Sequential Monte Carlo Methods in Practice”, edited by Nando de Freitas, Arnaud Doucet and Neil Gordon, to be published by Chapman and Hall, London, 1999.

1 Introduction

We model a time series $\{y_t, t = 1, \dots, n\}$ using a state space framework with the $\{y_t|\alpha_t\}$ being independent and with the state $\{\alpha_t\}$ assumed to be Markovian. The task will be to use simulation to estimate $f(\alpha_t|\mathcal{F}_t)$, $t = 1, \dots, n$, where \mathcal{F}_t is contemporaneously available information. We assume a known ‘measurement’ density $f(y_t|\alpha_t)$ and the ability to simulate from the ‘transition’ density $f(\alpha_{t+1}|\alpha_t)$. Sometimes we will also assume that we can evaluate $f(\alpha_{t+1}|\alpha_t)$.

Filtering can be thought of as the repeated application of the iteration

$$f(\alpha_{t+1}|\mathcal{F}_{t+1}) \propto f(y_{t+1}|\alpha_{t+1}) \int f(\alpha_{t+1}|\alpha_t) dF(\alpha_t|\mathcal{F}_t). \quad (1)$$

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

There have been numerous attempts to provide algorithms which approximate the filtering densities. Important recent work includes Kitagawa (1987), West (1992), Gerlach, Carter, and Kohn (1999) and those papers reviewed in West and Harrison (1997, Ch. 13 and 15). Here we use simulation to perform filtering following an extensive recent literature. Our approach is to extend the particle filter using an auxiliary variable, an idea which first appeared in Pitt and Shephard (1999a). The literature on particle filtering is reviewed extensively in previous Chapters of this book and will not be repeated here.

The outline of the paper is as follows. In Section 2 we analyse the statistical basis of particle filters and focus on its weaknesses. In Section 3 we review the main focus of the chapter, which is an auxiliary particle filter method. Section 4 discusses fixed lag filtering, while Section 5 uses stratified sampling to improve the performance of the algorithm.

2 Particle filters

2.1 The definition of particle filters

Particle filters are the class of simulation filters which recursively approximate the filtering random variable $\alpha_t | \mathcal{F}_t = (y_1, \dots, y_t)'$ by ‘particles’ $\alpha_t^1, \dots, \alpha_t^M$, with associated discrete probability masses π_t^1, \dots, π_t^M . Hence a continuous variable is approximated by a discrete one with random support. These discrete points are thought of as samples from $f(\alpha_t | \mathcal{F}_t)$. In our work we will always set the $\{\pi_t^M\}$ to be equal to $\frac{1}{M}$ where M is taken to be very large. Then we require that as $M \rightarrow \infty$, the particles can be used to increasingly well approximate the density of $\alpha_t | \mathcal{F}_t$.

Particle filters treat the discrete support generated by the particles as the true filtering density, then this is chained to produce a new density

$$\hat{f}(\alpha_{t+1} | \mathcal{F}_{t+1}) \propto f(y_{t+1} | \alpha_{t+1}) \sum_{j=1}^M f(\alpha_{t+1} | \alpha_t^j), \quad (2)$$

the ‘empirical filtering density’ as an approximation to the true filtering density (1). Generically particle filters then sample from this density to produce new particles $\alpha_{t+1}^1, \dots, \alpha_{t+1}^M$. This procedure can then be iterated through the data. We will call a particle filter ‘fully adapted’ if it produces independent and identically distributed samples from (2). There may be advantages in deliberately inducing (negative) correlations amongst the particles. This was first explicitly pointed out by Carpenter, Clifford, and Fearnhead (1998).

2.2 Sampling the empirical prediction density

One way of sampling from the empirical prediction density is to think of

$$\frac{1}{M} \sum_{j=1}^M f(\alpha_{t+1} | \alpha_t^j)$$

as a ‘prior’ density $\hat{f}(\alpha_{t+1} | \mathcal{F}_t)$ which 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 $\frac{1}{M}$ 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 this leaves us with three sampling methods to draw from $f(\alpha_{t+1} | \mathcal{F}_{t+1})$.

1. Sampling/importance resampling.
2. Acceptance sampling.
3. 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, in order to briefly describe these methods in this context.

2.2.1 Sampling/importance resampling (SIR)

This method (Rubin (1987) and Smith and Gelfand (1992)) draws $\alpha^1, \dots, \alpha^R$ from $f(\alpha)$ and then associates with each of these draws the weights π_j where

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

The weighted sample will converge, as $R \rightarrow \infty$, to a non-random sample from the desired posterior $f(\alpha|y)$ as $R^{-1} \sum_{i=1}^R w_i \xrightarrow{p} f(y)$. The non-random sample can be converted into a random sample of size M by resampling the $\alpha^1, \dots, \alpha^R$ using weights π_1, \dots, π_R . This requires $R \rightarrow \infty$ and $R \gg M$. The use of this method has been suggested in the particle filter framework by Gordon, Salmond, and Smith (1993), Kitagawa (1996), Berzuini, Best, Gilks, and Larizza (1997) and Isard and Blake (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 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 Adaption

The above SIR algorithm samples from $f(\alpha|y)$ by making *blind* proposals $\alpha^1, \dots, \alpha^R$ from the prior, ignoring the fact that we know the value of y . This is the main feature of the initial particle filter proposed by Gordon, Salmond, and Smith (1993). We say that a particle filter is adapted if we make proposals which take into account the value of y .

An adapted SIR based particle filter has the following general structure

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

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

2.2.3 Rejection and MCMC 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) and used on a univariate log-normal stochastic volatility model by Kim, Shephard, and Chib (1998). 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

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

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

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

The problem with this, as for the previous adapted SIR, is that evaluating $f(\alpha)$ is very expensive.

2.3 Particle filter's weaknesses

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 . Instead, the difficulty is, given that degree of accuracy, how to efficiently sample from (2)? We will show how to do this in the next section.

The second weakness holds in general for particle filters. As $R \rightarrow \infty$, so the weighted samples can be used to arbitrarily well approximate (2). However, the tails of $\frac{1}{M} \sum_{j=1}^M f(\alpha_{t+1}|\alpha_t^j)$ usually only poorly approximate the true tails of $\alpha_{t+1}|\mathcal{F}_t$ due to the use of the mixture approximation. As a result (2) can only ever poorly approximate the true $f(\alpha_{t+1}|\mathcal{F}_{t+1})$ when there is an outlier. Hence the second question is how do we improve the empirical prediction density's behaviour in the tails? This will be discussed in Section 4.

3 Auxiliary variable

3.1 The basics

A fundamental problem with conventional particle filters is that their mixture structure means that it is difficult to adapt the SIR, rejection or MCMC sampling methods

without massively slowing the running of the filter. Pitt and Shephard (1999a) have argued that many of these problems are reduced when we perform particle filtering in a higher dimension. In this section we review this argument.

Our task will be to sample from the joint density $f(\alpha_{t+1}, k | \mathcal{F}_{t+1})$, where k is an index on the mixture in (2). Let us define

$$f(\alpha_{t+1}, k | \mathcal{F}_{t+1}) \propto f(y_{t+1} | \alpha_{t+1}) f(\alpha_{t+1} | \alpha_t^k), \quad k = 1, \dots, M. \quad (3)$$

If we draw from this joint density and then discard the index we produce a sample from the empirical filtering density (2) as required. We call k an auxiliary variable as it is present simply to aid the task of the simulation. Generic particle filters of this type will be labelled auxiliary particle filters.

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

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

We have complete control over the design of $g(\cdot)$, which can depend on y_{t+1} and α_t^k , in order to make the weights even. Thus this method is adaptable and extremely flexible. In the next subsection we will 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 $\alpha_{t+1}^{(i+1)}, k^{(i+1)} \sim g(\alpha_{t+1}, k | \mathcal{F}_{t+1})$, where $g(\alpha_{t+1}, k | \mathcal{F}_{t+1})$ is some arbitrary density, then these moves are accepted with probability

$$\begin{aligned} & \min \left\{ 1, \frac{f(y_{t+1} | \alpha_{t+1}^{(i+1)}) f(\alpha_{t+1}^{(i+1)} | \alpha_t^{k^{(i+1)}})}{f(y_{t+1} | \alpha_{t+1}^{(i)}) f(\alpha_{t+1}^{(i)} | \alpha_t^{k^{(i)}})} \frac{g(\alpha_{t+1}^{(i)}, k^{(i)} | \mathcal{F}_{t+1})}{g(\alpha_{t+1}^{(i+1)}, k^{(i+1)} | \mathcal{F}_{t+1})} \right\} \\ &= \min \left(1, \frac{w_{i+1}}{w_i} \right). \end{aligned}$$

A special case of this argument has appeared in Berzuini, Best, Gilks, and Larizza (1997) who put $g(\alpha_{t+1}, k | \mathcal{F}_{t+1}) \propto f(\alpha_{t+1} | \alpha_t^k)$ which means their method is again blind.

3.2 A generic SIR based auxiliary proposal

3.2.1 The method

Here we will give a generic $g(\cdot)$ which can be broadly applied. We will base our discussion on the SIR algorithm, although we could equally have used an MCMC method. We approximate (3) by

$$g(\alpha_{t+1}, k | \mathcal{F}_{t+1}) \propto f(y_{t+1} | \mu_{t+1}^k) f(\alpha_{t+1} | \alpha_t^k), \quad k = 1, \dots, M,$$

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

$$g(k | \mathcal{F}_{t+1}) \propto \int f(y_{t+1} | \mu_{t+1}^k) dF(\alpha_{t+1} | \alpha_t^k) = f(y_{t+1} | \mu_{t+1}^k).$$

Thus we can sample from $g(\alpha_{t+1}, k | \mathcal{F}_{t+1})$ by simulating the index with probability $\lambda_k \propto g(k | \mathcal{F}_{t+1})$, 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 will simulate from particles which are associated with large predictive likelihoods. Having sampled the joint density of $g(\alpha_{t+1}, k | \mathcal{F}_{t+1})$ 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

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

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 which have high conditional likelihoods we reduce the costs of sampling many times from particles which 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. Then for a standard SIR based particle filter, for large R ,

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

where

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

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

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

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 $\sum_{k=1}^M \lambda_k \frac{1}{M} = \frac{1}{M} \leq \sum_{k=1}^M \lambda_k^2$. 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.

3.2.2 Example: a time series of angles

The model In this section we will 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, Salmond, and Smith (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 $\alpha_t = (x_t, vx_t, z_t, vz_t)'$,

$$\alpha_{t+1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \alpha_t + \sigma_\eta \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \\ 0 & 1 \end{pmatrix} u_t, \quad u_t \sim \text{NID}(0, 1). \quad (4)$$

In obvious notation x_t, z_t represent the ship's horizontal and vertical position at time t and vx_t, 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 source of state evolution error is due to the accelerations being 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 (4) describes the overall prior for the states.

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

$$f(y_t|\mu_t) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(y_t - \mu_t)}, \quad 0 \leq y_t < 2\pi, \quad 0 \leq \rho \leq 1. \quad (5)$$

ρ is termed the mean resultant length.

The simulated scenario In order 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, Salmond, and Smith (1993). They consider $\sigma_\eta = 0.001$ and $\sigma_\varepsilon = 0.005$, 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 $\alpha_1 = (-0.05, 0.001, 0.2, -0.055)'$. By contrast with Gordon, Salmond, and Smith (1993), 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 posterior density to be small enabling reasonably conclusive evidence to be formulated about the relative efficiency of the auxiliary method to the standard method. We therefore take $a_1 = \alpha_1$ and have a diagonal initial variance P_1 with the elements $0.01 \times (0.5^2, 0.005^2, 0.3^2, 0.01^2)$ on the diagonal.

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

Monte Carlo comparison The two methods are now compared using a Monte Carlo study of the above scenario with $T = 10$. The "true" filtered mean is calculated for each replication by using the auxiliary method with $M = 100,000$ and

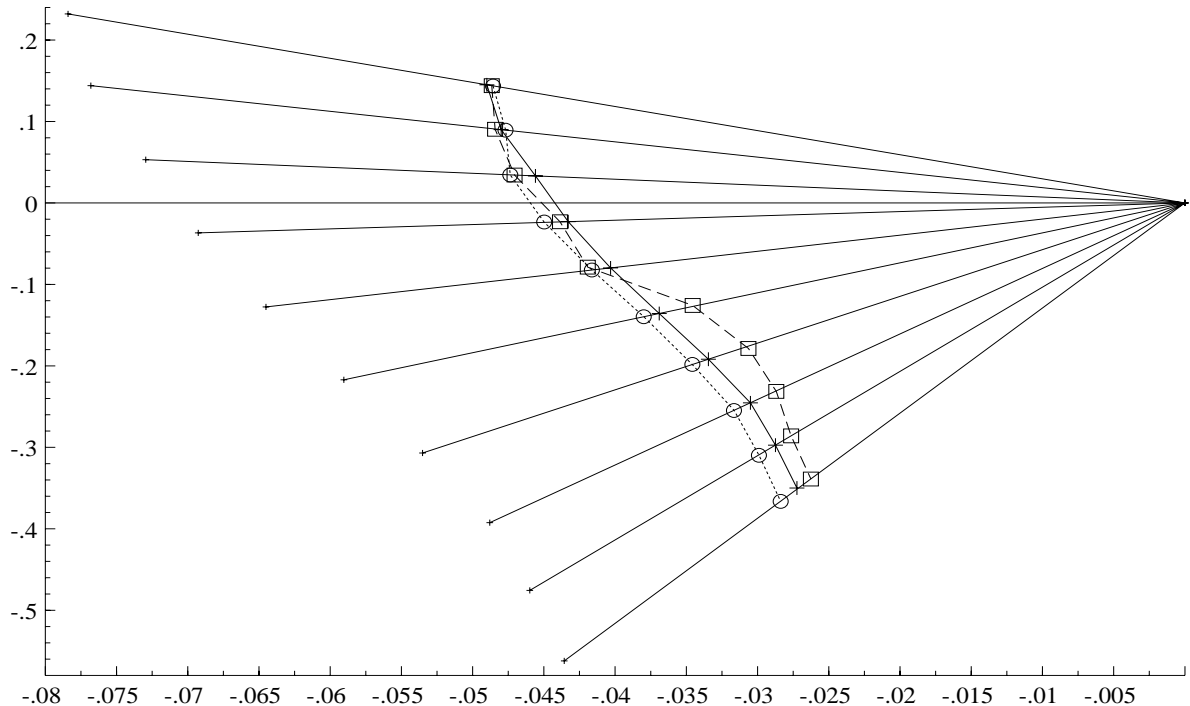


Figure 1: *Plot of the angular measurements from origin, the true trajectory (solid line, crosses), the particle filtered mean trajectory (dashed line, boxes) and the auxiliary particle mean trajectory (dotted line, circles). Ship moving South-East. $T = 10$, $M = 300$, $R = 500$.*

$R = 120,000$. Within each replication the mean squared error 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

$$MSE_{i,j,t}^P = \frac{1}{S} \sum_{s=1}^S (\bar{\alpha}_{t,j,s}^i - \tilde{\alpha}_{t,j}^i)^2,$$

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

$$LMSE_{j,t}^P = \log \frac{1}{REP} \sum_{i=1}^{REP} MSE_{i,j,t}^P. \quad (6)$$

The same operation is performed for the auxiliary method to deliver the corresponding quantity $LMSE_{j,t}^{AM}$. For this study we set use $REP = 40$ 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 $LMSE_{j,t}^{AM} - LMSE_{j,t}^P$ is plotted against time. Values close to 0 indicate that the two methods are broadly equivalent

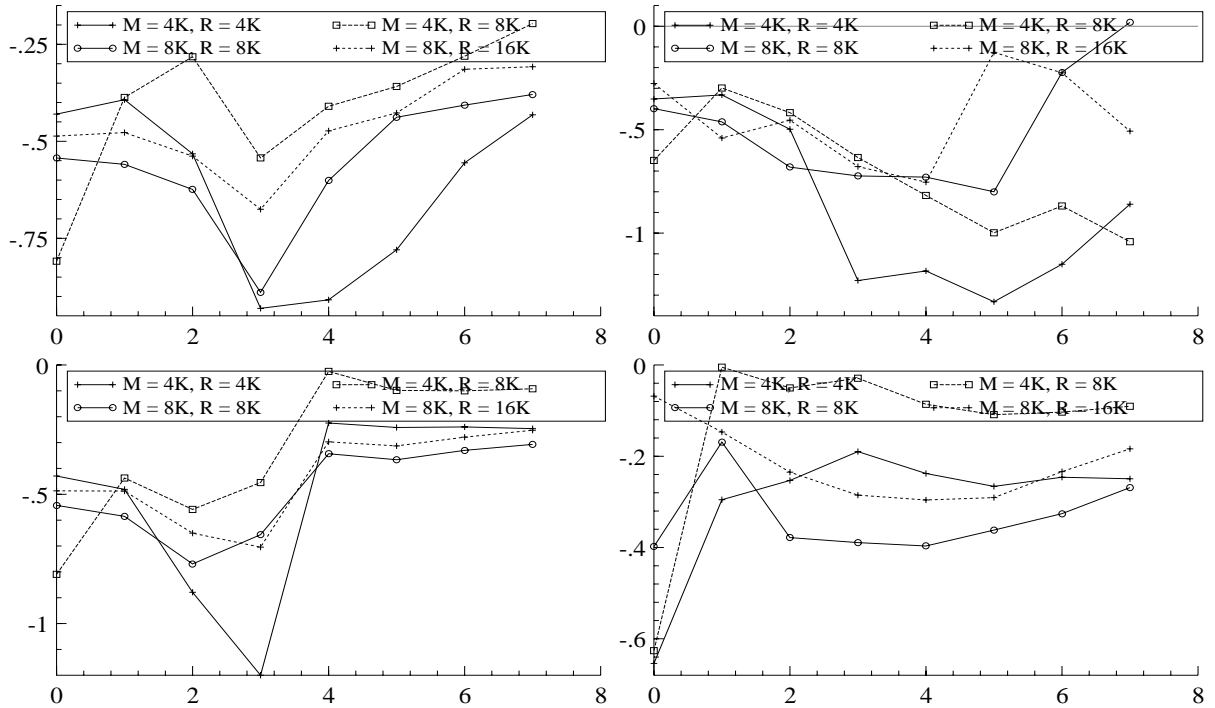


Figure 2: Plot of the relative mean square error performance (on the log-scale) of the particle filter and the auxiliary based particle filter for the bearings only tracking problem. Numbers below zero indicate a superior performance by the auxiliary particle filter. In these graphs $M = 4,000$ or $8,000$ while $R = M$ or $R = 2M$. Throughout SIR is used as the sampling mechanism. Top left: $\alpha_{t1} = x_t$, Bottom left: $\alpha_{t3} = z_t$, while Top right: $\alpha_{t2} = vx_t$ and Bottom right: $\alpha_{t4} = vz_t$.

in performance whilst 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.

3.3 Examples of adaption

3.3.1 Basics

Although the above generic scheme can usually reduce the variability of the second stage weights, there are sometimes other adaption schemes which use the specific structure of the time series model which allow us to achieve yet more even 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 i.i.d. samples from (2). This situation is particularly interesting as we are then close to the assumptions made by Kong, Liu, and Wong (1994) for their sequential importance sampler.

3.3.2 Non-linear Gaussian measurement model

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

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

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

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

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 the second stage weights are all equal.

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

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

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 Wadhvani (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 above 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

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

then

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

Rearranging, we can express this as

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

which means 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 for $j = 1, \dots, R$ with

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

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 $f(\alpha_{t+1}, k|\mathcal{F}_{t+1})$ 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

$$y_t = \epsilon_t \beta \exp(\alpha_t/2), \quad \alpha_{t+1} = \phi \alpha_t + \eta_t, \quad (7)$$

where ϵ_t and η_t are independent Gaussian processes with variances of 1 and σ^2 respectively. Here β has the interpretation as the modal volatility, ϕ 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, for instance, Hull and White (1987). MCMC methods have been used on this model by, for instance, Jacquier, Polson, and Rossi (1994), Shephard and Pitt (1997) and Kim, Shephard, and Chib (1998).

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

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

The implication is that

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

Likewise

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

Finally the log-probability of acceptance is

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

Notice that as σ^2 falls to zero so the acceptance probability goes to one.

Finally, the same argument goes through 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.

Simulation experiment The basic SV model was defined in Section 3.3.4. We construct 100 times the compound daily returns on the US Dollar against the UK Sterling from the first day of trading in 1997 and for the next 200 days of active trading. This data is discussed in more detail in Pitt and Shephard (1999b), where the parameters of the model were estimated using Bayesian methods. Throughout we take $\phi = 0.9702$, $\sigma_\eta = 0.178$ and $\beta = 0.5992$, 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, $f\{\beta \exp(\alpha_t/2)|\mathcal{F}_t\}$ computed using an auxiliary particle filter. Throughout the series we set $M = 5,000$, $R = 6,000$. 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 picture shows that the filtered volatility jumps up more quickly than it tends to go down. This reflects the fact that the volatility is modelled on the log scale.

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 mean square error for each value of t using the above model and again having $n = 50$. The data was simulated using the model parameters discussed above. The results are reported (using a log scale) in Figure 4. 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 use $REP = 40$ 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 around twice the time of the SIR based particle filter when $M = R$.

Figure 4 shows that the fully adapted particle filter is considerably more accurate than the other particle filters. It also has the advantage that it does not depend 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.

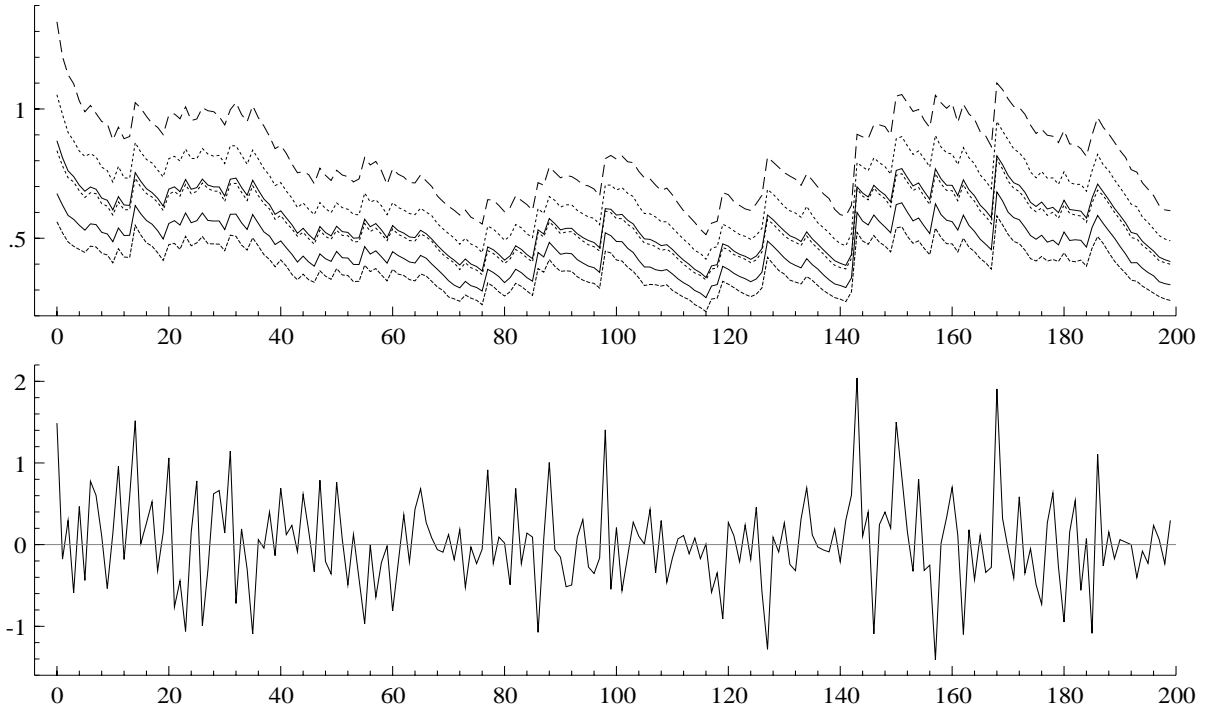


Figure 3: Bottom graph shows the daily returns on the Dollar against UK Sterling from the first day of trading in 1997 for 200 trading days. We display in the top graph the posterior filtered mean (heavy line) of $\beta \exp(\alpha_t/2) | Y_t$, together with the 5, 20, 50, 80, 95 percentage points of the distribution. Notice the median is always below the mean. $M = 5,000$, $R = 6,000$.

3.3.5 Mixtures of normals

Suppose $f(\alpha_{t+1} | \alpha_t)$ 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 $f(\alpha_{t+1}, k | \mathcal{F}_{t+1})$ by working with

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

Then we sample from $f(\alpha_{t+1}, k, j | \mathcal{F}_{t+1})$ 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 PM 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 for state space models with mixtures have been studied by, for example, Carter and Kohn (1994) and Shephard (1994).

4 Fixed lag filtering

The auxiliary particle filter method can also be used when we update the estimates of the states not by a single observation but by a block of observations. This idea

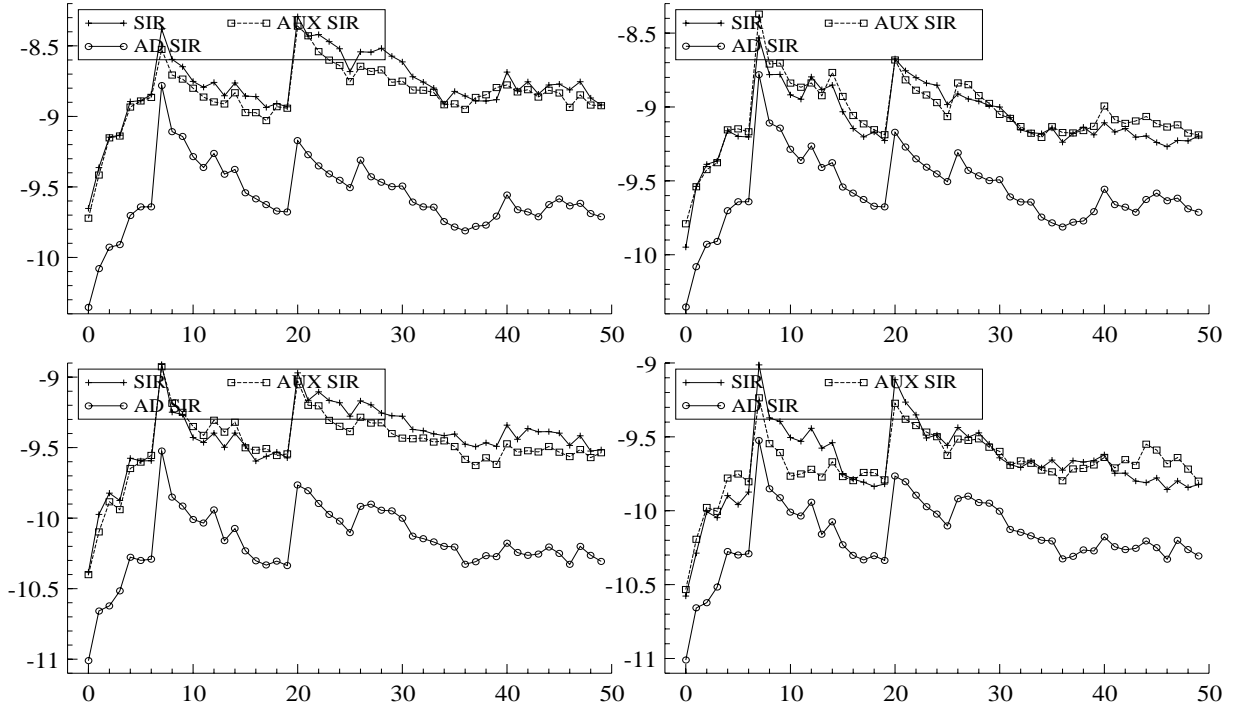


Figure 4: *Plot of the mean square error performance (on the log-scale) of the particle filter to the auxiliary based particle filter and an adapted particle filter. The lower the number the more efficient the method. Top graphs have $M = 2,000$, the bottom have $M = 4,000$. The Left graphs have $R = M$, while the right ones have $R = 2M$*

appeared in the initial working paper version of Pitt and Shephard (1999a) which was circulated in 1997, but was taken out of the paper which appeared in the journal.

Again suppose that we approximate the density of $\alpha_t | \mathcal{F}_t = (y_1, \dots, y_t)'$ by a distribution with discrete support at the points $\alpha_t^1, \dots, \alpha_t^M$, with mass $\{\frac{1}{M}\}$. Then the task will be to update this distribution to provide a sample from $\alpha_{t+1}, \dots, \alpha_{t+p} | \mathcal{F}_{t+p}$. At first sight this result seems specialized as it is not often that we have to update after the arrival of a block of observations. However, as well as solving this problem it also suggests a way of reducing the bias caused by using the empirical prediction density as an approximation to $f(\alpha_{t+1} | \mathcal{F}_t)$. Suppose that instead of updating p future observations simultaneously, we store $p - 1$ observations and update those observations together with an empirical prediction density for $f(\alpha_{t-p+2} | \mathcal{F}_{t-p+1})$. This would provide us with draws from $f(\alpha_{t+1} | \mathcal{F}_{t+1})$ as required. We call this fixed lag filtering. The hope is that the influence of the empirical prediction density will be reduced as it will have been propagated p times through the transition density. This may reduce the influence of outliers on the auxiliary method.

This can be carried out by using a straightforward particle filter using SIR, rejection sampling or MCMC, or by building in an auxiliary variable so that we sample from $\alpha_{t+1}, \dots, \alpha_{t+p}, k | \mathcal{F}_{t+p}$. Typically the gains from using the auxiliary approach is greater here for as p grows so naive implementations of the particle filter will become less and less efficient due to not being able to adapt the sampler to the measurement density.

To illustrate this general setup consider the use of an auxiliary particle filter where we take

$$\begin{aligned} g(k|\mathcal{F}_{t+p}) &\propto \int f(y_{t+p}|\mu_{t+p}^k)\dots f(y_{t+1}|\mu_{t+1}^k)dF(\alpha_{t+p}|\alpha_{t+p-1})\dots dF(\alpha_{t+1}|\alpha_t^k) \\ &= f(y_{t+p}|\mu_{t+p}^k)\dots f(y_{t+1}|\mu_{t+1}^k), \end{aligned}$$

and then sampling the index k with weights proportional to $g(k|\mathcal{F}_{t+p})$. Having selected the index k^j we then propagate the transition equation p steps to produce a draw $\alpha_{t+1}^j, \dots, \alpha_{t+p}^j$, $j = 1, \dots, R$. These are then reweighted according to the ratio

$$\frac{f(y_{t+p}|\alpha_{t+p}^j)\dots f(y_{t+1}|\alpha_{t+1}^j)}{f(y_{t+p}|\mu_{t+p}^{k^j})\dots f(y_{t+1}|\mu_{t+1}^{k^j})}.$$

This approach has three main difficulties. First it requires us to store p sets of observations and $p \times M$ mixture components. This is more expensive than the previous method as well as being slightly harder to implement. Second, each auxiliary variable draw now involves $3p$ density evaluations and the generation of p simulated propagation steps. Third, the auxiliary variable method is based on approximating the true density of $f(k, \alpha_{t-p+1}, \dots, \alpha_t|Y_t)$, and this approximation is likely to deteriorate as p increases. This suggests that the more sophisticated adaptive sampling schemes, discussed above, may be particularly useful at this point. Again however, this complicates the implementation of the algorithm.

We will illustrate the use of this sampler at the end of the next section on an outlier problem.

5 Reduced random sampling

5.1 Basic ideas

The generic auxiliary particle filter given in Section 3.2 has two sets of weighted bootstraps and that this introduces a large degree of randomness to the procedure. This is most stark in the case where $f(y_{t+1}|\alpha_{t+1})$ does not depend on α_{t+1} and so y_{t+1} is uninformative about α_{t+1} . For such a problem the first stage weights are $\lambda_k \propto g(k|\mathcal{F}_{t+1}) \propto 1$ while the second stage weights $w_j \propto 1$. The implication is that it would have been better simply to propagate every α_t^k through $f(\alpha_{t+1}|\alpha_t^k)$ once to produce a new α_{t+1}^k . This would produce a more efficient sample than our method which samples with replacement from these populations, twice killing interesting particles for no good reason.

This observation has appeared in the particle filtering literature on a number of occasions. Liu and Chen (1995) discuss carrying weights forward instead of resampling in order to keep alive particles which would otherwise die. Carpenter, Clifford, and Fearnhead (1998) think about the same issue using stratification ideas taken from sampling theory.

Here we use a method which is similar to an idea discussed by Liu and Chen (1998) in this context. We resample from a population $\alpha_t^1, \dots, \alpha_t^M$ with weights π_t^1, \dots, π_t^M to produce a sample of size R in the following way. We produce stratified uniforms

$\tilde{u}_t^1, \dots, \tilde{u}_t^R$ by writing

$$\tilde{u}_t^k = \frac{(k-1) + u_t^k}{R}, \quad \text{where } u_t^k \stackrel{iid}{\sim} UID(0, 1).$$

This is the scheme suggested by Carpenter, Clifford, and Fearnhead (1998). Then we compute the cumulative probabilities

$$\tilde{\pi}_t^r = \sum_{s=1}^r \pi_t^s, \quad r = 1, \dots, M.$$

We allocate n^k copies of the particle α_t^k to the new population, where n^k is the number of $\tilde{u}_t^1, \dots, \tilde{u}_t^R$ in the interval

$$\left(\sum_{s=1}^{k-1} \pi_t^s, \sum_{s=1}^k \pi_t^s \right].$$

The computation of each of $\{\tilde{u}_t^k\}$, $\{\tilde{\pi}_t^r\}$ and $\{n^k\}$ is straightforward and so this type of stratified sampling is very fast. In fact it is much faster than simple random sampling.

Although, as we noted above, this idea is not new it is particularly useful in the context of our generic auxiliary particle filter suggestion which has two weighted bootstraps — while a typical SIR based particle filter only has one. Hence we might expect the gains to be made over the original suggestion in Pitt and Shephard (1999a) would be particularly large.

5.2 Simple outlier example

We tried random and stratified sampling using fixed lag versions of SIR based particle and auxiliary particle filters on a difficult outlier problem where the analytic solution is available via the Kalman filter. We assume the observations arise from an autoregression observed with noise

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, 0.707^2) \\ \alpha_{t+1} &= 0.9702\alpha_t + \eta_t, & \eta_t &\sim NID(0, 0.178^2), \end{aligned} \tag{8}$$

where ε_t and η_t are independent processes. The model is initialised by α_t 's stationary prior. We added to the simulated $y_{n/2}$ a shock 6.5×0.707 , which represents a very significant outlier. Throughout we set $M = R = 500$ and measure the precision of the filter by the log mean square error criteria (6), taking $REP = 30$ and $S = 20$. As the problem is Gaussian the Kalman filter's MSE divided by M provides a lower bound on the mean square error criteria.

Figure 5 shows the results from the experiment, recording the mean square errors and bias of the various particle filters. It is important to note that the mean square errors are drawn on the log10 scale. The main features of the graphs are that: (i) when there is no outlier all the particle filters are basically unbiased with stratification being important. The use of the auxiliary variable does not have much impact in this situation (although it is still better). (ii) during an outlier, the ASIR methods dominate both in terms of bias and MSE. Stratification makes very little difference in this situation. (iii) after the outlier, stratified ASIR continues to work well while ASIR returns to being less effective than stratified SIR. (iv) The introduction of fixed

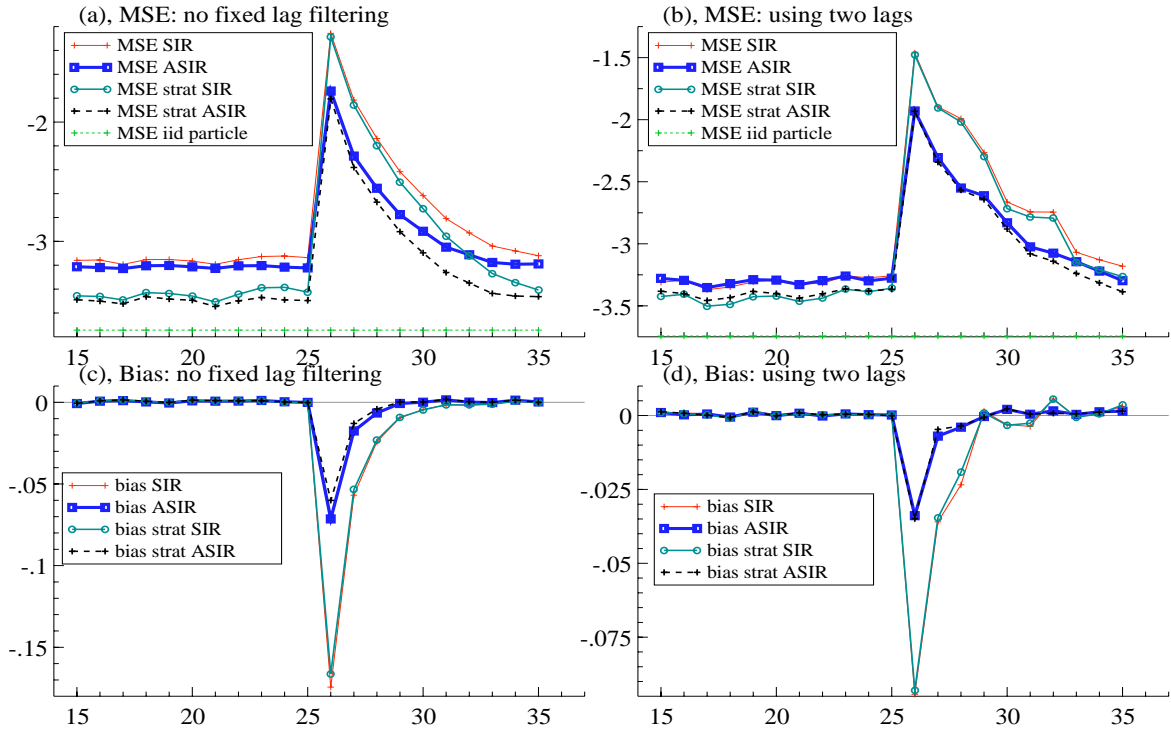


Figure 5: *The mean square error (MSE), using a log10 scale, and bias of four different particle filters using no and two filtered lag filters. The x-axis is always time, but we only graph results for $t = T/4, T/4 + 1, \dots, 3T/4$ in order to focus on the crucial aspects. The four particle filters are: SIR, ASIR, stratified SIR and stratified ASIR. The results are grouped according to the degree of fixed lag filtering. In particular: (a) shows the MSE when $p = 0$, (b) shows the MSE when $p = 2$. (c) shows the bias when $p = 0$, while (d) indicates the bias with $p = 2$. Throughout we have taken $M = R = 500$.*

lag filtering reduces the bias of all methods by an order of magnitude while the MSE reduces quite considerably.

In order to benchmark these results, we have repeated the same experiment but now with $M = R = 2500$. The results are given in Figure 6. This picture is remarkably similar to Figure 5 but with smaller bias and MSE. An important feature of this experiment is that the reduction in bias and MSE of a five fold increase in M and R produces around the same impact as the introduction of fixed lag filtering.

6 Conclusion

This Chapter has studied the weaknesses of the very attractive particle filtering method originally proposed by Gordon, Salmond, and Smith (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 introduce an auxiliary variable into the particle filter to overcome the first of these problems, providing a powerful framework which is as simple as SIR, but more flexible and reliable. We show that when it is possible to adapt the

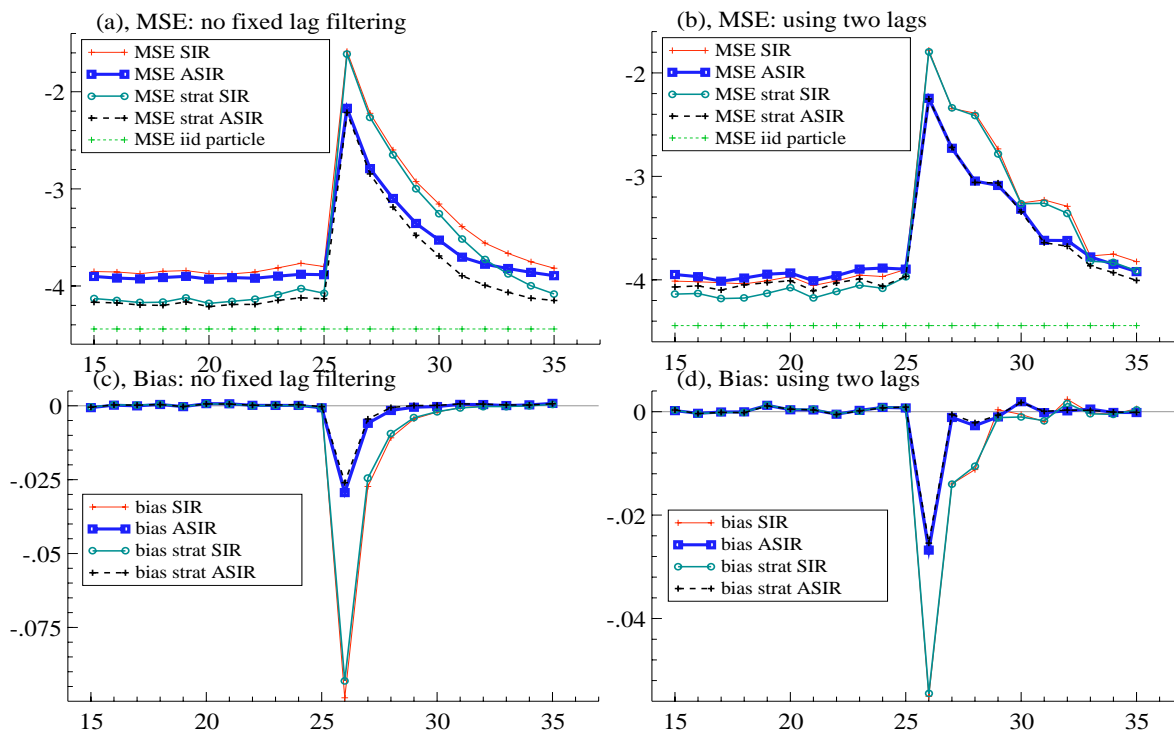


Figure 6: Repeat of Figure but with $M = R = 2500$.

particle filter then this can bring about large efficiency gains. The fixed-lag filter partially tackles the second problem, which suggests a possible real improvement in the reliability of these methods.

7 Acknowledgements

Michael K Pitt thanks the ESRC for their financial support through the grant “Modelling and analysis of econometric and financial time series.”

References

- Berzuini, C., N. G. Best, W. R. Gilks, and C. Larizza (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods. *J. American Statistical Association* 92, 1403-12.
- Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2959–3038. North-Holland.
- Carpenter, J. R., P. Clifford, and P. Fearnhead (1998). An improved particle filter for non-linear problems. Working paper, January 1998, Dept. of Statistics, University of Oxford, OX1 3TG, UK.
- Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541–53.

- Diebold, F. X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *J. Applied Econometrics* 4, 1–21.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.
- Gerlach, R., C. Carter, and R. Kohn (1999). Diagnostics for time series analysis. *J. Time Series Analysis* 21. Forthcoming.
- Gilks, W. K., S. Richardson, and D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.
- Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F* 140, 107–13.
- Harvey, A. C., E. Ruiz, and E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *J. Econometrics* 52, 129–158.
- Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* 42, 281–300.
- Hürzeler, M. and H. R. Künsch (1995). Monte Carlo approximations for general state space models. Unpublished paper: Seminar für Statistik, ETH.
- Isard, M. and A. Blake (1996). Contour tracking by stochastic propagation of conditional density. *Proceedings of the European Conference on Computer Vision, Cambridge* 1, 343–356.
- Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Business and Economic Statist.* 12, 371–417.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Economic Studies* 65, 361–93.
- King, M., E. Sentana, and S. Wadhvani (1994). Volatility and links between national stock markets. *Econometrica* 62, 901–933.
- Kitagawa, G. (1987). Non-Gaussian state space modelling of non-stationary time series. *J. American Statistical Association* 82, 503–514.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Computational and Graphical Statistics* 5, 1–25.
- Kong, A., J. S. Liu, and W. H. Wong (1994). Sequential imputations and Bayesian missing data problems. *J. American Statistical Association* 89, 278–88.
- Liu, J. (1996). Metropolized independent sampling with comparison to rejection sampling and importance sampling. *Statistics and Computing* 6, 113–119.
- Liu, J. and R. Chen (1995). Blind deconvolution via sequential imputation. *J. American Statistical Association* 90, 567–76.
- Liu, J. and R. Chen (1998). Sequential Monte Carlo methods for dynamic systems. *J. American Statistical Association* 93, 1032–1044.
- Pitt, M. K. and N. Shephard (1999a). Filtering via simulation based on auxiliary particle filters. *J. American Statistical Association* 94. Forthcoming.

- Pitt, M. K. and N. Shephard (1999b). Time varying covariances: a factor stochastic volatility approach (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6*. Oxford: Oxford University Press. Forthcoming.
- Rubin, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm. Discussion of Tanner and Wong (1987). *J. American Statistical Association* 82, 543–546.
- Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika* 81, 115–31.
- Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–67.
- Smith, A. F. M. and A. E. Gelfand (1992). Bayesian statistics without tears: a sampling-resampling perspective. *American Statistician* 46, 84–88.
- Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. American Statistical Association* 82, 528–50.
- West, M. (1992). Mixture models, Monte Carlo, Bayesian updating and dynamic models. *Computer Science and Statistics* 24, 325–333.
- West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (2 ed.). New York: Springer-Verlag.