The auxiliary particle filter

Michael Pitt

Department of Economics, University of Warwick, Coventry CV4 7AL

M.Pitt@warwick.ac.uk

September 20, 2001

Models:

General Form:

$$\begin{aligned} \text{meas eqn: } y_t &\sim f(y_t|\alpha_t) \\ \text{trans eqn: } \alpha_{t+1} &\sim f(\alpha_{t+1}|\alpha_t), \;\; t = 1, ..., n \end{aligned}$$

Filtering:

Require posterior $f(\alpha_t|Y_t)$, $t = 1, ..., n$. Require $O(n)$ methods.

Analytically – two stages; the prediction stage and the measurement update stage.

1. *Prediction stage:*

$$f(\alpha_{t+1}|Y_t) = \int f(\alpha_{t+1}|\alpha_t) f(\alpha_t|Y_t) d\alpha_t$$

2. *Measurement update stage:*

$$f(\alpha_{t+1}|Y_{t+1}) \propto f(y_{t+1}|\alpha_{t+1}) f(\alpha_{t+1}|Y_t)$$

In GSSF the above equations lead straightforwardly to the Kalman filter relations. For general densities above relations not analytically tractable.

References may be found in WestHarrison(1997, Ch 13); Harvey(1989, Ch 3).

Particle Filters:

Approximate $f(\alpha_t|Y_t)$ by cloud of points ("particles")

$$\begin{array}{lll} \text{points: } y_t & \alpha_t^1, ..., \alpha_t^M \\ \text{mass: } \alpha_{t+1} & \pi_t^1, ..., \pi_t^M \end{array}$$

$M \longrightarrow \infty$ arbitrarily close to truth. Usually $\pi_t^i = 1/M$.

Filtering Motivation:

1. May be interested in current state | current information for an *online* estimation problem, e.g. Bearings-only problem.

2. Give one-step ahead prediction density allowing:

   $\widehat{\Pr}(z < y_{t+1}|Y_t) \sim \mathsf{Uni}[0,1]$ if model and parameters true.

3. Gives likelihood for any set of parameters θ. Since

$$\widehat{f}(y_{t+1}|Y_t) = \sum_{i=1}^{M} f(y_{t+1}|\alpha_{t+1}^i)$$

   $\alpha_{t+1}^i \sim f(\alpha_{t+1}|Y_t)$ so we can get

$$f(y|\theta) = \prod_{t=1}^{n} f(y_t|Y_{t-1};\theta).$$

Simulation:

Main tool: Sampling-Importance Resampling; Rubin(1988); SmithGelfand(1992).

Spse want to simulate from $f(\alpha|y) \propto f(y|\alpha) \times f(\alpha)$.

1. Draw R samples $\alpha^1, ..., \alpha^R \sim f(\alpha)$.

2. Construct R weights

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

3. Convert to sample size M by sampling from above discrete distribution.

Sample $\sim f(\alpha|y)$ as $R \longrightarrow \infty$.

Properties:

- Efficient when weights as similar as possible –> flat likelihood.
- Requires can simulate from prior / evaluate lik.
- Similar objectives as A/R ; independence Metrop –>make proposal close to target.
- Get repeated samples.

SIR particle filter:

of GordonSalmondSmith(1993, IEE Transactions), GSS.

At time t: We have M samples $\alpha_t^j \sim f(\alpha_t|Y_t)$.

1. Prior at t+1:

   emprical prediction density:

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

2. Posterior t+1:

   emprical filtering density:

$$\widehat{f}(\alpha_{t+1}|Y_{t+1}) \propto f(y_{t+1}|\alpha_{t+1}) \times \sum_{j=1}^{M} f(\alpha_{t+1}|\alpha_t = \alpha_t^j) \qquad (0{\cdot}1)$$

   So GSS do:

1. Draw R samples $\alpha_{t+1}^1, ..., \alpha_{t+1}^R \sim \widehat{f}(\alpha_{t+1}|Y_t)$.

2. Get R weights $\pi_i \propto f(y_{t+1}|\alpha_{t+1}^i) \;\; i = 1, ..., R$

3. Sample M times from above discrete density giving approx $\alpha_{t+1}^{(k)} \sim \widehat{f}(\alpha_{t+1}|Y_{t+1})$

Auxiliary SIR:

PittShephard(1999, JASA).

–Require equally general $O(M+R)$ procedure: sim from state equation / evaluate meas dens.

emprical filtering density

$$f(\alpha_{t+1}|Y_{t+1}) \propto f(y_{t+1}|\alpha_{t+1}) \times \sum_{k=1}^{M} f(\alpha_{t+1}|\alpha_t^k)$$

Consider joint density of α_{t+1}, k :

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

First approximation:

$$\begin{aligned} &\simeq f(y_{t+1}|\mu_{t+1}^k) \times f(\alpha_{t+1}|\alpha_t^k) \\ &\propto g(\alpha_{t+1}, k), \text{ say} \end{aligned}$$

Sampling from $f(\alpha_{t+1}, k)$:

Easy by construction since

$$g(k) \propto f(y_{t+1}|\mu_{t+1}^k) \quad \text{and} \quad g(\alpha_{t+1}|k) = f(\alpha_{t+1}|\alpha_t^k).$$

Gives us R sets of samples $(k^j, \alpha_{t+1}^j) \quad j = 1, ..., R.$

IR weights now

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

- Weights now more even than for standard *SIR filter* $\Rightarrow$ can take R much smaller.

- Improvement over standard *SIR filter* becomes better as mixtures become tight = small error in trans density.

- Above prescription as general as *SIR filter*.

- Can adapt to make even closer to *target density* at price of generality.

Extensions

1. **Fixed lag filtering**

$\alpha_t|Y_t = (y_1, ..., y_t)'$ represented by a distribution with discrete support at the points $\alpha_t^1, ..., \alpha_t^M$, with probability mass of $1/M$ at each.

Then update this distribution to provide a sample from $\alpha_{t+1}, ..., \alpha_{t+p}|Y_{t+p}$. The ASIR method extends by computing the weights

$$\begin{aligned} g(k|Y_{t+p}) &\propto \int f(y_{t+p}|\mu_{t+p}^k)...f(y_{t+1}|\mu_{t+1}^k)f(\alpha_{t+p}|\alpha_{t+p-1}) \\ &\quad ...f(\alpha_{t+1}|\alpha_t^k)d\alpha_{t+1}...d\alpha_{t+p} \\ &= f(y_{t+p}|\mu_{t+p}^k)...f(y_{t+1}|\mu_{t+1}^k), \end{aligned}$$

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

$$\frac{f(y_{t+p}|\alpha^j_{t+p})...f(y_{t+1}|\alpha^j_{t+1})}{f(y_{t+p}|\mu^{kj}_{t+p})...f(y_{t+1}|\mu^{kj}_{t+1})}.$$

- The influence of the empirical prediction density reduced as it will have been propagated p times through the transition density. This may reduce the influence of outliers on the ASIR method.

- More computationally expensive (in memory).

- More variable weights for SIR.

- Can use as basis for MCMC proposal.

Example 1: Bearings-only tracking.

Following Gordon Salmond and Smith (1993, IEE Transactions). We have an observer, located in two dimensions, who observes only the bearing (subject to noise) of a moving object, $\alpha_t = (x_t, vx_t, z_t, vz_t)'$,

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

where x_t, z_t represent the object's horizontal and vertical position at time t and vx_t, vz_t represent the corresponding velocities, $\sigma_\eta = 0.001$ and $\sigma_\varepsilon = 0.005$. We have the observation model

$$f(y_t|\mu_t) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(y_t-\mu_t)}, \quad 0 \le y_t < 2\pi, \quad 0 \le \rho \le 1, \quad (0{\cdot}3)$$

$\mu_t = \tan^{-1}(z_t/x_t)$.

$T = 10$. The "true" filtered mean is calculated for each replication by using the auxiliary method with $M = 100,000$ and $R = 120,000$. Different random number seed, S times and recording the average of the resulting squared difference between the particle filter's estimated mean and the "true" filtered mean. Hence for replication i, state component j, at time t we calculate

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

where $\overline{\alpha}^i_{t,j,s}$ is the particle mean for replication i, state component j, at time t, for simulation s and $\widetilde{\alpha}^i_{t,j}$ 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

$$LMSE^P_{j,t} = \log \frac{1}{REP} \sum_{i=1}^{REP} MSE^P_{i,j,t}. \tag{0·4}$$

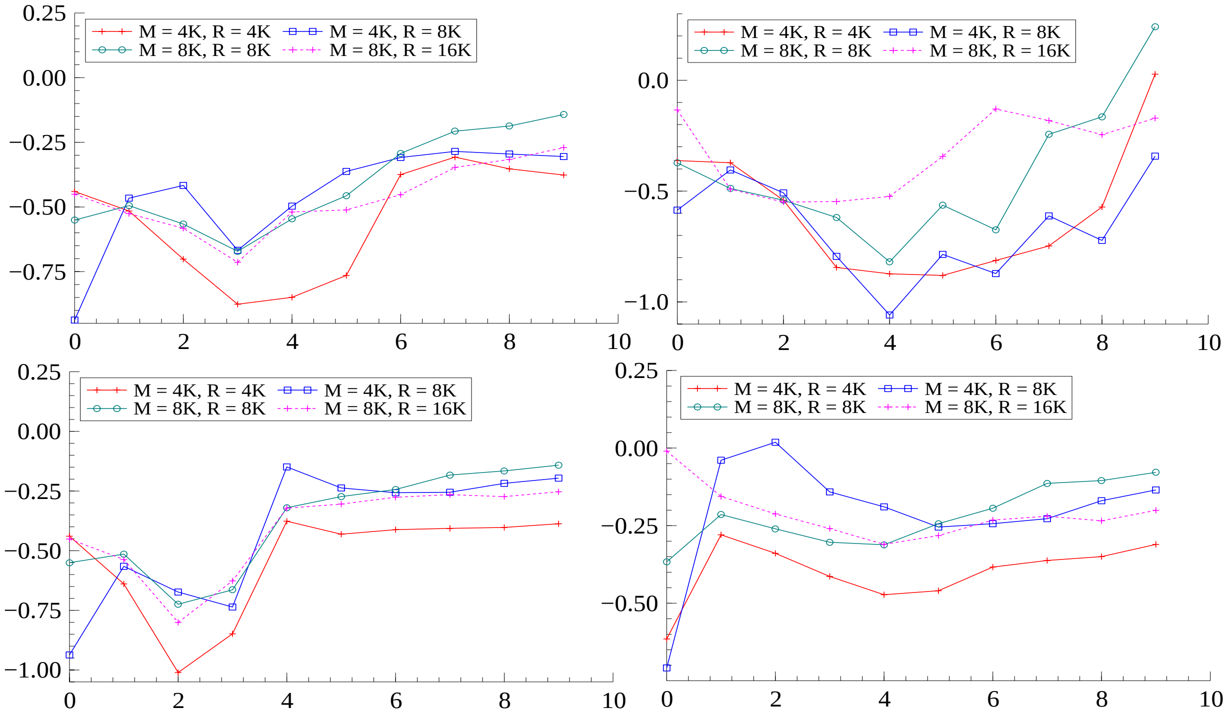

Figure 1: *Plot of the relative mean square error performance (on the log-scale) of the particle filter and the auxiliary based particle filter for the bearings only tracking problem. Numbers below zero indicate a superior performance by the auxiliary particle filter. In these graphs* $M = 4,000$ *or* $8,000$ *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:* $\alpha_{t4} = vz_t$.

2. **Adaption**

$$\begin{array}{llcll} \text{meas eqn:} & y_t & = & \epsilon_t \beta \exp(\alpha_t/2) & \epsilon_t \sim \mathsf{NID}(0,1) \\ \text{state eqn:} & \alpha_{t+1} & = & \phi\alpha_t + \eta_t, & \eta_t \sim \mathsf{NID}(0,\sigma_\eta^2) \end{array}$$

Let $l(\alpha_{t+1}) \equiv \log f(y_{t+1}|\alpha_{t+1})$

$$\begin{aligned} f(\alpha_{t+1}, k|Y_{t+1}) &\propto f(y_{t+1}|\alpha_{t+1}) \times f(\alpha_{t+1}|\alpha_t^k) \\ &\leq \exp\{l(\mu_{t+1}^k) + l'(\mu_{t+1}^k)(\alpha_{t+1} - \mu_{t+1}^k)\} \\ &\quad f(\alpha_{t+1}|\alpha_t^k) \\ &\propto g(k)g(\alpha_{t+1}|k) \end{aligned}$$

So can use proposal in accept-reject where

$$\log \Pr(\mathit{Accept}) = l(\alpha_{t+1}) - \{l(\mu_{t+1}^k) + l'(\mu_{t+1}^k)(\alpha_{t+1} - \mu_{t+1}^k)\} \simeq 0.$$

or in SIR method.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. To make the problem slightly more 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 around twice the time of the SIR based particle filter when $M = R$.

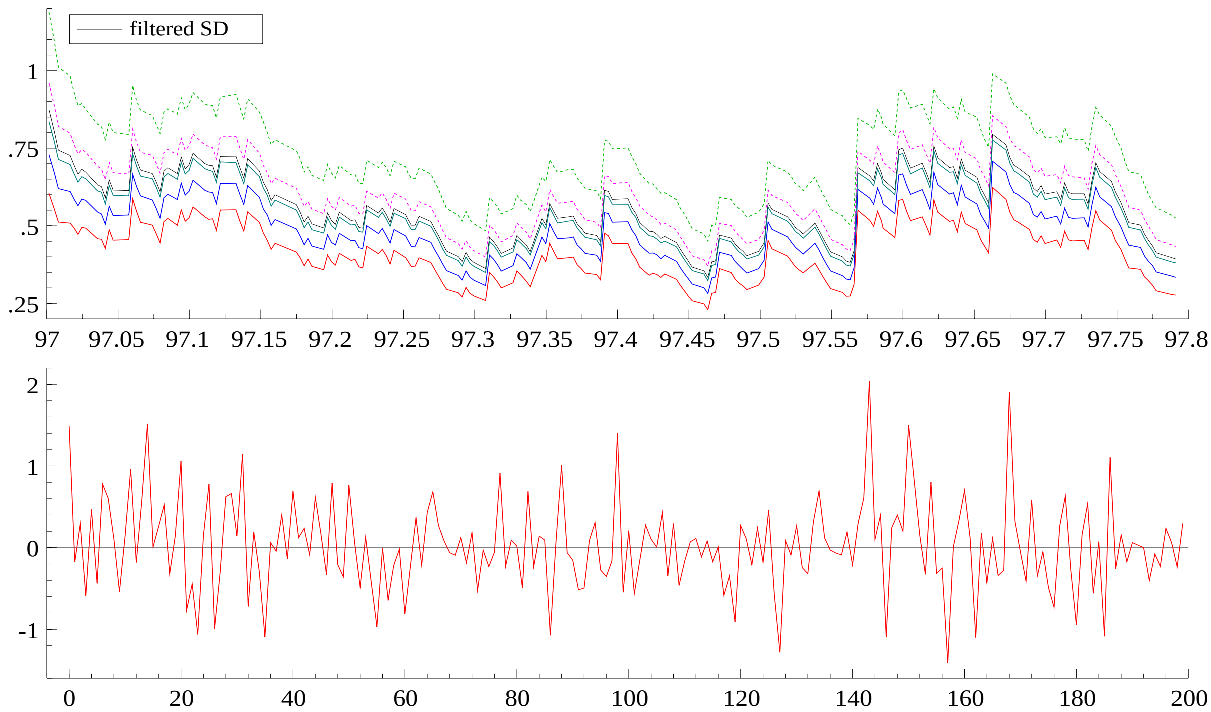

Figure 2: Bottom graph shows the daily returns on the Dollar against UK Sterling from the first day of trading in 1997 for 200 trading days. We display in the top graph the posterior filtered mean (heavy line) of $\beta \exp(\alpha_t/2)|Y_t$, together with the 5, 20, 50, 80, 95 percentage points of the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

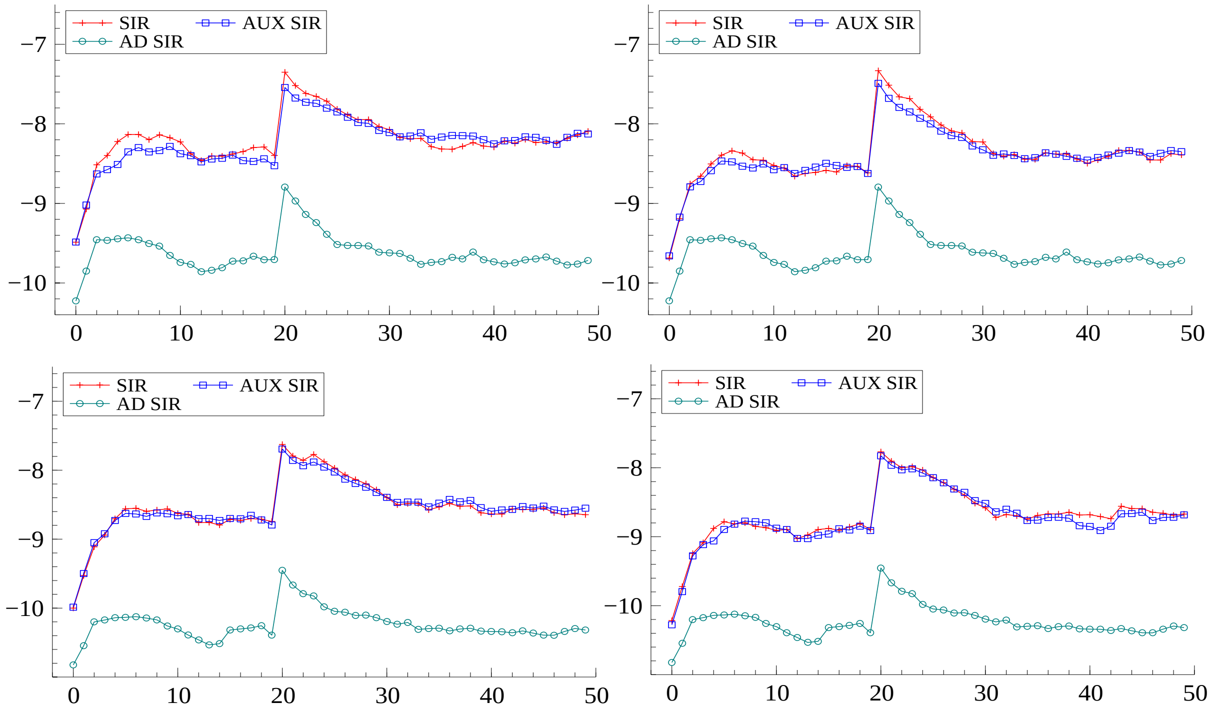

Figure 3: *Plot of the mean square error performance (on the log-scale) of the particle filter to the auxiliary based particle filter and an adapted particle filter. The lower the number the more efficient the method. Top graphs have $M = 2,000$, the bottom have $M = 4,000$. The Left graphs have $R = M$, while the right ones have $R = 2M$*

1. **Stratified sampling**:

This stratification scheme is briefly described in PittShephard(00) and uses the suggestion of CarpenterCliffordFearnhead(99). Explicitly, at the resampling stage, we produce stratified uniforms $\tilde{u}^1, ..., \tilde{u}^R$ by writing

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

That is we use a single uniform realisation to generate sorted stratified uniforms on $[0,1]$. An efficient method, see PS, for inverting the cdf is then used to produce the stratified sorted samples of our variables. CarpenterClifford-Fearnhead(99) justify using stratification ideas via sampling.

0·1 *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

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, 0.707^2) \\ \alpha_{t+1} &= 0.9702\alpha_t + \eta_t, & \eta_t &\sim NID(0, 0.178^2), \end{aligned} \qquad (0{\cdot}5)$$

where ε_t and η_t are independent processes. The model is initialised by α_t's stationary prior while we used $n = 35$. 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 (0·4), 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.

References

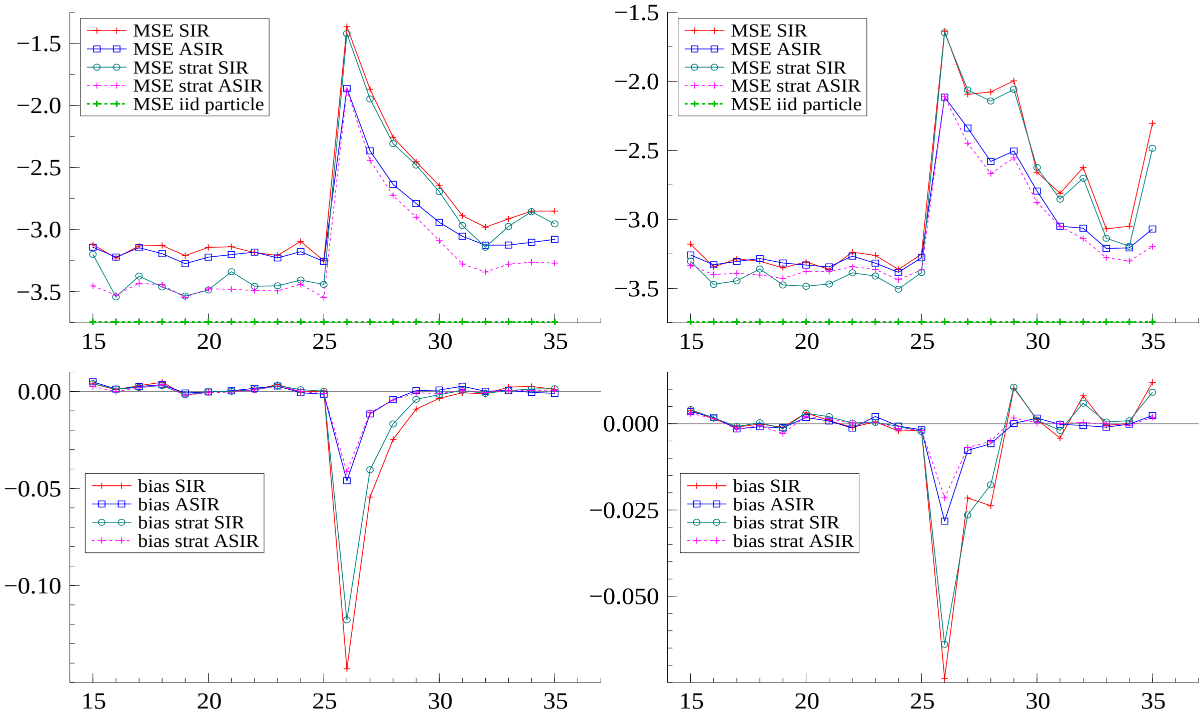

Figure 4: *The mean square error (MSE), using a log10 scale, and bias of four different particle filters using no and two filtered lag filters. The x-axis is always time, but we only graph results for* $t = T/4, T/4+1, ..., 3T/4$ *in order to focus on the crucial aspects. The four particle filters are: 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* $M = R = 500$*.*

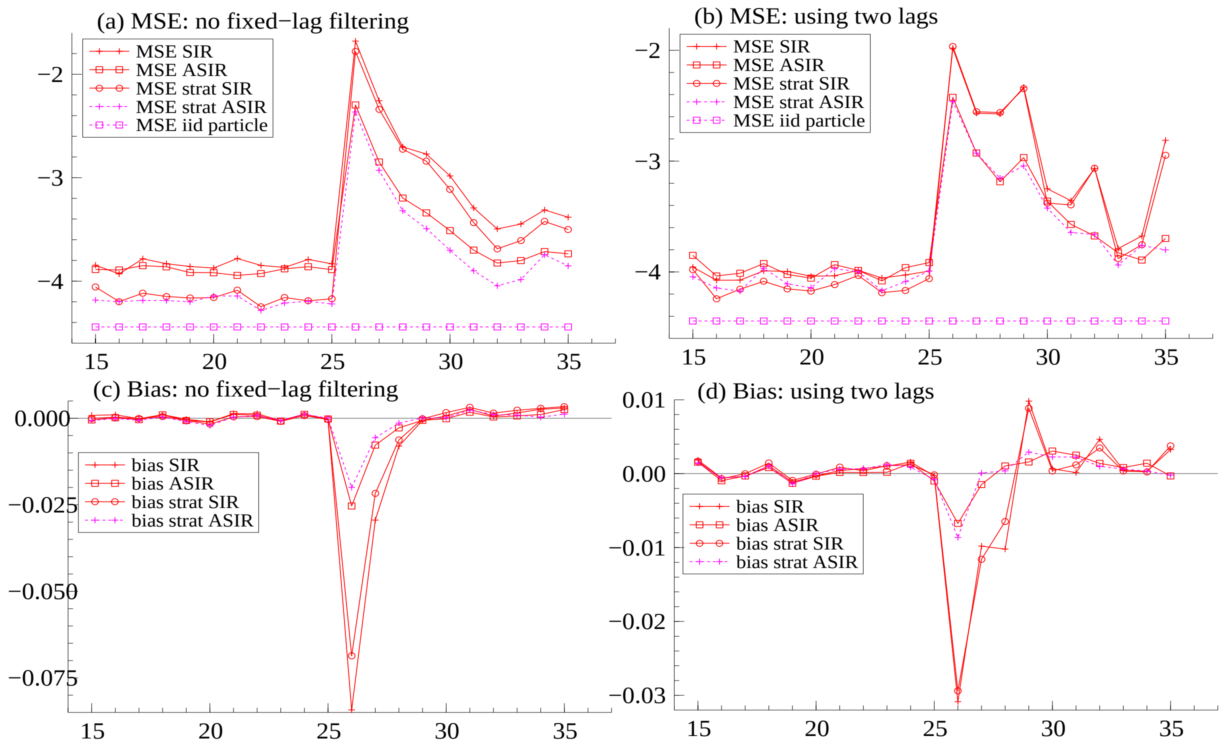

Figure 5: *Repeat of Figure but with $M = R = 2500$.*