Robust inference on parameters via particle filters and sandwich covariance matrices

Arnaud Doucet
Department of Statistics, University of Oxford
&
Oxford-Man Institute, University of Oxford,
doucet@stats.ox.ac.uk

Neil Shephard
Nuffield College, New Road, Oxford OX1 1NF, UK,
Department of Economics, University of Oxford
&
Oxford-Man Institute, University of Oxford,
neil.shephard@economics.ox.ac.uk

1st June, 2012

Abstract

Likelihood based estimation of the parameters of state space models can be carried out via a particle filter. In this paper we show how to make valid inference on such parameters when the model is incorrect. In particular we develop a simulation strategy for computing sandwich covariance matrices which can be used for asymptotic likelihood based inference. These methods are illustrated on some simulated data.

Keywords: quasi-likelihood; particle filter; sandwich matrix; sequential Monte Carlo.
JEL: C11, C15, C53, C58.

1 Introduction

The parameters of non-linear and non-Gaussian state space models are typically estimated using likelihood based methods, where the likelihood needs to be estimated using simulation. When a model is correctly specified the corresponding time series of individual scores are martingale difference sequences which obey a conditional version of the information equality. This allows simple inference using either Bayesian methods or the large sample asymptotics of maximum likelihood estimation.

Usually, however, these models are misspecified and so the above approach to inference could be entirely misleading. In such cases it may be better to make inference using the sandwich matrix. In this note we demonstrate how to reliably estimate the sandwich matrix using simulation for non-linear and non-Gaussian state space models.

Take a step back for a moment. This paper brings together two traditions. The first is the study of the asymptotic sampling behaviour of quasi-likelihoods. Early contributions include, for example, Cox (1961), Huber (1967), Eicker (1967), White (1982), Gallant and White (1988) and White (1994). Recent discussions, focusing from a Bayesian perspective, include Muller (2012) and Holmes and Walker (2012). This quasi-likelihood theme has been highly influential in modern statistics and econometrics. It involves the use of a "long-run covariance", or "heteroskedastic and autocorrelation consistent" (HAC), estimator of the variance of the score. The second is simulation based inference on non-linear and non-Gaussian state space models using sequential Monte Carlo or particle filters. Early contributions to this include, for example, Gordon, Salmond, and Smith (1993), Liu and Chen (1998), Pitt and Shephard (1999) and Doucet, de Freitas, and Gordon (2001). Modern surveys include, for example, Doucet and Johansen (2011) and Creal (2012). The paper which inspires much of this paper is Del Moral, Doucet, and Singh (2009), which we will discuss at some length.

There are many subjects where particle filters are widely applied. Particle filters are increasingly significantly used in economics, for example. Their first applications there appeared in Kim, Shephard, and Chib (1998) and Pitt and Shephard (1999) in the context of financial economics. Their use in macroeconomics appears in, for example, Fernandez-Villaverde and Rudio-Ramirez (2005), Fernandez-Villaverde, Rudio-Ramirez, and Santos (2006), Fernandez-Villaverde and Rudio-Ramirez (2007), An and Schorfheide (2007) and Hansen, Polson, and Sargent (2011). Their use in the analysis of auctions appears in, for example, Kim (2010). Their use in the study of structural microeconomic models with serially correlated latent state variables appears in, for example, Blevins (2011) and Gallant, Hong, and Khwaja (2010) and Gallant, Hong, and Khwaja (2011). Their more recent use on problems in finance include Johannes, Polson, and Stroud (2009). Creal (2012), Flury and Shephard (2011) and Durham and Geweke (2012) discuss their application to a variety of concrete economic problems.

The structure of this note is as follows. In section 2 we develop the background, including listing the model, recalling basic filtering results and discussing inference methods for unknown parameters. In Section 3 we detail a functional recursion for the time series of individual scores. In Section 4 we show how particle filter output can be used to estimate the time series of scores and derive the properties of such estimators. In Section 5 we derive the properties of the corresponding HAC estimator. In Section 6 we discuss particle methods for computing the Hessian. In Section 7 we report a Monte Carlo experiment into the performance of our particle estimator of the robust standard errors. In Section 8 we give some conclusions. Finally, the Appendix contains the proofs of two theorems given in the main part of the paper and recalls the details of a simple particle filter.

2 The background

2.1 A class of models

We will study non-linear and non-Gaussian state space models where there are observations y_t which conditioned on some states α_t are independent. The states are Markovian. We will write this model as

$$f\left(y_t|\alpha_t;\theta\right), \quad f(\alpha_t|\alpha_{t-1};\theta),$$

where θ is a $k\times 1$ dimensional vector of parameters. The former is labelled the measurement density, the latter the transition density. Throughout we will assume both are twice continuously differentiable with respect to θ. Reviews of the literature on "hidden Markov" or "state space" models include Harvey (1989), West and Harrison (1989) and Durbin and Koopman (2012). This type of model has been widely studied in modern applied science.

Our focus will be on filtering

$$f(\alpha_t|\mathcal{F}_t;\theta),$$

where $t\leq n$. Here $\mathcal{F}_t$ is the information available at and including time t. Also of interest is the prediction density

$$f(y_t|\mathcal{F}_{t-1};\theta).$$

Except in special cases (e.g. the linear Gaussian model and the case where α_t only has a small number of atoms of support) these two densities have to be estimated using simulation. The leading way of carrying this out is the particle filter or sequential Monte Carlo.

From the perspective of this paper, we do not need to relive the details of how the particle filter is configured or iterated through time. We will simply assume we have a weighted particle filter sample of size M from the $\alpha_t|\mathcal{F}_t;\theta$ which we write as

$$\left\{W_t^{(i)},\alpha_t^{(i)}\right\}.$$

We will need such a sample for every value of t, although we will see it is not necessary to cumulatively store the particles through time. To help readers less familiar with this area we give in Appendix 9.4 a particular particle filter. It is the one we will use in our examples.

A biproduct of the particle filter is the corresponding estimator

$$\widehat{f}\left(y_t|\mathcal{F}_{t-1};\theta\right),$$

which has the feature that $\prod_{t=1}^{n} \widehat{f}\left(y_t|\mathcal{F}_{t-1};\theta\right)$ is unbiased for $\prod_{t=1}^{n} f\left(y_t|\mathcal{F}_{t-1};\theta\right)$ (a result due to Moral (2004)) which in turn means it can be used inside a MCMC algorithm as if were the true prediction decomposition. The latter result is due to Andrieu, Doucet, and Holenstein (2010), see also Flury and Shephard (2011) for further discussion of it and Pitt, Silva, Giordani, and Kohn (2012) who prove unbiasedness of the auxiliary particle filter estimator of $f\left(y_t|\mathcal{F}_{t-1};\theta\right)$. Such Bayesian inference is complete if the model is correctly specified and the prior reflects the belief of the researcher, but if the quasi-likelihood has some form of misspecification we need to replace the measures of uncertainty by robust estimators. But how do we do this?

2.2 Inference from a quasi-likelihood

Our contribution is to consider an aspect of the parameter inference problem for these misspecified models. Consider the time series $y_{1:n} = y_1, ..., y_n$, then we can can write the model's joint density using a prediction decomposition

$$L_\theta = f(y_{1:n}|\mathcal{F}_0;\theta) = \prod_{t=1}^{n} f(y_t|\mathcal{F}_{t-1};\theta) = \exp\left(\sum_{t=1}^{n} l_{t,\theta}\right), \quad \text{where} \quad l_{t,\theta} = \log f(y_t|\mathcal{F}_{t-1};\theta).$$

Then the unbiased particle estimator of this will be

$$\widehat{L}_\theta = \prod_{t=1}^{n} \widehat{f}(y_t|\mathcal{F}_{t-1};\theta).$$

Example 1 *We suppose* $y_t|\alpha_t;\theta \sim N(\alpha_t, \sigma^2)$*,* $\alpha_t|\alpha_{t-1} \sim N(0, 0.1)$*,* $n = 100$ *and* $\theta = \log\sigma^2$*. We draw a single path* $y_{1:n}$*. Throughout the true value of* $\theta = 0$ *and plot* $\log L_\theta$ *and* $\log \widehat{L}_\theta$ *as a function of* θ*, using the filter in the Appendix. We vary* $M = 100$*,* 250*,* $1,000$ *and* $2,500$ *and show the results in Figure 1. Obviously* $\log \widehat{L}_\theta$ *is non-continuous in* θ *but improves as* M *increases as expected from well established particle filter theory.*

The sample score, if it exists, equals (sometimes, for compactness, we will drop the reference to θ in $S_{n,\theta}$ and $s_{t,\theta}$)

$$S_{n,\theta} = \frac{\partial \log L_\theta}{\partial\theta} = \sum_{t=1}^{n} s_{t,\theta}, \quad s_{t,\theta} = \frac{\partial l_{t,\theta}}{\partial\theta}.$$

Then a standard input into robust inference is to estimate

$$\mathcal{I}_{n,\theta} = \text{Var}\left(n^{-1/2}\sum_{t=1}^{n} s_{t,\theta}|\mathcal{F}_0\right)$$

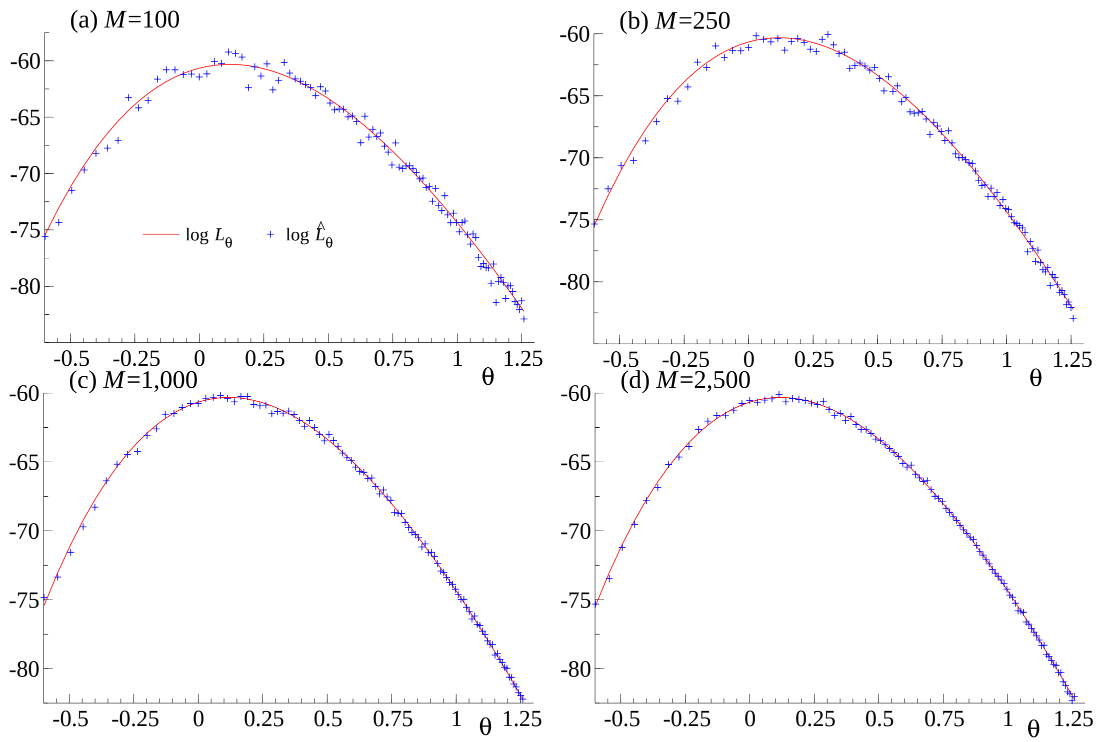

Figure 1: $\log L_\theta$ and $\log \widehat{L}_\theta$ as a function of $\theta = \log \sigma^2$. The true value of θ is zero. Graphs show the results for each value of M. The data $y_{1:n}$ is constant throughout, but each particle filter estimator of the likelihood is stochastically independent for each value of θ.

using a so-called "long-run covariance" or HAC statistic. The literature on this includes Parzen (1957), Gallant (1987), Newey and West (1987) and Andrews (1991). This is combined with an estimate of the Hessian

$$\mathcal{J}_{n,\theta} = -\mathrm{E}\left(n^{-1}\sum_{t=1}^{n}\frac{\partial^2 l_{t,\theta}}{\partial\theta\partial\theta'}|\mathcal{F}_0\right),$$

to deliver an estimator of the sandwich $\mathcal{J}_{n,\theta}^{-1}\mathcal{I}_{n,\theta}\mathcal{J}_{n,\theta}^{-1}$.

A typical HAC estimator takes on the form

$$\widehat{\mathcal{I}}_{n,\theta}(s) = \gamma_{n,\theta}(s;0) + \sum_{j=1}^{P} w(j/P)\left\{\gamma_{n,\theta}(s;j) + \gamma_{n,\theta}(s;j)'\right\},$$

where

$$\gamma_{n,\theta}(s;j) = \frac{1}{n}\sum_{t=j+1}^{n}\left(s_{t,\theta} - \overline{s}_\theta\right)\left(s_{t-j,\theta} - \overline{s}_\theta\right)', \quad \overline{s}_\theta = \frac{1}{n}\sum_{t=1}^{n} s_{t,\theta}.$$

Here w is a weight function, the most well known of which is the Bartlett weight where $w(j/P) = 1-j/(P+1)$. Throughout w is assumed chosen to satisfy $\int w(x)\exp(ix\lambda)dx \geq 0$ for all $\lambda \in R$, which means that the estimator is always positive semi-definite (e.g. Bochner's theorem and Andrews

(1991)). Then under very weak conditions, which involve P increasing very slowly with n discussed extensively in the above literature[1], $\widehat{\mathcal{I}}_{n,\theta}(s) - \mathcal{I}_{n,\theta} \overset{p}{\to} 0$ as $n \to \infty$.

Computationally the real challenge is to compute the time series $s_{t,\theta}$.

2.3 Estimating the individual scores

One approach to accessing the sample score is via the identity

$$\begin{aligned} S_{n,\theta} &= \sum_{t=1}^{n} \mathrm{E}_{\alpha_t|\mathcal{F}_n} \left\{ \frac{\partial \log f(y_t|\alpha_t;\theta)}{\partial\theta} \right\} + \sum_{t=2}^{n} \mathrm{E}_{\alpha_t,\alpha_{t-1}|\mathcal{F}_n} \left\{ \frac{\partial \log f(\alpha_t|\alpha_{t-1};\theta)}{\partial\theta} \right\} \\ &\quad + \mathrm{E}_{\alpha_1|\mathcal{F}_n} \left\{ \frac{\partial \log f(\alpha_1|\mathcal{F}_0;\theta)}{\partial\theta} \right\}. \end{aligned}$$

This appears in the Gaussian case in Koopman and Shephard (1992) but this does not give us the time series of individual scores for

$$\begin{aligned} s_{t,\theta} &= \frac{\partial \log f(y_t|\mathcal{F}_{t-1};\theta)}{\partial\theta} = S_{t,\theta} - S_{t-1,\theta} \\ &\neq \mathrm{E}_{\alpha_t|\mathcal{F}_n} \left\{ \frac{\partial l(y_t|\alpha_t)}{\partial\theta} \right\} + \mathrm{E}_{\alpha_t|\mathcal{F}_n} \left\{ \frac{\partial l(\alpha_t|\alpha_{t-1})}{\partial\theta} \right\}, \end{aligned}$$

where we have written for compactness

$$l(y_t|\alpha_t) = \log f(y_t|\alpha_t;\theta), \quad l(\alpha_t|\alpha_{t-1}) = \log f(\alpha_t|\alpha_{t-1};\theta).$$

Harvey (1989, pp. 142-3) report an involved recursion for the time series of individual scores for linear and Gaussian models, the case where the Kalman filter applies. We need this type of result for general state spaces in order to compute our HAC.

3 Estimating the time series of scores via simulation

3.1 Del Moral, Doucet, and Singh (2009) recursion for the sample score

Del Moral, Doucet, and Singh (2009) developed a particle filter version of recursive maximum likelihood estimation. The key ingredient of this is a sequential estimator of $S_{t,\theta}$. They derived properties of this estimator (note also the earlier Poyiadjis, Doucet, and Singh (2011)). Our target is somewhat different, $s_{t,\theta}$ and the corresponding sandwich estimator, but we will piggyback on their work.

Let us ignore the initial condition, then construct

$$\begin{aligned} u_t(\alpha_t, \alpha_{t-1}) &= \frac{\partial l(y_t|\alpha_t)}{\partial\theta} + \frac{\partial l(\alpha_t|\alpha_{t-1})}{\partial\theta}, \\ U_T(\alpha_{1:T}) &= \sum_{t=1}^{n} u_t(\alpha_t, \alpha_{t-1}), \end{aligned}$$

[1] Throughout our experiments we will take a conventional choice of using the Bartlett kernel and taking $P = \lfloor 4(n/100)^{2/9} \rfloor$.

$$S_t(\alpha_t) \quad = \quad \int U_t(\alpha_{1:t}) \mathrm{d}F(\alpha_{1:t-1}|\mathcal{F}_{t-1}, \alpha_t).$$

Now the sample score is (now dropping reference to θ)

$$S_t = \mathrm{E}_{\alpha_t|\mathcal{F}_t}\left\{S_t(\alpha_t)\right\}.$$

The significant insight of Del Moral, Doucet, and Singh (2009) is that

$$S_t(\alpha_t) = \mathrm{E}_{\alpha_{t-1}|\mathcal{F}_{t-1},\alpha_t}\left\{S_{t-1}(\alpha_{t-1}) + u_t(\alpha_t, \alpha_{t-1})\right\}. \tag{1}$$

Hence they suggested sequentially computing the time series of functionals $S_t(\alpha_t)$ and then off those S_t. They implement this approach using particles, but we step back on that for a moment.

3.2 Individual scores

The above argument means that the time series of individual scores (again dropping reference to θ)

$$s_t = \frac{\partial \log f(y_t|\mathcal{F}_{t-1};\theta)}{\partial \theta} = S_t - S_{t-1}.$$

Now define the functional

$$\begin{aligned}
s_t(\alpha_t) \quad &= \quad S_t(\alpha_t) - S_{t-1} \\
&= \quad \mathrm{E}_{\alpha_{t-1}|\mathcal{F}_{t-1},\alpha_t}\left\{u_t(\alpha_t, \alpha_{t-1}) + S_{t-1}(\alpha_{t-1})\right\} \\
&= \quad \mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}}\left[u_t(\alpha_t, \alpha_{t-1})\right] \\
&\quad\quad + \mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}}\left\{S_{t-1}(\alpha_{t-1}) - S_{t-2}\right\} + (S_{t-2} - S_{t-1}) \\
&= \quad \mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}}\left[u_t(\alpha_t, \alpha_{t-1}) + s_{t-1}(\alpha_{t-1}) - s_{t-1}\right].
\end{aligned} \tag{2}$$

Here

$$s_t = \mathrm{E}_{\alpha_t|\mathcal{F}_t}\left\{s_t(\alpha_t)\right\}.$$

So again this suggests sequentially computing the time series of functionals $s_t(\alpha_t)$ and then off those s_t.

3.3 Variation freeness

Our final detour before we move onto the particle implementations is to look at an important special case, which we nearly always see in practice, where

$$u_t(\alpha_t, \alpha_{t-1}) = \begin{pmatrix} \frac{\partial l(y_t|\alpha_t)}{\partial \psi} \\ \frac{\partial l(\alpha_t|\alpha_{t-1})}{\partial \lambda} \end{pmatrix}.$$

Then $s_t(\alpha_t)$ uncouples as

$$s_t(\alpha_t) = \begin{pmatrix} s_{t,\psi}(\alpha_t) \\ s_{t,\lambda}(\alpha_t) \end{pmatrix}, \quad s_t = \begin{pmatrix} s_{t,\psi} \\ s_{t,\lambda} \end{pmatrix}, \quad S_t(\alpha_t) = \begin{pmatrix} S_{t,\psi}(\alpha_t) \\ S_{t,\lambda}(\alpha_t) \end{pmatrix}, \quad S_t = \begin{pmatrix} S_{t,\psi} \\ S_{t,\lambda} \end{pmatrix},$$

with $\theta = (\psi', \lambda')'$. We suppose no connection between ψ and λ, that it they are variation free (e.g. Engle, Hendry, and Richard (1983)). Then the parallel recursions

$$\begin{aligned} s_{t,\psi}(\alpha_t) &= \frac{\partial l(y_t|\alpha_t)}{\partial \psi} + \left[\mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}} \{s_{t-1,\psi}(\alpha_{t-1})\} - s_{t-1,\psi}\right], \\ s_{t,\lambda}(\alpha_t) &= \mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}} \left\{ \frac{\partial l(\alpha_t|\alpha_{t-1})}{\partial \lambda} + s_{t-1,\lambda}(\alpha_{t-1}) - s_{t-1,\lambda} \right\}. \end{aligned}$$

Likewise

$$\begin{aligned} S_{t,\psi}(\alpha_t) &= \frac{\partial l(y_t|\alpha_t)}{\partial \psi} + \mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}} \{S_{t-1,\psi}(\alpha_{t-1})\}, \\ S_{t,\lambda}(\alpha_t) &= \mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}} \left\{ \frac{\partial l(\alpha_t|\alpha_{t-1})}{\partial \lambda} + S_{t-1,\lambda}(\alpha_{t-1}) \right\}. \end{aligned}$$

4 Particle implementation

4.1 Numerical implementation

We have, recalling (2),

$$\begin{aligned} s_t(\alpha_t) &= \mathrm{E}_{\alpha_{t-1}|\alpha_t,\mathcal{F}_{t-1}} \{s_{t-1}(\alpha_{t-1}) + u_t(\alpha_t, \alpha_{t-1})\} - s_{t-1} \\ &= \frac{\int f(\alpha_t|\alpha_{t-1}) \{s_{t-1}(\alpha_{t-1}) + u_t(\alpha_t, \alpha_{t-1})\} \mathrm{d}F(\alpha_{t-1}|\mathcal{F}_{t-1})}{\int f(\alpha_t|\alpha_{t-1}) \mathrm{d}F(\alpha_{t-1}|\mathcal{F}_{t-1})} - s_{t-1}. \end{aligned}$$

A particle estimator of this is thus

$$\widehat{s}_t(\alpha_t) = -\widehat{s}_{t-1} + \widehat{\mathrm{E}}_{j|\alpha_t} \left\{ s_{t-1}(\alpha_{t-1}^{(j)}) + u_t(\alpha_t, \alpha_{t-1}^{(j)}) \right\}, \tag{3}$$

where generically we write, for any function $H_{j,t-1}(\alpha_t)$,

$$\widehat{\mathrm{E}}_{j|\alpha_t} H_{j,t-1}(\alpha_t) = \frac{\sum_{j=1}^M W_{t-1}^{(j)} f(\alpha_t|\alpha_{t-1}^{(j)}) H_{j,t-1}(\alpha_t)}{\sum_{i=1}^M W_{t-1}^{(j)} f(\alpha_t|\alpha_{t-1}^{(j)})}.$$

Then

$$\widehat{s}_t = \sum_{j=1}^M W_t^{(j)} \widehat{s}_t(\alpha_t^{(j)}),$$

where $\left\{W_t^{(j)}, \alpha_t^{(j)}\right\}$ are the particles approximating $\mathrm{d}F(\alpha_t|\mathcal{F}_t)$. Once this is computed the lagged weighted particles $\left\{W_{t-1}^{(j)}, \alpha_{t-1}^{(j)}\right\}$ can be removed from memory if this is desirable, as they will never be used again.

In practice the above strategy is held back by the $O(M^2)$ calculation in (3), so it may make sense to carry out stratified resampling of the particles $\left\{W_{t-1}^{(j)}, \alpha_{t-1}^{(j)}\right\}$ and $\left\{W_t^{(j)}, \alpha_t^{(j)}\right\}$ to reduce the size of M for the score recursion by removing the small weighted particles. This reduction can be carried out without changing any aspect of the particle filter, in effect it is carried out as a post particle filtering computation.

Example 2 *(Continued from Example 1) We plot* $s_{t,\theta}$ *against* $\widehat{s}_{t,\theta}$ *as a function of* $t = 1, 2, ..., n$ *taking* $\theta = 0$*, the true value and* $n = 100$*. We vary* $M = 100,\ 250,\ 1,000$ *and* $2,500$ *and show the results in Figure 2. The results show a tightening of the estimator as* M *increases.*

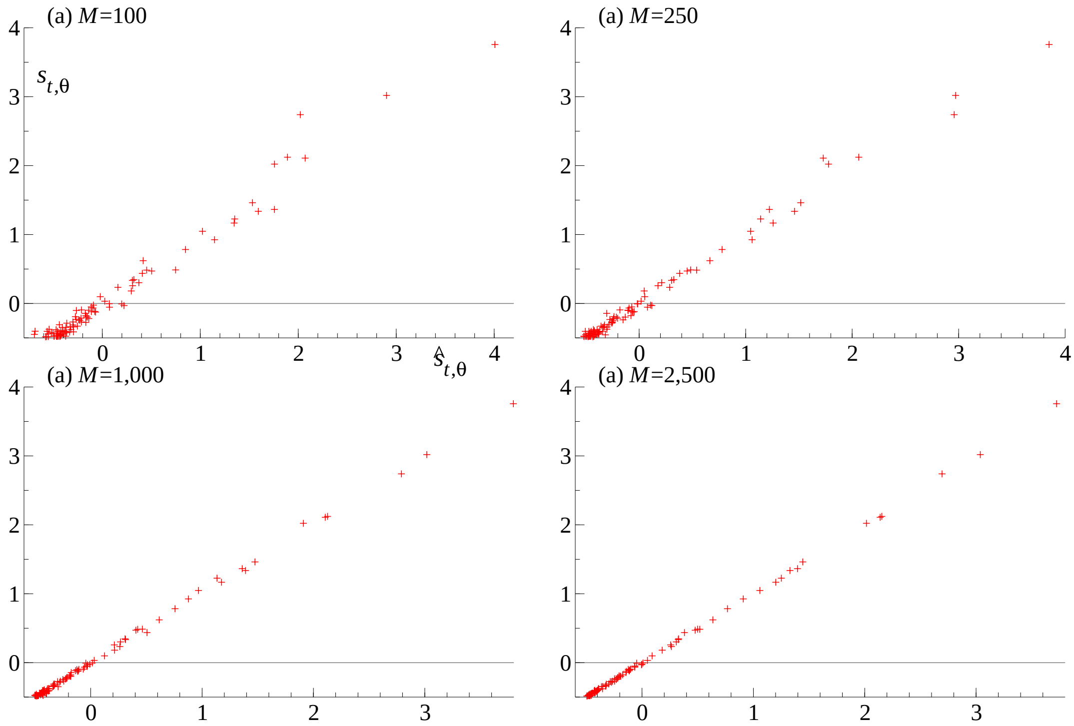

Figure 2: $s_{t,\theta}$ against $\widehat{s}_{t,\theta}$ as a function of $\theta = \log \sigma^2$. The true value of θ is zero. Graphs show the results for each value of M. The data $y_{1:n}$ is constant throughout.

4.2 The asymptotic properties of $\widehat{s}_{t,\theta}$

The following gives a uniform bound, over t, θ and y, on the error caused by the particle filter.

Theorem 1 *Assume Assumption A, given in the Appendix. For any* $r > 1$*, there exists a constant* $c_r < \infty$ *such that for any* $\theta \in \Theta$ *,* $y = \{y_t\}_{t \geq 0}$*,* $t \geq 0$*, that as* $M \to \infty$

$$\sqrt{M} \left\{ \mathrm{E}^M_{t,\theta} \left| \widehat{s}_t - s_t \right|^r \right\}^{1/r} < c_r.$$

Here the expectations are over the particles of size M *using the parameter* θ*, conditioning on* y*.*

Proof. Given in Appendix 9.2.

Extending the work of Del Moral, Doucet, and Singh (2011) it is worth noting it is possible to derive a Gaussian central limit theory for $\sqrt{M}\,(\widehat{s}_t - s_t)$, as $M \to \infty$. However, the resulting asymptotic variance is of not a great deal of practical importance, beyond the fact that the asymptotic variance is uniformly bounded.

Example 3 *(continued from Example 2). The corresponding results for $\widehat{s}_{t,\theta} - s_{t,\theta}$ is given in Figure 3. On the left column we show the 0.9 quantiles of the estimation errors of $s_{t,\theta} - \widehat{s}_{t,\theta}$. As we move across we go through using $M = 100,\ 250,\ 1,000$ and $2,500$. The right column shows the 0.9 quantiles of the scaled estimation errors $\sqrt{M}\,(s_{t,\theta} - \widehat{s}_{t,\theta})$. The key feature is that these are reasonably stable as M changes.*

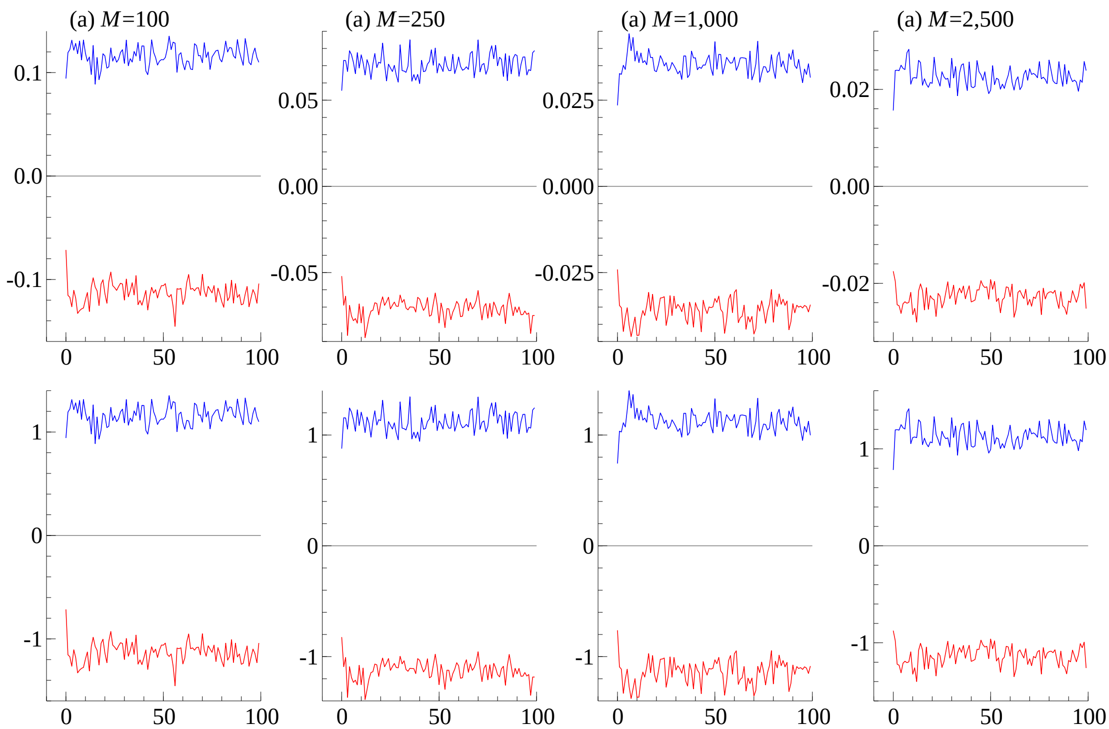

Figure 3: *Simulation based estimation of $s_{t,\theta}$ for each t. Top: 0.9 quantiles of the unscaled simulation errors $\widehat{s}_{t,\theta} - s_{t,\theta}$. Bottom: 0.9 quantiles of the scaled simulation errors $\sqrt{M}\,(\widehat{s}_{t,\theta} - s_{t,\theta})$.*

5 HAC based on estimated scores

We now move to our central concern, the estimation of the HAC of s_t, which we recall we write as $HAC(s)$. Our estimator of this is $HAC(\widehat{s})$.

Theorem 2 *Assume Assumption A, given in the Appendix. Then for all $\theta \in \Theta$, $y = \{y_t\}_{t \geq 0}$, as*

$M \to \infty$ so

$$HAC(\widehat{s}) - HAC(s) \overset{u.p.}{\to} 0$$

so long as $P/M \to 0$.

Proof. Given in the appendix.

This is a very conservative result and so in practice it is unlikely we need M to increase with P at all. Andrews (1991) showed that we typically need P to increase with $n^{1/5}$ to get consistency for the HAC.

Example 4 *(continued from Example 2). Figure 4 shows the corresponding results for the HAC estimator based upon the estimated score. This is based on 500 replications with $n = 100$. On the top it cross-plots $HAC_{\widehat{s}}$ against HAC_s, showing an increasing correlation as M increases and so reducing the simulation error. For small M, $HAC_{\widehat{s}}$ seems a very slightly upward biased estimator of HAC_s. The middle row compares $\sqrt{HAC_{\widehat{s}}}$ and $\sqrt{\widehat{\mathcal{H}}_n}$. This shows that as M increases there is little change in the difference between these two measures. The reason for this is that the HAC_s and $\widehat{\mathcal{H}}_n$ are not very close across the replications — although of course they converge to one another as n goes to infinity as the model is correct. The bottom row shows a histogram of $HAC_{\widehat{s}} - HAC_s$, again indicating that this difference reduces in size as M increases.*

6 Estimating the Hessian via particles

6.1 The basics

To estimate robust standard errors we have to estimate

$$\mathcal{J}_{n,\theta} = -\mathrm{E}\left(n^{-1} \sum_{t=1}^{n} \frac{\partial^2 l_{t,\theta}}{\partial\theta\partial\theta'} |\mathcal{F}_0 \right).$$

We do this by replacing the expectation with an average. The immediate task is thus to compute the Hessian

$$\Delta^2 l_{1:n} = \frac{\partial^2 \log p_\theta(y_{1:n})}{\partial\theta\partial\theta'}.$$

The approach we follow is numerically equivalent to Poyiadjis, Doucet, and Singh (2011) but removes some unnecessary computations and has a simpler derivation.

First we set

$$U_t = \frac{\partial \log p_\theta(y_{1:t}, \alpha_{1:t})}{\partial\theta} = U_{t-1} + u_t(\alpha_t, \alpha_{t-1}), \quad u_t(\alpha_t, \alpha_{t-1}) = \frac{\partial \left\{ l(y_t|\alpha_t) + l(\alpha_t|\alpha_{t-1}) \right\}}{\partial\theta},$$

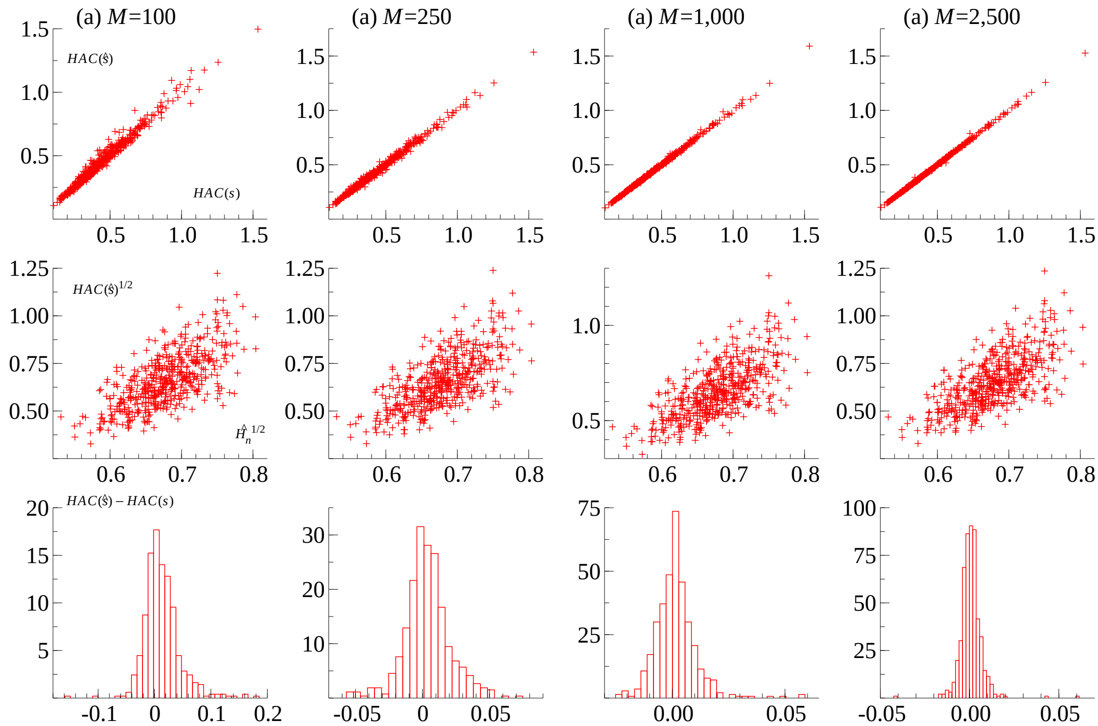

Figure 4: *Simulation based estimation of HAC estimator of the variance of the score when* $n = 100$. *There were 500 replications and were based on* $M =$ *100, 250, 1,000 and 2,500. Left hand column: cross plot of* $HAC(\widehat{s})$ *(on the y-axis) and* $HAC(s)$. *Middle column: cross plot of robust standard errors (on y-axis)* $HAC(\widehat{s})$ *and non-robust standard errors which solely use the Hessian matrix* $\widehat{\mathcal{H}}_n$. *Right hand column: histogram of* $HAC(\widehat{s}) - HAC(s)$.

$$V_t = \frac{\partial^2 \log p_\theta(y_{1:t}, \alpha_{1:t})}{\partial\theta\partial\theta'} = V_{t-1} + v_t(\alpha_t, \alpha_{t-1}), \tag{4}$$

$$v_t(\alpha_t, \alpha_{t-1}) = \frac{\partial^2 \left\{ l(y_t|\alpha_t) + l(\alpha_t|\alpha_{t-1}) \right\}}{\partial\theta\partial\theta'}. \tag{5}$$

Now note that

$$\Delta^2 l_{1:t} = \frac{\partial^2 \log p_\theta(y_{1:t})}{\partial\theta\partial\theta'} = \frac{\partial}{\partial\theta'} \left(\frac{1}{p_\theta(y_{1:t})} \frac{\partial p_\theta(y_{1:t})}{\partial\theta} \right) = B_t - S_t S_t',$$

where

$$B_t = \frac{1}{p_\theta(y_{1:t})} \frac{\partial^2 p_\theta(y_{1:t})}{\partial\theta\partial\theta'}, \quad S_t = \frac{\partial \log p_\theta(y_{1:t})}{\partial\theta},$$

and using the Louis (1982) formula,

$$B_t = H_t + K_t, \tag{6}$$

$$H_t = \int \left\{ U_t(\alpha_{1:t}) U(\alpha_{1:t})' \right\} \mathrm{d}F(\alpha_{1:t}|\mathcal{F}_t), \tag{7}$$

$$K_t = \int V(\alpha_{1:t}) \mathrm{d}F(\alpha_{1:t}|\mathcal{F}_t), \tag{8}$$

and recalling

$$S_t = \int U(\alpha_{1:t}) \mathrm{d}F(\alpha_{1:t}|\mathcal{F}_t).$$

So define the functionals, suppressing dependence on $\alpha_{1:t}$,

$$\begin{aligned} H_t(\alpha_t) &= \int \left(U_t U_t'\right) \mathrm{d}F(\alpha_{1:t-1}|\mathcal{F}_{t-1}, \alpha_t), \\ S_t(\alpha_t) &= \int U_t \mathrm{d}F(\alpha_{1:t-1}|\mathcal{F}_{t-1}, \alpha_t), \\ K_t(\alpha_t) &= \int V_t \mathrm{d}F(\alpha_{1:t-1}|\mathcal{F}_{t-1}, \alpha_t), \end{aligned}$$

delivers the desired[2]

$$H_t = \mathrm{E}_{\alpha_t|\mathcal{F}_t} \{H_t(\alpha_t)\}, \quad S_t = \mathrm{E}_{\alpha_t|\mathcal{F}_t} \{S_t(\alpha_t)\}, \quad K_t = \mathrm{E}_{\alpha_t|\mathcal{F}_t} \{K_t(\alpha_t)\}.$$

Obviously $U_t U_t' = (U_{t-1} + u_t)(U_{t-1} + u_t)'$ and $\mathrm{E}_{\alpha_{1:t-1}|\mathcal{F}_{t-1},\alpha_t} = \mathrm{E}_{\alpha_{t-1}|\mathcal{F}_{t-1},\alpha_t} \mathrm{E}_{\alpha_{1:t-2}|\mathcal{F}_{t-2},\alpha_{t-1}}$, so we run in parallel

$$\begin{aligned} H_t(\alpha_t) &= \mathrm{E}_{\alpha_{t-1}|\mathcal{F}_{t-1},\alpha_t} \left\{H_{t-1}(\alpha_{t-1}) + u_t u_t' + S_{t-1}(\alpha_{t-1}) u_t' + u_t S_{t-1}(\alpha_{t-1})'\right\}, \\ S_t(\alpha_t) &= \mathrm{E}_{\alpha_{t-1}|\mathcal{F}_{t-1},\alpha_t} \{S_{t-1}(\alpha_{t-1}) + u_t\}, \\ K_t(\alpha_t) &= \mathrm{E}_{\alpha_{t-1}|\mathcal{F}_{t-1},\alpha_t} \{K_{t-1}(\alpha_{t-1}) + v_t\}. \end{aligned}$$

6.2 Particle implementation

We will implement this using the existing weighted particles $\left\{W_{t-1}^{(j)}, \alpha_{t-1}^{(j)}\right\}$ (as before, in practice, it computationally makes sense to resample to make the weights equal before carrying this out, but we ignore that here). Write

$$\begin{aligned} H_{j,t-1} &= H_{t-1}(\alpha_{t-1}^{(j)}), \quad S_{j,t-1} = S_{t-1}(\alpha_{t-1}^{(j)}), \quad K_{j,t-1} = K_{t-1}(\alpha_{t-1}^{(j)}), \\ u_{j,t}(\alpha_t) &= u_t(\alpha_t, \alpha_{t-1}^{(j)}), \quad v_{j,t}(\alpha_t) = v_t(\alpha_t, \alpha_{t-1}^{(j)}), \quad f_{j,t}(\alpha_t) = f(\alpha_t|\alpha_{t-1}^{(j)}), \end{aligned}$$

and, again, generically

$$\widehat{\mathrm{E}}_{j|\alpha_t}(H_{j,t-1}) = \frac{\sum_{j=1}^M W_{t-1}^{(j)} f(\alpha_t|\alpha_{t-1}^{(j)}) H_{j,t-1}}{\sum_{j=1}^M W_{t-1}^{(j)} f(\alpha_t|\alpha_{t-1}^{(j)})}.$$

Then

$$\widehat{H}_t(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t} \left\{H_{j,t-1} + u_{j,t}(\alpha_t) u_{j,t}'(\alpha_t) + S_{j,t-1} u_{j,t}'(\alpha_t) + u_{j,t}(\alpha_t) S_{j,t-1}'\right\}, \quad (9)$$

[2]Note that $H_t(\alpha_t) \geq S_t(\alpha_t) S_t(\alpha_t)'$ (in the sense that the difference is positive semi-definite) for any α_t and t and so

$$H_t = \mathrm{E}_{\alpha_t|y_{1:t}} \{H_t(\alpha_t)\} \geq \mathrm{E}_{\alpha_t|y_{1:t}} \left\{S_t(\alpha_t) S_t(\alpha_t)'\right\} \geq \mathrm{E}_{\alpha_t|y_{1:t}} \{S_t(\alpha_t)\} \mathrm{E}_{\alpha_t|y_{1:t}} \left\{S_t(\alpha_t)'\right\} = S_t S_t'.$$

$$\widehat{S}_t(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}\left\{S_{j,t-1} + u_{j,t}(\alpha_t)\right\}, \tag{10}$$

$$\widehat{K}_t(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}\left\{K_{j,t-1} + v_{j,t}(\alpha_t)\right\}. \tag{11}$$

This drives

$$\widehat{H}_t = \sum_{j=1}^{M} W_t^{(j)} \widehat{H}_t(\alpha_t^{(j)}), \quad \widehat{S}_t = \sum_{j=1}^{M} W_t^{(j)} \widehat{S}_t(\alpha_t^{(j)}), \quad \widehat{K}_t = \sum_{j=1}^{M} W_t^{(j)} \widehat{K}_t(\alpha_t^{(j)}).$$

The resulting estimator is thus

$$\Delta^2 \widehat{l}_{1:t} = \widehat{H}_t - \widehat{S}_t \widehat{S}_t' + \widehat{K}_t.$$

In practice it is marginally computationally faster, and exactly equivalent, to combine the recursions for H and K, that is work with

$$\widehat{HK}_t(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}\left\{HK_{j,t-1} + u_{j,t}(\alpha_t)u_{j,t}'(\alpha_t) + S_{j,t-1}u_{j,t}'(\alpha_t) + u_{j,t}(\alpha_t)'_{j,t-1} + v_{j,t}(\alpha_t)\right\},$$

rather than $\widehat{H}_t + \widehat{K}_t$. This delivers $\Delta^2 \widehat{l}_{1:t} = \widehat{HK}_t - \widehat{S}_t \widehat{S}_t'$.

6.3 Variation freeness

In the variation free case, we have

$$u_t(\alpha_t, \alpha_{t-1}) = \begin{pmatrix} \frac{\partial l(y_t|\alpha_t)}{\partial \psi} \\ \frac{\partial l(\alpha_t|\alpha_{t-1})}{\partial \lambda} \end{pmatrix}, \quad v_t(\alpha_t, \alpha_{t-1}) = \begin{pmatrix} \frac{\partial^2 l(y_t|\alpha_t)}{\partial \psi \partial \psi'} & 0 \\ 0 & \frac{\partial^2 l(\alpha_t|\alpha_{t-1})}{\partial \lambda \partial \lambda'} \end{pmatrix},$$

and writing everywhere

$$S_t = \begin{pmatrix} S_{t,\psi} \\ S_{t,\lambda} \end{pmatrix}, \quad K_t = \begin{pmatrix} K_{t,\psi\psi} & K_{t,\lambda\psi} \\ K_{t,\psi\lambda} & K_{t,\lambda\lambda} \end{pmatrix}, \quad H_t = \begin{pmatrix} H_{t,\psi\psi} & H_{t,\lambda\psi} \\ H_{t,\psi\lambda} & H_{t,\lambda\lambda} \end{pmatrix},$$

we note that $K_{t,\psi\lambda} = 0$. Then

$$\widehat{S}_{t,\psi}(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}(S_{j,t-1,\psi}) + \frac{\partial l(y_t|\alpha_t)}{\partial \psi}, \quad \widehat{S}_{t,\lambda}(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}(S_{j,t-1,\lambda}) + \widehat{\mathrm{E}}_{j|\alpha_t}\left(\frac{\partial l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda}\right), \tag{12}$$

$$\widehat{K}_{t,\psi\psi}(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}(K_{j,t-1,\psi\psi}) + \frac{\partial^2 l(y_t|\alpha_t)}{\partial \psi \partial \psi'}, \quad \widehat{K}_{t,\lambda\lambda}(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}(K_{j,t-1,\lambda\lambda}) + \widehat{\mathrm{E}}_{j|\alpha_t}\left(\frac{\partial^2 l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda \partial \lambda'}\right) \tag{13}$$

Unfortunately H is a little more complicated. The easy piece is

$$\widehat{H}_{t,\psi\psi}(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}(H_{j,t-1,\psi\psi}) + \frac{\partial l(y_t|\alpha_t)}{\partial \psi}\frac{\partial l(y_t|\alpha_t)}{\partial \psi'} + \widehat{\mathrm{E}}_{j|\alpha_t}(S_{j,t-1,\psi})\frac{\partial l(y_t|\alpha_t)}{\partial \psi'} + \frac{\partial l(y_t|\alpha_t)}{\partial \psi}\widehat{\mathrm{E}}_{j|\alpha_t}(S_{j,t-1,\psi})'. \tag{14}$$

The slightly more complicated components are

$$\widehat{H}_{t,\lambda\lambda}(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}(H_{j,t-1,\lambda\lambda}) + \widehat{\mathrm{E}}_{j|\alpha_t}\left(\frac{\partial l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda}\frac{\partial l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda'}\right)$$

$$+\widehat{\mathrm{E}}_{j|\alpha_t}\left(\frac{\partial l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda}S'_{j,t-1,\lambda}\right)+\widehat{\mathrm{E}}_{j|\alpha_t}\left(S_{j,t-1,\lambda}\frac{\partial l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda'}\right) \quad (15)$$

$$\widehat{H}_{t,\psi\lambda}(\alpha_t) = \widehat{\mathrm{E}}_{j|\alpha_t}\left(H_{j,t-1,\psi\lambda}\right)+\widehat{\mathrm{E}}_{j|\alpha_t}\left(\frac{\partial l(y_t|\alpha_t)}{\partial \psi}\frac{\partial l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda'}\right)$$
$$+\widehat{\mathrm{E}}_{j|\alpha_t}\left(\frac{\partial l(y_t|\alpha_t)}{\partial \psi}S'_{j,t-1,\lambda}\right)+\widehat{\mathrm{E}}_{j|\alpha_t}\left(S_{j,t-1,\psi}\frac{\partial l(\alpha_t|\alpha_{t-1}^{(j)})}{\partial \lambda'}\right). \quad (16)$$

7 Monte Carlo assessment of the estimator of robust standard errors

We now move onto assessing the performance of this approach in cases where the model is incorrect.

Example 5 *Suppose our model is still* $y_t|\alpha_t;\theta \sim N(\alpha_t,\sigma^2)$ *and* $\alpha_t|\alpha_{t-1} \sim N(\alpha_{t-1},1)$ *and* $\theta = \log\sigma^2$*, but now the data generating process is taken to be*

$$y_t = \alpha_t + \varepsilon_t, \quad \varepsilon_t \overset{L}{=} \frac{\chi_p^2 - p}{\sqrt{2p}}, \quad p > 0,$$

where ε_t *is i.i.d. and* L *denotes "in law". Hence the measurement error is highly skewed for small* p*. As we are fitting a linear model the quasi-likelihood with deliver a consistent estimator of the pseudo-true value of* θ*, which is* $\theta^* = \log Var(\varepsilon_t) = 0$*. The values of* p *vary through* $\{1,2,3,4,5,10,25\}$*, where* $p = 25$ *delivers a law for* ε_t *which is close to being Gaussian. We taken* M *through* $\{100,250,1000,2500\}$ *and* $n = 100$ *and* $n = 250$*.*

Here we focus on estimating (i) the $\mathcal{I}_{n,\theta}^{1/2}$ and (ii) the sandwich $\mathcal{I}_{n,\theta}^{1/2}\mathcal{J}_{n,\theta}^{-1}$. But we need to take a step back for even if we had the true scores we would only have the estimates (i) the HAC estimator based $\widehat{\mathcal{I}}_{n,\theta}^{1/2}$, (ii) $\widehat{\mathcal{I}}_{n,\theta}^{1/2}\widehat{\mathcal{J}}_{n,\theta}^{-1}$. Our paper is about estimating these quantities by particle based (i) $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}$, (ii) $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}\widetilde{\mathcal{J}}_{n,\theta}^{-1}$.

It is not our intention to evaluate here the underlying quality of $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}$ or sandwich estimator $\widehat{\mathcal{I}}_{n,\theta}^{1/2}\widehat{\mathcal{J}}_{n,\theta}^{-1}$, but we also report a comparison of $\widetilde{\mathcal{J}}_{n,\theta}^{-1/2}$ to $\widehat{\mathcal{I}}_{n,\theta}^{1/2}\widehat{\mathcal{J}}_{n,\theta}^{-1}$, to give an impression of how much difference there is in the non-robust and robust standard errors in this case. Clearly the differences would be expected to decline as p increases.

So to summarise here we compare

- $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}$ to $\widehat{\mathcal{I}}_{n,\theta}^{1/2}$.
- $\widetilde{\mathcal{J}}_{n,\theta}^{-1/2}$ to $\widehat{\mathcal{I}}_{n,\theta}^{1/2}\widehat{\mathcal{J}}_{n,\theta}^{-1}$.
- $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}\widetilde{\mathcal{J}}_{n,\theta}^{-1}$ to $\widehat{\mathcal{I}}_{n,\theta}^{1/2}\widehat{\mathcal{J}}_{n,\theta}^{-1}$.

We will make these three comparisons in two different ways.

We look at their average levels, by reporting their medians over the replications in our simulations. For the first two comparisons they do not vary much with n, so we will only report in the Table of results the case where $n = 250$.

A significant focus for us is the percentage differences in the square root of the HACs

$$100\left|\frac{\widetilde{\mathcal{I}}_{n,\theta}^{1/2}}{\widehat{\mathcal{I}}_{n,\theta}^{1/2}}-1\right|,$$

and the estimating robust standard errors

$$100\left|\frac{\widetilde{\mathcal{I}}_{n,\theta}^{1/2}\widetilde{\mathcal{J}}_{n,\theta}^{-1}}{\widehat{\mathcal{I}}_{n,\theta}^{1/2}\widehat{\mathcal{J}}_{n,\theta}^{-1}}-1\right|.$$

The Table will record the Monte Carlo estimates of the 0.5 and 0.99 quantiles of the sampling distribution.

The first column of Table 1 shows the median of the estimated square root of the HAC $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}$, based upon the particle filter. This can be compared to the square root of the HAC based upon the true scores $\widehat{\mathcal{I}}_{n,\theta}^{1/2}$. The results show an upward bias in the estimated HAC for small M but this is removed by the time M reaches about 250.

The next block of columns looks at the percentage difference between the estimated square rooted HAC $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}$ and the true square root HAC $\widehat{\mathcal{I}}_{n,\theta}^{1/2}$. This is computed for every replication and we record the median and the .99 quantile. The latter is poorly estimated as our Monte Carlo experiment has a modest number of replications. The results show, broadly, that as n increases the percentage error tends to fall and that the percentage errors are much higher for the highly non-Gaussian cases where p is small. As M increases the percentage errors fall a great deal. By the time M hits $1,000$ the typical error is quite modest.

The next block of columns starts with the median standard error computed using the estimated scores $\widetilde{\mathcal{J}}_{n,\theta}^{-1/2}$ but wrongly assuming the model is correct. This is, subject to estimation noise, invariant over p and does not vary much with M. This is compared to the median of the robust standard errors using the particle filter approach. As p increases the two become very close, but for the highly non-Gaussian case the differences are marked.

The final block of columns again looks at percentage differences but this time it is the particle filter robust standard errors $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}\widetilde{\mathcal{J}}_{n,\theta}^{-1}$ compared to those based on the true scores $\widehat{\mathcal{I}}_{n,\theta}^{1/2}\widehat{\mathcal{J}}_{n,\theta}^{-1}$. The picture here is quite stable. The percentage differences are largest with small p, n and M. Increasing any of these three constants reduces the percentage errors. The most interesting feature is the improvement as n increases. Except in the most non-Gaussian cases, the percentage errors are quite modest.

		$\widehat{\mathcal{I}}_{n,\theta}^{1/2}$	$\widetilde{\mathcal{I}}_{n,\theta}^{1/2}$	$100\left\lvert\frac{\widehat{\mathcal{I}}_{n,\theta}^{1/2}}{\widetilde{\mathcal{I}}_{n,\theta}^{1/2}}-1\right\rvert$				$\widetilde{\mathcal{J}}_{n,\theta}^{-1/2}$	$\widehat{\mathcal{J}}_{n,\theta}^{-1}\widehat{\mathcal{I}}_{n,\theta}^{1/2}$	$100\left\lvert\frac{\widehat{\mathcal{J}}_{n,\theta}^{-1}\widehat{\mathcal{I}}_{n,\theta}^{1/2}}{\widetilde{\mathcal{J}}_{n,\theta}^{-1}\widetilde{\mathcal{I}}_{n,\theta}^{1/2}}-1\right\rvert$			
quantile		.5	.5	.5	.5	.99	.99	.5	.5	.5	.5	.99	.99
n		250	250	100	250	100	250	250	250	100	250	100	250
	M												
$p=1$	100	1.43	1.31	9.7	9.9	39.7	48.2	0.10	0.20	6.2	3.6	49.6	14.3
	250	1.31	1.31	5.1	5.6	27.3	39.2	0.10	0.20	4.6	2.5	36.1	8.2
	1,000	1.32	1.31	3.2	2.2	18.0	26.9	0.10	0.20	2.2	1.3	31.1	4.3
	2,500	1.32	1.31	1.4	1.8	9.2	27.2	0.10	0.20	1.3	0.9	15.6	4.3
$p=2$	100	0.88	0.86	9.1	7.6	32.7	18.7	0.10	0.16	5.4	3.4	47.8	12.0
	250	0.86	0.86	5.4	3.9	15.7	16.7	0.10	0.16	3.1	2.0	27.8	5.1
	1,000	0.86	0.86	2.9	2.0	13.3	17.1	0.10	0.16	1.7	1.1	12.7	3.2
	2,500	0.85	0.86	1.5	1.3	6.5	6.2	0.10	0.16	1.1	0.6	10.2	2.3
$p=3$	100	0.77	0.76	8.1	7.2	35.1	35.8	0.10	0.15	5.6	3.3	21.9	8.3
	250	0.77	0.76	4.8	3.9	16.9	10.5	0.10	0.14	3.0	2.0	11.9	4.8
	1,000	0.75	0.76	2.2	2.0	8.7	8.8	0.10	0.15	1.7	1.1	5.3	2.9
	2,500	0.76	0.76	1.4	1.0	6.2	6.3	0.10	0.15	0.9	0.7	3.6	1.8
$p=4$	100	0.67	0.64	9.6	6.3	32.7	26.6	0.09	0.13	5.2	3.3	21.0	6.6
	250	0.66	0.64	5.0	3.3	23.4	10.1	0.10	0.13	2.8	1.9	14.4	4.1
	1,000	0.64	0.64	2.4	1.7	9.5	5.8	0.10	0.13	1.6	1.0	8.3	3.9
	2,500	0.64	0.64	1.7	0.9	11.1	3.6	0.10	0.13	0.9	0.7	5.1	1.4
$p=5$	100	0.61	0.59	7.2	7.6	29.3	23.6	0.10	0.13	4.5	3.2	14.5	9.7
	250	0.59	0.59	4.7	3.8	17.9	12.1	0.10	0.13	3.1	2.0	7.2	5.3
	1,000	0.59	0.59	2.0	1.7	6.5	5.7	0.10	0.13	1.6	1.0	3.9	3.9
	2,500	0.59	0.59	1.4	1.1	5.0	3.4	0.10	0.13	1.0	0.7	2.5	1.9
$p=10$	100	0.48	0.47	8.8	6.9	38.0	34.7	0.10	0.11	5.2	3.3	12.4	6.8
	250	0.47	0.47	4.7	3.5	19.6	10.9	0.10	0.11	3.3	2.0	21.0	6.9
	1,000	0.47	0.47	2.2	1.9	7.2	8.0	0.10	0.11	1.4	1.0	4.7	2.5
	2,500	0.47	0.47	1.4	1.1	4.4	4.7	0.10	0.11	0.9	0.7	5.2	1.9
$p=25$	100	0.42	0.41	8.3	7.6	35.7	22.1	0.10	0.10	5.2	3.0	11.3	6.7
	250	0.41	0.41	4.7	3.4	20.5	9.4	0.10	0.10	3.2	2.1	7.2	4.5
	1,000	0.41	0.41	2.5	1.7	7.8	5.1	0.10	0.10	1.5	1.0	3.3	2.9
	2,500	0.41	0.41	1.3	1.0	3.1	3.8	0.10	0.10	1.1	0.6	2.3	2.6

Table 1: Monte Carlo estimates of various quantiles of the distribution of information measures and standard errors. The results are based on 250 replications. n is either 100 or 250. M ranges over 100, 250, 1000 and 2,500. $\widehat{\mathcal{I}}_{n,\theta}^{1/2}$ is the particle filter estimator of the true HAC $\widetilde{\mathcal{I}}_{n,\theta}^{1/2}$ which in turn is based on the true scores. The columns $100\left\lvert\frac{\widehat{\mathcal{I}}_{n,\theta}^{1/2}}{\widetilde{\mathcal{I}}_{n,\theta}^{1/2}}-1\right\rvert$ record the percentage error of induced by the particle filter. $\widetilde{\mathcal{J}}_{n,\theta}^{-1/2}$ provides asymptotic standard error using the true scores under the assumption the model is correct. $\widehat{\mathcal{J}}_{n,\theta}^{-1}\widehat{\mathcal{I}}_{n,\theta}^{1/2}$ provides robust standard errors using estimated scores. $100\left\lvert\frac{\widehat{\mathcal{J}}_{n,\theta}^{-1}\widehat{\mathcal{I}}_{n,\theta}^{1/2}}{\widetilde{\mathcal{J}}_{n,\theta}^{-1}\widetilde{\mathcal{I}}_{n,\theta}^{1/2}}-1\right\rvert$ is the percentage error induced by estimating the scores.

8 Conclusion

For non-linear and non-Gaussian models the time series of individual scores are not directly available, but have to be estimated by simulation. In this note we derived the properties of these estimators. The estimated individual scores are used to calculate a HAC estimator of the long-run variance of the score and in turn inside an estimator of a robust covariance matrix. Here we have studied the properties of the resulting simulation based estimators using asymptotic theory and Monte Carlo experiments. The results are encouraging suggesting that these standard errors are good approximations to the standard errors which we would have seen if the individual scores were directly computable.

This analysis continues an important line of work. In Andrieu, Doucet, and Holenstein (2010) they showed how to carry out Bayesian inference using only the output from a particle filter. Here we extend that journey, allowing robust inference to be joined with the Bayesian estimator. This provides a complete analysis of inference for these compelling but computationally challenging models.

A disappointment of our approach is that it requires us to be able to compute both $f(y_t|\alpha_t, \theta)$ and $f(\alpha_t|\alpha_{t-1}, \theta)$ (as well as their log derivatives with respect to θ, although these latter terms could be computed by numerical differentiation). The main reason for the limitation is that there are a considerable number of models where it is difficult to access the form of $f(\alpha_t|\alpha_{t-1}, \theta)$ even though we can simulate from $\alpha_t|\alpha_{t-1}, \theta$. At the moment our methods cannot handle such models.

9 Appendix

9.1 Assumptions

Assumption A.

Uniformly over y, x, x' and θ, we assume $\partial f(y|x;\theta)/\partial\theta$ and $\partial f(x'|x;\theta)/\partial\theta$ exist and there exists constants $0 < \delta, \rho, c < \infty$ such that

$$\begin{aligned} \rho^{-1} &\leq f(y|x;\theta) \leq \rho, \\ \delta^{-1} &\leq f(x'|x;\theta) \leq \delta, \end{aligned}$$

and

$$\left|\frac{\partial f(y|x;\theta)}{\partial\theta}\right| \vee \left|\frac{\partial f(x'|x;\theta)}{\partial\theta}\right| < c.$$

9.2 Proof of Theorem 1

First introduce the generic notation

$$\nu[\varphi] = \int \varphi(x) \mathrm{d}\nu(x).$$

Then recall that

$$\begin{aligned} s_t &= \frac{1}{f(y_t|\mathcal{F}_{t-1};\theta)} \frac{\partial f(y_t|\mathcal{F}_{t-1};\theta)}{\partial\theta} \\ &= \frac{1}{f(y_t|\mathcal{F}_{t-1};\theta)} \left\{ \int \frac{\partial f(y_t|\alpha_t;\theta)}{\partial\theta} f(\alpha_t|\mathcal{F}_{t-1};\theta)\mathrm{d}\alpha_t + \int f(y_t|\alpha_t;\theta) \frac{\partial f(\alpha_t|\mathcal{F}_{t-1};\theta)}{\partial\theta} \mathrm{d}\alpha_t \right\} \\ &= \frac{1}{f(y_t|\mathcal{F}_{t-1};\theta)} \left\{ f(\alpha_t|\mathcal{F}_{t-1};\theta)[\frac{\partial f(y_t|\alpha_t;\theta)}{\partial\theta}] + \frac{\partial f(\alpha_t|\mathcal{F}_{t-1};\theta)}{\partial\theta}[f(y_t|\alpha_t;\theta)] \right\}. \end{aligned}$$

Our particle filter approach above is the same as using a particle filter to estimate $f(\alpha_t|\mathcal{F}_{t-1};\theta)$ and the functional recursion above to estimate $\partial f(\alpha_t|\mathcal{F}_{t-1};\theta)/\partial\theta$.

Now let us write

$$\eta_{t,\theta}^M = \widehat{f}(\alpha_t|\mathcal{F}_{t-1};\theta), \quad \eta_{t,\theta} = f(\alpha_t|\mathcal{F}_{t-1};\theta), \quad \zeta_{t,\theta}^M = \frac{\partial\widehat{f}(\alpha_t|\mathcal{F}_{t-1};\theta)}{\partial\theta}, \quad \zeta_{t,\theta} = \frac{\partial f(\alpha_t|\mathcal{F}_{t-1};\theta)}{\partial\theta},$$

Then

$$f(y_t|\mathcal{F}_{t-1};\theta)\sqrt{M}\left(\widehat{s}_t - s_t\right) = \sqrt{M}\left(\eta_{t,\theta}^M - \eta_{t,\theta}\right)[\frac{\partial f(y_t|\alpha_t;\theta)}{\partial\theta}] + \sqrt{M}\left(\zeta_{t,\theta}^M - \zeta_{t,\theta}\right)[f(y_t|\alpha_t;\theta)].$$

Now under Assumption A Moral (2004) and Del Moral and Rio (2011) have established that for a

$$\sqrt{M}\left\{\mathrm{E}^M \left|\left(\eta_{t,\theta}^M - \eta_{t,\theta}\right)[\varphi]\right|^r\right\}^{1/r} \leq d_{1,r}$$

and Del Moral, Doucet, and Singh (2009) have established under Assumption A that

$$\sqrt{M}\left\{\mathrm{E}^M \left|\left(\zeta_{t,\theta}^M - \zeta_{t,\theta}\right)[\varphi]\right|^r\right\}^{1/r} \leq d_{2,r}.$$

Minkowski inequality means that

$$\begin{aligned} \sqrt{M}\left\{\mathrm{E}^M \left|f(y_t|\mathcal{F}_{t-1};\theta)\left(\widehat{s}_t - s_t\right)\right|^r\right\}^{1/r} &\leq \sqrt{M}\left\{\mathrm{E}^M \left|\left(\eta_{t,\theta}^M - \eta_{t,\theta}\right)[\frac{\partial f(y_t|\alpha_t;\theta)}{\partial\theta}]\right|^r\right\}^{1/r} \\ &\quad + \sqrt{M}\left\{\mathrm{E}^M \left|\left(\zeta_{t,\theta}^M - \zeta_{t,\theta}\right)[f(y_t|\alpha_t;\theta)]\right|^r\right\}^{1/r} \\ &\leq d_{1,r} + d_{2,r}. \end{aligned}$$

Then

$$\sqrt{M}\left\{\mathrm{E}^M \left|\sqrt{M}\left(\widehat{s}_t - s_t\right)\right|^r\right\}^{1/r} \leq \frac{d_{1,r} + d_{2,r}}{f(y_t|\mathcal{F}_{t-1};\theta)} \leq c_r,$$

given the measurement density $f(y_t|\alpha_t)$ is bounded from below.

9.3 Proof of Theorem 2

As the HAC is non-negative, a sufficient condition for this to happen is that

$$HAC(\widehat{s}-s)\overset{u.p.}{\to}0,$$

a matrix of zeros. Recall that

$$EHAC(\widehat{s}-s)=\gamma_{n,\theta}(\widehat{s}-s;0)+\sum_{j=1}^{P}w(j/P)\left\{\gamma_{n,\theta}(\widehat{s}-s;j)+\gamma_{n,\theta}(\widehat{s}-s;j)'\right\}.$$

Now for a conformable k vector we have, if the weight function is everywhere positive,

$$\begin{aligned}
k'HAC(\widehat{s}-s)k &= HAC(k'(\widehat{s}-s)) \\
&\leq \gamma_{n,\theta}(z;0)\left(1+2\sum_{j=1}^{P}w(j/P)\right) \\
&\leq (2P+1)\left(\frac{1}{n}\sum_{t=1}^{n}z_t^2\right), \quad z_t=k'(\widehat{s}-s).
\end{aligned}$$

Now using Theorem 1, we have that

$$\mathrm{E}_{\theta}^{M}\left(\frac{1}{n}\sum_{t=1}^{n}z_t^2\right)=\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}_{t,\theta}^{M}\left(z_t^2\right)\leq\frac{c_2^2}{M}.$$

A similar result can be obtained for higher order moments. Hence we have uniform convergence so long as $P/M\to 0$.

9.4 A simple particle filter

For simplicity in our examples we have used a basic "bootstrap" particle filter, whose modern use was developed by Gordon, Salmond, and Smith (1993). Of course it is possible to design more efficient particle filters using stratification, auxiliary particle filters (e.g. Pitt and Shephard (1999) and Whiteley and Johansen (2011)), irregularly spaced resampling based upon effective sample sizes (e.g. Kong, Liu, and Wong (1994)) as well as other important techniques. See the reviews in, for example, Doucet and Johansen (2011) and Creal (2012). Such improvements can be harnessed by the developments given here.

The basic bootstrap filter has the following structure

1. Set $t=1$ and assume we have a sample $\alpha_{t-1}^{(1)},...,\alpha_{t-1}^{(M)}$ from $\alpha_{t-1}|\mathcal{F}_{t-1},\theta$.

2. For each particle $\alpha_{t-1}^{(j)}$ generate R children from drawing from $\alpha_t|\alpha_{t-1},\theta$. Write the children as $\alpha_t^{(j,1)},...,\alpha_t^{(j,R)}$ and compute weights

$$w_t^{(j,k)}=f(y_t|\alpha_t^{(j,k)},\theta).$$

Record

$$\widehat{p}(y_t|\mathcal{F}_{t-1},\theta) = \frac{1}{MR}\sum_{j=1}^{M}\sum_{k=1}^{R} f(y_t|\alpha_t^{(j,k)},\theta).$$

3. Resample with replacement $\alpha_t^{(j,k)}$ with probability proportional to $w_t^{(j,k)}$ a total of M times. Write this new population as $\alpha_t^{(1)},...,\alpha_t^{(M)}$ which are roughly from $\alpha_t|\mathcal{F}_t,\theta$.

4. Use the two populations of particles $\alpha_{t-1}^{(1)},...,\alpha_{t-1}^{(M)}$ and $\alpha_t^{(1)},...,\alpha_t^{(M)}$ to estimate the score and Hessian at time t. One can then remove $\alpha_{t-1}^{(1)},...,\alpha_{t-1}^{(M)}$ from memory if this is desired.

5. Increment t by one and go to step 2.

In our experiments we have taken $R = 1$.

References

An, S. and F. Schorfheide (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews 26*, 113–172.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B 72*, 1–33.

Blevins, J. R. (2011). Sequential Monte Carlo methods for estimating dynamic microeconomic models. Department of Economics, Ohio State University.

Cox, D. R. (1961). Tests of seperate families of hypotheses. *Proceedings of the Berkeley Symposium 4*, 105–123.

Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews 31*, 245–296.

Del Moral, P., A. Doucet, and S. S. Singh (2009). Forward smoothing using sequential Monte Carlo. Unpublished paper: Cambridge University Engineering Department, CUE/F-INFENG/TR 638.

Del Moral, P., A. Doucet, and S. S. Singh (2011). Uniform stability of a particle approximation of the optimal filter derivative. Unpublished paper: Cambridge University Engineering Department, CUE/F-INFENG/TR 668.

Del Moral, P. and E. Rio (2011). Concentration inequalities for mean field particle models. *Annals of Applied Probability 21*, 1017–1052.

Doucet, A., N. de Freitas, and N. J. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

Doucet, A. and A. Johansen (2011). A tutorial on particle filtering and smoothing: fifteen years later. In *The Oxford Handbook of Nonlinear filtering*. Oxford University Press.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2 ed.). Oxford: Oxford University Press.

Durham, G. and J. Geweke (2012). Massively parallel sequential monte carlo for Bayesian inference. Unpublished paper: Leeds School of Business, University of Colorado.

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, pp. 59–82. Berkeley: University of California.

Engle, R. F., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Fernandez-Villaverde, J. and J. F. Rudio-Ramirez (2005). Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *Journal of Applied Econometrics 20*, 891–910.

Fernandez-Villaverde, J. and J. F. Rudio-Ramirez (2007). Estimating macroeconomics models: a likelihood approach. *Review of Economic Studies 74*, 1059–1087.

Fernandez-Villaverde, J., J. F. Rudio-Ramirez, and M. S. Santos (2006). Convergence properties of the likelihood of computed dynamic models. *Econometrica 74*, 93–119.

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

Gallant, A. R. (1987). *Nonlinear Statistical Models.* New York: John Wiley.

Gallant, A. R. and H. White (1988). *A Unified Theory of Estimation and Inference for nonlinear Dynamic Models.* Oxford: Basil Blackwell.

Gallant, R. A., H. Hong, and A. Khwaja (2010). Dynamic entry with cross product spillovers: An application to the generic drug industry. Unpublished paper: Department of Economics, Yale University.

Gallant, R. A., H. Hong, and A. Khwaja (2011). Bayesian estimation of a dynamic game with endogenous, partially observed, serially correlated state. Unpublished paper: Department of Economics, Duke University.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F 140*, 107–113.

Hansen, L., N. Polson, and T. Sargent (2011). Nonlinear filtering and learning dynamics. *Journal of Business and Economic Statistics*.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Holmes, C. and S. G. Walker (2012). A general framework for the updating of belief distributions. Unpublished paper: Department of Statistics, University of Oxford.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 221–233.

Johannes, M., N. Polson, and J. Stroud (2009). Optimal filtering of jump-diffusions: Extracting latent states from asset prices. *Review of Financial Studies 22*, 2259–2299.

Kim, D.-H. (2010). Flexible Bayesian analysis of first price auctions using simulated likelihood. Chapter 2, unpublished Ph.D. thesis, University of Arizona.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Kong, A., J. S. Liu, and W. H. Wong (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association 89*, 278–88.

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*, 823–6.

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

Louis, T. A. (1982). Finding observed information using the EM algorithm. *Journal of the Royal Statistical Society, Series B 44*, 98–103.

Moral, P. D. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications.* New York: Springer.

Muller, U. (2012). Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. Unpublished paper: Department of Economics, Princeton University.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Parzen, E. (1957). On the consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics 28*, 329–348.

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

Pitt, M. K., R. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*. Forthcoming.

Poyiadjis, G., A. Doucet, and S. Singh (2011). Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika 98*, 65–80.

West, M. and J. Harrison (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.

Whiteley, N. and A. Johansen (2011). Recent developments in auxiliary particle filtering. In D. Barber, A. T. Cemgil, and S. Chiappa (Eds.), *Inference and Learning in Dynamic Models*. Cambridge University Press.