

Second Order Filter Distribution  
Approximations for Financial Time  
Series with Extreme Outliers

J. Q. Smith

Department of Statistics, University of Warwick, United Kingdom (J.Q.Smith@warwick.ac.uk)

António A. F. Santos

Faculty of Economics, Monetary and Financial Research Group (GEMF), University of Coimbra, Portugal (aasantos@fe.uc.pt)

## Abstract

Particle Filters are now regularly used to obtain the filter distributions associated with state space financial time series. Most commonly used nowadays is the auxiliary particle filter method in conjunction with a first order Taylor expansion of the log-likelihood. We argue in this paper that for series such as stock returns, which exhibit fairly frequent and extreme outliers, filters based on this first order approximation can easily break down. However, an auxiliary particle filter based on the much more rarely used second order approximation appears to perform well in these circumstances. To detach the issue of algorithm design from problems related to model misspecification and parameter estimation, we demonstrate the lack of robustness of the first order approximation and the feasibility of a specific second order approximation using simulated data.

**Key words:** Bayesian inference; Importance sampling; Particle filter; State space model; Stochastic volatility.

# 1. INTRODUCTION

One of the two most often reported characteristics associated with financial returns time series is the fat tails in the unconditional distribution of returns. More observations appear in the tails than for Gaussian processes, giving rise to high kurtosis. The second is volatility clustering, indicating the need to model the variance evolution of the series. Empirical and theoretical investigations have both clearly established that for short term financial time series, variances as measures of volatility are time varying but present some degree of predictability (Bollerslev, Engle and Nelson 1994; Taylor 1994; Diebold and Lopez 1995; Engle 1995; Campbell, Lo and MacKinlay 1997; Diebold, Hickman, Inoue and Schuermann 1998; Ait-Sahalia 1998; Andersen, Bollerslev and Lange 1999; Christoffersen and Diebold 2000). Variance is used as a measure of risk in a variety of situations: Value-at-Risk (VaR) calculations, portfolio allocation and pricing options.

To model variance dynamics, nonlinear models have to be used (Gallant, Rossi and Tauchen 1993; Hsieh 1993; Bollerslev et al. 1994; Campbell et al. 1997), which, in turn requires numerical algorithms for estimation and prediction. The two most common classes of models used in financial time series are

the AutoRegressive Conditional Heteroskedastic (ARCH) and Stochastic Volatility (SV) models. Understanding and predicting the evolution of volatility has been a key issue faced by people who have to take decisions in financial markets. These two classes of models have been widely used by academics and practitioners. However, new challenges have appeared and more sophisticated algorithms allowing them to deal with real time decisions are needed. Nowadays, data have become more and more abundant, and one mainstream of research uses intraday data, which enable us to take other characteristics of financial time series into account and enables us to measure and estimate other quantities such as *realized volatility* and *integrated volatility*, respectively. Examples of this kind of research can be found in Andersen Diebold and Ebens (2001), Andersen, Diebold and Labys (2001, 2003) and Andersen, Bollerslev and Meddahi (2004). Certainly, taking these new characteristics and measures into account will pose new challenges to the models mentioned above, and the development of well-adapted algorithms, which is the main object of this paper, is important to the development of models used to characterize financial time series.

For the reasons given above, any state space model of financial returns needs to be nonlinear. The SV model (Taylor 1986)

is the simplest nonlinear state space model. Financial returns  $y_t$  are related to unobserved states which are serially correlated. Thus, we may write

$$y_t = \beta \exp\left(\frac{\alpha_t}{2}\right) \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad (1)$$

$$\alpha_t = \phi \alpha_{t-1} + \sigma_\eta \eta_t, \quad \eta_t \sim N(0, 1), \quad (2)$$

where  $\alpha_t$  are the states of the process for  $t = 1, \dots, n$ . Note that the model is characterized by the vector of parameters  $\theta = (\beta, \phi, \sigma_\eta)$ . Generalizations of this model can be considered. In this paper we consider that the innovation process in the measurement equation follows a Student-t distribution.

In this paper, we assume that the parameters are known or have been previously estimated, for example, using Markov chain Monte Carlo (MCMC) techniques (Jacquier, Polson and Rossi 1994). Our aim is to present modifications to certain recent particle filter methods to improve predictions in the process defining variance evolution. We have adopted a Bayesian state space approach where predictions are expressed through the posterior density of states,  $f(\alpha_t|\theta, D_t)$ , and the predictive density of returns,  $f(y_{t+1}|\theta, D_t)$ , rather than through point predictions, where  $D_t = \{y_1, \dots, y_t\}$  represents the available information at time  $t$ . All densities considered in this paper are conditioned by the set of parameters  $\theta$ , although later to simplify the notation,

we do not make this conditioning explicit. The modifications of the more conventional algorithms proposed here are easy to implement, but they nevertheless appear to improve the predictive performance of particle filter methods dramatically.

In general, a Bayesian analysis will deliver the posterior density of the state  $f(\alpha_t|\theta, D_t)$ , on the unobservable state random variable  $\alpha_t$ ,  $t = 1, \dots, n$ . This summarizes all information available at time  $t$  relevant for predicting future values of the series. As new information arrives, for example  $y_{t+1}$ , the density of the state is updated to  $f(\alpha_{t+1}|\theta, D_{t+1})$ . This forms the basis of a recursion where, as new information arrives, at each given time point the state probability densities are sufficient for all predictive statements to be updated.

This paper focuses on predicting variance evolution in SV models. The method used here is the Particle Filter as developed in Gordon, Salmond and Smith (1993), Kong, Liu and Wong (1994), Fearnhead (1998), Liu and Chen (1998), Carpenter, Clifford and Fearnhead (1999), de Freitas (1999), Doucet, Godsill, and Andrieu (2000), Doucet, de Freitas and Gordon (2001), Liu (2001) and Godsill, Doucet, and West (2004). In it, a distribution that is difficult to analyze algebraically is approximated by a discrete set of points (particles), each with an associated weight. The particles and their associated weights

are updated at each time step according to the state dynamics and to take account of the information in the observation. The standard methods for updating the particles are based on importance sampling.

In this paper, we demonstrate that particle filter algorithms based on a first order approximation of the log-likelihood, when used with the SV model (1)–(2), can give very unsatisfactory results. We then propose a set of simple extensions to these algorithms that can improve the performance of such filters dramatically. As well as making the filters more robust, our proposed method greatly simplifies the construction of the filter in comparison with competing algorithms. Our focus here is on the performance of the filter rather than on the estimation of the model. The algorithms we propose are robust to models that present likelihoods that are not log-concave. This characteristic is missing from the algorithm that serves as a benchmark for this paper: the one based on the first order approximation of the log-likelihood proposed by Pitt and Shephard (1999, 2001).

To attempt to isolate the issue of algorithm performance so that it is not confounded with issues related to model misspecification and methods used to estimate the parameters, in the empirical section of this paper, we use two simulated series from two models. The first is the simplest SV model (1)–(2), and

the second is its analogue, which uses Student- $t$  innovations, with parameters taken from a well-known source (Jacquier et al. 1994). This enables us to demonstrate that the filter performance issue discussed here is largely independent of model misspecification and parameter estimation techniques. We can therefore demonstrate that the main issue here is determining the relative efficacy of different filters, not primarily because the model is misspecified, but because of the lack of robustness of more basic particle filter algorithms to extreme outliers. Throughout this paper we use the term extreme outlier, associated with a given series and a run of a particle filter, for any observation  $y_t$  in that series which lies outside the range of particles used to approximate its predictive distribution.

## 2. FIRST VS SECOND ORDER APPROXIMATIONS IN A PARTICLE FILTER IMPLEMENTATION

Bayes' rule allows us to assert that the posterior density  $f(\alpha_t|D_t)$  of states is related to the density  $f(\alpha_t|D_{t-1})$  prior to  $y_t$  and the density  $f(y_t|\alpha_t)$  of  $y_t$  given  $\alpha_t$  by

$$f(\alpha_t|D_t) \propto f(y_t|\alpha_t)f(\alpha_t|D_{t-1}). \quad (3)$$



The predictive density of  $y_{t+1}$  given  $D_t$  is

$$f(y_{t+1}|D_t) = \int f(y_{t+1}|\alpha_{t+1})f(\alpha_{t+1}|D_t) d\alpha_{t+1}. \quad (4)$$

Instead of estimating these integrals numerically, particle filter methods approximate these densities using a simulated sample. A particle filter method approximates the posterior density of interest,  $f(\alpha_t|D_t)$ , through a set of  $m$  “particles”  $\{\alpha_{t,1}, \dots, \alpha_{t,m}\}$  and their respective weights  $\{\pi_{t,1}, \dots, \pi_{t,m}\}$  where  $\pi_{t,j} \geq 0$  and  $\sum_{j=1}^m \pi_{t,j} = 1$ . To implement these filters, we must be able to sample from possibly nonstandard densities. It is possible to develop simulation procedures to approximate the distribution of interest and to calculate certain statistics that characterize the distribution. We must be able to implement these procedures sequentially, as states evolve over time and new information becomes available. This implementation needs to be efficient and the approximations need to remain good as we move through the sequence of states.

From a sequential perspective, the main aim is to update the particles at  $t-1$ , and their respective weights,  $\{\alpha_{t-1,1}, \dots, \alpha_{t-1,m}\}$  and  $\{\pi_{t-1,1}, \dots, \pi_{t-1,m}\}$ . These are the particles and respective weights that approximate a given density, usually a continuous density function. In this context, the target density is often hard to sample from, so we must use an approximating density.

We can sample using this density and then resample as a way of approximating better the target density. This is the procedure associated with the sampling importance resampling (SIR) algorithm. However, Pitt and Shephard (1999, 2001) point out that using  $f(\alpha_t|\alpha_{t-1})$  as a density approximating  $f(\alpha_t|D_t)$  is not generally efficient because it constitutes a *blind* proposal that does not take into account the information contained in  $y_t$ . To improve efficiency, we include this information in the approximating density. The nonlinear/non-Gaussian component of the measurement equation then starts to play an important role, and certain algebraic manipulations need to be carried out in order to use standard approximations. This can be accomplished by sampling from a higher dimensional density. First an index  $k$  is sampled, which defines the particles at  $t - 1$  that are propagated to  $t$ , thus defining what the authors call an auxiliary particle filter. This corresponds to sampling from

$$f(\alpha_t, k|D_t) \propto f(y_t|\alpha_t)f(\alpha_t|\alpha_{t-1})\pi_k, \quad k = 1, \dots, m, \quad (5)$$

where  $\pi_k$  represents the weight given to each particle. We can sample first from  $f(k|D_t)$ , and then from  $f(\alpha_t|k, D_t)$ , obtaining the sample  $\{(\alpha_{t,j}, k_j); j = 1, \dots, m\}$ . The marginal density  $f(\alpha_t|D_t)$  is obtained by dropping the index  $k$ . If information contained in  $y_t$  is included, this resolves the problem of too many

states with negligible weight being carried forward, thereby improving numerical approximations. Now the target distribution becomes  $f(\alpha_t, k|D_t)$  and the information in  $y_t$  is carried forward by  $\pi_k$ . The next step is to define a density approximating  $f(\alpha_t, k|D_t)$ . One of the simplest approaches, described in Pitt and Shephard (1999, 2001), is to define

$$f(\alpha_t, k|D_t) \simeq g(\alpha_t, k|D_t) \propto f(y_t|\mu_{t,k})f(\alpha_t|\alpha_{t-1})\pi_k, \quad (6)$$

where  $\mu_{t,k}$  is the mean, mode or a highly probable value associated with  $f(\alpha_t|\alpha_{t-1})$ .

Outliers are commonly observed in financial series and for such datum the information in the prior is very different from that contained in the likelihood. This means that only very few particles used to approximate the filter density at  $t - 1$  are propagated to approximate the filter density at  $t$ . This gives rise to sample impoverishment.

Let  $g(\cdot|\cdot)$  represent any density approximating the target density  $f(\cdot|\cdot)$ . If the likelihood is log-concave, with a first order approximation, it can be easily ascertained that  $g(y_t|\alpha_t, \mu_{t,k}) \geq f(y_t|\alpha_t)$  for all values of  $\alpha_t$ , where  $g(y_t|\alpha_t, \mu_{t,k})$  constitutes the first order approximation of  $f(y_t|\alpha_t)$  around  $\mu_{t,k}$ . This means that with the approximating density, in this context we can define a perfect envelope for the target density and a rejection sam-

pler can be implemented (see, Pitt and Shephard 1999, 2001). But we can demonstrate here that this algorithm is not robust to extreme observations when the aim is to update the filter density within the model (1)–(2). We need a better approximation of the likelihood function to define the required approximating density.

The main modification considered is the definition of a second order, instead of a first order approximation, that is taken around a different point  $\alpha_t^*$  from that proposed by Pitt and Shephard (1999, 2001). The details of the algebra applied to the model used in this paper are given later, but in general we are defining a second order approximation around  $\alpha_t^*$  for the log-likelihood,  $\log f(y_t|\alpha_t) = l(\alpha_t)$ , and we designate it as

$$\log g(y_t|\alpha_t, \alpha_t^*) = l(\alpha_t^*) + l'(\alpha_t^*)(\alpha_t - \alpha_t^*) + \frac{1}{2}l''(\alpha_t^*)(\alpha_t - \alpha_t^*)^2. \quad (7)$$

Because we propose using a second order approximation, we cannot guarantee that the approximating density constitutes an envelope to the target density, as we can with the first order approximation, and we need to specify the algebra to implement the sampling importance resampling algorithm. This algebra depends on the point used to perform the Taylor expansion of the log-likelihood. Pitt and Shephard (1999, 2001) used  $\alpha_t^* = \phi\alpha_{t-1}$ , and suggested other possible points such as the mode

of the posterior distribution or a point between posterior and prior mode. Here, our main concern is to choose an expansion point to, as far as possible, avoid a given filter degenerating, as happens, for example, when a distribution with a continuous support is approximated by a single point.

A second order approximation defines a Gaussian approximating density and the variance of the approximating density is defined through the second derivative of the log-likelihood. When the likelihood is not log-concave, it is not always possible to define meaningful Gaussian approximations for all possible points considered. To overcome this problem, and also to obtain a less complicated algebra, we consider a second order approximation around the point that maximizes the log-likelihood. Assuming the regularity conditions that guarantee that the likelihood has a maximum, we designate by  $\alpha_t^*$  the point that maximizes the log-likelihood, and we have  $l'(\alpha_t^*) = 0$ . In this case, the equation (7) becomes

$$\log g(y_t | \alpha_t, \alpha_t^*) = l(\alpha_t^*) + \frac{1}{2} l''(\alpha_t^*) (\alpha_t - \alpha_t^*)^2, \quad (8)$$

which resembles the log-kernel of a Gaussian density with mean  $\alpha_t^*$  and variance  $-1/l''(\alpha_t^*)$ . When applied to the model in (1)–(2), we find that it simplifies the algebra considerably and, as we are assuming that the log-likelihood has a maximum, we have

$l''(\alpha_t^*) < 0$ , thus defining a meaningful Gaussian approximation. The component  $l(\alpha_t^*)$  is absorbed into the normalizing constant, and the remainder is the log-kernel of a Gaussian density, and then we need  $l''(\alpha_t^*) < 0$  because  $-1/l''(\alpha_t^*)$  defines the variance of the corresponding distribution.

In this paper, we claim that the particle filter algorithm implemented by Pitt and Shephard (1999, 2001), based on a first order approximation of the likelihood function, is not robust to extreme outliers. We demonstrate, however, that a second order approximation (suggested but not implemented by these authors), used in conjunction with an appropriate expansion point, gives meaningful results supported by straightforward algebra. The calculation of appropriate formulae is presented in the next section. The algebra for specifying the first order filter has already been calculated by Pitt and Shephard (1999, 2001), and the algebra for the second order filter using a generic point  $\mu_{t,k}$  is given in the Appendix.

### **3. APPROXIMATIONS BASED ON MAXIMUM LIKELIHOOD POINTS**

For likelihoods associated with extreme outliers and the usual classes of SV models, it can be shown theoretically that the ex-

pansion point  $\mu_{t,k} = \phi\alpha_{t-1,k}$ , suggested by Pitt and Shephard (1999, 2001) is not where we should expect the posterior density to centre its weight (Dawid 1973). For the class of stochastic volatility models the weight should be more closely centred around the maximum of the likelihood function. In a standard SV model, the calculation of this quantity is straightforward, and we find that

$$\alpha_t^* = \log\left(\frac{y_t^2}{\beta^2}\right). \quad (9)$$

Therefore, we propose using the Taylor series approximation defined in (7) with  $\alpha_t^*$  above. There are two main advantages to using this approximation. First, the algebra needed to implement the algorithm is greatly simplified. Second, this procedure can be extended to include the cases where the likelihood is no longer strictly log-concave. We will focus here on the first advantage. The algebra is simpler because we are combining the log-kernel of two Gaussian densities, one given by the transition density, and the other given by  $\frac{1}{2}l''(\alpha_t^*)(\alpha_t - \alpha_t^*)^2$ , which is the log-kernel of a Gaussian density with mean  $\alpha_t^*$  and variance  $-1/l''(\alpha_t^*) = 2$ .

In this setting, we have to take into account the first stage weights, which are the ones that define which particles approximating the filter density at  $t-1$  will be carried forward to define the filter density at  $t$ . Using the notation of the equation (5)

these are denoted by  $\pi_k$ , and carry the information contained in  $y_t$ . The second order approximation of the log-likelihood combined with the log-prior gives an approximating density

$$g(\alpha_t, k|D_t) \propto g(y_t|\alpha_t^*)g(\alpha_t|\alpha_{t-1,k}, \alpha_t^*), \quad (10)$$

where

$$g(y_t|\alpha_t^*) = \exp\left(-\frac{\alpha_t^*}{2}\left(1 + \frac{\alpha_t^*}{2}\right) + \frac{\mu_{t,k}^{*2}}{4} + \frac{\mu_{t,k}^{*2} - \mu_{t,k}^2}{2\sigma_\eta^2}\right) \quad (11)$$

and

$$g(\alpha_t|\alpha_{t-1,k}, \alpha_t^*) = N(\mu_{t,k}^*, \sigma_{t,k}^2), \quad (12)$$

where

$$\mu_{t,k}^* = \frac{2\mu_{t,k} + \sigma_\eta^2\alpha_t^*}{2 + \sigma_\eta^2} \quad (13)$$

and

$$\sigma_{t,k}^2 = \frac{2\sigma_\eta^2}{2 + \sigma_\eta^2}. \quad (14)$$

The particles at  $t - 1$  that are used to define the approximating density are sampled using first stage weights, now defined through equation (11), which depend on the information in  $y_t$  as  $\alpha_t^*$  depends on  $y_t$ . When sampling the index  $k$  from a distribution proportional to (11), the particle  $\alpha_{t-1,k}$  is chosen, and the density, assuming the role of prior density, assumes a Gaussian form with mean  $\mu_{t,k} = \phi\alpha_{t-1,k}$  and variance  $\sigma_\eta^2$ . This is combined with a Gaussian density with mean  $\alpha_t^*$  and variance  $-1/l''(\alpha_t^*) = 2$ . After the particles have been sampled,



they must be resampled in order to take into account the target density. They are resampled using the second stage weights

$$\log w_j = -\frac{\alpha_{t,j}}{2} - \frac{y_t^2}{2\beta^2 \exp(\alpha_{t,j})} + \frac{(\alpha_{t,j} - \alpha_t^*)^2}{4} \quad (15)$$

$$\pi_{t,j} = \frac{w_j}{\sum_{j=1}^m w_j}, \quad j = 1, \dots, m. \quad (16)$$

Following the resampling stage, an approximation of the target posterior distribution of the states at  $t$  is available, which is used as a prior distribution to update the states at  $t + 1$ , and the process continues recursively as new information arrives.

To summarize, the particles at  $t - 1$  propagated to update the distribution of the states at  $t$  are chosen randomly according to the weights defined by (11). These weights are influenced by the information contained in  $y_t$ . By conditioning on each particle chosen through the first stage weights, new particles are sampled. As these come from an approximating distribution, a second step is needed. The particles are resampled using the weights defined by (15)-(16). Our modification, outlined above, makes this second order auxiliary particle filter (APF) straightforward and quick to implement.

#### 4. A MODEL EXTENSION

It has long been known that there are better models than the

standard SV model for financial returns series. Two of the best known modifications of the SV model use Student- $t$  instead of Gaussian innovations and allow for the possibility of leverage effects. Consider then that the innovation process in (1) follows a scaled Student- $t$  distribution with  $v$  degrees of freedom,

$$\varepsilon_t \sim \sqrt{\frac{v-2}{v}} \mathbf{t}_v. \quad (17)$$

Second order approximations around the proposed point in the last section allow us to produce results analogous to (9)–(16) above, and the choice of the expansion point greatly simplifies calculations in this case. Now, the point used to perform the approximation is different and the second derivative of the log-likelihood function, used to define the variance of the approximating distribution, is also different.

The algorithms described above can be extended quite straightforwardly by defining the new approximation for the log-likelihood function. In this case, the log-likelihood can be written as

$$l(\alpha_t) \propto -\frac{v+1}{2} \log\left(1 + \frac{y_t^2}{\beta^2(v-2)\exp(\alpha_t)}\right) - \frac{\alpha_t}{2}. \quad (18)$$

The first derivative is given by

$$l'(\alpha_t) = \frac{vy_t^2 - (v-2)\beta^2 \exp(\alpha_t)}{2(v-2)\beta^2 \exp(\alpha_t) + 2y_t^2} \quad (19)$$

and the second derivative is

$$l''(\alpha_t) = -\frac{(v^2 - v - 2)\beta^2 \exp(\alpha_t)y_t^2}{2((v-2)\beta^2 \exp(\alpha_t) + y_t^2)^2} < 0. \quad (20)$$

Using these results, we can easily define the value of  $\alpha_t$  that maximizes the likelihood function. If we solve the equation  $l'(\alpha_t) = 0$ , we obtain

$$\alpha_t^* = \log\left(\frac{vy_t^2}{\beta^2(v-2)}\right), \quad (21)$$

which is a maximum, because, for  $v > 2$ , we have  $l''(\alpha_t) < 0$  for all values of  $\alpha_t$ . The other component that needs to be calculated is the value of the second derivative on the point which maximizes the log-likelihood. It is straightforward to show that

$$s_t^2 = -\frac{1}{l''(\alpha_t^*)} = 2 + \frac{2}{v}. \quad (22)$$

In this context, the equation (11) is reformulated and we obtain

$$g(y_t|\alpha_t^*) = \exp\left(-\frac{\alpha_t^*}{2} - \frac{\alpha_t^*}{2s_t^2} + \frac{\mu_{t,k}^{*2}}{2\sigma_{t,k}^2} - \frac{\mu_{t,k}^2}{2\sigma_\eta^2}\right), \quad (23)$$

where  $\alpha_t^*$  is given by (21) and  $\mu_{t,k}^*$  and  $\sigma_{t,k}^2$  are defined below. The analogue for the density defined in (12) is still a Gaussian density function, but now,  $\mu_{t,k}^*$  and  $\sigma_{t,k}^2$  are defined as

$$\mu_{t,k}^* = \frac{(2v+2)\mu_{t,k}^2 + v\sigma_\eta^2\alpha_t^*}{2v+2+v\sigma_\eta^2} \quad (24)$$

and

$$\sigma_{t,k}^2 = \frac{(2v+2)\sigma_\eta^2}{2v+2+v\sigma_\eta^2}. \quad (25)$$

After the approximating draws are obtained,  $\alpha_{t,j}$ , for  $j = 1, \dots, m$ ,

the resampling step is performed using the weights

$$\log w_j = -\frac{\alpha_{t,j}}{2} - \frac{v+1}{2} \log \left( 1 + \frac{y_t^2}{(v-2)\beta^2 \exp(\alpha_{t,j})} \right) \quad (26)$$

$$+ \frac{v(\alpha_{t,j} - \alpha_t^*)^2}{4v+4} \quad (27)$$

$$\pi_{t,j} = \frac{w_j}{\sum_{i=1}^m w_i}, \quad j = 1, \dots, m. \quad (28)$$

With the Student- $t$  extension, the implementation of the algorithm is still uncomplicated, because, as in the previous case, we have a likelihood that is log-concave and a maximum that can be calculated analytically.

## 5. AN EMPIRICAL DEMONSTRATION

To demonstrate the efficiency of our second order filter over its first order competitors, independent of estimation considerations, we will first analyze a set of simulated series using the correct filter, then compare filters when we know that they have been misspecified.

[TABLE 1: about here]

We simulated series from two different models, the standard SV model and the SV model with Student- $t$  innovations. We used the parameters obtained by Jaquier et al. (1994) as the

estimates of the parameters associated with IBM and TEXACO stocks. The parameters, translated to the parametrization used in this paper, are given in Table 1. We chose these two stocks because one seems to exhibit a lower degree of persistence (IBM) and the other a higher degree of persistence (TEXACO). We simulated 1000 observations for each stock using the standard SV model, and also performed simulations for the same parameters, but assuming that the innovations follow a scaled Student- $t$  distribution with 5 degrees of freedom. We then applied the filters to the four series using the correct model specification. We also ran simulations where the filter design for the Gaussian distribution is applied to the Student- $t$  and vice-versa. Of course, in this case, we obtained poor approximations when compared with the ones obtained using the correct filter although the main point is still relevant, using the first order approximation, we can get extreme sample impoverishment.

[FIGURE 1: about here]

[FIGURE 2: about here]

The series are depicted in Figure 1, and we can see that they present two of the essential characteristics commonly found in

any real data financial time series, i.e., volatility clustering and extreme observations, which can be modeled by a process that allow time varying variances with persistence and unconditional distribution with fat tails. The information contained in the extreme observations seen in these series and models cannot be dealt with effectively using standard first order approximations within an APF.

[FIGURE 3: about here]

We ran the filters, using first and second order approximations, for the two simulated series obtained through a standard SV model, and we found that for TEXACO, there was one filter distribution that could not be approximated properly, and 9 in the case of the IBM stock. This is due to the fact that these observations assume extreme values for which the first order filter cannot accommodate the information contained in them.

In Figures 2 and 3, we have depicted smoothing and then filtered state density approximations around two potentially problematic observations for the filters, the first to appear in each series, the 919<sup>th</sup> for TEXACO and the 24<sup>th</sup> for IBM. In the first row we present the smoothing densities of the states. These provide a rough benchmark for the true distribution of the state,

although they are obtained by conditioning in a different information set, so they simply give an indication of the type of distribution we might expect to see, albeit with some shift in the mean and variance. The approximating smoothing densities were obtained using Markov chain Monte Carlo simulation techniques. It is clear that when we apply the first order filter to update the information contained in the problematic observations, we obtain meaningless results. The density function in a given range, which we know from theoretical considerations must be continuous, is approximated by a small number of distinct particles, and in extreme cases, by a single particle. On the other hand, this problem clearly does not emerge when we apply a second order filter to approximate the densities. In the second row of Figures 2 and 3, using the first order approximation, the filter densities associated with the information in the problematic observations are obtained. To make the comparisons clearer, we also present the respective densities for the preceding and succeeding observations of those considered problematic. The densities based on the second order approximation can be compared with these and the smoothing densities, and in this case they exhibit a more sensible configuration.

When a model is misspecified, the forecasts obviously become worse. We ran the filter based on a standard SV model

for a data set obtained from a model with Student- $t$  innovations. As might be expected, the Student- $t$  distribution produces series with more extreme observations. If we use the filter assuming Gaussian innovations, due to model misspecification, we therefore obtain significantly worse results compared with the situation where we use the true model for all filters and their associated approximations.

However, the first order approximation gave far more degeneracy when the filter based on a model with Gaussian innovations was applied to a model in which the true innovations are Student- $t$ . When using the first order filter without the misspecification of the model, we had one and nine problematic observations for TEXACO and IBM respectively, when we apply the filter to the series with Student- $t$  innovations, we found 5 observations for TEXACO and 16 observations for IBM in which the filter could not update the information. In contrast, although the second order filter assuming Gaussian innovations gave worse results than one assuming the true Student- $t$  predictions, we did not experience the degenerate breakdown apparent for the first order filter.

Figure 4 compares the performance of the filters empirically by showing first and second order filter Gaussian approximations, and second order filter Student- $t$  approximations for two



typical extreme observations. Note that the errors associated with model misspecification are of a smaller order than those associated with the failure of the filter approximation. These results are entirely representative of other comparisons we made.

[FIGURE 4: about here]

To demonstrate even further the lack of robustness of a particle filter based on a first order approximation, we again used the series of IBM based on the standard SV model, and for the first extreme observation, the 24<sup>th</sup>, we approximated the distribution of the estimate to the mean of the state in the corresponding observation. The estimated mean is sometimes obtained using only a small number of particles, and as has been demonstrated, for example in Liu and Chen (1998) and Pitt and Shephard (1999), as we use less even weights the variance of our estimates tends to increase.

[TABLE 2: about here]

[FIGURE 5: about here]

The second order particle filter performs considerably better than the first order filter. In the simulation the filters were run 1000 times and the estimated mean of the filter distribution at observation 24 was recorded. Table 2 gives the descriptive statistics associated with the two approximating distributions, which are also depicted in Figure 5. The estimated means of the state mean do not differ much. However, there is a much greater uncertainty associated with the estimate yielded by the first order filter. Because of parameter and model uncertainty, we are usually more interested in the approximation of an entire density than just a simple statistic. Then, although we can sometimes obtain a sensible value for the estimated mean for the filter distribution of the state, most of the time, this estimate is based on a very small number of distinct values. For example, in the extreme case of sample impoverishment, the mean can be obtained using just a single particle. The estimate based on the first order filter can give very imprecise results. Using the results in Table 2, we can see that the coefficient of variation, in this simulation, is reduced by more than 90% when we use the second order instead of the first order filter, as we might have expected.

## 6. CONCLUSION

We have demonstrated that it is possible to develop APFs based on a second order Taylor series approximation, which, unlike their first order analogues, perform well for series with extreme observations, which are fairly common in financial time series. We are now developing analogous procedures for time series whose likelihood is not log-concave. Our preliminary results are encouraging and will be presented in a future paper.

## ACKNOWLEDGMENTS

We are grateful to the Editor, Associate Editor and the anonymous referees for their valuable comments.

## APPENDIX: SECOND ORDER APPROXIMATION BASED ON THE PRIOR MEAN

Using the points used by Pitt and Shephard (1999,2001) to define the first order filter,  $\mu_{t,k} = \phi\alpha_{t-1,k}$ , we present the algebra associated with the second order filter. In this case, apart from

the formulas, the main difference is that only the SIR can be used. The approximation  $g(y_t|\alpha_t, \mu_{t,k})$  is defined through the second order Taylor approximation of  $\log f(y_t|\alpha_t)$  around  $\mu_{t,k}$ ,

$$\begin{aligned}\log g(y_t|\alpha_t, \mu_{t,k}) &\propto l(\mu_{t,k}) + l'(\mu_{t,k})(\alpha_t - \mu_{t,k}) \\ &+ \frac{1}{2}l''(\mu_{t,k})(\alpha_t - \mu_{t,k})^2 \\ &= \text{const} - \frac{\alpha_t}{2} - \frac{y_t^2 A_{t,k}}{2\beta^2 \exp(\mu_{t,k})},\end{aligned}$$

where

$$A_{t,k} = (\alpha_t - \mu_{t,k}) + \frac{(\alpha_t - \mu_{t,k})^2}{2} - 1.$$

Using this second order approximation, the density  $g(\alpha_t, k|D_t)$  is factorized in

$$g(\alpha_t|\alpha_{t-1,k}, y_t, \mu_{t,k}) = N(\mu_{t,k}^*, \sigma_{t,k}^2)$$

and

$$\begin{aligned}g(y_t|\mu_{t,k}) &= \exp\left(\frac{1}{2}\left(\frac{1}{\sigma_\eta^2} + \frac{y_t^2}{2\beta^2 \exp(\mu_{t,k})}\right)(\mu_{t,k}^* - \mu_{t,k})\right) \\ &\times \exp\left(-\frac{y_t^2(1 + \mu_{t,k})}{2\beta^2 \exp(\mu_{t,k})}\right)\end{aligned}$$

where

$$\mu_{t,k}^* = \left(\frac{1}{\sigma_\eta^2} + \frac{y_t^2}{2\beta^2 \exp(\mu_{t,k})}\right)^{-1} \left(\frac{y_t^2(1 + \mu_{t,k})}{2\beta^2 \exp(\mu_{t,k})} + \frac{\mu_{t,k}}{\sigma_\eta^2} - \frac{1}{2}\right)$$

and

$$\sigma_{t,k}^2 = \frac{2\beta^2 \sigma_\eta^2}{2\beta^2 + \exp(-\mu_{t,k}) \sigma_\eta^2 y_t^2}.$$

As we sample from  $g(\alpha_t, k|D_t)$ , an approximating sample, the elements in it must be resampled in order to obtain a sample that gives a better approximation of the target density  $f(\alpha_t, k|D_t)$ . The weights used in this resampling step are

$$\begin{aligned}\log w_j &= -\frac{y_t^2}{2\beta^2 \exp(\alpha_{t,j})} + \frac{y_t^2 B_{t,j,k}}{2\beta^2 \exp(\mu_{t,k})} \\ \pi_j &= \frac{w_j}{\sum_{i=1}^m w_i}, \quad j = 1, \dots, m\end{aligned}$$

where

$$B_{t,j,k} = \left( 1 - \alpha_{t,j} \left( 1 - \frac{\alpha_{t,j}}{2} + \mu_{t,k} \right) + \left( \mu_{t,k} + \frac{\mu_{t,k}^2}{2} \right) \right).$$

These are the so-called second stage weights that allow the modification of the approximating distribution towards the target distribution. Obviously, these weights must be more evenly distributed than those from the first order approximation, because the second order approximation allows a better approximation of the target distribution.

## REFERENCES

Ait-Sahalia, Y. (1998), “Dynamic equilibrium and volatility in financial asset markets,” *Journal of Econometrics*, 84, 93-127.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001), “The Distribution of Realized Stock Return volatility,” *Journal of Financial Economics*, 61, 43-76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), “The Distribution of Exchange Rate Volatility,” *Journal of the American Statistical Association*, 96, 42-55.

— (2003), “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 579-625.

Andersen, T. G., Bollerslev, T., and Lange, S. (1999), “Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast horizon,” *Journal of Empirical Finance*, 6, 457-477.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2004), “Analytic Evaluation of Volatility Forecasts,” *International Economic Review*, 45, 1079–1110.

Bollerslev, T., Engle, R. F., and Nelson, D. (1994), “ARCH models,” in *Handbook of Econometrics*, eds. Engle, R. and McFadden, D., Amsterdam: Elsevier, vol. 4.

Campbell, J. Y., Lo, A. W., and MacKinlay, C. (1997), *The Econometrics of Financial Markets*, New Jersey: Princeton.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999), "An Improved Filter for Non-linear Problems," *IEE Proc.-Radar, Sonar Navig.*, 146, 2-7.

Christoffersen, P. F. and Diebold, F. X. (2000), "How Relevant Is Volatility Forecasting for Financial Risk Management?" *The Review of Economic and Statistics*, 82, 12-22.

Dawid, A. P. (1973), "Posterior means for large observations," *Biometrika*, 60, 664-666.

de Freitas, J. F. G. (1999), "Bayesian Methods for Neural Networks," Ph.D. thesis, Engineering Department, University of Cambridge.

Diebold, F. X., Hickman, A., Inoue, A., and Schuermann, T. (1998), "Scale Models," *Risk*, 104-107.

Diebold, F. X. and Lopez, J. A. (1995), "Modeling Volatility Dynamics," in *Macroeconomics: Developments, Tensions, and*

*Prospects*, ed. Hoover, K., Boston: Kluwer Academic Press, pp. 427-472.

Doucet, A., Godsill, S., and Andrieu, C. (2000), “On Sequential Simulation-Based Methods for Bayesian Filtering,” *Statistics and Computing*, 10, 197-208.

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

Engle, R. F. (1995), *ARCH: Selected Readings*, Oxford: Oxford University Press.

Fearnhead, P. (1998), “Sequential Monte Carlo methods in filter theory,” Ph.D. thesis, University of Oxford.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993), “Nonlinear Dynamics Structures,” *Econometrica*, 61, 871-907.

Godsill, S. J., Doucet, A., and West, M. (2004), “Monte Carlo Smoothing for Non-Linear Time Series,” *Journal of the American Statistical Association*, 99, 156-168.



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

Hsieh, D. A. (1993), "Implications of Nonlinear Dynamics for Financial Risk Management," *Journal of Financial and Quantitative Analysis*, 28, 41-64.

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

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

Liu, J. S. (2001), *Monte Carlo Strategies in Scientific Computing*, New York : Springer.

Liu, J. S. and Chen, R. (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93, 1032-1044.

Pitt, M. K. and Shephard, N. (1999), “Filtering via Simulation: Auxiliary Particle Filters,” *Journal of the American Statistical Association*, 94, 590-599.

— (2001), “Auxiliary variable based particle filters,” in *Sequential Monte Carlo Methods in Practice*, eds. Doucet, A., de Freitas, J. F. G., and Gordon, N. J., New York: Springer, pp. 273-293.

Taylor, S. J. (1986), *Modelling Financial Time Series*, Chichester: Wiley.

— (1994), “Modelling Stochastic Volatility: A Review and Comparative Study,” *Mathematical Finance*, 4, 183-204.

Table 1: Simulation parameters

Parameters	IBM	TEXACO
$\beta$	2.9322	2.2371
$\phi$	0.83	0.95
$\sigma$	0.4	0.23

NOTE: This table gives the values of the parameters used to simulate the four series used in this section. These parameters are translated from Jaquier et al. (1994) as the authors used another parametrization.

Table 2: First vs second order filter comparison

Statistic	$\hat{E}(\alpha_{24} D_{24})_f$	$\hat{E}(\alpha_{24} D_{24})_s$
mean	0.7651	0.8077
stand. dev.	0.2477	0.0172
coef. var.	0.3238	0.0213
min	-0.6649	0.7626
max	1.9847	0.8908

NOTE: This table presents the summary of statistics from a simulation where the distribution of the mean estimates of the states associated with the 24<sup>th</sup> observations in the IBM stock are calculated. The estimates of the first order filter that is designated by the subscript  $f$  and the second order filter with the subscript  $s$  are compared.

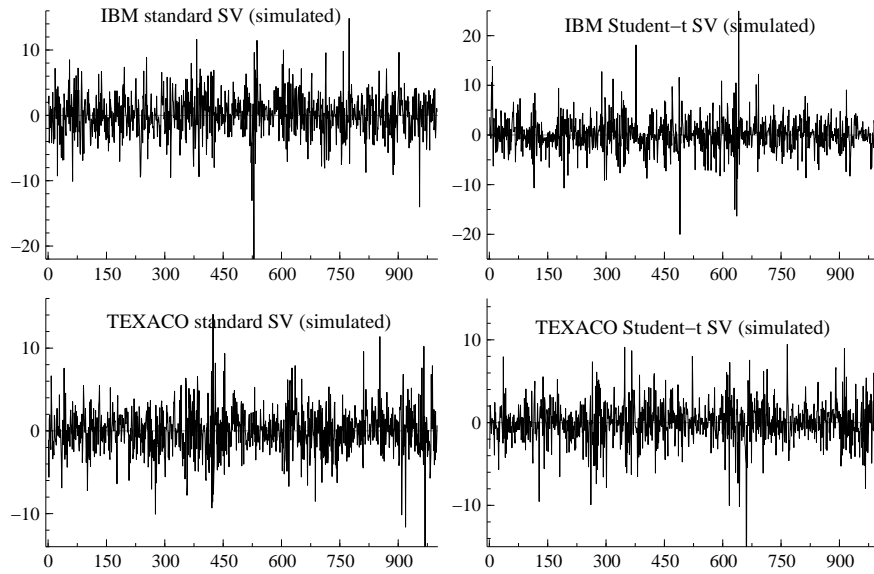


Figure 1: Simulated data using the parameters of Table 1. The left-hand side series are obtained assuming the model in (1)-(2). When we consider Student-t innovations another parameter is added, the parameter  $\nu$ , which represents the degrees of freedom. Here we use  $\nu = 5$ , and in this way we simulated the series on the right-hand side.

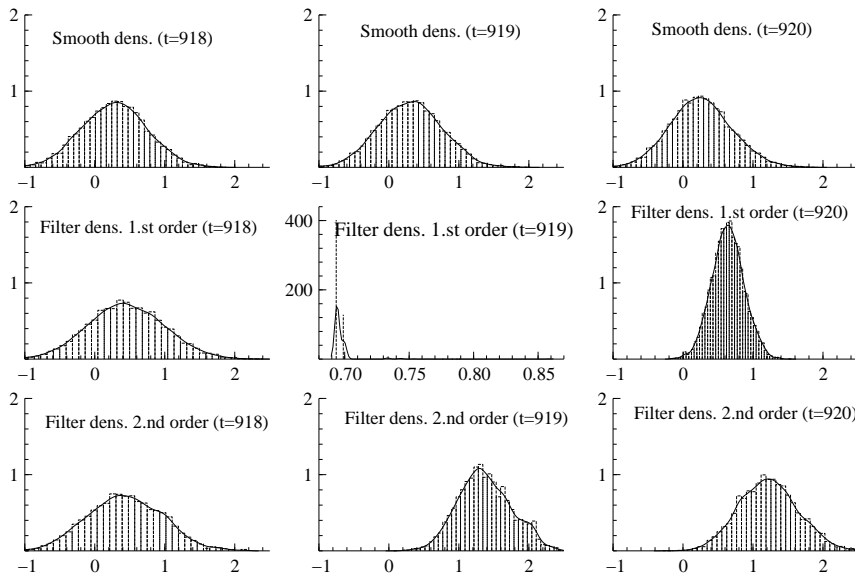


Figure 2: Density estimation comparison for TEXACO for reference observation 919. The first row shows the approximating smooth densities for this observation, as well for its predecessor and its successor. The second row gives the approximating filter densities obtained through a first order particle filter, and the third row gives the approximating filter densities using the second order particle filter.

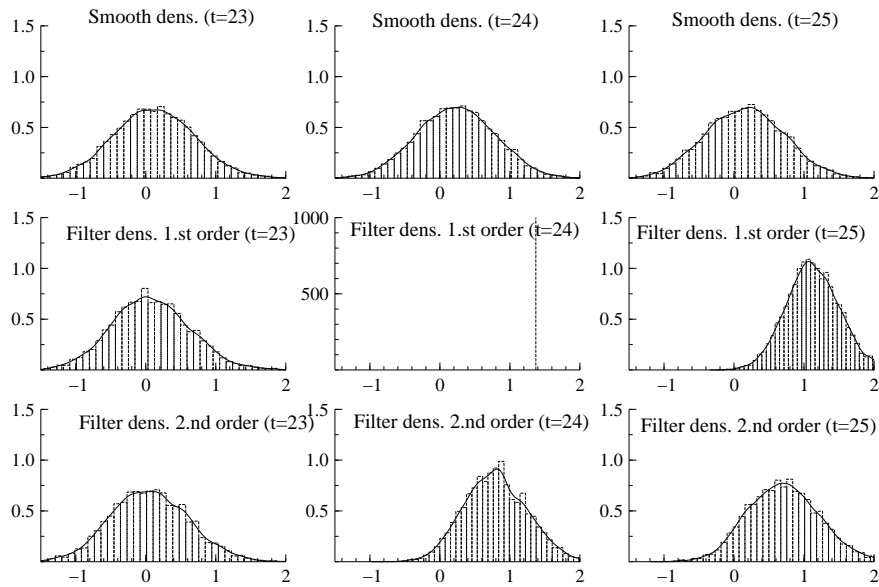


Figure 3: Density estimation comparison for IBM for reference observation 24. The first row are shows the approximating smooth densities for this observation, as well its predecessor and its successor. The second row presents the approximating filter densities obtained through a first order particle filter, and the third row gives the approximating filter densities using the second order particle filter.

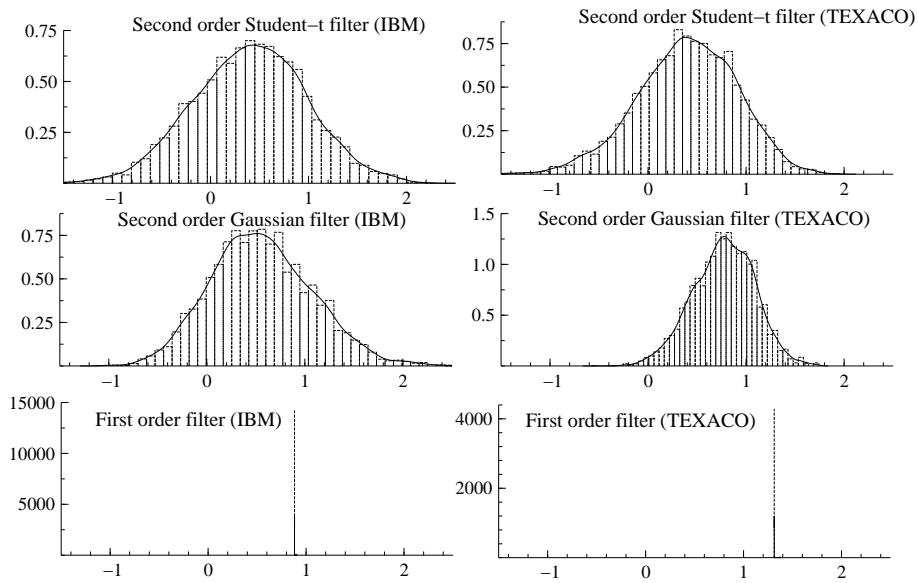


Figure 4: Filter densities associated with the distribution of the states at observation 79 for the IBM stock and 128 for TEXACO, simulated through an SV model with Student-t innovations. The first row represents the filter applied without model misspecification, whereas the second and third rows are from a filter assuming Gaussian innovations, the second row using a second order filter and the third a first order filter.



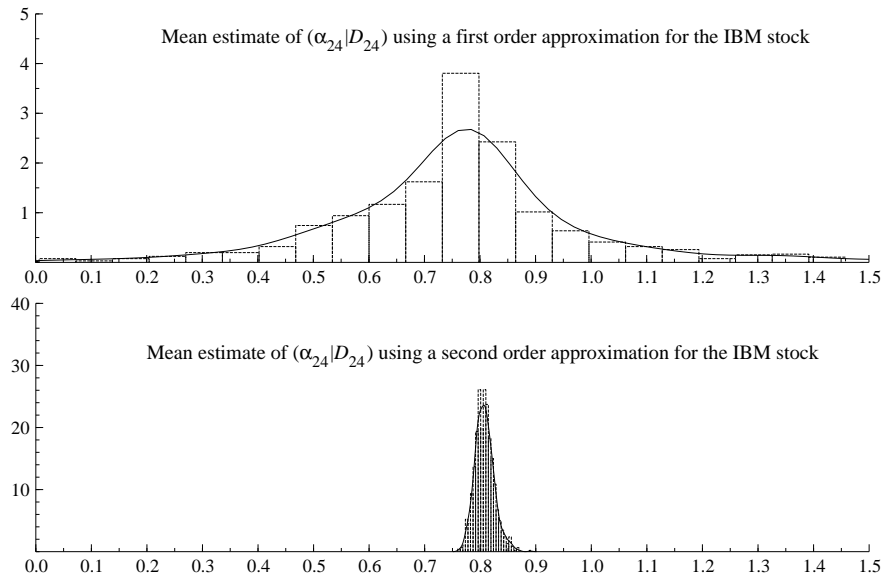


Figure 5: This Figure presents the approximating densities associated with the mean estimate of the states at observation 24 for the IBM stock, using the standard SV model. The first was obtained using the first order filter, while the second used the second order filter.