Volatility, Jumps and Predictability of Returns: a Sequential Analysis

Silvano Bordignon[1] and Davide Raggi[2*†]

[1] Department of Statistics
University of Padova
via Cesare Battisti 241-243, 35121 Padova - Italy
email: bordigno@stat.unipd.it

[2] Department of Economics
University of Bologna
Piazza Scaravilli, 2 40126 Bologna - Italy
email: davide.raggi@unibo.it

May 28, 2008

Abstract

In this paper we propose a sequential Monte Carlo algorithm to estimate a stochastic volatility model with leverage effects and non constant conditional mean and jumps. We are interested in estimating the time invariant parameters and the non-observable dynamics involved in the model. Our idea relies on the auxiliary particle filter algorithm mixed together with Markov Chain Monte Carlo (MCMC) methodology. Adding an MCMC step to the auxiliary particle filter prevents numerical degeneracies in the sequential algorithm and allows sequential evaluation of the fixed parameters and the latent processes. Empirical evaluation on simulated and real data is presented to assess the performance of the algorithm.

Keywords: Stochastic volatility with jumps, leverage, return's predictability, Bayesian estimation, auxiliary particle filters, MCMC

*Corresponding author

†Financial support from the MIUR under grant PRIN 2005 Prot. N. 2005132539 and Prot. N. 2002135473 are gratefully acknowledged. We wish to thank Stuart Coles, Andrea Torsello and the participants at the Workshop *New Directions in Monte Carlo Methods* held in Fleurance (France) on June 2007. All errors are our own.

1 Introduction

In this paper we propose a methodology to analyze the sequential parameter learning problem for a stochastic volatility model with jumps and a predictable component, i.e., the conditional mean. We aim at updating the estimates of the parameters of interest together with the states continuously, following the flow of information arriving in the markets. There are various reasons why we think sequential methods are appealing, both from a practical and a theoretical point of view. Sequential procedures seem suitable when we are interested in real time applications, where we need to update our estimates regularly. For example, economic agents need to produce estimates and forecasts in real time, meaning that we need to adapt our estimates every time a new observation is available. One of the most compelling advantages of sequential Monte Carlo methods is their reduced computational burden compared with other Monte Carlo procedures such as MCMC, which require that for each new observation we have to restart the inferential procedure from scratch.

Our procedure builds on the particle filtering algorithm of Liu and West (Liu & West 2001) in which we include an MCMC step to prevent the algorithm degenerating after a number of iterations. The use of MCMC together with particle filters has been proposed in Gilks & Berzuini (2001) and Berzuini & Gilks (2001) and has been proved to be an effective combination between the computational advantages of sequential algorithms and the statistical efficiency of MCMC methods. The introduction of the MCMC step is particularly useful when dealing with long time series, since it sensibly reduces the degeneration difficulties connected with sequential Monte Carlo methods.

We apply our methodology in a stochastic volatility context. Time varying conditional variance modelling represents an important topic for financial applications, and a large literature has grown up on describing financial time series using stochastic volatility models (see Ghysels et al. 1996 for a review). Furthermore, the introduction of a jump component has been proved to give an improved fit to data, both in relation to the model's ability to describe the return's behavior (Eraker et al. 2003), as well as for the pricing of financial derivatives (see Bakshi et al. 1997, Pan 2002 and Eraker 2004 amongst other).

Several variants of ARCH and SV models have been proposed so far to account for the empirical regularities of financial time series. In particular, in this paper we deal with three such regularities within a stochastic volatility framework. First, we consider the leverage effect between returns and conditional variances; second, we model the conditional mean, that is the predictable component of the returns; finally, we take into account a jump's dynamics to describe extreme and rare events such as crashes on the market. The leverage effect has been thoroughly investigated in the GARCH setting in Nelson (1991), whereas in a stochastic volatility framework this issue has been tackled in Yu (2005). This characteristic describes the relationship between returns and conditional variances. It is in fact reasonable to think that bad news in the markets, (e.g., the price decreases), leads to a boost on the variance, which is a measure of the financial risks. On the other hand, episodes of high volatility induce expectations of lower future returns, hence, the negative correlation between these shocks. Furthermore, Hull & White (1987) noted how financial leverage is also important for option pricing inference.

In financial applications there is substantial evidence of some predictability on the returns. This finding has been noticed since the early works of Merton (1971), that gave a theoretical justification for this behavior. In applications related to optimal portfolio choices, it is important to take into account this predictable component. In fact, economic theory shows that an investor gains from market predictability and volatility timing, even if the impact of these benefits is difficult to quantify. This is why it is interesting to explicitly model the conditional expected value of the returns together with the dynamics of the volatilities.

Finally, in the recent literature, there is also evidence in favor of jumps on returns and volatilities. In fact, a diffusive behavior of these two processes seems to be inadequate to describe the underlying dynamics (Eraker et al. 2003, Raggi 2005). Furthermore, if we consider the asset allocation problem in which the risky asset follows a jump diffusion process, there is some evidence that an extreme and rare event influences the conditional mean and the volatility, thus implying a modification on the optimal portfolio weights (Liu et al. 2003).

The remainder of the paper is organized as follows. The basic model is described in Section 2. Our inferential solution for that class of models is outlined in Section 3. Finally, some empirical results based on simulated and real data are illustrated in section 4.

2 The Model

A stochastic volatility model for the observable return process is usually specified as

$$y_{t+1} = \mu_t + \exp\{v_t/2\}\epsilon_{t+1} + \kappa_{t+1}J_{t+1} \quad (1)$$

$$v_{t+1} = \mu + \phi v_t + \sigma_\eta \eta_{t+1} \quad (2)$$

$$\mu_{t+1} = \alpha + \beta\mu_t + \sigma_\mu \zeta_{t+1}. \quad (3)$$

Returns are defined as $y_{t+1} = 100 \times (\log p_{t+1} - \log p_t)$, where p_t is the asset price. In this framework we assume that the error term ϵ_{t+1} is standardized Gaussian white noise. The conditional mean μ_{t+1} and the logarithm of the conditional variance or volatility v_{t+1} are described by two non observable processes. The autoregressive specification of the conditional variance is an approximation of the Euler discretization of the continuous time dynamics proposed in Hull & White (1987) and in Heston (1993). We assume that the initial state v_0 is distributed according to

$$N\left(\frac{\mu}{1-\phi}; \frac{\sigma_\eta^2}{1-\phi^2}\right),$$

which is the invariant law of the autoregressive model, identified by the first two marginal moments of the log-volatility process. The parameter ϕ is the persistence of the volatility that describes the volatility clustering. In empirical applications this parameter is close to 1 even though it is assumed that $|\phi| < 1$. This condition implies the stationarity of returns and volatilites. The parameter

μ is the drift component and σ_η can be interpreted as the volatility of the volatility. We assume that the error η_{t+1} is a Gaussian white noise. In order to describe the leverage effect, we assume $\text{Cov}(\epsilon_{t+1}, \eta_{t+1}) = \rho$. This parameter in general describes a negative relation between returns and risks even though, in some application such as in the analysis of exchange rates data, its estimate is usually close to zero.

In order to properly describe extreme events such as crashes in the markets, a useful extension is to introduce a jump component in the returns and in the volatilities. Duffie et al. (2000) for instance propose a model based on a stochastic differential equation with jumps driven by a marked point process. In the discrete time model, these discontinuities are governed by a sequence of independent Bernoulli random variables J_{t+1} with fixed intensity[1] λ. A Gaussian random variable κ_{t+1} with mean μ_y and variance σ_y^2 describes the size or mark associated to each jump.

We also directly model the conditional mean via an unobservable autoregressive process μ_{t+1}. Chernov et al. (2003) suggest that some serial dependence on μ_{t+1} can be motivated by the effect of non-synchronous trading and unexpected stochastic dividends. This dependence is assumed to be mean reverting. Similar dynamics for the conditional mean have been studied recently in Johannes et al. (2002*b*). The conditional mean at time 0 is distributed as

$$\mu_0 \sim N\left(\frac{\alpha}{1-\beta}; \frac{\sigma_\mu^2}{1-\beta^2}\right).$$

In this paper we assume that the noise ζ_{t+1} is uncorrelated with ϵ_{t+1} and

η_{t+1} even if there are no theoretical reasons to impose this constraint.

We need also to define the prior distribution for the parameters vector $\boldsymbol{\theta}$. Our choice is consistent with Kim et al. (1998) and with Eraker et al. (2003). We thus hypothesize the following prior distributions: $\mu \sim \mathrm{N}(0;10)$, $\phi \sim \mathrm{Beta}(25;2)$, $\sigma_\eta^2 \sim \mathrm{IG}(2.5, 0.05)$, $\rho \sim \mathrm{U}_{(0,1)}$, $\alpha \sim \mathrm{N}(0;4)$, $\beta \sim \mathrm{Beta}(25;2)$, $\sigma_\zeta^2 \sim \mathrm{IG}(2.5;0.05)$, $\lambda \sim \mathrm{Beta}(2;100)$, $\mu_y \sim \mathrm{N}(0;20)$, $\sigma_y^2 \sim \mathrm{IG}(2.5;0.05)$, where, in particular, IG denotes the inverse of a Gamma distribution.

3 Sequential Parameter and States Learning

Since their introduction, stochastic volatility models have been an interesting benchmark for many estimation techniques. Some of these rely on the Efficient Method of Moments of Gallant & Tauchen (1996), others on the Implied-State Generalized Method of Moments (IS-GMM) of Pan (2002). Estimation through Maximum Likelihood has been carried out in Aït-Sahalia (2002), by approximating analytically the transition density through Hermite polynomials. Recently, many simulation based methods have been implemented in order to approximate the likelihood. Simulated maximum likelihood methods have been proposed in Brandt & Santa-Clara (2002), Durham & Gallant (2002) and Koopman & Hol-Uspensky (2002) among others. Filtering techniques to evaluate the likelihood have been implemented in Johannes et al. (2002*a*) and in Pitt (2002).

In the recent literature, Monte Carlo algorithms have provided a flexible yet powerful tool for inference on complex models possibly with non observable components. MCMC methods have been introduced in Jacquier et al. (1994)

and in Kim et al. (1998). Applications to models with jumps have been developed in Chib et al. (2002) and in Eraker et al. (2003). Furthermore, MCMC methods for inference on continuous time models have been implemented in Eraker (2001) and in Elerian et al. (2001). MCMC methods provide efficient and accurate estimates when applied to off-line applications, but seem to be inadequate when dealing with real time applications where we need to update regularly our estimates at each time step.

Particle filter algorithms, introduced in Gordon et al. (1993), have been successfully used in a variety of fields such as engineering, econometrics and biology. They provide a sub-optimal but feasible solution to the Bayesian filtering problem. A detailed review on adaptive sequential algorithms is given in Liu & Chen (1998) and in Doucet et al. (2001), whereas an useful tutorial is Arulampalam et al. (2002).

We first describe the mechanics of these algorithms when the parameters are known. We then extend our solution to the parameter learning problem. Consider, for example, the general state-space model

$$y_{t+1} = h_m(\boldsymbol{x}_{t+1}, \epsilon_{t+1}) \tag{4}$$

$$\boldsymbol{x}_{t+1} = h_s(\boldsymbol{x}_t, \eta_{t+1}) \tag{5}$$

where (4) and (5) are respectively the measurement and the state equations. Here $\boldsymbol{x}_{t+1}$ is the so called state sequence, y_{t+1} is the observed process, $(\epsilon_{t+1}, \eta_{t+1})$ is a white noise and $h_s(\cdot)$ and $h_m(\cdot)$ are possibly nonlinear functions. Our goal is to estimate the distribution $p(\boldsymbol{x}_{t+1}|y_{1:t+1})$ given $p(\boldsymbol{x}_t|y_{1:t})$ in which

$y_{1:t} = (y_1, \ldots, y_t)$ is the past history of the observable process up to time t.

To implement the filter, we require the knowledge of the initial distribution $p(\boldsymbol{x}_0)$, of the transition distribution $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t),\ t \geq 0$ and of the measurement distribution $p(y_{t+1}|\boldsymbol{x}_{t+1}),\ t \geq 1$. The key idea is to approximate the filtering density $p(\boldsymbol{x}_{t+1}|y_{1:t+1})$ by a discrete cloud of points called particles $\boldsymbol{x}^j_{t+1},\ j = 1, \ldots N$, and a set of weights ω^j_{t+1} as follows

$$\hat{p}(\boldsymbol{x}_{t+1}|y_{1:t+1}) = \sum_{j=1}^{N} \omega^j_{t+1} \delta(\boldsymbol{x}_{t+1} - \boldsymbol{x}^j_{t+1}), \tag{6}$$

where $\delta(\cdot)$ is an indicator function. The cloud of points at time $t+1$ can be generated from a proposal distribution $q(\boldsymbol{x}_{t+1}|\boldsymbol{x}^i_t, y_t)$ and then weighted according to

$$\omega^i_{t+1} \propto \omega^i_t \frac{p(y_{t+1}|\boldsymbol{x}_{t+1}) p(\boldsymbol{x}^i_{t+1}|\boldsymbol{x}^i_t)}{q(\boldsymbol{x}^i_{t+1}|\boldsymbol{x}^i_t, y_t)} \qquad i = 1, \ldots N \tag{7}$$

With this setup, it can be proved that the variance of the weights increases systematically over t with the consequence that we eventually associate unit weight to one particle and zero to the others. For this reason a resampling step is added to this simple scheme in order to avoid numerical degeneracies by getting rid of the points with low probability.

An important variant of the basic filter is the auxiliary particle filter suggested by Pitt & Shephard (1999) in which the proposal depends on the whole stream of particles through an auxiliary variable J that is an index for the past trajectories (more details on this method are provided in Liu & Chen 1998 and in Godsill & Clapp 2001). In practice, the probability ω_{t+1} is corrected by an adjustment multiplier that should diversify the particles. In general this factor is taken to be dependent on a likely value of $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t^j)$ such as the mean or the mode. In many applications this extension helps to generate particles that are likely to be close to the filtering distribution.

Monte Carlo filtering techniques provide a viable and efficient solution to the filtering problem when the parameters are known. However, inference for the parameters is a challenging question. Recently a number of papers have tackled the problem of estimating the fixed parameters in a sequential context. For example Storvik (2002) proposes a filter in which the parameters are sequentially updated by simulating from their conditional distribution $p(\boldsymbol{\theta}|y_{1:t+1})$ through MCMC. A different approach, named the practical filter by Johannes et al. (2006), is based on the idea that $p(\boldsymbol{x}_{t+1}, \boldsymbol{\theta}|y_{1:t+1})$ can be expressed as a mixture of lag-filtering distributions. The estimate is then based on a rolling-window MCMC algorithm. In the context of stochastic volatility models, however, these methods seem to provide unstable results for some parameters[2]. Furthermore, a common practice is to artificially define an autoregressive dynamics for the parameters, say $\boldsymbol{\theta}_{t+1}$, and then include it in an augmented state vector $(\boldsymbol{x}_{t+1}, \boldsymbol{\theta}_{t+1})$ (see Gordon et al. 1993 and Kitagawa 1998 for example). The main point against this approach is that it leads to time varying and not to fixed parameter estimates. To correct for this artificial evolution, West (1993) and Liu & West (2001) propose to approximate the posterior distribution $p(\boldsymbol{\theta}|y_{1:t+1})$ by a smooth kernel density, leading to

$$p(\boldsymbol{\theta}|y_{1:t+1}) \approx \sum_{i=1}^{N} \omega_t^i N(\boldsymbol{m}_{t+1}^i; h^2 \boldsymbol{\Sigma}_{t+1}). \tag{8}$$

The quantity $\boldsymbol{m}_{t+1}^i = a\boldsymbol{\theta}_{t+1}^i + (1-a)\bar{\boldsymbol{\theta}}_{t+1}$ is the kernel location for the i-th component of the mixture whereas the matrix $\boldsymbol{\Sigma}_{t+1}$ and the vector $\bar{\boldsymbol{\theta}}_{t+1}$ are respectively estimates of the variance-covariance matrix and of the mean of the posterior distribution at time $t+1$. Furthermore, $\boldsymbol{\theta}_{t+1}^i, i = 1, \ldots, N$ is a sample from $p(\boldsymbol{\theta}|y_{1:t+1})$. The constants h and a, which measure the extent of the shrinkage and the degree of overdispersion of the mixture, are given by $h^2 = 1 - ((2\delta - 1)/2\delta)^2$ and $a = \sqrt{1-h^2}$, whereas the discount factor δ ranges between 0.95-0.99. It can be proved that the variance of the mixture approximation in (8) is $\boldsymbol{\Sigma}_{t+1}$ and the mean is obviously $\bar{\boldsymbol{\theta}}_{t+1}$. According to this setup, at time $t+1$, a reasonable proposal for the posterior is then

$$\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t \sim \mathrm{N}\left(a\boldsymbol{\theta}_t + (1-a)\bar{\boldsymbol{\theta}}_t, h^2\boldsymbol{\Sigma}_t\right). \tag{9}$$

This methodology has been successfully used in Liu & West (2001) in a dynamic factor stochastic volatility context and in Carvalho & Lopes (2006) in a switching regime stochastic volatility framework.

For the stochastic volatility model with jumps defined in eq (1)-(3) we found that the basic setup described above perform poorly. The major drawback with this algorithm is that the estimated posterior variance-covariance matrix $\boldsymbol{\Sigma}_{t+1}$ collapses to zero after a few hundred iterations. This problem is probably due to the sample impoverishment phenomenon caused by the resampling procedure and also by the discontinuous nature of the jump process. In fact, particles with high probability are selected many times causing a loss of diversity in the cloud of points. This effect is severe when the noise of the latent process is small[3]. A possible remedy is to choose an efficient resampling scheme that keeps low the

Monte Carlo variance. The *residual sampling* proposed in Liu & Chen (1995) is a useful alternative. Instead of resampling N particles with replacement, this strategy first takes $\lfloor N\omega_{t+1}^j \rfloor$ copies of $\boldsymbol{x}_{t+1}^j$ and then samples the remaining according to a probability proportional to $N\omega_{t+1}^j - \lfloor N\omega_{t+1}^j \rfloor$, where the symbol $\lfloor z \rfloor$ refers to the greatest integer less or equal to z. The procedure can be synthesized as follows

`Residual Sampling`

- Retain $k_j = \lfloor N\omega_{t+1}^j \rfloor$ copies of x_{t+1};
- Sample the remaining $N - \sum_{i=1}^N k_i$ with probability proportional to $N\omega_{t+1}^j - \lfloor N\omega_{t+1}^j \rfloor$;
- Reset the weights to $\frac{1}{N}$.

Another approach to increase the sample variability is to resort to MCMC moves. This should also help to reduce the correlation between particles after resampling. This idea has been recently developed in Gilks & Berzuini (2001) and in Berzuini & Gilks (2001). In practice, calling $\tilde{\boldsymbol{x}}_{t+1} = (\boldsymbol{x}_{t+1}, \boldsymbol{\theta})$, the particles $\tilde{\boldsymbol{x}}_{t+1}^i$ approximating $p(\boldsymbol{\theta}, \boldsymbol{x}_{t+1}|y_{1:t+1})$, can be moved to a different location $\tilde{\boldsymbol{x}}_{t+1}'^i$ according to a Markov transition kernel $T(\tilde{\boldsymbol{x}}_{t+1}, \tilde{\boldsymbol{x}}_{t+1}')$, that is invariant with respect to the same filtering distribution. For this reason, a burn-in period for the MCMC step is not necessary.

More formally, given the posterior distribution $p(\tilde{\boldsymbol{x}}_{t+1}|y_{1:t+1})$, the importance weights $\omega_{t+1}(\tilde{\boldsymbol{x}}_{t+1})$ and the proposal $q(\tilde{\boldsymbol{x}}_{t+1}|\tilde{\boldsymbol{x}}_{t+1}, y_t)$, it is easy to check that

$$p(\tilde{\boldsymbol{x}}_{t+1}|y_{1:t+1}) = \int \omega_{t+1}(\tilde{\boldsymbol{x}}_{t+1})q(\tilde{\boldsymbol{x}}_{t+1}|\tilde{\boldsymbol{x}}_t, y_t)T(\tilde{\boldsymbol{x}}_{t+1}, \tilde{\boldsymbol{x}}'_{t+1})\ d\tilde{\boldsymbol{x}}_{t+1}$$
$$= p(\tilde{\boldsymbol{x}}'_{t+1}|y_{1:t+1}). \quad (10)$$

In other words, we move all the particles $(\boldsymbol{x}^i_{t+1}, \boldsymbol{\theta}^i)$, that approximate the posterior, through $T(\cdot,\cdot)$ thus obtaining a further approximation of the filtering distribution based on the weighted sample $(\boldsymbol{\theta}'^i, \boldsymbol{x}'^i_{t+1}, \omega^i_{t+1})$.

Our proposal is to apply the MCMC correction to the parameter learning methodology proposed in Liu & West (2001). We now provide the details of the algorithm considering the version we implement for the model described in eq. (1)-(3). Using the notation introduced in Johannes et al. (2002a), we write the vector of the states as $\boldsymbol{x}_{t+1} = (v_t, \mu_t, J_{t+1}, \kappa_{t+1})$ and we estimate the posterior distribution $p(v_t, \mu_t, J_{t+1}, \kappa_{t+1}, \boldsymbol{\theta}|y_{1:t+1})$.
In order to perform the MCMC step we need to keep track of the whole trajectory of each particle. A useful way to store all of these information is through a set of sufficient statistics S_t (Fearnhead 2002). For our model, the sufficient statistics up to time t are

$$S_t = \Bigg(v_0, \sum_{i=1}^t v_i, \sum_{i=1}^t v_{i-1}, \sum_{i=1}^t v_{i-1}^2, \sum_{i=1}^t v_i^2, \sum_{i=1}^t v_i v_{i-1},$$
$$\sum_{i=1}^t a_i b_i, \sum_{i=1}^t a_i b_i v_{i-1}, \sum_{i=1}^t a_i b_i v_i, \sum_{i=1}^t a_i^2 b_i^2, \mu_0, \sum_{i=1}^t \mu_i,$$
$$\sum_{i=1}^t \mu_{i-1}, \sum_{i=1}^t \mu_i^2, \sum_{i=1}^t \mu_{i-1}^2, \sum_{i=1}^t J_i, \sum_{i=1}^t \kappa_i, \sum_{i=1}^t \kappa_i^2 \Bigg).$$

where $a_i = y_i - \mu_{i-1} - \kappa_i J_i$ and $b_i = \exp\{-v_{i-1}/2\}$. It can be noticed that the sufficient statistics may depend on v_t and μ_t that belong to $\boldsymbol{x}_{t+1}$. In this case we estimate these quantities by simulating them from their dynamics. The amount of computer memory required is, thus, sensibly reduced. The resulting algorithm is summarized as follows

`Parameter learning algorithm`

0. Simulate N particles from the prior $p(\boldsymbol{\theta})$, from $p(v_0)$ and from $p(\mu_0)$, $J_0 = 0$ and $\kappa_0 = 0$ with equal weights;

   For $t = 1$ to T:

1. Given $\boldsymbol{x}_t^j = (v_{t-1}^j, \mu_{t-1}^j, J_t^j, \kappa_t^j, \boldsymbol{\theta}_t^j)$ and ω_t^j, $j = 1, \ldots, N$, compute

$$\begin{aligned}\bar{v}_t^j &= E[v_t | v_{t-1}^j, \boldsymbol{\theta}_t^j] \\ \bar{\mu}_t^j &= E[\mu_t | \mu_{t-1}^j, \boldsymbol{\theta}_t^j] \\ \boldsymbol{m}_t^j &= a\boldsymbol{\theta}_t^j + (1-a)\bar{\boldsymbol{\theta}}_t \\ \bar{J}_{t+1}^j &= 0\end{aligned}$$

2. Draw an integer τ from $\tau \in \{1, \ldots, N\}$ using residual sampling with probabilities

$$g_{t+1}^j \propto \omega_t^j p(y_{t+1} | \bar{v}_t^j, \bar{\mu}_t^j, \bar{J}_{t+1}, \boldsymbol{m}_t^j)$$

3. Update $\boldsymbol{\theta}_{t+1}$ from $N(\boldsymbol{m}_t^\tau, h^2 \boldsymbol{\Sigma}_t)$

4. Update v_t from $p(v_t | v_{t-1}^\tau, \boldsymbol{\theta}_{t+1}^\tau)$

5. Update μ_t from $p(\mu_t | \mu_{t-1}^\tau, \boldsymbol{\theta}_{t+1}^\tau)$

6. Update J_{t+1} from $p(J_{t+1}|\boldsymbol{\theta}^{\tau}_{t+1})$

7. Update κ_{t+1} from $p(\kappa_{t+1}|\boldsymbol{\theta}^{\tau}_{t+1})$

8. Update the sufficient statistics according to the draws in step 3 to 7.

9. Compute $\omega^{\tau}_{t+1} \propto \frac{p(y_{t+1}|v^{\tau}_t,\mu^{\tau}_t\boldsymbol{\theta}^{\tau}_{t+1})}{p(y_{t+1}|\bar{\mu}^{\tau}_t,\bar{v}^{\tau}_t,\boldsymbol{m}^{\tau}_t)}$

10. Repeat step (2)-(9) N times. Record $\boldsymbol{x}^j_{t+1} = (v^j_t, \mu^j_t, J^j_{t+1}, \kappa^j_{t+1}, \boldsymbol{\theta}^j_{t+1})$.

11. (Optional) Move the former particles according to MCMC with invariant distribution equal to the posterior and update the sufficient statistics according to the former MCMC move.

We perform the MCMC step through a Gibbs sampler. In this way, we update the parameters $\boldsymbol{\theta}$ every 50 iteration of the algorithm, whereas J_{t+1} and κ_{t+1} are updated systematically. This choice provides a reasonable compromise between statistical precision and computational burden. It is also convenient to use some transformation of the parameters $\boldsymbol{\theta}$ in order to extend their support to the real line. In fact the posterior is approximated by a mixture of Normals, and then a convenient reparameterization of the model is in terms of parameters lying on the real line. This is important in order to perform *step 3* of the algorithm. We then consider the transformed parameter $\phi^* = \log\phi - \log(1-\phi)$ and $\beta^* = \log\beta - \log(1-\beta)$. We also define $\rho^* = \log(1+\rho) - \log(1-\rho)$. For the same reason we consider the logarithm of σ_η, σ_μ, σ_ζ and of the intensity λ.

4 Empirical Results

In this section we provide some illustrative examples to show the performance of the algorithm. More precisely we apply our parameter learning procedure to simulated and real data, i.e., daily Standard's & Poor 500 index returns and daily 3-months Treasury bill. All the calculations are based on software written using the Ox©3.2 language of Doornik (2001).

4.1 Simulated Data

We simulate a time series of length $T = 2000$ from the model described by equations (1)-(3). The true parameters, consistent with empirical findings on similar stochastic volatility models with jumps, are the following

- Volatility process: $\mu = 0.06$, $\phi = 0.95$, $\sigma_\eta = 0.15$, $\rho = -0.5$;
- Conditional mean: $\alpha = 0.001$, $\beta = 0.90$, $\sigma_\mu = 0.1$;
- Jump Process: $\lambda = 0.01$, $\mu_y = -4$, $\sigma_y = 2$.

We approximate the posterior distributions of interest through a cloud of 25,000 particles. Figure 1 reports the sequential learning process for the parameters, i.e., the evolution of the posterior mean together with the 2.5 and the 97.5 percent posterior quantiles.

Our algorithm provides accurate estimates for the parameters of the log-volatility process and, in fact, the posterior means of ϕ, σ_η and ρ quickly converge to their true values. In particular, the algorithm provides very precise estimates of the leverage ρ and of the persistence ϕ. It is also interesting to

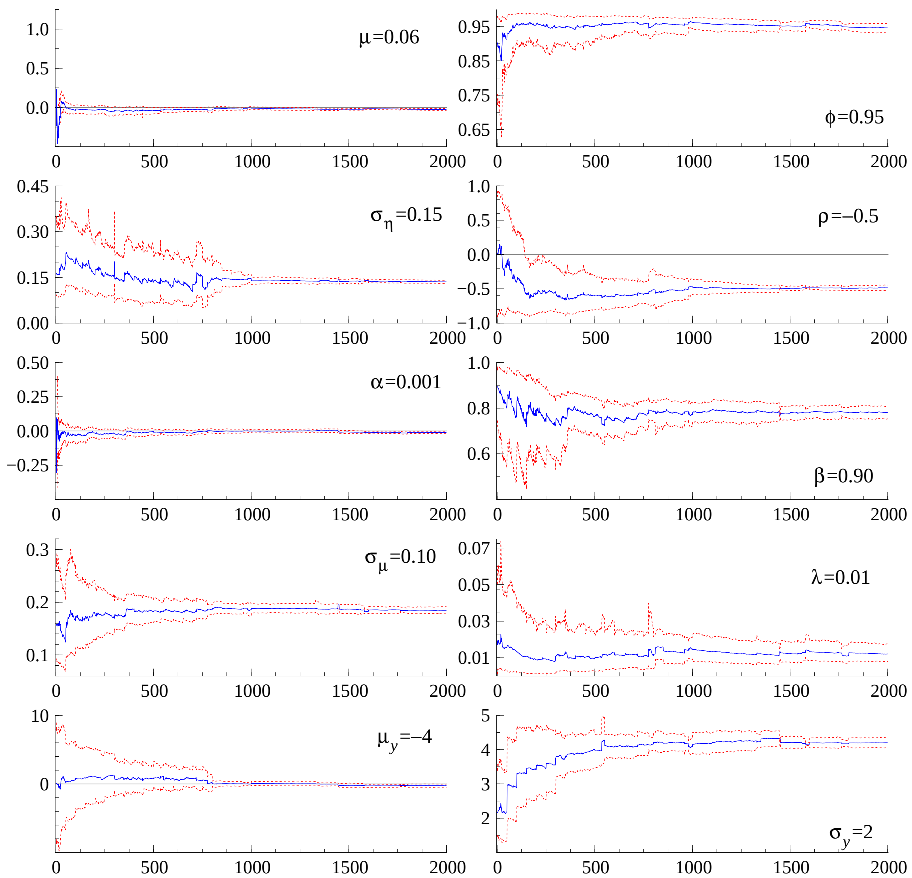

Figure 1: Estimated parameters together with the 2.5 and the 97.5 percent posterior quantiles

note the accuracy obtained for the volatility of volatilities parameter σ_η. This is surprising since, in the sequential literature, this parameter seems really sensitive to outliers, (see Johannes et al. 2006 for further comments on this point). The top panel of Figure 3 shows that the estimated log-volatility closely follows the true process.

More difficulties arise with the conditional mean parameters. Even though Figure 3 suggests that the true trajectory of μ_t is well approximated by its estimate, we find that the persistence parameter β is slightly under-estimated, while the estimate of σ_μ is slightly bigger than its true value. However, we note that these estimates are of a similar magnitude as the true values. We think that this effect can be reduced by introducing a non null correlation between y_{t+1} and μ_{t+1} in order to strengthen the bonds between the observable and the latent processes. This adjustment should make the observed data more informative for the conditional mean's parameters.

It is interesting to note that the algorithm detects the jumps accurately. This feature is displayed in Figure 2. In a few other cases we have noted an occasional inability of the algorithm to distinguish between outliers and actual jumps. This is especially evident when an extreme return is observed at the beginning of the series and when the jump size is small. However, Figure 2 suggests that the algorithm is very accurate in detecting expected size and timing. In some occasions difficulties arise when estimating the parameters related to the jump process, in which case some care has to be taken in the empirical analysis. The reason for these occasional pitfalls is most likely due to the rare nature of the

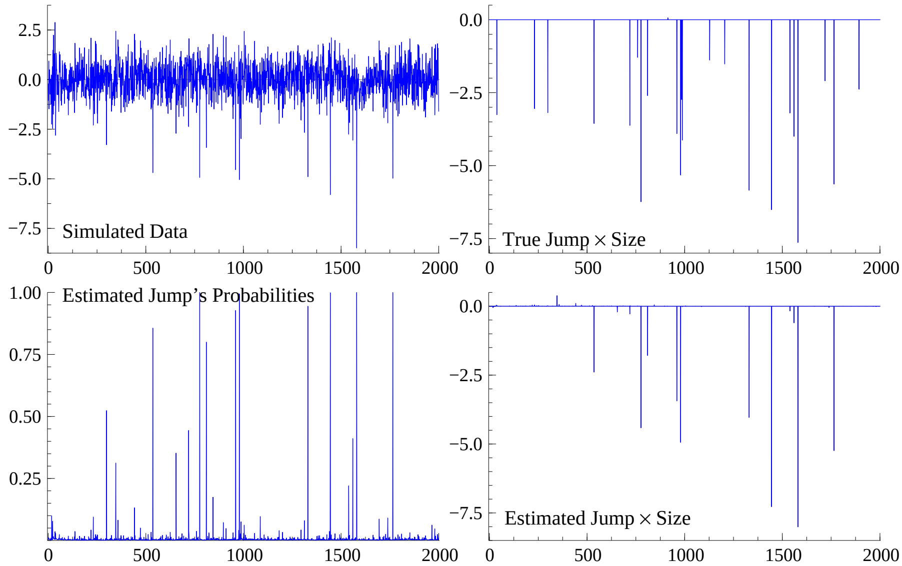

Figure 2: The simulated data are in the top left panel; in the bottom left panel the estimated probabilities of jumps; on the right the true and estimated impact of a jump event.

jumps. It is thus difficult to identify the parameters describing κ_t, i.e., μ_y and σ_y. Difficulties related to the lack of identification of jump models are however a common problem in this field and have also been noticed in Chib et al. (2002) and in Eraker et al. (2003). The algorithm, however, provides a precise estimate for λ.

As a final experiment, we consider a simulated time series with $T = 2500$, the same true parameters as before, but in which we add some positive jumps in order to check whether the algorithm is able to detect extreme observations with heterogeneous sizes. More precisely we add jumps of size $+5\%$ at $t = 1150$ and at $t = 2095$. The jump at $t = 1150$ corresponds to a positive jump in a period of quiet (no jumps immediately before that observation) whereas the

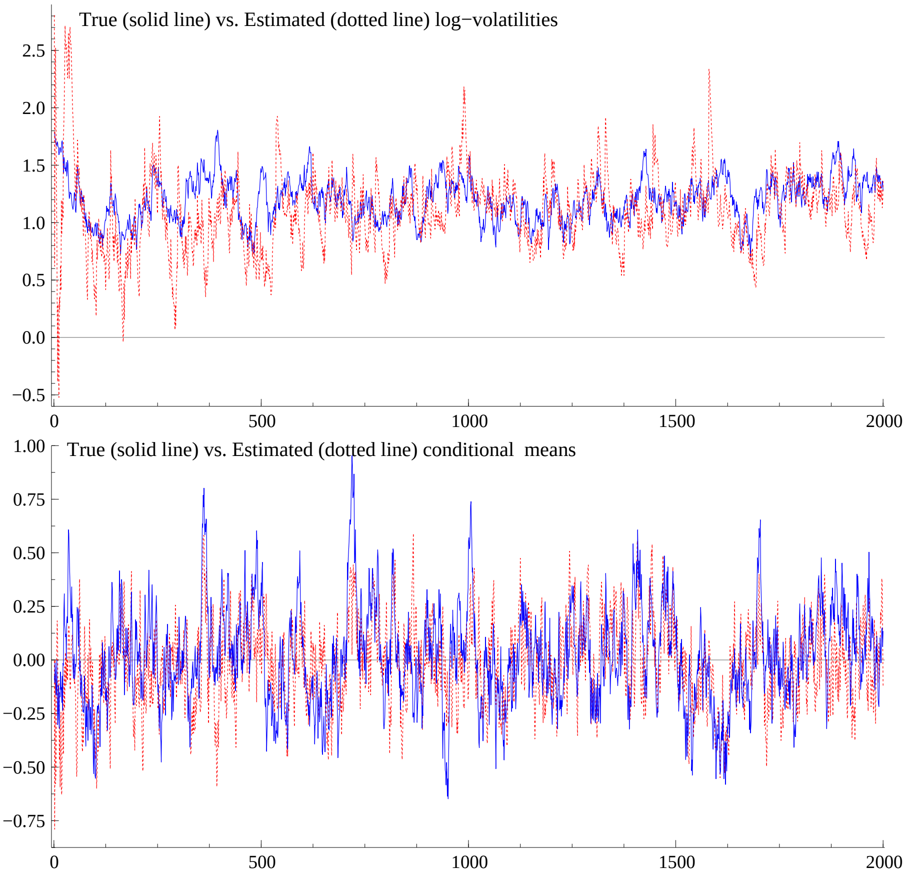

Figure 3: True vs. Estimated log-Volatilities (upper panel) and conditional means (lower panel)

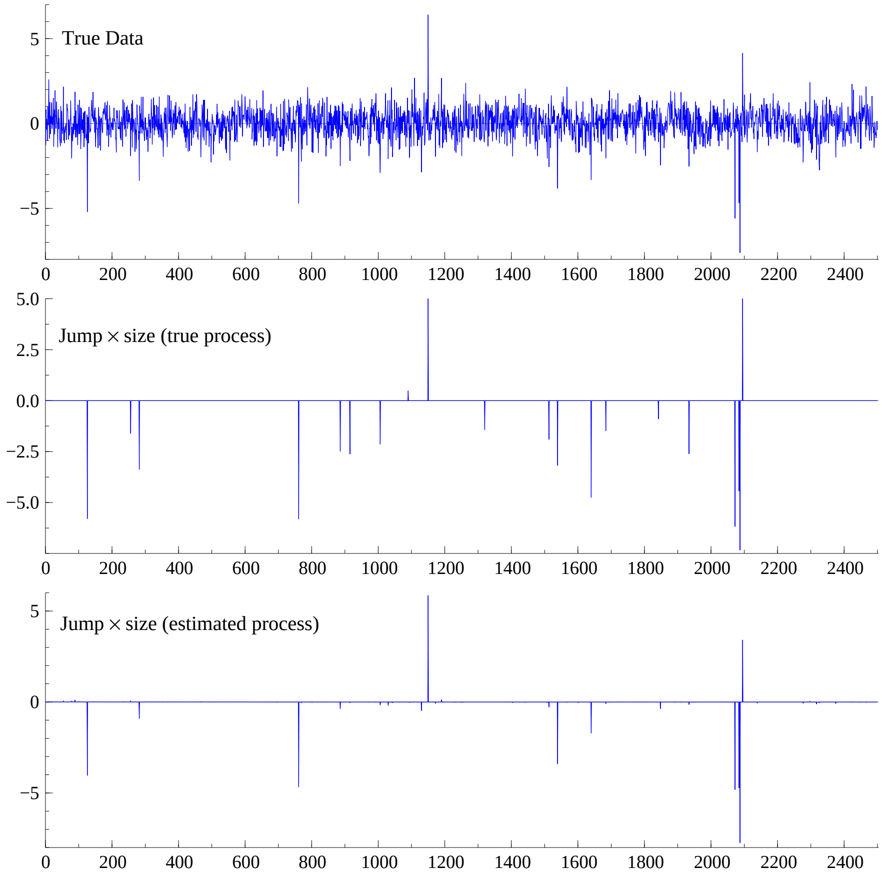

Figure 4: In the top panel we report the true data, whereas in the middle and in the bottom there are the true expected jump's impact and the estimated one.

second follows a sequence of negative jumps. The results are reported in Figure 4.

It is evident that the algorithm still detects all of the major jumps, including the two with positive size. For the first one we also obtain an accurate estimate of its expected size. We estimate the jump at $t = 1150$ with probability 1 and size 5.85%. We detect the second jump with a probability of about 96%

although the expected size is lower than the true value at around 3.56%. It is worth noting that the parameter estimates are in line with the results of the first Monte Carlo experiment.

4.2 S&P 500 Index

In this section we report some empirical results based on the S&P 500 index observed daily from January 1985 to July 2003. The data set has been downloaded from `Datastream`. As usual, the returns are defined as $y_{t+1} = 100 \times (\log p_{t+1} - \log p_t)$. We estimate the model by approximating the distributions of interest through 50,000 particles, though, halving this number leads to an analysis with similar results. The output is summarized in Figures from 5 to 8.

Figure 5 provides the plot of the observed time series together with the estimates of the latent processes. For the log-volatility and the conditional mean we also give 95% confidence bands. It is remarkable to note that associated with each spike on the original data set is an estimated high probability of jump. This is particularly evident for the crash observed during October 1987. Furthermore, it seems that other jumps observed in the last six years are properly estimated. Together with the jumps, it is easy to note that the log-volatility bursts every time a jump is detected, which is a reasonable feature since an extreme and negative event leads to a sudden and huge increase on the variability of the financial asset. The impact of the jump process on explaining the total marginal variance is about 20.4 percent. This estimate provides further evidence on the importance of the jumps to explain the variability of the returns.

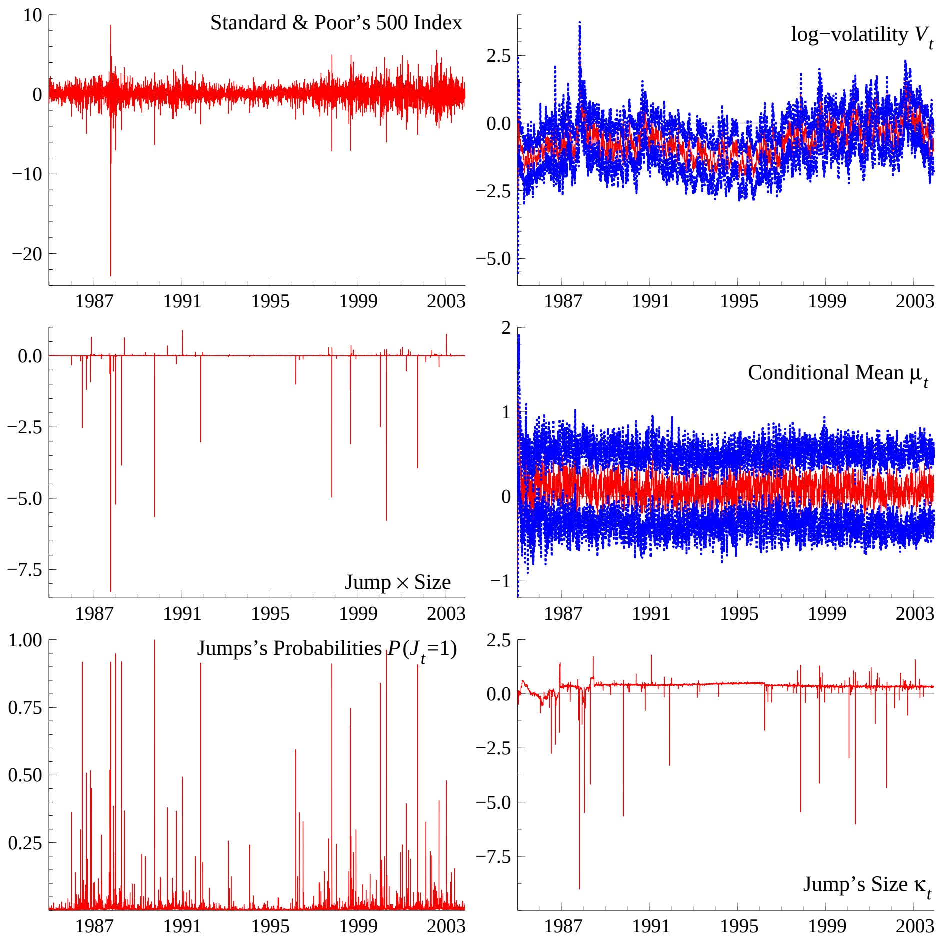

Figure 5: On the top left panel we report the original dataset. Filtered estimates of the unobservable processes are reported on the other panels. For volatility and conditional means, the figures display the 2.5% and the 97.5% confidence bands. For the jump times and sizes we report posterior means.

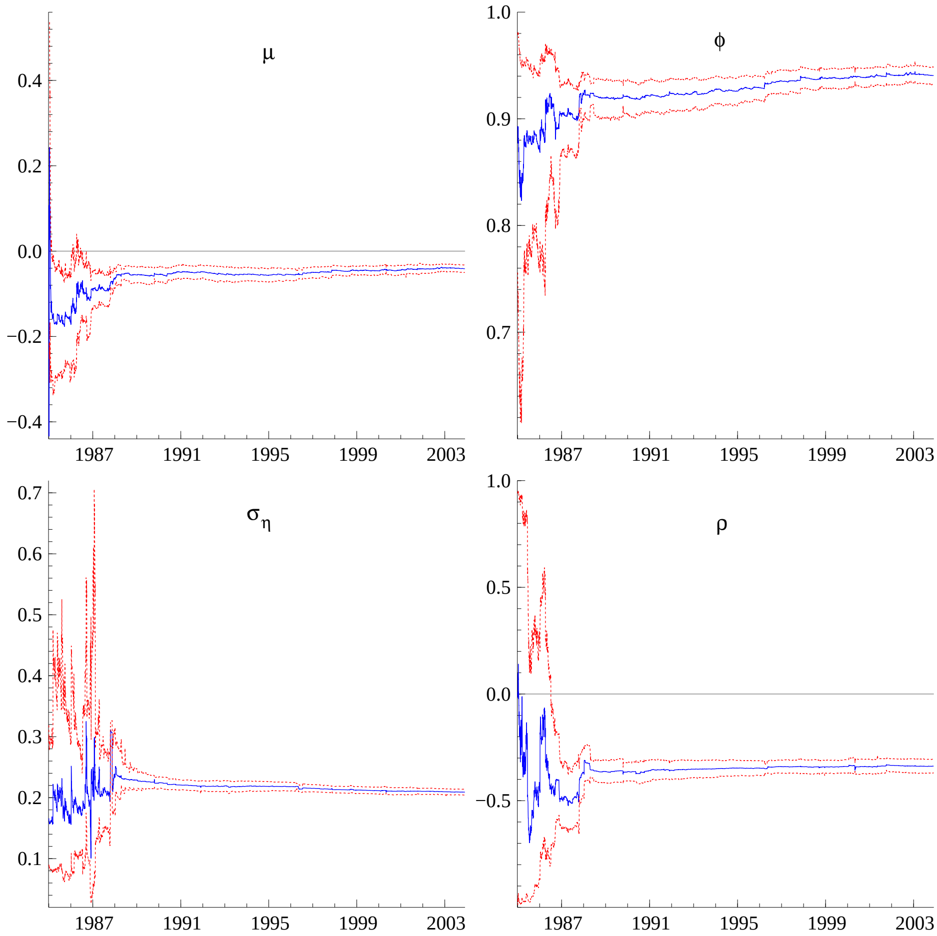

Figure 6: Estimated parameters of the volatility dynamic $\mu, \phi, \sigma_\eta, \rho$ (solid line) and the 95% confidence bounds (dotted line).

In Figure 6 we show the sequential evolution of the parameters involved in equation 2. The estimate of ρ is approximately -0.33 and confirms a marked leverage effect, since it is negative and substantially different from zero. The log-volatility process is persistent since ϕ is greater than 0.92. We found that ϕ tends to increase slightly in time, but this behavior can be explained by the rising volatility observed during the last four years. The parameter σ_η is approximately 0.21, which is slightly higher than the MCMC estimate obtained with the simpler stochastic volatility model with no jumps and no time varying conditional mean[4].

The analysis of μ_t provides evidence about the predictability of the returns. The intercept α is positive but close to zero and the persistence parameter β converges to 0.76. This high estimate of β clearly implies a non null autocorrelation of μ_t and suggests that the effect of a jump is persistent over time, thus influencing future returns. We think it is important to notice this feature, since in the current literature jumps are often taken to be independent with a transient impact on returns. This is one of the reasons why jumps are usually added to the volatility process.

Finally, the parameter estimates related to J_t and κ_t are plotted in Figure 8. The intensity λ suggests that the model detects about three extreme events per year. However, this estimate is about 6 times larger during the 1987 crisis. During that period, in fact, there are a number of small jumps close to the main one dated 19th of October. Concerning μ_y and σ_y, the expected size and the variability of κ_t, we obtain that $\mu_y \approx 0.33$ and $\sigma_y \approx 3.63$. This high value

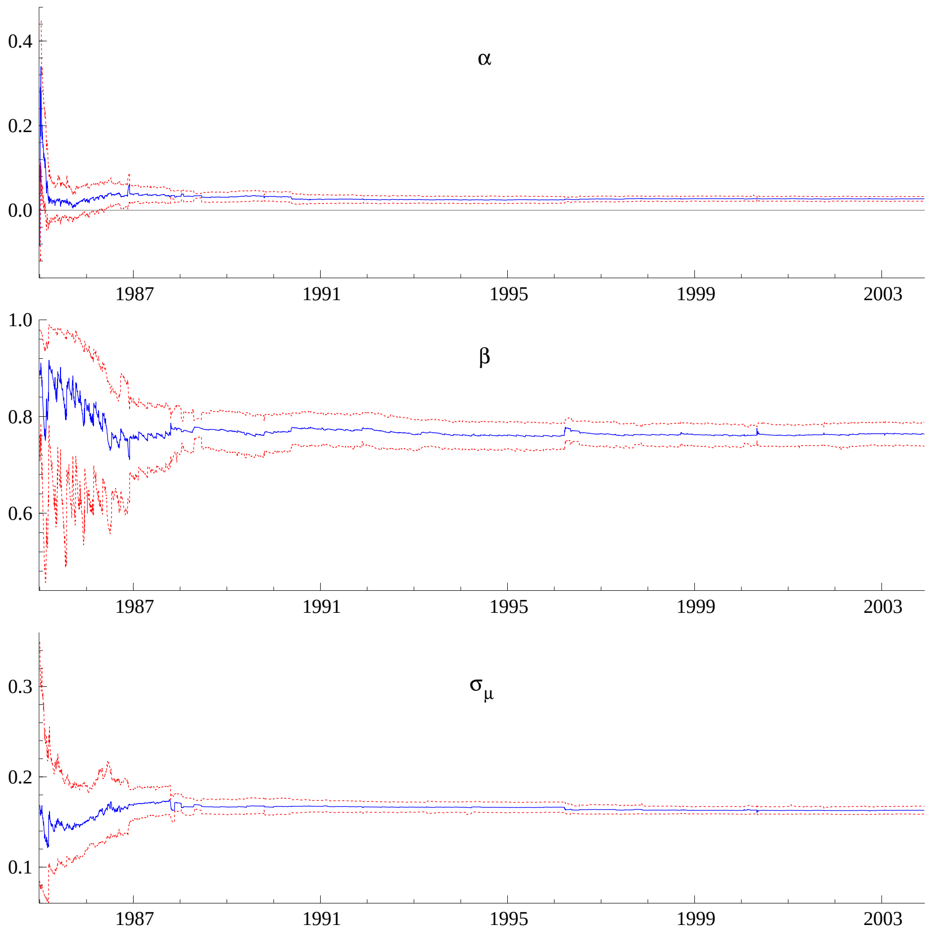

Figure 7: Estimated parameters of the conditional mean process, $\alpha, \beta, \sigma_\zeta$ (solid line) and the 95% confidence bounds (dotted line).

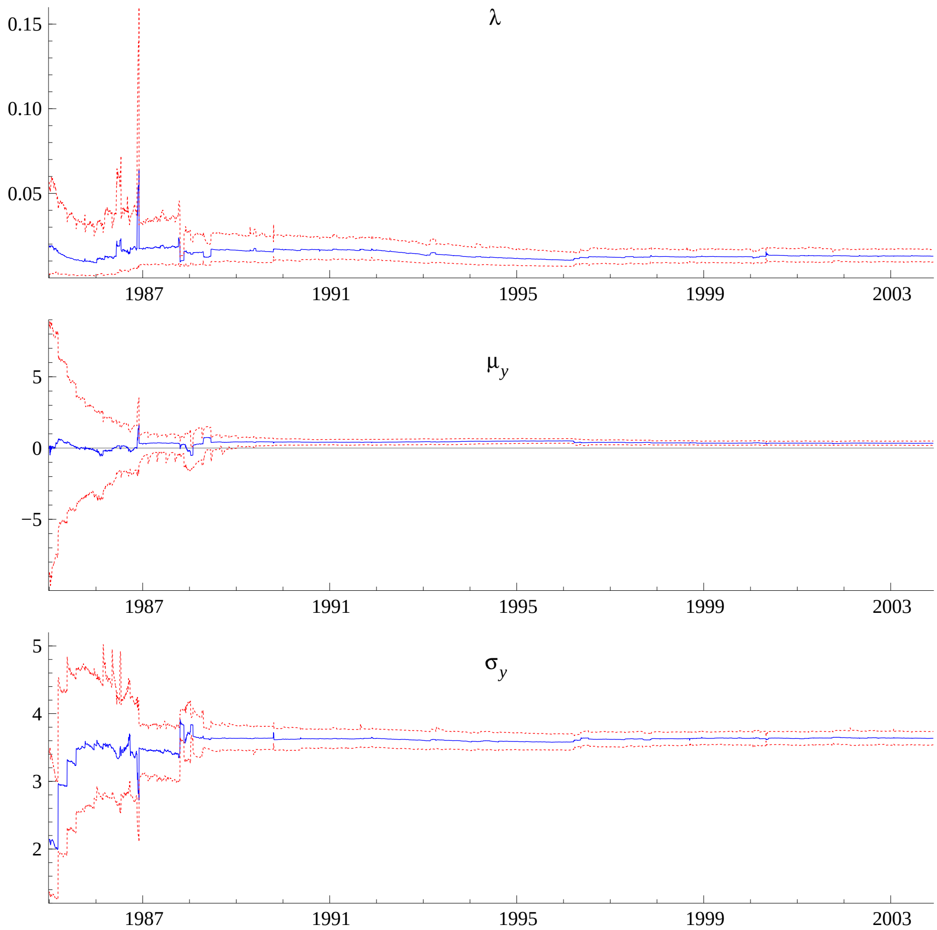

Figure 8: Estimated parameters of the jump and the size processes λ, μ_y, σ_y (solid line) and the 95% confidence bounds (dotted line).

of σ_y implies that the impact of jumps on the returns is heterogeneous. More precisely, it seems that the model accurately describes the timing of the jumps, but their effect is quite variable. The estimates reported, in fact, indicate that κ_t likely ranges between ± 7 percent.

This analysis suggests that the model can be generalized to allow for a time dependent intensity λ_t. On closer inspection, Figure 5 suggests that jumps arrive in clusters. For example, we estimate many jumps between 1986 and 1991, none in the subsequent five years and then several jumps again in the final period. It is also easy to note that jumps with high size are more frequent in periods with high volatility, thus suggesting that the intensity λ and the jump's size κ_t may be time varying and dependent on the volatility.

4.3 Short-term interest rates

We now apply the stochastic volatility model with jumps to short-term interest rates data. Recently, Johannes (2004) and Andersen et al. (2004) argued that the introduction of jumps on the interest rates dynamics should provide a better description of the statistical characteristic of the data and of the term structure of interest rates. Johannes (2004) also develops a test to detect the presence of the jumps dynamics based on the ability of a model that describes the kurtosis of the data. It is clear from that framework that pure diffusive models are unable to properly describe higher moments of the data.

From an economic point of view, Johannes (2004) suggests that large movements on interest rates are motivated by the need to describe the impact of some unexpected macroeconomic announcement. In fact, interest rates are not

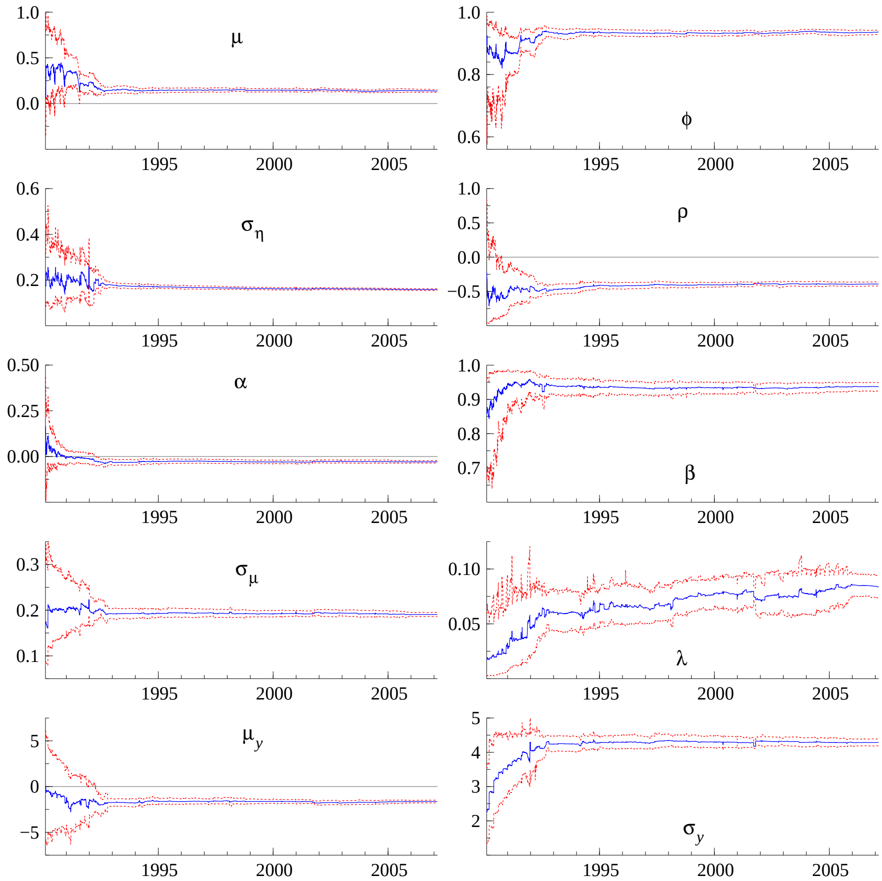

Figure 9: Interest rates data. Estimated parameters together with the 2.5 and 97.5 percent posterior quantiles

influenced by macroeconomic news, but rather by the surprise effect induced by the news themselves. We perform our analysis on the daily series of the 3-months Treasury bill (T-bill) r_t, from January 1990 to the 22nd of February 2007, downloaded from the H.15 release of the Federal Reserve System. In this analysis we consider the movements of the interest rates in basis point, that is, $y_t = 100 \times (r_{t+1} - r_t)$. The results are displayed in Figure 9 and in Figure 10.

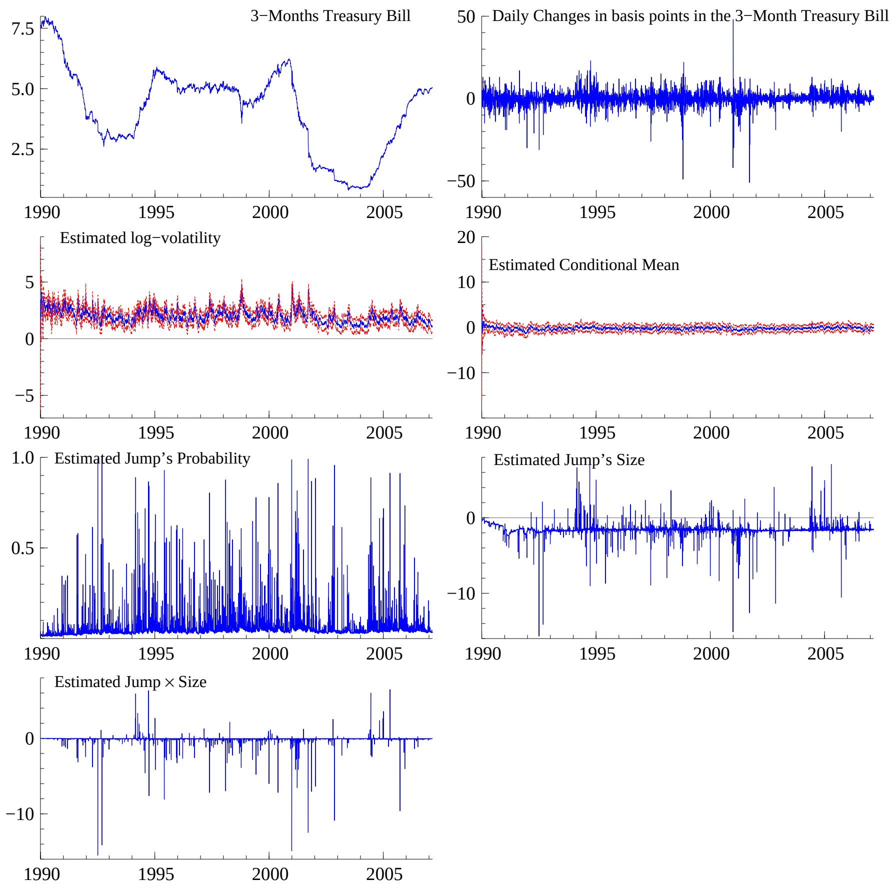

Figure 10: Interest rates data. Real data and estimated processes

There is strong evidence that the conditional means and conditional variances are persistent. In fact both the parameter estimates ϕ and β are greater than 0.93. It is interesting to compute the half-life of the two autoregressive processes, defined as the number of periods required for the impulse response to a unit shock to a time series to dissipate by half. In practice, if the persistence parameter is ϕ, the half-life is defined as $\frac{\log 0.5}{\log \phi}$. The half-life for the log-volatility process is about 10.57 whereas for the conditional mean it is 10.64. These quantities imply that it takes about two weeks for the two processes to absorb 50% of a shock.

In this application we find that ρ is significantly negative and is about -0.42. This is quite different from the results of Andersen et al. (2004) in which this parameter is set to 0. This difference is probably due to the different choice of the drift term of y_t. Similar findings have been reported in Raggi (2005) on a study of equity returns through affine models.

The parameter λ describes the intensity of the jump process. According to its estimate at time T we expect $0.08363 \times 250 \approx 21$ jumps per year. The expected size of the jumps is negative ($\mu_y \approx -1.63$) and σ_y is approximately 4.29. These estimates implies that a reasonable range for the jumps size lies between -10.70% and 6.8%.

The introduction of the jump factor is also useful on explaining the second moment of the interest rates process. We compute the ratio $\frac{\text{Var}[(J_t \kappa_t)]}{\text{Var}[y_t]}$ that expresses the percentage of the total variance due to jumps. In our analysis we find that jumps explain 15.89% of the total variance. This result is consistent with the findings reported Eraker et al. (2003) for their analysis on equity indexes.

5 Conclusions and Further Developments

Monte Carlo sequential methods represent a valuable and reliable methodology to estimate non linear and non gaussian state-space models. Their application also seem to be useful to the analysis of stochastic volatility models. In this paper we have proposed an algorithm based on the kernel smoothing approximation of the posterior suggested in Liu & West (2001) in which an MCMC step is incorporated in order to reduce sampling impoverishment problems related to sequential Monte Carlo strategies. Furthermore, in our empirical applications, we noticed that the algorithm also provides consistent and stable results with longer time series that are typical in financial econometrics.

An interesting economic issue to explore is to quantify how an extreme event has an impact on the optimal portfolio weights. In an affine jump diffusive framework (see Duffie et al. 2000 for a theoretical treatment for these models), Liu et al. (2003) prove that these optimal weights can be computed through the solution of an ordinary differential equation. We believe it would be interesting to estimate sequentially these quantities immediately before and after a crash, taking into account the parameters and states uncertainty related to the inferential procedure.

References

Aït-Sahalia, Y. (2002), 'Maximum Likelihood Estimation of Discretely Sampled Diffusions: a Closed-Form Approximation Approach', *Econometrica* **70**, 223–262.

Andersen, T. G., Benzoni, L. & Lund, J. (2004), Stochastic Volatility, Mean Drift, and Jumps in the Short-Term Interest Rate, Technical report. Working Paper.

Arulampalam, M. S., Maskell, S., Gordon, N. & Clapp, T. (2002), 'A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking', *IEEE Transactions on Signal Processing* **50**, 174–188.

Bakshi, G., Cao, C. & Chen, Z. (1997), 'Empirical Performance of Alternative Option Pricing Models', *Journal of Finance* **52**, 2003–2049.

Berzuini, C. & Gilks, W. (2001), RESAMPLE-MOVE Filtering with Cross Model Jump, *in* A. Doucet, N. de Freitas & N. Gordon, eds, 'Sequential Monte Carlo Methods in Practice', Springer, New York, pp. 117–138.

Brandt, M. & Santa-Clara, P. (2002), 'Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets', *Journal of Financial Economics* **63**, 161–210.

Carvalho, C. & Lopes, H. (2006), 'Simulation-Based Sequential Anlysis of Markov Switching Stochastic Volatility Models', *Computational Statistics & Data Analysis*.

Chernov, M., Gallant, A., Ghysels, E. & Tauchen, G. (2003), 'Alternative Models for Stock Price Dynamics', *Journal of Econometrics* **116**, 225–257.

Chib, S., Nardari, F. & Shephard, N. (2002), 'Markov Chain Monte Carlo Methods for Stochastic Volatility Models', *Journal of Econometrics* **108**, 281–316.

Doornik, J. (2001), *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

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

Duffie, D., Pan, J. & Singleton, K. (2000), 'Trasform Analysis and Asset Pricing for Affine Jump-Diffusions', *Econometrica* **68**, 1343–1376.

Durham, G. & Gallant, R. (2002), 'Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes (with comments)', *Journal of Business & Economic Statistics* **20**, 297–337.

Elerian, O., Chib, S. & Shephard, N. (2001), 'Likelihood Inference for Discretely Observed Non-Linear Diffusions', *Econometrica* **69**, 959–993.

Eraker, B. (2001), 'MCMC Analysis of Diffusion Processes with Application to Finance', *Journal of Business & Economic Statistics* **19**, 177–191.

Eraker, B. (2004), 'Do Equity Prices and Volatility Jumps? Reconciling Evidence from Spot and Option Prices', *Journal of Finance* **59**, 1367–1403.

Eraker, B., Johannes, M. & Polson, N. (2003), 'The Impact of Jumps in Volatility and Returns', *Journal of Finance* **58**, 1269–1300.

Fearnhead, P. (2002), 'Markov Chain Monte Carlo, Sufficient Statistics and Particle Filters', *Journal of Computational and Graphical Statistics* **11**, 848–862.

Gallant, A. & Tauchen, G. (1996), 'Which Moments To Match?', *Econometric Theory* **12**, 657–681.

Ghysels, E., Harvey, A. & Renault, E. (1996), Stochastic Volatility, *in* C. Rao & G. Maddala, eds, 'Statistical Methods in Finance', North Holland, Amsterdam, pp. 119–191.

Gilks, W. & Berzuini, C. (2001), 'Following a Moving Target - Monte Carlo Inference for Dynamic Bayesian Model', *Journal of the Royal Statistical Society, Ser. B* **63**, 127–146.

Godsill, S. & Clapp, T. (2001), Improvement Strategies for Monte Carlo Particle Filters, *in* A. Doucet, N. de Freitas & N. Gordon, eds, 'Sequential Monte Carlo Methods in Practice', Springer, New York, pp. 139–158.

Gordon, N., Salmond, D. & Smith, A. (1993), 'Novel Approach to Non-linear and Non-Gaussian Bayesian State Estimation', *IEE Proceedings-F* **140**, 107–113.

Heston, S. (1993), 'A Closed-Form Solution of Options with Stochastic Volatility with Applications to Bond and Currency Options', *Review of Financial Studies* **6**, 327–343.

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

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

Johannes, M. (2004), 'The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models', *Journal of Finance* **59**, 227–260.

Johannes, M., Polson, N. & Stroud, J. (2002*a*), Nonlinear Filtering of Stochastic Differential Equations with Jumps, Working paper, Graduate School of Business, University of Chicago.

Johannes, M., Polson, N. & Stroud, J. (2002*b*), Sequential Optimal Portfolio Performance: Market and Volatility Timing, Working paper, Columbia University.

Johannes, M., Polson, N. & Stroud, J. (2006), Sequential Parameter Estimation in Stochastic Volatility Models with Jumps, Working paper, Graduate School of Business, University of Chicago.

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

Kitagawa, G. (1998), 'Self Organizing State Space Models', *Journal of the American Statistical Association* **93**, 1203–1215.

Koopman, S. & Hol-Uspensky, E. (2002), 'The Stochastic Volatility in Mean model: Empirical evidence from international stock markets', *Journal of Applied Econometrics* **17**, 667–689.

Liu, J. & Chen, R. (1995), 'Blind Deconvolution via Sequential Imputations', *Journal of the American Statistical Associacion* **90**, 567–576.

Liu, J. & Chen, R. (1998), 'Sequential Monte Carlo Methods for Dynamic Systems', *Journal of the American Statistical Associacion* **93**, 1032–1998.

Liu, J. & West, M. (2001), Combined Parameter and State Estimation in Simulation-Based Filtering, *in* A. Doucet, N. de Freitas & N. Gordon, eds, 'Sequential Monte Carlo Methods in Practice', Springer-Verlag, New York, pp. 197–223.

Liu, J., Longstaff, F. & Pan, J. (2003), 'Dynamic Asset Allocation with Event Risk', *Journal of Finance* **63**, 231–259.

Merton, R. C. (1971), 'Optimum Consumption and Portfolio Rule in Continuous Time', *Journal of Economic Theory* **3**, 373–413.

Nelson, D. (1991), 'Conditional Heteroskedasticity in Asset Pricing: a New Approach', *Econometrica* **59**, 347–370.

Pan, J. (2002), 'The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study', *Journal of Financial Economics* **63**, 3–50.

Pitt, M. (2002), Smooth Particle Filters for Likelihood Evaluation and Maximization, Working paper, University of Warwick.

Pitt, M. K. & Shephard, N. (1999), 'Filtering via Simulation: Auxiliary Particle Filters', *Journal of the American Statistical Association* **94**, 590–599.

Raggi, D. (2005), 'Adaptive MCMC Methods for Inference on Affine Stochastic Volatility Models with Jumps', *Econometrics Journal* **8**, 235–250.

Storvik, G. (2002), 'Particle Filters for State-Space Models with the Presence of Unknown Static Parameters', *IEEE Transactions on Signal Processing* **50**, 281–289.

West, M. (1993), 'Approximating Posterior Distribution by Mixture', *Journal of the Royal Statistical Society, Ser. B* **55**, 409422.

Yu, J. (2005), 'On Leverage in a Stochastic Volatility Model', *Journal of Econometrics* **127**, 165–178.