

# Adaptive MCMC methods for inference on affine stochastic volatility models with jumps

DAVIDE RAGGI

*Dipartimento di Scienze Economiche, Università degli Studi di Verona, Verona, Italy*  
E-mail: raggi@stat.unipd.it

Received: May 2004

**Summary** In this paper we propose an efficient Markov chain Monte Carlo (MCMC) algorithm to estimate stochastic volatility models with jumps and affine structure. Our idea relies on the use of adaptive methods that aim at reducing the asymptotic variance of the estimates. We focus on the Delayed Rejection algorithm in order to find accurate proposals and to efficiently simulate the volatility path. Furthermore, Bayesian model selection is addressed through the use of reduced runs of the MCMC together with an auxiliary particle filter necessary to evaluate the likelihood function. An empirical application based on the study of the Dow Jones Composite 65 and of the FTSE 100 financial indexes is presented to study some empirical properties of the algorithm implemented.

**Keywords:** *Adaptive MCMC, Auxiliary particle filter, Bayes factor, Jump diffusions.*

## 1. INTRODUCTION

Stochastic volatility modelling represents an important topic for financial applications (Ghysels *et al.* 1996 for a general treatment). A generalization for these dynamics is represented by the introduction of jumps, as it seems natural to add this component to account for extreme and rare events that happen in the markets. Empirical evidence shows that the use of jumps and stochastic volatility leads to an improved fit of the data, both in relation to their descriptive power (Eraker *et al.* 2003) as well as for the pricing of derivatives (Bakshi *et al.* 1997; Pan, 2002; Eraker, 2004). A jump diffusion model for the state vector  $\mathbf{X}_t$  adapted to the filtration  $\{\mathcal{F}_t\}_{t \geq 0}$  can be described in general terms<sup>1</sup>

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t) dt + \boldsymbol{\Sigma}(\mathbf{X}_t) d\mathbf{W}_t + d\mathbf{J}_t, \quad (1)$$

where  $\mathbf{W}_t$  is a multi-dimensional Brownian motion, the vector  $\boldsymbol{\mu}(\mathbf{X}_t)$  is the drift,  $\boldsymbol{\Sigma}(\mathbf{X}_t)$  is a squared matrix and  $\mathbf{J}_t$  is a multi-variate marked point process with intensity  $\boldsymbol{\lambda}(\mathbf{X}_t)$ . A recent survey on jump diffusion models for financial applications can be found in Scott (1997) and in Runggaldier (2003).

<sup>1</sup>The vector  $\mathbf{X}_t$  includes the dynamics of the prices and of the volatilities for example.

The aim of this paper is to propose and implement an efficient Markov chain Monte Carlo (MCMC) procedure in order to estimate the affine jump diffusion models recently introduced in Duffie *et al.* (2000). We also aim at assessing some empirical properties of our algorithm by comparing it with other solutions proposed in the literature. Furthermore, we implement a sequential Monte Carlo strategy based on the framework introduced in Johannes *et al.* (2002) to evaluate the Bayes factor.

The remainder of the paper is organized as follows. The Duffie *et al.* (2000) framework is described in Section 3. Our inferential solution for that class of models is outlined in Section 3. Finally some empirical results based on the study of financial indexes are illustrated in Section 4.

## 2. AFFINE JUMP DIFFUSION MODELS

Affine jump diffusion models can be described by equation (1) in which the functions  $\mu(\mathbf{X}_t)$ ,  $\Sigma(\mathbf{X}_t)\Sigma'(\mathbf{X}_t)$  and  $\lambda(\mathbf{X}_t)$  are linear with respect to the state  $\mathbf{X}_t$ . Duffie *et al.* (2000) show that, in the affine jump diffusion framework, a general version of the Fourier transform for  $\mathbf{X}_t$  conditional on the  $\sigma$ -algebra  $\mathcal{F}_t$  is

$$\mathcal{F}(\mathbf{u}, \mathbf{X}_t, t, T) = E \left[ \exp \left( - \int_t^T R(\mathbf{X}_s) ds \right) e^{\mathbf{u}' \mathbf{X}_T} \middle| \mathcal{F}_t \right] = e^{\alpha(t) + \beta(t)' \mathbf{X}_t}, \tag{2}$$

where  $\alpha(t)$  and  $\beta(t)$  are solutions to suitable ordinary differential equations and  $\mathbf{u} \in \mathbb{C}^n$ . The process  $R(\mathbf{X}_s)$  is the discount factor that is also affine. In general,  $\alpha(t)$  and  $\beta(t)$  can be computed numerically even though in many cases it is possible to derive an analytical expression for them.

An interesting application of the affine jump diffusion theory is the following stochastic volatility model in which log-prices  $Y_t$  and volatilities  $V_t$  are described by

$$dY_t = \left( \mu - \frac{1}{2} V_t \right) dt + \sqrt{V_t} dW_{1,t} + H_t^y dN_t^y, \tag{3}$$

$$dV_t = \kappa(\theta - V_t) dt + \sigma_v \sqrt{V_t} dW_{2,t} + H_t^v dN_t^v, \tag{4}$$

where  $(W_{1,t}, W_{2,t})$  is a bivariate Brownian motion with correlated components. The correlation parameter  $\rho$  is often interpreted as the leverage effect and induces a relationship between volatilities and returns. For an extensive study of this topic see Yu (2005). The parameter  $\mu$  is the drift of the log prices. The parameters  $\kappa$ ,  $\theta$  and  $\sigma_v$  are, respectively, the mean reversion, the long run mean and the standard error of the volatility process. The jumps are modelled by marked point processes  $H_t^i dN_t^i, i = y, v$ . A marked point process is a random sum  $J_t = \sum_{j=1}^{N_t} H_j^i$  in which  $N_t$  represents the number of arrivals of a Poisson process in the interval  $(0, t]$  whereas  $H_j^i$  is the random size, or mark, corresponding to each arrival. In this paper the intensities  $\lambda_i$  are constant over time. In order to avoid problems with the negativity of the volatilities, common practice is to consider the logarithmic transform  $\log(V_t) = Z_t$ . A general version of the Itô's formula gives

$$dZ_t = \left[ \kappa(\theta e^{-Z_t} - 1) - \frac{1}{2} \sigma_v^2 e^{-Z_t} \right] dt + \sigma_v e^{-\frac{1}{2} Z_t} dW_{2,t} + \log(1 + H_t^v e^{Z_t-}) dN_t^v, \tag{5}$$

in which  $Z_{t-}$  is the left limit of a trajectory. The model in (3) and (4) can incorporate various types of jumps. As an example, in models with independent jumps both of the equations are allowed to jump but the two jump processes and their sizes are independent. The process  $N_t^v$  has intensity  $\lambda_v$

and jump size  $H_t^v \sim \exp(\mu_v)$  while  $N_t^y$  with intensity  $\lambda_y$  has jump size given by  $H_t^y \sim N(\mu_y, \sigma_y^2)$ . By imposing  $\lambda_v = 0$  the model proposed in Bates (1996) is defined. Alternatively, the model can exhibit contemporaneous jumps, i.e. there exists one jump component in the model and one intensity parameter, but the sizes are different and correlated. In this version  $H_t^v \sim \exp(\mu_v)$  and  $H_t^y | H_t^v \sim N(\mu_y + \rho_j H_t^v, \sigma_y^2)$ . Closed form solutions to the characteristic function are available for all the stochastic volatility models described above. In this way, the price of an European derivative can be computed through numerical integration methods.

### 3. INFERENCE

In this section we develop an efficient MCMC procedure for estimating the stochastic volatility dynamics introduced before. From a statistical point of view, inference for continuous time models is a challenging task for mainly two reasons. First, the trajectories of the processes are continuous. Second, there are some non-observable factors that affect the likelihood function. The main difficulty in deriving inferences for continuous time stochastic processes discretely sampled consists in the evaluation of the transition probability  $p(X_{t+1}|X_t)$ , which in general is not available in closed form. However, various computational intensive methods have been proposed to approximate it. Some of these rely on the Efficient Method of Moments of Gallant and Tauchen (1996), others on the Implied-State 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 methods based on simulation techniques have been implemented in order to approximate the likelihood. These procedures are based on adding some latent data in-between each pair of observations. These auxiliary points are then integrated out through Monte Carlo simulations. Simulated maximum likelihood methods have been proposed in Brandt and Santa-Clara (2002) and in Durham and Gallant (2002) among others whereas filtering techniques to evaluate the likelihood have been implemented in Johannes *et al.* (2002) and in Pitt (2002).

We focus on MCMC methods that have been introduced for discrete stochastic volatility models in Jacquier *et al.* (1994) and in Kim *et al.* (1998). An application to models with jumps has been implemented in Chib *et al.* (2002) and in Eraker *et al.* (2003). Furthermore, MCMC methods for inference on continuous time models have been proposed in Eraker (2001) and in Elerian *et al.* (2001).

The starting point of our inferential procedure involves an Euler discretization. The model is thus evaluated on a set of discrete times  $\{\tau_i : i = 1, \dots, N\}$  such that  $\tau_{i+1} - \tau_i = \Delta$ . Johannes *et al.* (2002) and Eraker *et al.* (2003) assert that in practice the use of daily intervals leads to a small bias with respect to the continuous model. In this paper, we got similar findings through some Monte Carlo experiments. Daily intervals imply that  $\Delta = \tau_{i+1} - \tau_i = (t + 1) - t = 1$ . We get

$$Y_{t+1} - Y_t = \left( \mu - \frac{1}{2} V_t \right) + \sqrt{V_t} \epsilon_{t+1}^y + H_{t+1}^y J_{t+1}^y, \quad (6)$$

$$V_{t+1} - V_t = \kappa(\theta - V_t) + \sigma_v \sqrt{V_t} \epsilon_{t+1}^v + H_{t+1}^v J_{t+1}^v, \quad (7)$$

in which  $(\epsilon_{t+1}^y, \epsilon_{t+1}^v)$  is a bivariate Normal with correlation  $\rho$  and  $J_{t+1}$  is Binomial  $Bi(1, \lambda)$ .

It is important to note that this approximation is appropriate if both sizes and intensities are independent of returns and volatilities. However, if we consider the model in terms of

log-volatilities as in equation (5), more general procedures are needed. We apply a version of the method proposed in Glasserman and Merener (2004) that gives

$$Y_{t+1} = \mu - \frac{1}{2}e^{Z_t} + e^{\frac{1}{2}Z_t}\epsilon_{t+1}^y + H_{t+1}^y J_{t+1}, \tag{8}$$

$$Z_{t+1^-} = Z_t + \kappa(\theta e^{-Z_t} - 1) - \frac{1}{2}\sigma_v^2 e^{-Z_t} + \sigma_v e^{-\frac{1}{2}Z_t}\epsilon_{t+1}^v, \tag{9}$$

$$Z_{t+1} = Z_{t+1^-} + \log(1 + H_{t+1}^v e^{Z_{t+1}^-}) J_{t+1}. \tag{10}$$

In this paper, equations (6) and (7) have been used to carry out MCMC inference. The approximation given in (8)–(10) is convenient to evaluate the likelihood function through the particle filter. To clarify the notation used for the rest of the paper, we define  $\mathbf{Y} = \{Y_1, \dots, Y_T\}$  as the complete vector,  $\mathbf{Y}_t = \{Y_1, \dots, Y_t\}$  indicates the observations up to time  $t$  whereas  $\mathbf{Y}_{-t} = \{Y_1, \dots, Y_{t-1}, Y_{t+1}, \dots, Y_T\}$  comprises all of  $\mathbf{Y}$  except  $Y_t$ .

Since volatilities and jumps are not observed, they are treated as missing data in the MCMC setting where the target distribution is  $\pi(\boldsymbol{\theta}, \mathbf{V}, \mathbf{J}, \mathbf{H}^y, \mathbf{H}^v | \mathbf{Y})$ . The vector  $\mathbf{V}$  represents the stochastic volatilities,  $\mathbf{J}$  is the jump process with jump sizes  $\mathbf{H}^y$  and  $\mathbf{H}^v$  while  $\boldsymbol{\theta}$  are the parameters. MCMC aims at estimating the distribution  $p(\boldsymbol{\theta} | \mathbf{Y})$  simply by recursively simulating the latent processes and the parameters from their full conditional distributions (see Robert and Casella, 1999 chap. 6–7). Estimates of the latent factors are given by averaging over the realization of the chain, i.e.  $\hat{V}_t = n^{-1} \sum_{j=1}^n V_t^j$  and  $\hat{P}_t(J_t = 1) = n^{-1} \sum_{j=1}^n J_t^j$ , respectively.

The priors for the parameters<sup>2</sup> are  $\mu \sim N(0, 25)$ ,  $\kappa \sim N(0, 1)$ ,  $\kappa\theta \sim N(0, 1)$ ,  $\sigma_v^2 \sim U_{(0,1)}$ ,  $\rho \sim U_{(-1,1)}$ ,  $\lambda \sim \text{Beta}(2, 40)$ ,  $\rho_j \sim N(0, 4)$ ,  $\mu_v \sim \text{IG}(20, 10)$ ,  $\sigma_y^2 \sim \text{IG}(5, 20)$ ,  $\mu_y \sim N(0, 100)$ . This choice leads to a closed form for the full conditionals for each but two parameters. The only exceptions are for  $\rho$  and  $\sigma_v^2$ . To update them we apply an adaptive Metropolis–Hastings algorithm (ARMS). It is easy to show that the full conditional distributions of the jump sizes  $H^y$  and  $H^v$  are Gaussian. Finally, the jump times  $J_t$  are a sequence of i.i.d. Bernoulli random variables. The conditional posterior is still Bernoulli with probability of success given by

$$P_c(J_t = 1) = \frac{P(J_t = 1 | \boldsymbol{\theta}, \mathbf{Y}, \mathbf{V}, \mathbf{H}^v, \mathbf{H}^y, \mathbf{J}_{-t})}{\sum_{j=0}^1 P(J_t = j | \boldsymbol{\theta}, \mathbf{Y}, \mathbf{V}, \mathbf{H}^v, \mathbf{H}^y, \mathbf{J}_{-t})}, \tag{11}$$

and in which  $P(J_t = j | \boldsymbol{\theta}, \mathbf{Y}, \mathbf{V}, \mathbf{H}^v, \mathbf{H}^y, \mathbf{J}_{-t})$ ,  $j = 0, 1$  is known up to a proportionality term. The main problem with this approach is to efficiently update the volatilities since it is not possible to simulate directly from their full conditional distributions. Solutions based on a basic Metropolis–Hastings algorithm seem to be not reliable and are also inefficient. For these reasons, we propose and investigate the behaviour of an adaptive procedure in order to select a precise proposal and to update the latent process.

### 3.1. Adaptive MCMC strategies

The aim of MCMC is to generate a Markov chain with given invariant distribution  $\pi(x)$  and estimating  $E_\pi[g(X)] \simeq \bar{g}_n$  by averaging over the trajectories obtained. A useful criterion to

<sup>2</sup> The symbol  $\text{IG}(\alpha, \beta)$  indicates an Inverse Gamma random variable.

evaluate the goodness of an estimator is based on its asymptotic variance

$$V(g, P) = \lim_{n \rightarrow \infty} n \text{var}_{\pi}[\bar{g}_n] = \text{var}_{\pi}[g(X_0)] + 2 \sum_{n=1}^{\infty} \text{cov}_{\pi}[g(X_0), X_n]. \quad (12)$$

Suppose that the realization of the chain is approximately an AR(1) process with autocorrelation  $\rho$ . Then the asymptotic variance of  $\bar{g}_n$  is  $N^{-1}\sigma^2(1 + \rho)/(1 - \rho)$ , where  $\sigma^2$  is the variance of  $g(X)$  under  $\pi(x)$ . A small  $\rho$  leads to an improved estimate in terms of its efficiency. A number of techniques, known as adaptive, have been proposed in literature to reduce  $\rho$  in the MCMC setting. One possible way to do adaptation consists in exploiting the entire history of the chain (see Gilks *et al.* 1998; Gåsemir 2003, for example). Unfortunately, these solutions relax the Markovian assumption and make it difficult to assess the conditions required for the convergence. A possible improvement to the random walk algorithm can be obtained by locally adapting the proposal densities to the target distribution. This can be done for example through the Metropolis Adjusted Langevin Approximation of Besag (1994). A different solution exploits the information given by previous rejections. An example is the Adaptive Rejection Metropolis Sampling (Gilks *et al.* 1995, ARMS) where the rejected candidates are used to build an accurate envelope function for the conditional distribution of interest. The main drawback in ARMS is that it is computationally demanding since the conditional distribution has to be evaluated at many points on the domain at each sweep of the algorithm. In the following we focus on the Delayed Rejection algorithm proposed in Tierney and Mira (1999).

### 3.2. Delayed rejection algorithm

A general notion of efficiency with respect to (12) is the well-known Peskun optimality (Peskun 1973; Tierney 1998). Given two Markov chains  $P$  and  $Q$  with the same invariant distribution  $\pi(\cdot)$ , we say that  $P$  dominates  $Q$  in the Peskun sense if

$$P(x, A) \geq Q(x, A), \quad x \notin A, \quad \forall A \in \mathcal{F}. \quad (13)$$

This condition states that the kernel generated by  $P$  dominates the transition defined by  $Q$  off the diagonal, i.e., the probability to stay at the same position at the next iteration is lower with  $P$  for each  $x$ . In the Metropolis–Hastings framework, this means that  $P$  allows less rejections than  $Q$  for every set  $A$  given the present state  $x$ . Following this intuition, Tierney and Mira (1999) propose the Delayed Rejection (DR) algorithm. The aim of the DR is to increase the rate of acceptance at each sweep of the algorithm.<sup>3</sup> In case we observe a rejection, the idea is to sample a new candidate for the chain from a different proposal, taking into account the information given by the previous rejections. In practice, if during the run of the chain a candidate  $y$  for  $X_t$  is rejected, a further Metropolis–Hastings step according to a new proposal is appended. To be more precise, the proposal at the  $i$ th step depends on the former rejections  $y_1, \dots, y_{i-1}$ , i.e.  $q_i(x, y_1, y_2, \dots, y_i)$ . It is worth pointing out that this method allows for a lot of flexibility, since there are no constraints for the choice of the proposal. Furthermore, the Markov property of the process is met by construction. In the general framework, Tierney and Mira (1999) and Mira (2001) prove that for

<sup>3</sup>Note that taking into account just the number of rejection at each sweep of the algorithm can be misleading. See Roberts *et al.* (1997) for a study on the optimal scaling.

each sub-step of the MH algorithm the acceptance probability  $\alpha_i(x, y_1, \dots, y_i)$  that guarantees the reversibility of the Markov chain is

$$\alpha_i(x, y_1, \dots, y_i) = \min \left\{ \frac{\pi(y_i)q_1(y_i, y_{i-1})q_2(y_i, y_{i-1}, y_{i-2}) \dots q_i(y_i, y_{i-1}, \dots, x)}{\pi(x)q_1(x, y_1)q_2(x, y_1, y_2) \dots q_i(x, y_1, y_i)} \times \frac{[1 - \alpha_1(y_i, y_{i-1})][1 - \alpha_2(y_i, y_{i-1}, y_{i-2})] \dots [1 - \alpha_{i-1}(y_i, \dots, y_1)]}{[1 - \alpha_1(x, y_1)][1 - \alpha_2(x, y_1, y_2)] \dots [1 - \alpha_{i-1}(x, y_1, y_{i-1})]}, 1 \right\}. \quad (14)$$

Unfortunately the probability  $\alpha_i$  is not increasing in  $i$ . For this reason a maximum number of rejections has to be defined. The algorithm can be summarized as follows.

*Delayed Rejection algorithm*

1. If  $y \sim p_1(x, y)$  is rejected in MH.
2. For  $i = 2, \dots, k$ .
3. Sample  $y_i$  from  $q_i(x, y, y_1, \dots, y_i)$ .
4. Accept  $y_i$  with probability  $\alpha_i(x, y_1, \dots, y_i)$  given by (14).
5. If rejection and  $i < k$ , then  $i = i + 1$  otherwise  $X_{i+1} = y_i$ .
6. End for.
7.  $X_{i+1} = y_i$  or  $X_{i+1} = X_i$ .

The convergence of the chain is guaranteed by checking the sufficient conditions at each Metropolis–Hastings sub-step. These conditions basically state that the proposals have to be strictly positive for each  $x$  in which  $\pi(x) > 0$ . Some of these conditions are stated in Tierney (1994) and in Mengersen and Tweedie (1996).

It is also interesting to assess in some way the speed of convergence of the algorithm. Up to now, it is not clear whether the DR provides an improvement with respect to other sampling schemes. However, following Mira (2001) there are some special cases in which DR cannot be worse than the independence algorithm. In general, the speed of convergence depends on the choice of the proposal distribution. A practical solution for improving the convergence is to choose a proposal scheme with good asymptotic properties until the chain converges and then eventually to append a DR step once the convergence to the invariant distribution is reached in order to reduce the asymptotic variance.

In this paper, we propose a DR algorithm based on three stages in order to update the volatility process. In our empirical analysis, it seems that the algorithm mixes properly.<sup>4</sup> At the same time, our choice guarantees a good computational speed. In our setting, at the first step the proposal distribution does not depend on the current state of the chain (independence proposal). We exploit the idea suggested in Eraker (2001) based on the concept of Brownian bridge<sup>5</sup> through  $V_{t-1}$  and  $V_{t+1}$ . It is easy to prove that a good approximation to the conditional posterior, or at least to the diffusive part of it, can be obtained by shrinking the length of  $\Delta$ . The proposal  $q(x, y)$  is then

$$q(V_t | V_{t-1}, V_{t+1}, \theta) \sim \mathcal{N}(\mu_t, \sigma_t^2), \quad (15)$$

<sup>4</sup>We say that a chain is mixing well if the autocorrelation in the chain is low. In general, this is a good indicator of efficiency of the algorithm.

<sup>5</sup>Different random walk proposals based on the Langevin approximation of the full conditionals have also been taken into account. However they seemed to perform poorly with respect to the one described above.

where  $\mu_t = (V_{t+1} + V_{t-1})/2$  and  $\sigma_t^2 = \sigma_v^2(1 - \rho^2)V_{t-1}/2$ . At the second step we consider a random walk proposal, that is,  $y = x + \epsilon$  where  $\epsilon$  is Gaussian with the same variance than before and given the current state  $x$ . Finally, at the third step we add another random walk step. We would like to stress that the acceptance probabilities between the second and the third step are different due to the adaptive nature of the algorithm.

The combination of the random walk and the independence proposals allows us to exploit the advantages of both. In fact, if the independence proposal is a good approximation of  $\pi(x)$  then the number of rejections will be small. But in case of a rejection, that is a poor approximation of the target distribution, a random walk proposal gives a control on this bad behaviour.

### 3.3. Model ranking and Bayes factor calculations

In Bayesian statistics it is a common practice to use the marginal likelihood as an indicator for the goodness-of-fit. This quantity is defined as

$$m(\mathbf{Y}|\mathcal{M}_i) = \int p(\mathbf{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\mathcal{M}_i) d\boldsymbol{\theta}_i, \tag{16}$$

where  $(\mathcal{M}_i, \boldsymbol{\theta}_i)$  indicate the  $i$ th model and its parameters. The functions  $p(\mathbf{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)$  and  $\pi(\boldsymbol{\theta}_i|\mathcal{M}_i)$  are, respectively, the likelihood and the prior distributions. The Bayes factor is evaluated as the ratio  $B_{ij} = m(\mathbf{Y}|\mathcal{M}_i)/m(\mathbf{Y}|\mathcal{M}_j)$ .

It is often difficult to evaluate the marginal likelihood, and many techniques have been proposed in the literature.<sup>6</sup> In many cases, it is useful to recur to the following decomposition

$$m(\mathbf{Y}|\mathcal{M}_i) = \frac{p(\mathbf{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\mathcal{M}_i)}{\pi(\boldsymbol{\theta}_i|\mathbf{Y}, \mathcal{M}_i)}, \forall \boldsymbol{\theta}_i \in \text{supp}(\boldsymbol{\theta}_i). \tag{17}$$

In order to evaluate this ratio, it is necessary to estimate  $\pi(\boldsymbol{\theta}_i|\mathbf{Y}, \mathcal{M}_i)$  and  $p(\mathbf{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)$  at a fixed point.  $\pi(\boldsymbol{\theta}_i|\mathbf{Y}, \mathcal{M}_i)$  is computed by using the algorithm of Chib and Jeliazkov (2001). Nicholls and Mira (2004) noticed that this is a special case of the bridge sampling setup of Meng and Wong (1996). The method consists in dividing the parameters vector  $\boldsymbol{\theta}_i = (\theta_1, \dots, \theta_K)$  in blocks, and at each step of the algorithm associate with the  $i$ th block, a given value  $\theta_i^*$ . In this way, we split the vector into two parts, say  $\boldsymbol{\psi}_i^* = (\theta_1^*, \dots, \theta_{i-1}^*)$  and  $\boldsymbol{\psi}^i = (\theta_{i+1}, \dots, \theta_K)$ . The estimate of the posterior at  $\theta^*$  is then given by

$$\hat{\pi}(\boldsymbol{\theta}_i|\mathbf{Y}, \mathcal{M}_i) = \prod_{i=1}^K \hat{\pi}(\theta_i^*|\mathbf{Y}, \theta_1^*, \dots, \theta_{i-1}^*), \tag{18}$$

where

$$\hat{\pi}(\theta_i^*|\mathbf{Y}, \theta_1^*, \dots, \theta_{i-1}^*) = \frac{\hat{E}_1 [\alpha(\theta_i, \theta_i^*|\mathbf{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)q(\theta_i, \theta_i^*|\mathbf{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)]}{\hat{E}_2 [\alpha(\theta_i^*, \theta_i|\mathbf{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)]}. \tag{19}$$

Here the numerator is the expected value taken with respect to  $\pi(\theta_i, \boldsymbol{\psi}^i|\mathbf{Y}, \boldsymbol{\psi}_i^*)$  and the denominator with respect to  $\pi(\boldsymbol{\psi}^i|\mathbf{Y}, \boldsymbol{\psi}_i^*)q(\theta_i^*, \theta_i|\mathbf{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)$ . Finally the quantities  $q(\theta_i, \theta_i^*|\mathbf{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)$  and  $\alpha(\theta_i^*, \theta_i|\mathbf{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)$  are, respectively, the proposal and the acceptance probability for the MH

<sup>6</sup>A review on some Monte Carlo alternatives is given in Han and Carlin (2001).

algorithm. It is then possible to exploit the output of the reduced MH procedure to estimate these expected values, using the reduced full conditional distributions described on the right-hand side of equation (18).

### 3.4. Particle filtering and likelihood evaluation

In many applications, the likelihood  $p(\mathbf{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)$  is known, so that the Bayes factor can be evaluated once the posterior is estimated. Unfortunately, in models with latent factors, the non-observable variables  $X$  have to be integrated out. Chib *et al.* (2002) propose a particle filter to evaluate the likelihood for a basic stochastic volatility model with jumps. In this paper we adopt a version of the auxiliary particle filter proposed in Johannes *et al.* (2002).

Particle filters aim at estimating the latent processes through sequential Monte Carlo simulations. Important recent works on this topic include Gordon *et al.* (1993) and Kita-gawa (1996) whereas a useful introduction is given in Arulampalam *et al.* (2002). To run the filter, it is required to have the knowledge of  $p(X_0)$ , the transition distribution  $p(X_{t+1}|X_t), t \geq 0$  and the measurement distribution  $p(Y_t|X_t), t \geq 1$ . The key idea is to approximate the filtering density  $p(X_{t+1}|\mathbf{Y}_{t+1})$  by a discrete cloud of points,  $\{X_{t+1}^j: j = 1, \dots, M\}$ , called particles. This density is estimated as

$$\hat{p}(X_{t+1}|\mathbf{Y}_{t+1}) = \sum_{j=1}^M \omega_{t+1}^j \delta(X_{t+1} - X_{t+1}^j), \quad (20)$$

where  $\omega_{t+1}^j$  are suitable weights and  $\delta(\cdot)$  is an indicator function. Obtaining a sample from (20) is easy if we recur to the importance sampling method in which the proposal  $q(X_{t+1}|\mathbf{Y}_{t+1})$  can be set to  $p(X_{t+1}|X_t)$  with weights  $\omega_{t+1}^j \propto \omega_t^j p(Y_{t+1}|X_{t+1}^j)$ .

In many practical applications, the plain particle filter described above is really sensible to outliers. The introduction of auxiliary variables (Pitt and Shephard 1999) improves the choice of the new particles by reducing the variability of the importance weights  $\omega_t$ . In many cases, this kind of generalizations allow to select particles that are more concentrated around the true state. A systematic description of the auxiliary particle filter method can be found in Doucet *et al.* (2001) [chap. 7] and in Arulampalam *et al.* (2002).

We apply the auxiliary particle filter to the affine jump diffusion model for the log-volatilities as in equations (8)–(10). This choice solves the problems related to the sampling of negative particles.

In the following, we will adopt the notation introduced in Johannes *et al.* (2002) in which the state vector is  $X_t = (Z_{t-1}, J_t, H_t^y, H_t^v)$ . We also take into account the leverage effect between processes. The introduction of  $\rho$  induces a dependence between the state vector  $X_{t+1}$  and the observation  $Y_t$ . This is called feedback effect, in which the transition distribution depends on  $Y_t$ , i.e.  $p(X_{t+1}|X_t, Y_t)$ . This complication can be accounted for by decomposing the covariance matrix through its Cholesky decomposition. In practice, it is possible to prove that it does not affect the filtering procedure.

Once the states are filtered, it is easy to evaluate the likelihood function by  $p(\mathbf{Y}_T|\boldsymbol{\theta}) = \prod_{t=1}^T \hat{p}(Y_t|\mathbf{Y}_{t-1})$  where  $p(Y_t|\mathbf{Y}_{t-1}) = \int p(Y_t|X_t, \boldsymbol{\theta})p(X_t|\mathbf{Y}_{t-1})dX_t$  is estimated simulating the state  $X_t$  from  $p(X_t|X_{t-1}^i), i = 1, \dots, M$  and in which  $X_{t-1}^i$  is the output of the filter at time  $t - 1$ .

We carried out a Monte Carlo experiment in order to check whether the introduction of the auxiliary variables improves the efficiency of the filter. We found some evidence that the auxiliary particle filter outperforms the plain Sampling Importance Resampling filter (SIR). We performed



100 Monte Carlo experiments on simulated series of size 500. For the auxiliary particle filter we set  $M = 10,000$  resampled  $R = 20,000$  times at each step. For the SIR filter we used 20,000 particles in order to maintain the same computational burden between the two algorithms. For the Heston's stochastic volatility model, for example, the Root Mean Square Error (RMSE) is 0.035 for the auxiliary filter whereas it is 0.078 for the SIR. If we perform the same computations for the complete model, we noticed that there is a systematic improvement when we use the auxiliary particle filter even though the difference is not so remarkable than before.

We also performed a Monte Carlo experiment to check the sensitivity of the discretization bias to the introduction of auxiliary points in-between each pair of observations. We found that there is just a slight improvement if we add 4–10 data points to daily observations. Furthermore, the RMSE remains basically unchanged with respect to the case with  $\Delta = 1$  and no missing point inserted. These findings are consistent with the Monte Carlo results presented in Johannes *et al.* (2002).

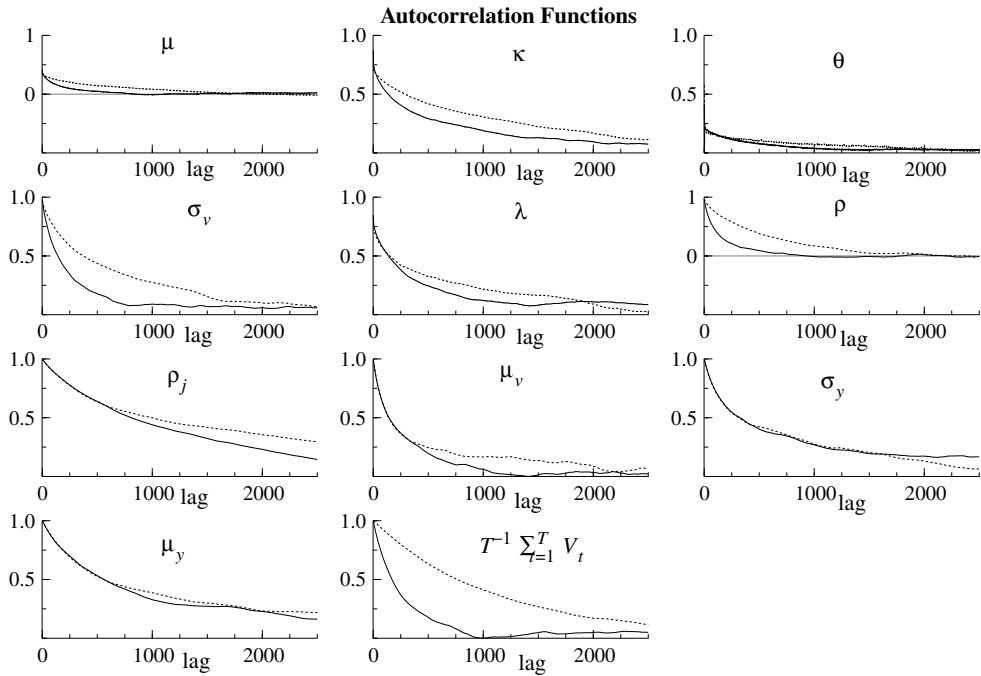
## 4. EMPIRICAL RESULTS

Our empirical application is based on the study of financial indexes, that is, the Dow Jones 65 Composite index and FTSE 100 observed on a daily basis during the last 15 years from July 1988 to July 2003. The data set has been downloaded from *Datastream*. All the calculations made in this paper are based on software written using the  $\text{Ox}^{\text{®}}$  3.0 language of Doornik (2001). By letting  $Y_t$  be the observed log-prices, the returns are defined as  $100 \times [Y_t - Y_{t-1}]$ .

### 4.1. Posterior analysis

Inference has been performed for the general model of Duffie *et al.* (2000, SVCJ) described in equations (6) and (7). Nested specifications proposed in Bates (1996, SVJ) and in Heston (1993, SV) have been estimated as well. The analysis is implemented through MCMC, running the DR algorithm for 250,000 iterations and discarding the first 50,000. In our experience this choice for the burn-in is more than adequate, even though our aim is to completely remove the effect of the initial values. The Monte Carlo standard error (MCSE) has been computed through the use of a kernel estimator to take into account the dependencies due to the Markov nature of the algorithm. Since draws from the posterior distributions are not independent, the reported MCSEs estimate  $2\pi$  times the spectral density matrix at frequency zero computed by standard time series method. In particular, the estimator is based on a VAR(1) prewhitening, than  $2\pi$  times the spectral density matrix at frequency zero of VAR residuals is estimated by smoothing methods using the Parzen kernel and automatic bandwidth selection. Recolouring provides an estimate of  $2\pi$  times the spectral density matrix at frequency zero of interest.

We first analyze the behaviour of the Delayed Rejection algorithm for these models. We compare the procedures based on three steps (Independence proposal plus two random walk steps) with the random walk algorithm proposed in Eraker *et al.* (2003). The adaptive procedure reduces dramatically the rejection rate that drop from 30 to 40% to about 5%. This finding has been noticed through an extensive simulation study for real and simulated data. This reduction of the rejection rate improve the mixing properties of the chains. In fact we can observe a sensible reduction of the autocorrelation among the draws of the volatility process. In general we notice that the introduction of the DR steps reduces the autocorrelations for all the parameters of the

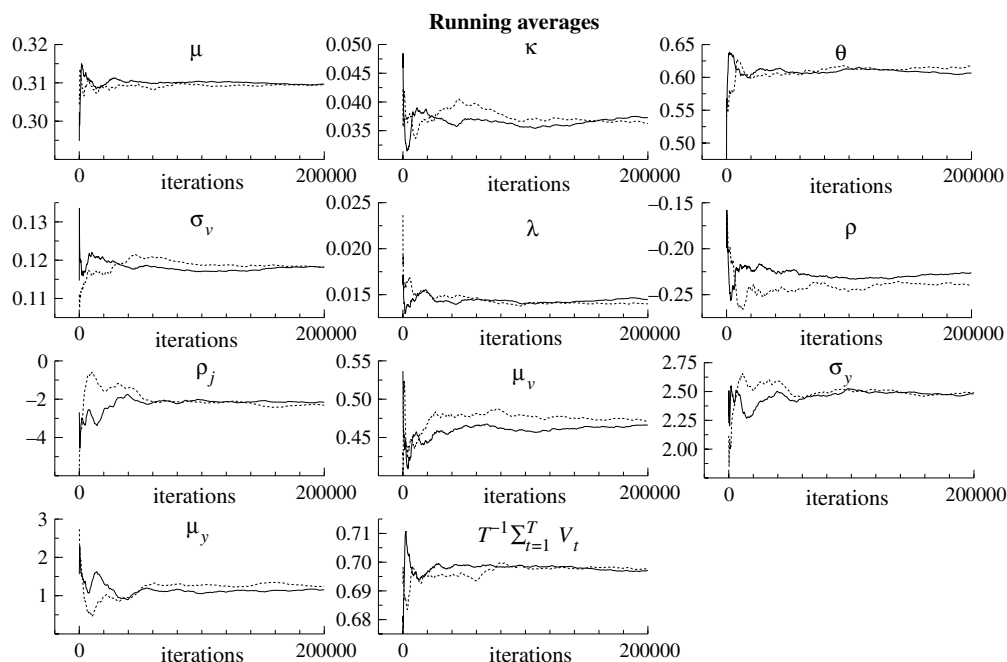


**Figure 1.** Dow Jones Composite 65 – Sample autocorrelation functions for each parameter and the averaged volatility. Random walk (dotted line) versus Delayed Rejection (solid line).

model. The plot of the autocorrelations for the SVCJ model for the Dow Jones series is shown in Figure 1.

It is interesting to note that the mixing for the volatilities shown in the last box is definitely improved. In order to obtain a similar autocorrelation function with the random walk algorithm, it is in general necessary to record just one on five draws. Furthermore, the improvement obtained for the volatilities induces some benefits on the mixing of the other parameters. The reduction is in fact systematic for each parameter as shown in Figure 1. We are now interested on the speed of convergence of the Delayed Rejection with respect to the Random walk. Of course, it is impossible to assess convergence properties of an MCMC algorithm through the study of just few realization of the chains. However a common practice is to check for the convergence of the empirical averages of the draws (see Robert and Casella 1999, chap. 8 for a survey on this topic. We can say that a chain converges rapidly if evolution of its empirical averages stabilizes after a few iterations. The results for the Dow Jones series are plotted in Figure 2.<sup>7</sup> In the last box of Figure 2 we plot the evolution of the averages for  $T^{-1} \sum_{t=1}^T V_t$ . It is quite evident that the running averages for the DR algorithm for this variable stabilize sooner than the corresponding plot for the random walk. This evidence can be interpreted as an empirical proof for the speed of convergence. In our experience, it seems that the random walk algorithm needs about 50,000 more iteration to converge to its correct expected value.

<sup>7</sup>The plots obtained for other real and simulated time series are quite similar and are available upon request.



**Figure 2.** Dow Jones Composite 65 – Evolution of the empirical averages for each parameter and for the expected volatility updated at each iteration of the MCMC algorithm. Random Walk (dotted line) versus Delayed Rejection (solid line).

In general we observe that both algorithms converge to the same expected value. However, apart for the average volatilities discussed before, it is not clear whether the adaptive method proposed allows for a systematic improvement of the speed of convergence for the other parameters.

The results for the Dow Jones series are reported in Table 1. We show the mean of the posterior and the Monte Carlo standard errors for the SV, the SVJ and the SVCJ models. As a benchmark, in the last column we report the output of the random walk algorithm. We also report at the bottom the marginal likelihoods useful to compare the fitting of the three models studied.

For the basic stochastic volatility model in column 2, the annualized spot volatility is  $\sqrt{260 \times \theta}$ . Its estimate is 15.17, that has to be compared with the sample volatility<sup>8</sup> 15.09. We observe that the estimated expected volatility and its sample counterpart are really close. The parameter  $\kappa = 0.023$  is known as mean reversion, that is the time needed for the process to return to its expected level. The leverage effect is described by  $\rho$  and is negative, even though it is different in magnitude with respect to the results usually stylized in literature (see Eraker *et al.* 2003). The difference we found is explained by the choice of the drift for the returns, that in this paper includes the volatility.  $\sigma_v$ , that is 0.1251, determine the variance of the volatilities defined as  $\sigma_v^2 V_t$ .

<sup>8</sup>Annualized volatilities has been computed by multiplying the sample standard error by  $\sqrt{260}$ .

**Table 1.** Dow Jones Composite 65 – Parameter estimates.

	SV <sub>DR</sub>	SVJ <sub>DR</sub>	SVCJ <sub>DR</sub>	SVCJ <sub>RW</sub>
$\mu$	0.3268 (0.0002)	0.2979 (0.0004)	0.3097 (0.0003)	0.3095 (0.0005)
$\kappa$	0.0234 (0.0002)	0.0177 (0.0001)	0.0372 (0.0005)	0.0362 (0.0005)
$\theta$	0.8857 (0.0017)	0.7969 (0.0018)	0.6067 (0.0031)	0.6177 (0.0041)
$\sigma_v$	0.1251 (0.0005)	0.1012 (0.0002)	0.1181 (0.0004)	0.1182 (0.0005)
$\rho$	-0.1788 (0.0020)	-0.2361 (0.0035)	-0.2264 (0.0031)	-0.2401 (0.0045)
$\lambda$		0.0274 (0.0004)	0.0145 (0.0002)	0.0140 (0.0002)
$\rho_j$			-2.1556 (0.1141)	-2.3188 (0.1131)
$\mu_v$			0.4661 (0.0040)	0.4710 (0.0042)
$\mu_y$		0.3317 (0.0103)	1.1519 (0.0575)	1.2469 (0.0581)
$\sigma_y$		2.2389 (0.0128)	2.4809 (0.0251)	2.4892 (0.0251)
Marg. Lik.	-5533.9	-5530.4	-5497.1	-

The Monte Carlo standard Errors are presented in brackets.

If the first jump component is taken into account as in column 3, the annualized volatility decreases to 14.39. Furthermore we observed a similar behaviour for the parameter  $\sigma_v$ . This reduction is due to the introduction of the jumps on the returns. In fact, this factor explains part of the volatility that in the previous model has been described just by the diffusive process. In a neighbourhood of a jump, it is in fact intuitive to understand that a diffusive process tends to overestimate the volatility because of the continuity of its trajectories. The intensity of the jump process  $\lambda$ , that represents the frequency of the jump events, is 0.027. The model then expects about seven jumps per year. The annualized spot volatility in this model becomes  $\sqrt{260(\theta + \mu_v\lambda)}/\kappa$  and reduces from the 15.17 of the SV model to the 14.39. Furthermore, the annualized total volatility, the mean square error of the returns 15.61 percent. This quantity<sup>9</sup> is computed noticing that the conditional expected value of the squared returns is  $V_t + \lambda E[H^{y^2}]$  where  $E[H^{y^2}] = \mu_y^2 + 2\mu_y\mu_v\rho_j + \rho_j^2\mu_v^2 + \sigma_y^2$ . In percentage terms, the average effect of the jump component on the total volatility is approximately 15%. Finally, we analyze the model with two jump factors. The introduction of the second jump component increases the parameter  $\kappa$ , so that the mean reversion phenomenon sensibly speeds up. The parameter  $\lambda$  halves.

The general model expects just 3.6 jumps per year. The annualized spot volatility sensibly decreases with respect to the other two parametrizations and it drops at the 13.4 percent. On the other side, the total volatility moves to 14.2 and then the impact of the jumps on the returns is about the 17.3%. There is then evidence of a reduction of the expected volatility with respect to the SVJ model. In fact, the second jump factor explicitly describes a shift on the volatility process from  $\sqrt{V_{t-}}$  to  $\sqrt{V_{t-} + H_t^v}$ . However, this effect is milder than some empirical findings described in Eraker *et al.* (2003) and in this work is estimated approximately at the 4%. These differences are likely explained by the different sample interval considered for our empirical analysis. Eraker *et al.* (2003) include the date of October 19, 1987 on which date a huge crash in the markets was observed. Likely, that single observation heavily influence all the estimates.

<sup>9</sup>Eraker *et al.* (2003) derive other conditional moments for these models.

**Table 2.** FTSE – Parameter estimates.

	$SV_{DR}$	$SVJ_{DR}$	$SVCJ_{DR}$	$SVCJ_{RW}$
$\mu$	0.3763 (0.0003)	0.3456 (0.0006)	0.3560 (0.0004)	0.3563 (0.0007)
$\kappa$	0.0159 (0.0001)	0.0165 (0.0001)	0.0276 (0.0003)	0.0265 (0.0003)
$\theta$	1.0279 (0.0015)	0.9609 (0.0016)	0.7642 (0.0036)	0.7842 (0.0043)
$\sigma_v$	0.1053 (0.0003)	0.1038 (0.0002)	0.1110 (0.0002)	0.1115 (0.0003)
$\rho$	-0.1041 (0.0022)	-0.0970 (0.0025)	-0.1386 (0.0038)	-0.1384 (0.0053)
$\lambda$		0.0234 (0.0004)	0.0115 (0.0002)	0.0106 (0.0002)
$\rho_j$			-0.8266 (0.0715)	-0.8722 (0.0832)
$\mu_v$			0.5299 (0.0053)	0.5262 (0.0053)
$\mu_y$		1.1230 (0.0252)	2.1477 (0.0581)	2.4716 (0.0699)
$\sigma_y$		1.8250 (0.0118)	1.9692 (0.0195)	1.9097 (0.0183)
Marg. Lik.	-5333.7	-5321.6	-5316.7	-

The Monte Carlo standard Errors are presented in brackets.

Finally, in Table 2 we report the results for the FTSE index.

These results are in line with the ones observed for the Dow Jones series. The introduction of the jump process reduces the expected volatility due to the diffusive part of the model. As usual this is shown by the parameters  $\theta$  and  $\sigma_v$  and by the computation of the total and of the spot annualized volatilities. As before, the introduction of the jumps sensibly increases the mean reversion  $\kappa$  and reduces the expected volatility  $\theta$ . In general, for all the three models the annualized spot volatility is close to 16% whereas the sampling spot volatility is 16.6%. The parameter  $\lambda$  sensibly reduces when the two models with jumps are chosen and the estimate is 0.023 for the SVJ model and is 0.011 for the SVCJ. On the other side, the expected size of the jumps in the returns increases from 1.12 in the SVJ to 1.70 in the SVCJ model. In practice, the SVCJ model detects less jumps per year but with higher impact. However, for the SVJ model the jumps on the returns explain about the 10% of the total volatility while for the SVCJ model the impact is 7.8%. Finally the contribution of the second jump component, i.e.  $J_t H_t^v$  on explaining the volatility is the 3.8 percent.

Estimates of the marginal likelihoods are reported at the bottom of Tables 1 and 2. According to the thresholds defined in Kass and Raftery (1995), in general the complete model is strongly preferred to the others. For the Dow Jones series, there is strong evidence that the complete model SVCJ is preferred to the SVJ. Furthermore, the SVJ model gives a better fit to the data with respect to the SV model even though the log-Bayes factor are close to 3. For the FTSE series, the ordering among the three models is remarkable. The logarithm of the Bayes factors is always larger than 4 and then there is no uncertainty about the ranking. In general, we can state that there is a natural ordering among models.

## 5. CONCLUSION

In this paper we propose an efficient MCMC method to estimate the adaptive stochastic volatility model with jumps. First, empirical results evidence a sensible improvement of the DR algorithm

with respect to the random walk. It is evident that the mixing (low autocorrelation) of the chain based on the DR is better than the one obtained by the use of the simple random walk approach. Furthermore, there is evidence suggesting an improved speed of convergence for the volatilities to their invariant distributions. However there are not strong theoretical results assessing the speed of convergence of this algorithm so far. Therefore further research is needed in this respect.

Second, the particle-filtering approach provides a reliable and fast method to compute the likelihood and to estimate the latent processes. Some Monte Carlo experiments suggest that the auxiliary particle filter improves the plain SIR. This finding is systematic for all the models considered. The introduction of missing data to reduce the approximation bias of the discretized model does not seem to be very useful with daily data. However more research is needed to provide more conclusive evidence regarding this point.

Finally, from the point of view of the modelling, an interesting extension of the results obtained in this work includes the study of models for the pricing of derivatives based on more general stochastic dynamics. For this purpose, the use of sequential Monte Carlo methods for estimating both states and parameters seems to be appropriate and faster with respect to MCMC to carry out inference mainly if we aim at estimating the risk neutral parameters.

### ACKNOWLEDGEMENTS

Financial support from the ‘G. Ferrari Fund’ and from MIUR, Grant 2002135473, is gratefully acknowledged. I would like to thank Bjørn Eraker, Erich Battistin, Fabio Rigat, Silvano Bordignon, Nunzio Cappuccio, Diego Lubian, Alessandro Palandri, Alessandro Platania and all the participants at the Financial Econometric Lunch Group at Duke University for helpful comments. I also thank an anonymous referee and the Editor whose comments have resulted in many improvements. All errors are my own.

### REFERENCES

- Ait-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* 70, 223–62.
- Arulampalam, M. S., S. Maskell, N. Gordon and T. Clapp (2002). A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50, 174–88.
- Bakshi, G., C. Cao and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance* 52, 2003–49.
- Bates, D. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies* 9, 69–107.
- Besag, J. (1994). Comments on “Representations of Knowledge in Complex Systems” by U. Grenander and M.I. Miller. *Journal of the Royal Statistical Society*, 56 B, 569–603.
- Brandt, M. and P. Santa-Clara (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics* 63, 161–210.
- Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association* 96, 270–81.
- Chib, S., F. Nardari and N. Shephard (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*, London: Timberlake Consultants Press.

- Doucet, A., N. de Freitas and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*, New York: Springer.
- Duffie, D., J. Pan and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–76.
- Durham, G. and R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with comments). *Journal of Business & Economic Statistics* 20, 297–337.
- Elerian, O., S. Chib and N. Shephard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica* 69, 959–93.
- Eraker, B. (2001). MCMC analysis of diffusion processes with application to finance. *Journal of Business & Economic Statistics* 19, 177–91.
- Eraker, B. (2004). Do equity prices and volatility jumps? Reconciling evidence from spot and option prices. *Journal of Finance* 59, 1367–403.
- Eraker, B., M. Johannes and N. Polson (2003). The impact of jumps in volatility and returns. *Journal of Finance* 58, 1269–300.
- Gallant, A. and G. Tauchen (1996). Which moments to match? *Econometric Theory* 12, 657–81.
- Gåsemir, J. (2003). On an adaptive version of the Metropolis–Hastings algorithm with independent proposal distribution. *Scandinavian Journal of Statistics* 30, 159–63.
- Ghysels, E., A. Harvey and E. Renault (1996). Stochastic volatility. In C. Rao and G. Maddala, (eds.), *Statistical Methods in Finance* pp. 119–91, Amsterdam: North Holland.
- Gilks, W., N. Best and K. Tan (1995). Adaptive rejection metropolis sampling within gibbs sampling. *Applied Statistics* 44, 455–72.
- Gilks, W., G. Roberts and S. Shau (1998). Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association* 93, 1045–54.
- Glasserman, P. and N. Merener (2004). Convergence of a discretization scheme for jump-diffusion processes with state-dependent intensities. *Proceedings of the Royal Society* 460 A, 111–27.
- Gordon, N., D. Salmond and A. Smith (1993). Novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEE Proceedings-F* 140, 107–13.
- Han, C. and B. Carlin (2001). MCMC methods for computing Bayes factors: A comparative review. *Journal of the American Statistical Association* 96, 1122–32.
- Heston, S. (1993). A closed-form solution of options with Stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327–43.
- Jacquier, E., N. Polson and P. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business & Economic Statistics* 12, 371–417.
- Johannes, M., N. Polson and J. Stroud (2002), Nonlinear filtering of stochastic differential equations with jumps, Working paper, Graduate School of Business, University of Chicago.
- Kass, R. and A. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773–95.
- Kim, S., N. Shephard and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–93.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian non-linear state space models. *Journal of Computational and Graphical Statistics* 5, 1–25.
- Meng, X.-L. and W. Wong (1996). Simulating ratios of normalising constants via a simple identity: A theoretical explanation. *Statistica Sinica* 6, 831–60.
- Mengersen, K. and R. Tweedie (1996). Rates of convergence of the hastings and metropolis algorithms. *Annals of Statistics* 24, 101–21.
- Mira, A. (2001). On Metropolis–Hastings algorithms with delayed rejections. *Metron* 59, 231–41.
- Nicholls, G. and A. Mira (2004). Bridge estimation of the probability density at a point. *Statistica Sinica* 14, 603–12.

- Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3–50.
- Peskun, P. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika* 60, 607–12.
- Pitt, M. (2002). Smooth particle filters for likelihood evaluation and maximization, Working paper, University of Warwick.
- Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association* 94(446), 590–99.
- Robert, C. and G. Casella (1999), *Monte Carlo Statistical Methods*, Berlin: Springer.
- Roberts, G., A. Gelman and W. Gilks (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *Annals of Applied Probability* 7, 110–20.
- Runggaldier, W. (2003). Jump Diffusion Models. In S. Rachev, (Ed.), *Handbook of Heavy Tailed Distributions in Finance*, Amsterdam: North Holland.
- Scott, L. (1997). Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Applications of fourier inversion methods. *Mathematical Finance* 7, 413–24.
- Tierney, L. (1994). Markov chain for exploring posterior distributions (with discussion). *Annals of Statistics* 22, 1701–62.
- Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Annals of Applied Probability* 8, 1–9.
- Tierney, L. and A. Mira (1999). Some adaptive Monte carlo methods for Bayesian inference. *Statistics in Medicine* 18, 2507–15.
- Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165–178.