Working Paper Series, N. 1, January 2004

Adaptive MCMC Methods for Inference on Discretely Observed Affine Jump Diffusion Models

Davide Raggi
Department of Statistical Sciences
University of Padua
Italy

Abstract: In the present paper we generalize in a Bayesian framework the inferential solution proposed by Eraker, Johannes & Polson (2003) for stochastic volatility models with jumps and affine structure. We will use an adaptive sampling methodology known as *Delayed Rejection* suggested in Tierney & Mira (1999) in a Markov Chain Monte Carlo settings in order to reduce the asymptotic variance of the estimates. Furthermore, the use of a particle filtering procedure allows to compute the Bayes factor.

Keywords: Jump Diffusion, Adaptive MCMC, Particle Filters, Bayes factor.

Final version (2004-01-16)

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Davide Raggi
tel: +39 049 827 4168
daviragg@stat.unipd.it
http://www.stat.unipd.it/~name

Adaptive MCMC Methods for Inference on Discretely Observed Affine Jump Diffusion Models

Davide Raggi
Department of Statistical Sciences
University of Padua
Italy

1 Introduction

In the last 20 years the work of Black & Scholes (1973) led to a growing interest in the studying of continuous time stochastic processes for financial applications. More precisely, the relevance of stochastic calculus for the pricing of financial derivatives has fully emerged. The model proposed in Black & Scholes (1973) describes the behaviour of the underlying asset as

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{1}$$

The main advantage of this framework is that it allows to easily handle the derivative pricing task. The main drawback is that the volatility of the model is described using a single parameter constant over time. Empirical studies show that the latter assumption is not realistic and that volatilities tend to change over time (see Ghysels, Harvey & Renault 1996 or Taylor (1994) for a comprehensive study). Thus, more flexible models have been proposed in literature such as Hull & White (1987) where asset price changes are expressed as:

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_{1,t} \tag{2}$$

$$dV_t = aV_t dt + b_t V_t dW_{2,t} \tag{3}$$

and Heston (1993) where dS_t is modeled as:

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_{1,t} \tag{4}$$

$$dV_t = \kappa(\vartheta - V_t)dt + \sigma\sqrt{V_t} dW_{2,t}. \tag{5}$$

where $W_{1,t}, W_{2,t}$ are possibly correlated Brownian motions in $\mathbb{R}^2$. Both of the above models allow the pricing of some derivative assets such as European options. The solution proposed in Hull & White (1987) for pricing is based on simulation techniques whereas in Heston (1993) an analytical representation of the characteristic function for the marginal S_t is derived, exploiting the affine[1] structure of its coefficients. Furthermore in the latter model semi-closed forms for the option prices are

[1] The coefficients of the model are linear plus the constant with respect the variable. See Björk (1998)[chap. 17] for an introduction on affine term structure concepts.

obtained as opposed to Hull & White (1987).
A further generalization of stochastic volatility models involves the presence of jumps. It seems natural indeed to add a jump component in the return equation in order to describe rare events like crashes in the market. But it is not intuitive to understand whether the volatility process jumps or not. Empirical evidence anyway show that taking into account jumps, together with stochastic volatility leads to an improved fit of the data as stressed in Bakshi, Cao & Chen (1997). It is possible to note that the volatility process hardly follows a diffusive behaviour and tend to sharply increase when a jump is observed in the return series.
A survey on some recent mathematical results in that field can be found in Runggaldier (2003).
Following this direction, Duffie, Pan & Singleton (2000) propose an affine diffusion model with jumps that leads to a semi-closed derivative price form, thus generalizing the results obtained by Heston (1993).
The reminder of this paper is then organized as follows. In Section 2 the Duffie et al. (2000) framework is described. The inferential solution proposed here is outlined in Section 3. Finally in section 4 empirical results both with real and simulated data are illustrated.

2 Affine Jump Diffusion models

An affine process $\{X_t : t \geq 0\}$ is a Markov process such that, for every time t, its characteristic function of X_t is an exponential-affine function of the initial state X_0. These models are widely used in financial applications, due to their analytical tractability (see Duffie, Filipović & Schachermayer 2003 for general results). In general the model is written in state space form for the n-dimensional vector of equations X as:

$$dX_t = \mu\left(X_{t-}\right)dt + \sigma\left(X_{t-}\right)dW_t + dZ_t \tag{6}$$

where W_t is a multi-dimensional Brownian motion and Z_t is a marked point process. The drift $\mu(x)$, the elements of the covariance matrix $\sigma(x)\sigma(x)'_{ij}, i, j = 1, \ldots, n$, the intensity of the jump processes $\lambda(x)$ and the interest rate $R(x)$ are assumed to be affine in X. It is possible to show that (Duffie et al. 2000)

$$\psi(u, X_t, t, T) = E\left[\exp\left(-\int_t^T R(X_s)ds\right)e^{uX_T}\Big|\mathcal{F}_t\right] = e^{\alpha(t)+\beta(t)x_t} \tag{7}$$

where $\alpha(t)$ and $\beta(t)$ are solution of suitable ordinary complex valued differential equations and $u \in \mathbb{C}^n$. In general it is possible to calculate $\alpha(t)$ and $\beta(t)$ using numerical techniques such as the Runge-Kutta method. Sometimes the functions $\alpha(t)$ and $\beta(t)$ can be derived analytically. The knowledge of (7) allows to evaluate the price of an European option (Scott 1997). Suppose in fact that we are dealing with European call option with strike price K and payoff $(S_t - K)^+$. It is possible to guess a solution for its price of the form

$$C_t = S_t\left(P_1 - \frac{K}{S_t}P_2\right) \tag{8}$$

as suggested in Heston (1993) where

$$P_1 = \frac{\psi(1, X_t, t, T)}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}(\psi(1-iu, X_t, t, T))e^{iu(\ln k)}}{u}du \tag{9}$$

and

$$P_2 = \frac{\psi(0, X_t, t, T)}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}(\psi(-iu, X_t, t, T))e^{iu(\ln k)}}{u}du \tag{10}$$

with $k = K/S_t$, are obtained by inverting the characteristic function through the Lévy inversion formula[2]. It is important to stress that the affine jump structure allows pricing several types of derivatives. Results in (7) can be applied to more general payoffs function of the form $v_0 + v_1 X_T e^{uX_T}$ making more flexible the pricing of different financial instruments such like bond derivatives, quantos and asian options for example. Efficient numerical procedures to invert the characteristic function have been proposed in Pan (2002) and Carr & Madan (1999). In practical applications the use of quadrature methods appear reliable as well.

2.1 The Stochastic Volatility Model

An interesting application of the affine jump diffusion theory is the following stochastic volatility model in which it is assumed that $Y_t = \ln(S_t)$ is described by

$$dY_t = (\mu - \frac{1}{2}V_{t-})dt + \sqrt{V_{t-}}dW_{1,t} + H_t^y dN_t^y \tag{11}$$

$$dV_t = \kappa(\theta - V_{t-})dt + \sqrt{V_{t-}}\left(\rho\sigma_v dW_{1,t} + \sigma_v\sqrt{1-\rho^2}dW_{2,t}\right) + H^v dN_t^v \tag{12}$$

where $(W_{1,t}, W_{2,t})$ is a Brownian motion in $\mathbb{R}^2$ with independent components. X_{t-} indicates the left limit. The jump components $H_t^i dN_t^i$, $i = y, v$ are marked point processes. In a more formal way, such a processes can be described as a random sum $Z_t = \sum_{j=1}^{N(t)} H_j^i$ in which $N(t)$ represent the number of arrival of a Poisson process in $(0, t]$. In this paper we assume that the intensities for both processes are constant which can be relaxed into an affine scheme. The measure of the jumps are described by the random variables $H^i, i = y, v$. It is immediate to note that Heston's model can be obtained simply equating to zero the intensities.
The model of (11) and (12) 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^v has intensity λ_v and jump size $H^v \sim \mathcal{E}xp(\mu_v)$ while N^y with intensity λ_y has jump size given by $H^y \sim \mathcal{N}(\mu_y, \sigma_y^2)$. Constraining $\lambda_v = 0$ leads to the model proposed in Bates (1996). 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 $H^v \sim \mathcal{E}xp(\mu_v)$ and $H^y|H^v \sim \mathcal{N}(\mu_y + \rho_j H^v, \sigma_y^2)$. It is easy to show (Eraker et al. 2003) that for the contemporaneous jump model the conditional instantaneous moments are expressed by

$$\lim_{\Delta\to 0}\mathbb{E}\left[(\log(S_{t+\Delta}/S_t))^2|V_t, S_t\right] = V_t + \lambda\mathbb{E}\left[(H^y)^2\right] \tag{13}$$

[2]See for example Williams (1991) for a treatment of this problem.

$$\lim_{\Delta\to 0}\mathbb{E}\big[(V_{t+\Delta}-V_t))^2|V_t,S_t\big]=\sigma_v^2 v_t+\mu_v\lambda \tag{14}$$

$$\lim_{\Delta\to 0}\mathbb{E}\big[(V_{t+\Delta}-V_t)\log(S_{t+\Delta}/S_t)|V_t,S_t\big]=\rho\sigma_v V_t+\rho_j\mu_v^2 \tag{15}$$

$$\mathbb{E}\left[(H^y)^2\right]=\mu_y^2+2\mu_y\mu_v\rho_j+\rho_j^2\mu_v^2+\sigma_y^2. \tag{16}$$

It is also easy to prove that the expected volatility is $\mathbb{E}[V_t]=\theta+\lambda\mu_v/\kappa$. The expressions in eq. (13) and in eq. (14) stated above represent respectively the conditional variance of the returns and the conditional variance of the volatility while (15) is the conditional correlation among the processes composing the model. It is important to stress that, for all the models described are available analytical forms for the characteristic function and then the option price is obtainable, up to a numerical integration. In the reminder of this paper we analyze the latter model.

3 Inference

From a statistical point of view inference is a challenging problem for mainly two reasons. The first one is that the trajectories of the processes are continuous. The second one is that not all the processes involved are observable.
Many techniques have been proposed to solve the inferential problem. Some methods rely on the efficient method of moments (Gallant & Tauchen 1996), others are based on the indirect inference principle (Gourieroux, Monfort & Ranault 1993), others on filtering techniques (Johannes, Polson & Stroud 2002, Durham & Gallant 2002). We adopt a Markov chain Monte Carlo approach that gives good results for discretely specified stochastic volatility models (Jacquier, Polson & Rossi 1994, Kim, Shephard & Chib 1998). Eraker (2001) and Elerian, Chib & Shephard (2001) propose a strategy to infer the parameters of continuous time stochastic processes using MCMC algorithms.
The main difficulty in deriving inferences for continuous time stochastic processes discretely sampled is to evaluate the transition probability $p(X_{\tau_{i+1}}\mid X_{\tau_i})$. In fact, the sampling distance between τ_i and τ_{i+1} sometimes cannot give a good approximation of the real process. Furthermore, as pointed out in Aït-Sahalia (2002) among others, for many processes proposed in the economic literature, it is not even possible to obtain a closed form for the density of the transition probability. A possible solution is approximating $p(X_{\tau_{i+1}}\mid X_{\tau_i})$ by numerical techniques. An overview for these methods is presented in Durham & Gallant (2002). Once evaluated the transition density $\widehat{p}(\tau_{i+1}|X_{\tau_i})$ it is possible to estimate the log-likelihood function

$$\ell_n(\theta)=\sum_{i=0}^{n}\ln\widehat{p}(X_{\tau_{i+1}}\mid X_{\tau_i};\boldsymbol{\theta}) \tag{17}$$

It is reasonable to expect that if the sampling times are close to each other, then the approximated transition probability should be closed to the real one. In order to reduce the bias due to the length of the interval a promising idea is to augment the state space with high frequencies data, filling the interval (τ_i,τ_{i+1}) with missing data points. The first step of this inference procedure is to represent the continuous

model with a discrete time approximation. Then it is possible to approximate the true transition probability integrating out the non-observed variables. Calculating this integral using standard numerical integration techniques can be cumbersome, therefore we recur to Monte Carlo strategies.

The starting point for the inferential procedure involves a discretization that, in this paper is based on the scheme of Euler-Maruyama. See at this proposal Kloeden & Platen (1995)[chap. 9] for standard diffusion processes and Glasserman & Merener (2001) for an application to jump diffusions. The model is evaluated on a set of discrete times $\{\tau_i : i = 1, \ldots, n\}$. That leads to

$$Y_{\tau_{i+1}} - Y_{\tau_i} = (\mu - \frac{1}{2}V_{\tau_i})\Delta + \sqrt{V_{\tau_i}}\epsilon^y_{\tau_{i+1}} + H^y_{\tau_{i+1}}J^y_{\tau_{i+1}} \tag{18}$$

$$V_{\tau_{i+1}} - V_{\tau_i} = \kappa(\vartheta - V_{\tau_i})\Delta + \sigma_v\sqrt{V_{\tau_i}}\epsilon^v_{\tau_{i+1}} + H^v_{\tau_{i+1}}J^v_{\tau_{i+1}} \tag{19}$$

in which $\tau_{i+1} - \tau_i = \Delta$, $Y_{\tau_{i+1}} = \log(S_{\tau_{i+1}})$ and $J_{\tau_{i+1}}$ is described by a Binomial distribution $Bi(1, \lambda\Delta)$. The error terms are such that $(\epsilon^y_{\tau_{i+1}}, \epsilon^v_{\tau_{i+1}})$ are correlated standard normal. To fix the notation $\boldsymbol{Y} = \{Y_{\tau_1}, \ldots, Y_{\tau_n}\}$, whereas $\boldsymbol{Y}_{\tau_i} = \{Y_1, \ldots, Y_{\tau_i}\}$ indicate past observations up to time τ_i. It is important to note that this scheme is appropriate in this case because both sizes and intensities are independent from the states of the model, that is, returns and volatilities. In case there exists such a dependence, a more general scheme should be adopted (see at this proposal Glasserman & Merener 2001 or Cyganowski & Kloeden 2000). An example of the more general scheme will be given in Section 3.2.1 when a particle filter procedure is implemented for estimating the likelihood of the model.
In practice the use of daily intervals seems to produce small biases with respect to the continuous model as stated in Eraker et al. (2003).

To perform inference for these models we consider a Bayesian approach. More precisely we introduce in the context of stochastic volatility the use of a methodology known as *Delayed Rejection Metropolis-Hastings* proposed in Tierney & Mira (1999). In general a Metropolis-Hastings algorithm works in the following way:

`Metropolis-Hastings algorithm`

1. Sample y from a proposal $q(x, y)$.
2. Define $\alpha(x, y) = \begin{cases} \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right) & \text{if } \pi(x)q(x, y) > 0 \\ 1 & \text{otherwise} \end{cases}$
3. Sample u from $U(0, 1)$.
4. If $u \leq \alpha(x, y)$

   then $X_{t+1} = y$

   otherwise $X_{t+1} = x$.

The algorithm produces a Markov chain that converges, under suitable conditions to a given stationary distribution function $\pi(x)$ (see Robert & Casella 1999, chapter 6-7). In empirical studies, a challenging point is to find a good proposal distribution $q(x, y)$ since it heavily influences the convergence properties of the entire Markov chain. If the proposal does not depend on the present state of the chain, that is $q(x, y) = f(y)$, then it generates a so called *Independence Chain*. On the other hand, if $Y = x + Z$ with $Z \sim f(z)$ implying $q(x, y) = f(y - x)$ we deal with a *Random Walk Chain*. If $X = (X_1, \dots, X_k)$, it is possible to update the chain one component at a time, using for each component a one-dimensional Metropolis-Hastings scheme. The *Gibbs sampler* is an algorithm of this type where the proposal for every subcomponent of X is the full conditional distribution $\pi(X_i|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k) = \pi(X_i|X_{-i})$. It is immediate to check that using the full conditional as proposal the acceptance probability is $\alpha(x, y) = 1$.
The key idea for the application of MCMC techniques to volatility models is basically due to Tanner & Wong (1987) in a context of missing data. This framework is also known as *data augmentation*. Since volatilities and jumps are not observed, they are treated as missing data in an MCMC algorithm where the objective distribution is $\pi(\boldsymbol{\theta}, \boldsymbol{V}, \boldsymbol{J}|\boldsymbol{Y})$. The vector $\boldsymbol{V}$ represents the stochastic volatilities, $\boldsymbol{J}$ is the jump process and $\boldsymbol{\theta}$ is the set of the parameters. The technique aims at estimating the distribution $p(\boldsymbol{\theta}|\boldsymbol{Y})$.
Unfortunately direct inference is not possible since the processes $\boldsymbol{V}$ and $\boldsymbol{J}$ are latent and an high dimensional integration scheme should be applied in order to evaluate the above density. However, by generating the process $\boldsymbol{V}$ and $\boldsymbol{J}$ from $p(\boldsymbol{V}|\boldsymbol{Y})$ and $p(\boldsymbol{J}|\boldsymbol{Y})$ respectively, we can evaluate $p(\boldsymbol{\theta}|\boldsymbol{Y})$ as the sample expectation of $\pi(\boldsymbol{\theta}|\boldsymbol{V}, \boldsymbol{J}, \boldsymbol{Y})$. Obviously there exists a mutual dependence among $p(\boldsymbol{V}|\boldsymbol{Y})$, $p(\boldsymbol{J}|\boldsymbol{Y})$ and $p(\boldsymbol{\theta}|\boldsymbol{Y})$. One possible way to take this dependence into account is to use a "one component at a time Metropolis-Hastings" scheme, simulating recursively from the distribution $p(V_{\tau_i}|\boldsymbol{\theta}, \boldsymbol{V}_{-\tau_i}, \boldsymbol{J}, \boldsymbol{Y})$, $p(J_{\tau_i}|\boldsymbol{\theta}, \boldsymbol{V}, \boldsymbol{J}_{-\tau_i}, \boldsymbol{Y})$, $i = 1, \dots, n$, and the distribution $p(\boldsymbol{\theta}|\boldsymbol{V}, \boldsymbol{J}, \boldsymbol{Y})$.
This approach was introduced by Carlin, Polson & Stoffer (1992) in state space modeling and applied to stochastic volatility models by Jacquier et al. (1994), Jacquier, Polson & Rossi (1999) and Kim et al. (1998). The main inconvenient of the method is the high dimensionality of the latent process. It is then necessary to design an efficient and fast scheme in order to update it effectively.

3.1 Adaptive strategies and Delayed Rejection

Goal of MCMC is to generate a Markov Chain with given invariant distribution $\pi(x)$ and then estimate $\mathbb{E}_\pi[g(X)]$ averaging along the trajectories obtained. A possible criterion used to evaluate the goodness of an estimator is based on the asymptotic variance

$$V(g, P) = \lim_{n\to\infty} n\mathrm{Var}_\pi[\overline{g}_n] = \mathrm{Var}_\pi[g(X_0)] + 2\sum_{n=1}^{\infty} \mathrm{Cov}_\pi[g(X_0, X_n)] \qquad (20)$$

It is evident that dependencies due to the Markov structure of the process can cause troubles on evaluating the expected value stated above. To give a simple example,

suppose that the realization of the chain can be approximated by an AR(1) process with autocorrelation ρ. Then the asymptotic variance of the sample average is $\frac{\sigma^2}{N}\frac{1+\rho}{1-\rho}$, where σ^2 is the variance of $g(X)$ under $\pi(x)$. It is evident that reducing the autocorrelation of $\{X_t : t = 1, \ldots, N\}$ leads to an improvement on the efficiency of the estimates. A challenging task consists on defining an appropriate scheme in order to do it. It is not easy to decide what strategy to adopt. This is because it is not always easy to define in a unique way an optimality criterion that can help in the choice. In fact different rules can lead to different decisions. For example there are situations in which the algorithm quickly converge to the limit distribution and at the same time the estimate has high asymptotic variance (and vice versa). To be more precise, suppose for example a discrete state space with transition probability described by a matrix P. Small eigenvalues of P indicate small asymptotic variance. On the other side, small eigenvalues in absolute value are associated to high speed of convergence. It is evident that eigenvalues close to -1 indicate that the chain produce estimate with low asymptotic variance, but contemporaneously converge badly to the target distribution.
In literature there are different criteria used to order different Markov chains with common invariant law. The asymptotic variance of the estimator defined in eq. (20) is one option.
Suppose there are two Markov chains with transition probabilities P and Q but with the same invariant law π. A strong ordering claims that P is preferred to Q if

$$V(g, P) \leq V(g, Q) \qquad \forall g \in L^2(\pi). \tag{21}$$

A weaker condition is

$$V(g, P) \leq V(g, Q) \tag{22}$$

for a given function g. A different notion of optimality has been introduced in Peskun (1973) for chains with discrete states and extended to general state spaces in Tierney (1998). It is the so called Peskun optimality that says that P dominate Q if

$$P(x, A) \geq Q(x, A), \qquad x \notin A, \qquad \forall A \in \mathcal{F} \tag{23}$$

This condition basically 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. In the Metropolis-Hastings framework, this can be intuitively intended that P allows less rejections than Q for every set A and given the present state x. Lemma 3 and Theorem 4 in Tierney (1998) prove that Peskun optimality imply the ordering defined in eq. (21).
Following the intuition, a good idea for checking the Peskun optimality is to consider the number of time the chain get stuck at the same position. It is in fact intuitive to think that if a chain rejects just few proposals, then it should converge rapidly to π. Unfortunately, taking into account the number of rejection as benchmark can be misleading. If an independence proposal is used, then an high number of acceptances is appreciated, due to the close relation between independence chain and importance sampler. On the other side, if a random walk proposal is suggested, then an optimal

rate of acceptance should fall between $20 - 40\%$ (Roberts, Gelman & Gilks 1997)[3]. In practice, for many empirical applications, it is not possible to identify a practical test that orders different algorithms. For this reason an ex-post comparison among them has to be done.

In order to improve the efficiency of the simple Metropolis-Hastings procedure, a number of techniques have been proposed in literature. They are in general called *adaptive methods*. One possible way to do adaptation is exploiting the entire history of the chain. Gilks, Roberts & Shau (1998) propose to update the shape of the proposal at every regeneration time of the process. Following a similar philosophy, Gåsemir (2003) use the burn in period to find a proposal that minimize the distance to the target distribution. In both cases the chain is no more Markov, since in general the new parameterization for the proposal depends on the entire past history. These approaches suffer for some problems. It is in fact hard to verify the sufficient conditions that guarantee the convergence. The other drawback is that it is not easy to extend the methodology to multivariate problems.
A different way to do adaptation is to use the information given by previous rejections during the run of the algorithm. Examples are the Adaptive Rejection Sampling (Gilks & Wild 1992, ARS) and the Adaptive Rejection Metropolis Sampling (Gilks, Best & Tan 1995, ARMS) where the rejected candidates are used to build a more accurate envelope function for the conditional distribution. The main drawback in ARS and ARMS is that they are cumbersome from a computational point of view, mainly because they need to evaluate the conditional distribution at many points on the domain. This can lead to very slow codes if the procedure has to be replicated thousand of time at each sweep of the algorithm. This is what happens for stochastic volatility models.

From a computational point of view the *Delayed Rejection* method seems very promising. Goal of the Delayed Rejection is to increase the rate of acceptance at each sweep of the algorithm in order to reach the Peskun optimality.
The basic idea is, in case of a rejection, to resample the new state of the chain from a different proposal, taking into account the information contained at the previous stage. 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 interesting to note that the method allows a lot of flexibility, since there are not constraints for the choice of the proposal. Furthermore this way of operating maintains the Markov property of the chain. In the general case Tierney & Mira (1999) and Mira (2002) prove that for each sub-step of the MH algorithm the acceptance probability $\alpha_i(x, y_1, \dots, y_i)$ that maintain the

[3]In that paper the concept of optimal signify that minimize the asymptotic variance.

reversibility of the Markov chain is

$$\alpha_i(x, y_1, \ldots, y_i) = \left\{ \frac{\pi(y_i) q_1(y_i, y_{i-1}) q_2(y_i, y_{i-1}, y_{i-2}) \ldots q_i(y_i, y_{i-1}, \ldots, x)}{\pi(x) q_1(x, y_1) q_2(x, y_1, y_2) \ldots q_i(x, y_1, y_i)} \right.$$
$$\left. \frac{[1 - \alpha_1(y_i, y_{i-1})][1 - \alpha_2(y_i, y_{i-1}, y_{i-2})] \ldots [1 - \alpha_{i-1}(y_i, \ldots, y_1)]}{[1 - \alpha_1(x, y_1)][1 - \alpha_2(x, y_1, y_2)] \ldots [1 - \alpha_{i-1}(x, y_1, y_{i-1})]} \right\} \wedge 1. \quad (24)$$

Unfortunately, there is no way to prove that the sequence of the α_i is increasing in i. For this reason a maximal number of trial has to be set. It can be deterministic or stochastic. The Delayed Rejection algorithm can be synthesized in the following way

`Delayed Rejection algorithm`

1. $y \sim p_1(x, y)$ is rejected in `MH`
2. For $i = 2, \ldots, k$
   - 2a. Withdraw a new candidate y_i from $q_i(x, y, y_1, \ldots y_k)$.
   - 2b. Accept y_i with probability $\alpha_i(x, y_1, \ldots, y_i)$ given by (24).
   - 2c. If rejection and $i < k$, then $i = i + 1$ otherwise $X_{t+1} = y_i$.
3. $X_{t+1} = y_i$ or $X_{t+1} = X_t$.

In this paper we generate the volatility path with a procedure based on 3 stages. In our empirical analysis it seems that the algorithm mix properly. At the same time the choice guarantees a good computational speed.
At the first step the proposal distribution is based on an Independence Chain. We exploited the idea suggested in Eraker (2001) that use the concept of Brownian bridge through $V_{\tau_{i-1}}$ and $V_{\tau_{i+1}}$. It is easy to prove that shrinking the length of the discretization interval make possible to well approximate the conditional posterior, or at least the diffusive part of it. This allows to write $q(x, y)$ as

$$q(V_{\tau_i} | V_{\tau_{i-1}}, V_{\tau_{i+1}}, \boldsymbol{\theta}) \sim \mathcal{N}(\mu_{\tau_i}, \sigma^2_{\tau_i}) \quad (25)$$

where $\mu_{\tau_i} = (V_{\tau_{i+1}} + V_{\tau_{i-1}})/2$ and $\sigma^2_{\tau_i} = \sigma^2_v(1 - \rho^2)V_{\tau_{i-1}}/2$. At the second step we decide to consider a random walk proposal with the same variance used at the previous step. Finally at the third step we use another random walk step. We would like to stress that the acceptance probabilities among the second and the third step are different due to the adaptive nature of the algorithm.
Combining the strategies allows 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 rejection, that means a poor approximation of the right distribution, a random walk proposal gives a control on this bad behavior.
To guarantee the convergence of the chain it is sufficient to check sufficient conditions for each stage. Some of that conditions are stated in Tierney (1994) and in Mengersen & Tweedie (1996) for example. The procedure adopted can be synthesized as follows

Data Augmentation through Metropolis-Hastings

1. Update $V_{\tau_i}, i = 1, \dots, n$ from $p(V_{\tau_i}|\boldsymbol{J}, \boldsymbol{\theta}, \boldsymbol{V}_{-\tau_i}, \boldsymbol{H}^v, \boldsymbol{H}^y)$ via DR.
2. Update the jump times J_{τ_i}, $i = 1, \dots, n$ from $p(J_{\tau_i}|\boldsymbol{V}, \boldsymbol{\theta}, \boldsymbol{H}^v, \boldsymbol{H}^y)$.
3. Update the sizes $H^v_{\tau_i}$, $i = 1, \dots, n$ from $p(H^v_{\tau_i}|\boldsymbol{V}, \boldsymbol{\theta}, \boldsymbol{J}_t, \boldsymbol{H}^y)$.
3. Update the sizes $H^y_{\tau_i}$, $i = 1, \dots, n$ from $p(H^y_{\tau_i}|\boldsymbol{V}, \boldsymbol{\theta}, \boldsymbol{J}, \boldsymbol{H}^v)$.
4. Update $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{-k}, \boldsymbol{V}, \boldsymbol{J}, \boldsymbol{H}^y, \boldsymbol{H}^v)$.

In particular for the complete model $\boldsymbol{\theta} = (\mu, \kappa, \theta, \sigma_v, \rho, \lambda, \mu_y, \mu_v, \rho_J, \sigma_y)$. Details for the full conditional distributions are showed in Appendix A.

It is natural to evaluate the latent processes as a by-product of the algorithm. An estimate of the volatility path is given by

$$\widehat{V}_{\tau_i} = \mathbb{E}[V_{\tau_i}|\boldsymbol{Y}] \approx \frac{1}{N}\sum_{j=1}^{N} V^j_{\tau_i}.$$

The same arguing is applicable to the jump times process

$$\widehat{Pr}(J_{\tau_i} = 1) = \mathbb{E}[J_{\tau_i}|\boldsymbol{Y}] \approx \frac{1}{N}\sum_{j=1}^{N} J^j_{\tau_i}$$

where $V^i_{\tau_i}$ and $J^i_{\tau_i}$ are realization of the Markov chain excluding the burn-in-period.

3.2 Model ranking and Bayes factor calculations

Duffie, Pan and Singleton's model, say $(\mathcal{M}_1, \boldsymbol{\theta}_1)$, generalize Bates's, $(\mathcal{M}_2, \boldsymbol{\theta}_2)$, and Heston's, $(\mathcal{M}_3, \boldsymbol{\theta}_3)$. A natural criterion to select a model in Bayesian statistics is through the use of the Bayes factor, defined as the ratio of the posterior to prior odds

$$B_{ij} = \frac{m(\boldsymbol{Y}|\mathcal{M}_i)}{m(\boldsymbol{Y}|\mathcal{M}_j)} = \frac{p(\mathcal{M}_i|\boldsymbol{Y})/p(\mathcal{M}_j|\boldsymbol{Y})}{p(\mathcal{M}_i)/p(\mathcal{M}_j)} \tag{26}$$

where the distribution

$$m(\boldsymbol{Y}|\mathcal{M}_i) = \int p(\boldsymbol{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\mathcal{M}_i)d\boldsymbol{\theta}_i. \tag{27}$$

is the marginal likelihood. The functions $p(\boldsymbol{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)$ and $\pi(\boldsymbol{\theta}_i|\mathcal{M}_i)$ are respectively the likelihood and the prior for the i-th model and $\boldsymbol{\theta}_i$ is the model-specific parameter vector. It is often difficult to evaluate the integral in eq. (27) and many estimation techniques have been proposed in literature. A review of the various alternatives proposed is showed in Kass & Raftery (1995) whereas a survey on Monte Carlo simulations methods is given in Han & Carlin (2001). Since this paper is focused on simulation methods, it seems more coherent to adopt the latter alternative. Furthermore, the use of asymptotic approximations is difficult to apply for models with latent factors.

An extremely flexible and powerful tool is the reversible jump proposed by Green (1995), that allows to move from a model to another through a general Metropolis-Hastings algorithm. The probability $m(\mathcal{M}_i|\boldsymbol{Y})$ can be estimated as the ratio between the number of times the algorithm hit the i-th model and the total number of iteration.
Another appealing strategy is to evaluate the marginal likelihood exploiting the same strategy adopted for the inference. We decided to use the latter approach, because it is computationally efficient. The marginal likelihood can be written as

$$m(\boldsymbol{Y}|\mathcal{M}_i) = \frac{p(\boldsymbol{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\mathcal{M}_i)}{\pi(\boldsymbol{\theta}_i|\boldsymbol{Y}, \mathcal{M}_i)}, \ \forall\ \boldsymbol{\theta}_i \in \text{supp}(\boldsymbol{\theta}_i). \tag{28}$$

In order to evaluate this ratio, it is necessary to find an estimate of $\pi(\boldsymbol{\theta}_i|\boldsymbol{Y}, \mathcal{M}_i)$ and of $p(\boldsymbol{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)$. It is possible to compute $\pi(\boldsymbol{\theta}_i|\boldsymbol{Y}, \mathcal{M}_i)$ by using the method proposed by Chib & Jeliazkov (2001). As stressed in Nicholls & Mira (2003), this is a particular case of the bridge sampling framework (Meng & Wong 1996, Meng & Shilling 2002). The method consists in dividing the parameters vector $\boldsymbol{\theta}_i = (\theta_1, \ldots, \theta_K)$ in blocks and at each step of the algorithm associate to the i-th block, a given value θ_i^*. In this way it is possible to divide the parameter vector into two parts, say $\boldsymbol{\psi}_i^* = (\theta_1^*, \ldots, \theta_{i-1}^*)$ and $\boldsymbol{\psi}^i = (\theta_{i+1}, \ldots, \theta_K)$. At this point Chib & Jeliazkov (2001) estimate

$$\hat{\pi}(\boldsymbol{\theta}_i|\boldsymbol{Y}, \mathcal{M}_i) = \prod_{i=1}^{K} \hat{\pi}(\theta_i^*|\boldsymbol{Y}, \theta_1^*, \ldots, \theta_{i-1}^*) \tag{29}$$

where

$$\hat{\pi}(\theta_i^*|\boldsymbol{Y}, \theta_1^*, \ldots, \theta_{i-1}^*) = \frac{\hat{\mathbb{E}}_1\left[\alpha(\theta_i, \theta_i^*|\boldsymbol{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i) q(\theta_i, \theta_i^*|\boldsymbol{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)\right]}{\hat{\mathbb{E}}_2\left[\alpha(\theta_i^*, \theta_i|\boldsymbol{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)\right]} \tag{30}$$

Here the numerator is the expected value with respect to $\pi(\theta_i, \boldsymbol{\psi}^i|\boldsymbol{Y}, \boldsymbol{\psi}_i^*)$ and the denominator with respect to $\pi(\boldsymbol{\psi}^i|\boldsymbol{Y}, \boldsymbol{\psi}_i^*) q(\theta_i^*, \theta_i|\boldsymbol{Y}, \boldsymbol{\psi}_i^*, \boldsymbol{\psi}^i)$. The quantities $q(\theta, \theta'|\boldsymbol{Y}, \boldsymbol{\psi}_i, \boldsymbol{\psi}^i)$ and $\alpha(\theta, \theta'|\boldsymbol{\psi}_i, \boldsymbol{\psi}^i)$ are respectively the proposal and the acceptance probability used in a standard Metropolis-Hastings scheme. The expected values can be estimated exploiting the output of a run of the MH, using the *reduced* full conditional distributions described on the right hand side of eq. (29).

3.2.1 Particle filter and likelihood evaluation.

In many problems the likelihood function $p(\boldsymbol{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)$ is known in closed form and then the Bayes factor can be easily evaluated once the posterior is estimated. Unfortunately, in models with latent factors $\boldsymbol{X}$, the non observable variables have to be integrated out. In fact the likelihood is

$$p(\boldsymbol{Y}_T|\boldsymbol{\theta}_i, \mathcal{M}_i) = \prod_{t=1}^{T} p(Y_t|\boldsymbol{Y}_{t-1}, \mathcal{M}_i) \tag{31}$$

where $p(Y_t|\boldsymbol{Y}_{t-1}, \mathcal{M}_i) = \int_{\boldsymbol{X}} p(Y_t|X_t, \boldsymbol{\theta}_i, \mathcal{M}_i)p(X_t|\boldsymbol{Y}_{t-1}, \mathcal{M}_i)dX_t$.
Chib, Nardari & Shephard (2002) propose the use of an auxiliary particle filtering procedure (Pitt & Shephard 1997) for a stochastic volatility model with jumps.
A good introduction to sequential Monte Carlo methods can be found in Doucet, de Freitas & Gordon (2001). See also Maskell & Gordon (2001) for a tutorial style explanation. Aim of a filter is to perform an on-line inference for a latent process, also called state or signal.

In general, to apply the procedure, the model has to be described by the initial state distribution $p(X_{\tau_0})$, by the transition equation $p(X_{\tau_{i+1}}|X_{\tau_i}),\ i \geq 0$ and by the measurement equation $p(Y_{\tau_i}|X_{\tau_i}),\ t \geq 1$.

The key idea is to approximate the filtering density $p(X_{\tau_{i+1}}|\boldsymbol{Y}_{\tau_{i+1}})$ by a discrete cloud of points, $\{X^j_{\tau_{i+1}} : j = 1, \dots M\}$, called particles. This allows to estimate $p(X_{\tau_{i+1}}|\boldsymbol{Y}_{\tau_{i+1}})$ by

$$\hat{p}(X_{\tau_{i+1}}|\boldsymbol{Y}_{\tau_{i+1}}) = \sum_{j=1}^{M} \omega^j_{\tau_{i+1}} \delta(X_{\tau_{i+1}} - X^j_{\tau_{i+1}}) \tag{32}$$

where $\omega^j_{\tau_{i+1}}$ are suitable weights and $\delta(\cdot)$ is a kernel function. All the necessary to do at this point is to draw a sample from (32).

The simpler way is to recur to an importance sampling procedure where the proposal $q(X_{\tau_{i+1}}|\boldsymbol{Y}_{\tau_{i+1}})$ can be set equal to $q(X_{\tau_{i+1}}|X_t, Y_{\tau_{i+1}})$ for example. Given a sample $X^j_{\tau_i},\ j = 1, \dots, M$ from $\hat{p}(X_{\tau_i}|\boldsymbol{Y}_{\tau_i})$ it is easy to prove (see for example Maskell & Gordon 2001) that the weights are

$$\omega^j_{\tau_{i+1}} \propto \omega^j_{\tau_i}\ \frac{p(Y_{\tau_{i+1}}|X^j_{\tau_{i+1}})p(X^j_{\tau_{i+1}}|X^j_{\tau_i})}{q(X^j_{\tau_{i+1}}|X^j_{\tau_i}, Y_{\tau_{i+1}})}, \qquad j = 1, \dots, M \tag{33}$$

It is often convenient to choose $q(X_{\tau_{i+1}}|X_t, Y_{\tau_{i+1}}) = p(X_{\tau_{i+1}}|X_{\tau_i})$ that simplify equation (33) in $\omega^j_{\tau_{i+1}} \propto \omega^j_{\tau_i}\ p(Y_{\tau_{i+1}}|X^j_{\tau_{i+1}})$.

Sometime the simple particle filter is not flexible enough in practical applications and the introduction of auxiliary variables seems to improve the performance of the algorithm.
In the *Auxiliary Particle Filter* (Pitt & Shephard 1997) a new filtering density is introduced. In practice, the goal of the auxiliary filter is to estimate the *augmented* state (X_{t+1}, i) in which i indexes the particle at time τ_i. Since the index variable is just instrumental, it is discarded after its use. The new filtering density is $p(X_{t+1}, i|\boldsymbol{Y}_{\tau_{i+1}})$. The augmented distribution can be expressed, up to proportionality, as

$$p(X_{\tau_{i+1}}, i|\boldsymbol{Y}_{\tau_{i+1}}) \propto p(Y_{\tau_{i+1}}|X_{\tau_{i+1}})p(X_{\tau_{i+1}}|X^i_{\tau_i})\omega^i_{\tau_i}. \tag{34}$$

The proposal associated to (34) is assumed to be $q(X_{\tau_{i+1}}, i|\boldsymbol{Y}_{\tau_{i+1}})$ and in general can be decomposed as

$$q(X_{\tau_{i+1}}, i|Y_{\tau_{i+1}}) \propto p(Y_{\tau_{i+1}}|\bar{X}^i_{\tau_{i+1}})p(X_{\tau_{i+1}}|X^i_{\tau_i})\omega^i_{\tau_i} \tag{35}$$

where $\bar{X}^i_{\tau_{i+1}}$ is some characterization of $p(X_{\tau_{i+1}}|X^i_{\tau_i})$ and can be a likely value for the distribution, the expectation or a draw for example. It is possible to

prove that the weights associated to $(X^j_{\tau_{i+1}}, j^i),\ j = 1, \ldots, M$ are proportional to $p(Y_{\tau_{i+1}}|X^j_{\tau_{i+1}})/p(Y_{\tau_{i+1}}|\bar{X}^{ij}_{\tau_{i+1}})$. An i.i.d. sample is obtained by adding a resampling step (see Smith & Gelfand 1992 for an introduction). This imply that $\omega^j_{\tau_{i+1}} = 1/M$.

We now apply this methodology to the affine jump diffusion model described by the system (11)-(12) in its version with contemporaneous jumps.

In order to avoid negativity troubles, the logarithmic transform of the volatility is considered, i.e. $\log(V_t) = Z_t$. A general version of Itô's formula for multivariate semimartingales (see Protter 1990, Theorem 33, Ch.2) gives

$$dZ_t = \left[\kappa(\theta e^{-Z_{t-}} - 1) - \frac{1}{2}\sigma_v^2 e^{-Z_{t-}}\right] dt + \sigma_v e^{-\frac{1}{2}Z_t} dW_{2,t} + \log\left(1 + H^V_t e^{Z_{t-}}\right) dN_t.$$

With this precaution, the jump size of the marked point process becomes dependent on the state Z_{t-}. For this reason a more general Euler scheme has to be adopted to approximate the continuous trajectories. To be more precise, a particular case of the scheme proposed in Glasserman & Merener (2001) has been applied. The discretely sampled volatility process is

$$Z_{\tau^-_{i+1}} = Z_{\tau_i} + f_0\left(Z_{\tau_i}\right)(\tau_{i+1} - \tau_i) + f_1\left(Z_{\tau_i}\right)(W_{2,\tau_{i+1}} - W_{2,\tau_i}) \tag{36}$$

$$Z_{\tau_{i+1}} = Z_{\tau^-_{i+1}} + \log\left(1 + H^v_{\tau_{i+1}} e^{Z_{\tau^-_{i+1}}}\right)(N_{\tau_{i+1}} - N_{\tau_i}) \tag{37}$$

where $f_0(Z_{\tau_i}) = \kappa(\theta e^{-Z_{\tau_i}} - 1) - \frac{1}{2}\sigma_v^2 e^{-Z_{\tau_i}}$ and $f_1(Z_{\tau_i}) = \sigma_v e^{-\frac{1}{2}Z_{\tau_i}}$. The discrete time return process remain unchanged, apart for the reparameterization of the volatility, that is

$$Y_{\tau_{i+1}} = \mu - \frac{1}{2}e^{Z_{\tau_i}} + e^{\frac{1}{2}Z_{\tau_i}}(W_{1,\tau_{i+1}} - W_{1,\tau_i}) + H^y_{\tau_{i+1}}(N_{\tau_{i+1}} - N_{\tau_i}) \tag{38}$$

The Brownian increments are correlated and $(N_{\tau_{i+1}} - N_{\tau_i})$ is approximately a Binomial random variable $J_{\tau_{i+1}}$ with intensity λ. In practice, a daily approximation has been adopted, that is $\Delta = \tau_{t+1} - \tau_t = (t+1) - t = 1$.

Exploiting the notation introduced in Johannes et al. (2002), the vector of the states is $X_t = (Z_{t-1}, J_t, H^y_t, H^v_t)$. For the purposes of this application, in order to simplify the structure of the state vector, it is convenient to integrate out the jump process. This is not difficult, since a jump at time t is a random variable that assume value $0-1$. One of the differences between the model used in Johannes et al. (2002) and the one analyzed in here, is that the Brownian motions in this paper are taken to be correlated.

The introduction of the correlation parameter induce a dependence between the state vector X_{t+1} and the observations Y_t. In the particle filtering literature this cause the so called feedback effect that is identified by the transition $p(X_{t+1}|X_t, \boldsymbol{Y}_t)$ instead of the usual $p(X_{t+1}|X_t)$. This is a small complication, and in practice it is possible to prove that it does not affect the filtering procedure. In this application the feedback effect is caused by the single observation Y_t and not by the whole history, that is, $p(X_{t+1}|X_t, Y_t)$.

The auxiliary particle filter becomes

`Auxiliary Particle filter`

For $t = 1$ to T:

1. Given a sample $\{(Z_{t-1}, H_t^y, H_t^v)^i : i = 1, \dots, M\}$ from $p(X_t|\boldsymbol{Y}_t)$
2. Calculate $\overline{X}_{t+1}^i$
   $\overline{Z}_t^i = \mathbb{E}[Z_t|Z_{t-1}^i, H_t^{y\ i}, H_t^{v\ i}]$.
   $\overline{H}_{t+1}^{y\ i} = \overline{H}_{t+1}^{v\ i} = 0$.
3. Calculate $\omega_i = \frac{p(Y_{t+1}|\overline{X}_{t+1}^i)}{\sum_{i=1}^M \omega_i}$
4. Resample R times the indexes $1, \dots, M$ with weights ω_i, obtaining $\{m_j : j = 1, \dots R\}$.
5. Draw X_{t+1}^{*j} from $p(X_{t+1}|X_t^{m_j})$.
6. Calculate $\hat{\omega}_i = \frac{p(Y_{t+1}|X_{t+1}^{*j})/p(Y_{t+1}|\overline{X}_{t+1}^{m_j})}{\sum_{i=1}^M \hat{\omega}_i}$.
7. Resample M times the indexes $1, \dots, R$ with weights $\hat{\omega}_i$
8. Store $\{X_{t+1}^i : i = 1, \dots, M\}$ from $p(X_{t+1}|\boldsymbol{Y}_{t+1})$
9. Back to 2.

Once the states are filtered, it is immediate to evaluate the likelihood function by

$$p(\boldsymbol{Y}_T|\boldsymbol{\theta}) = \prod_{t=1}^{T} \hat{p}(Y_t|\boldsymbol{Y}_{t-1}) \tag{39}$$

where $p(Y_t|\boldsymbol{Y}_{t-1}) = \int p(Y_t|X_t, \boldsymbol{\theta}) p(X_t|\boldsymbol{Y}_{t-1}) dX_t$ can be estimated through a Monte Carlo procedure, 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 outcome of the filtering procedure at time $t-1$.
The estimated likelihood is needed to compute the Bayes factor.

4 An Application to Financial Indexes

The empirical application is based on financial indexes, observed on a daily basis during the last 15 years. The datasets have been downloaded from `Datastream`. The series are the FTSE 100 index, the Standard & Poor's 500 composite and the Dow Jones 65 composite (see Table 1). As before, letting S_t the observed price, the returns are defined as $100 \times [\log S_t - \log S_{t-1}]$. The descriptive statistics in Table 2 report the per cent annualized means and volatilities. They are obtained by multiplying the sample mean and the sample standard error by 252 and $\sqrt{252}$ respectively. The annualized volatility for all the indexes lie between the 15 and the 17 percent. Furthermore, it is evident that all the series sensibly reject the hypothesis of normality. Daily returns are displayed in Figures 1-3.
All the calculations made in this paper are based on software written by using the `Ox`©3.0 language of Doornik (2001).

4.1 Posterior Analysis

Inference has been performed for the general model described by eq. (18)-(19). Of course, the important particular cases such as the model proposed by Bates (Bates 1996) and the model introduced by Heston (Heston 1993) have been estimated.
In order to check the fairness of the algorithm proposed, a simulated data set has been estimated. It is a time series of $2,000$ observations generated by the contemporaneous jumps stochastic volatility model. Results are reported in Table 3. In general the posterior means are close to the true values. Anyway it seems that the parameters related to the jump sizes are not accurately estimated. This is probably due to the fact that a jump is a rare event and then the number of observations affected by this happening are not many in the entire time series.
In order to control an eventual bad behaviour of the algorithm in some area of the support of the parameters, the chain generated has been perturbed by random and deterministic shocks. Some graphical analysis showed that after the shock, the chain return to its regular paths in few iterations.
Figure 4 evidences the decreasing number of rejections observed when using different proposals and the delayed rejection method. It is evident that the number of rejections for the random walk chain and the independence chain are sensibly higher with respect to the delayed rejection. The estimate of the latent processes fit fairly well the ones generated. This is showed in Figures 5-6.

The analysis on real data is implemented through MCMC. The chain has been run for $50,000$ iterations with a burn-in of $10,000$. This seems an appropriate choice for the models considered. 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 are an estimate of 2π 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 2π 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 2π times the spectral density matrix at frequency zero of interest. Tables 4-6 report the posterior means, together with the Monte Carlo standard error and and the 95% confidence intervals evaluated using the percentiles of the empirical posterior distribution.
From a computational point of view it is interesting to note how the Delayed Rejection algorithm performs. The introduction of the random walk steps sensibly reduce the autocorrelation induced by the Markov structure of the algorithm. The number of rejections is similar to the one observed for the simulated data and basically reduce to the 5% using a Delayed Rejection based on 3 steps. Figures 7 and 9 show the improvement between the two methods analyzed. The effect is impressive and some empirical studies evidence that it is possible to obtain similar autocorrelations recording just one draw every four or five.

Table 4 evidences the results for the FTSE series. The second column shows the re-

sults obtained for the stochastic volatility model, while the third and the fifth show the results for the models with jumps. For comparative purposes, results based on the random walk algorithm are showed in the fourth column. For the simple stochastic volatility model, the average annualized volatility is $\sqrt{252 \times \theta}$. The estimate of the quantity is 16.9, that is really close to the sample volatility 16.66, as evidenced in Table 2. The parameter κ is the mean reversion of the volatility equation and is 0.024 and basically represents the time the process come back to its expected value level. The leverage effect between processes is mild: this finding is slightly different with previous results stylized in literature (see for example Eraker et al. 2003), but on the other side the period of reference considered here is different. The variance of the volatility process is $\sigma_v^2 V_t$ and then the parameter σ_v, that is 0.131, is important to asses the *volatility of the volatility* behaviour. If the first jump component is taken into account, the volatility decreases, and the same thing happens to the parameter σ_v. This is because the introduction of a jump component explains part of the volatility that in the previous model has been described just by a diffusive process. The parameter λ is 0.02. This means that the model expects 5.3 jumps per year. The annualized spot volatility, i.e. $\sqrt{252(\theta + \mu_v\lambda)/\kappa}$ reduces to 16.1 percent and the annualized total volatility, i.e. the mean square error of the returns that take into account the jump component is 16.94 percent. This latter statistic is computed according to eq. (13) and eq. (16) of Section 2. In percentage terms, the average effect of the jump component on the total volatility is approximately the 8.8 percent. Finally the more general model is analyzed. The introduction of the second jump component sensibly increase the parameter κ and then the mean reversion phenomenon sensibly speeds up. The parameter λ halves. The general model expects just 2.8 jumps per year. The annualized spot volatility sensibly decreases with respect to the other models and is the 15.6 percent. On the other side the total volatility is the 16.2 percent and then the return jump increases it of about 7.4 percent. It is evident that there exists a reduction with respect to the SVJ model. This is due to the extra jump component that itself explains part of the volatility behavior of the complete model. In fact, the contribution of the jump on volatilities shift $\sqrt{V_{t-}}$ to $\sqrt{V_{t-} + H^v}$. The effect is mild with respect to the findings showed in Eraker et al. (2003) and in this work is approximately the 3.3 percent. These differences can be explained by the different time interval considered for the empirical analysis. Eraker et al. (2003) in fact include October, the 19-th 1987 in which a huge crash in the markets has been observed. It is really likely that the single observation heavily influence all the estimates. Figures 10 and 11 show the estimates of the latent processes for the various models. At a first sight they seem equal, but a more detailed analysis evidence the findings stressed by the statistics on the volatilities described before.

Results for S&P 500 are reported in Table 5. For this series, the mean reversion parameter κ increases sensibly by moving from the basic stochastic volatility model to the model with contemporaneous jumps. For the SV and the SVJ it is 0.011 while for the SVCJ is 0.026. At the same time θ, that represents the average of the volatility process, decrease from 1.19 for the SV to 1.4 for SVJ and drop to 0.70 in the SVCJ. The spot annualized volatility for the Heston's model is close to the 22 percent. In the Bates' model this quantity diminish to the 19 percent. The total

volatility for the returns is about the 20.5 percent. This means that the jump effect explains about the 14 percent of the whole volatility. In the complete model the diffusive volatility drops to the 16 percent. The total volatility increases of just 1 point. Even in this case the average volatility jump's size is negligible and is equal to 1.6 percent. It is interesting to note that for this data set the sensible change for λ. It moves from 0.06 to 0.02. The estimated volatilities are showed in Figure 12 while the probabilities of jump are showed in Figure 13.

Finally results for the Dow Jones index are reported in Table 6. The findings for this time series are similar to the results obtained for the others. The introduction of the jump process reduce the diffusive volatility component. As usual this is evidenced by the parameters θ and σ_v. As before the introduction of the jumps slightly increases the mean reversion κ. This series anyway seems more regular than the others. The parameter λ does not change much when the two models with jumps are chosen and the estimate is 0.003 for the Bates' model and is 0.004. In general, for all the three models the volatility is low (6.5 percent) and it seems it is not affected by the jumps. The estimates of the latent processes are showed in Figures 14-15.

4.2 Model ranking

The contemporaneous jumps stochastic volatility model proposed in Duffie et al. (2000), say $(\mathcal{M}_3, \boldsymbol{\theta}_3)$ encompasses the model proposed in Bates (1996), i.e. $(\mathcal{M}_2, \boldsymbol{\theta}_2)$ and the model proposed in Heston (1993), that is $(\mathcal{M}_1, \boldsymbol{\theta}_1)$. A standard practice is to compute the Bayes factor in order to rank the various competing models. As stressed in Section 3.2 the ratio is defined as

$$B_{ij} = \frac{p(\boldsymbol{Y}|\mathcal{M}_i)}{p(\boldsymbol{Y}|\mathcal{M}_j)} \qquad i = 1, 2, 3 \tag{40}$$

where

$$m(\boldsymbol{Y}|\mathcal{M}_i) = \frac{p(\boldsymbol{Y}|\mathcal{M}_i, \boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i|\mathcal{M}_i)}{\pi(\boldsymbol{\theta}_i|\boldsymbol{Y}, \mathcal{M}_i)}, \ \forall\ \boldsymbol{\theta}_i \in \text{supp}(\boldsymbol{\theta}_i). \tag{41}$$

The prior among models is uniform, i.e. $p(\mathcal{M}_i) = p(\mathcal{M}_j) = 1/3, \forall i, j$. The details of the procedure adopted here for the models considered has been showed in Section 3.2.1.

The estimates of the Bayes factors are reported in Tables 7-8. According to the thresholds defined in Kass & Raftery (1995), in general the complete model is strongly preferred to the others.
For the FTSE series, the ordering among the three models is remarkable. The model $(\mathcal{M}_3, \boldsymbol{\theta}_3)$ is strongly preferred to the others. Again, $(\mathcal{M}_2, \boldsymbol{\theta}_2)$ is better than the simple stochastic volatility $(\mathcal{M}_1, \boldsymbol{\theta}_1)$. The logarithm of the Bayes factors are always superior to 4 and then there is no uncertainty about the ranking.

For the S&P500, SVCJ is preferred to the SVJ, but the evidence is not so strong. In fact the log-Bayes factor is 1.56. Furthermore, the stochastic volatility with jump on the returns is systematically preferred to the plain stochastic volatility.

References

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

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

Bates, D. (1996), 'Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options', *Review of Financial Studies* **9**, 69–107.

Björk, T. (1998), *Arbitrage Theory in Continuous Time*, Oxford University Press, New York.

Black, F. & Scholes, M. (1973), 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy* **81**, 637–654.

Carlin, B., Polson, N. & Stoffer, D. (1992), 'A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling', *Journal of the American Statistical Association* **87**, 493–500.

Carr, P. & Madan, D. (1999), Option Valuation Using the Fast Fourier Transform, Working paper, University of Maryland.

Chib, S. & Jeliazkov, I. (2001), 'Marginal Likelihood from the Metropolis-Hastings Output', *Journal of the American Statistical Association* **96**, 270–281.

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

Cyganowski, S. & Kloeden, P. (2000), MAPLE Schemes for Jump-Diffusion Stochastic Differential Equations, Technical report, Proc. 16th IMACS World Congress, Lausanne 2000, M. Deville and R. Owens eds.

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

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

Duffie, D., Filipović, D. & Schachermayer, W. (2003), 'Affine Processes and Applications in Finance', *Annals of Applied Probability* **13**, 984–1053.

Duffie, D., Pan, J. & Singleton, K. (2000), 'Trasform Anlysis 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., Johannes, M. & Polson, N. (2003), 'The Impact of Jumps in Volatility and Returns', *Journal of Finance* . (forthcoming).

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

Gåsemir, J. (2003), 'On an Adaptive Version of the Metropolis-Hastings Algorithm with Independent Proposal Distribution', *Scandinavian Journal of Statistics* **30**, 159–163.

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

Gilks, W., Best, N. & Tan, K. (1995), 'Adaptive Rejection Metropolis Sampling within Gibbs Sampling', *Applied Statistics* **44**, 455–472.

Gilks, W., Roberts, G. & Shau, S. (1998), 'Adaptive Markov Chain Monte Carlo Through Regeneration', *Journal of the American Statistical Association* **93**, 1045–1054.

Gilks, W. & Wild, P. (1992), 'Adaptive Rejection Sampling for Gibbs Sampling', *Applied Statististics* **43**, 337–348.

Glasserman, P. & Merener, N. (2001), Convergence of a Discretization Scheme for Jump-Diffusion Processes with State-Dependent Intensities, Working paper, Columbia University.

Gourieroux, C., Monfort, A. & Ranault, E. (1993), 'Indirect Inference', *Journal of Applied Econometrics* **85**, 85–118.

Green, P. (1995), 'Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination', *Biometrika* **82**, 711–732.

Han, C. & Carlin, B. (2001), 'MCMC Methods for Computing Bayes Factors: A Comparative Review', *Journal of the American Statistical Association* **96**, 1122–1132.

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.

Jacquier, E., Polson, N. & Rossi, P. (1999), Stochastic Volatility: Univariate and Multivariate Extensions, Working Paper, Scientific Series 99s-26, CIRANO.

Johannes, M., Polson, N. & Stroud, J. (2002), Nonlinear Filtering of Stochastic Differential Equations with Jumps, Working paper, Wharton School, University of Pennsylvania.

Kass, R. & Raftery, A. (1995), 'Bayes Factors', *Journal of the American Statistical Association* **90**, 773–795.

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

Kloeden, P. & Platen, E. (1995), *Numerical Solution to Stochastic Differential Equations*, 2 edn, Springer-Verlag, New York.

Maskell, S. & Gordon, N. (2001), A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking, Working paper, QuinetiQ Ltd.

Meng, X.-L. & Shilling, S. (2002), 'Warp Bridge Sampling', *Journal of Computational & Graphical Statistics* **11**, 552–586.

Meng, X.-L. & Wong, W. (1996), 'Simulating Ratios of Normalising Constants Via a Simple Identity: a Theoretical Explaination', *Statistica Sinica* **6**, 831–860.

Mengersen, K. & Tweedie, R. (1996), 'Rates of Convergence of the Hastings and Metropolis Algorithms', *Annals of Statistics* **24**, 101–121.

Mira, A. (2002), On Metropolis-Hastings Algorithms with Delayed Rejections, Working paper, Università dell'Insubria, Varese.

Nicholls, G. & Mira, A. (2003), Bridge Estiamtion of the Probability Density at a Point, Technical report n. 7, University of Insubria, Varese, Italy. To appear in: Statistica Sinica.

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**, 607–612.

Pitt, M. & Shephard, N. (1997), Filtering Via Simulation: Auxiliary Particle Filter, Working paper, Nuffield College.

Protter, P. (1990), *Stochastic Integration and Differential Equations: a New Approach*, Springer-Verlag, Berlin.

Robert, C. & Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer, Berlin.

Robert, C. P. (2001), *The Bayesian Choice*, 2 edn, Springer-Verlag, New York.

Roberts, G., Gelman, A. & Gilks, W. (1997), 'Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms', *Annals of Applied Probability* **7**, 110–120.

Runggaldier, W. (2003), Jump Diffusion Models, *in* S. Rachev, ed., 'Handbook of Heavy Tailed Distributions in Finance', 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**, 413–424.

Smith, A. & Gelfand, A. (1992), 'Bayesian Statistics without Tears: A Sampling-Resampling Perspective', *American Statistician* **46**, 84–88.

Tanner, M. & Wong, W. (1987), 'Calculation of Posterior Distribution by Data Augmentation', *Journal of the American Statistical Association* **82**, 528–540.

Taylor, S. (1994), 'Modelling Stochastic Volatility: a Review and Comparative Study', *Mathematical Finance* **4**, 183–204.

Tierney, L. (1994), 'Markov Chain for Exploring Posterior Distributions (with discussion)', *Annals of Statistics* **22**, 1701–1762.

Tierney, L. (1998), 'A Note on Metropolis-Hastings Kernels for General State Spaces', *Annals of Applied Probability* **8**, 1–9.

Tierney, L. & Mira, A. (1999), 'Some Adaptive Monte Carlo Methods for Bayesian Inference', *Statistics in Medicine* **18**, 2507–2515.

Williams, D. (1991), *Probability with Martingales*, Cambridge University Press.

A Conditional Posterior Distributions.

We will give the expression of the conditional distribution obtained for the model. In general, to make the notation simpler, I write $a_t = y_t - \mu - H_t^y J_t$, $b_t = V_t - V_{t-1} - \kappa(\theta - V_{t-1}) - H_t^v J_t$ and $\omega_t = \sigma_v^2(1-\rho^2)V_{t-1}$. y_t is the log-return at time t derived as $\log(S_t) - \log(S_{t-1})$. For details on conjugate families see Robert (2001).
The main problem, as widely stressed in previous sections is to simulate the volatility process for which doesn't exist a closed form for the conditional posterior. Its expression is

$$p(V_t|rest) \propto \frac{1}{V_t}\exp\left\{-\frac{\sigma_v^2 a_t^2 + b_t^2 - 2\rho\sigma_v a_t b_t}{2\omega_t} - \frac{\sigma_v^2 a_{t+1}^2 + b_{t+1}^2 - 2\rho\sigma_v a_{t+1} b_{t+1}}{2\omega_{t+1}}\right\} \quad (42)$$

The jump time J_t is a sequence of Bernoulli random variables. In order to calculate the conditional posterior it is sufficient to evaluate $P(J_t = 1|rest)$ and $P(J_t = 0|rest)$ in the following way,

$$P_c(1) = P(J_t = 1|rest) \propto \lambda \exp\left\{\frac{\sigma_v^2 a_{t,1}^2 + b_{t,1}^2 - 2\rho\sigma_v a_{t,1} b_{t,1}}{\omega}\right\} \quad (43)$$

$$P_c(0) = P(J_t = 0|rest) \propto (1-\lambda) \exp\left\{\frac{\sigma_v^2 a_{t,0}^2 + b_{t,0}^2 - 2\rho\sigma_v a_{t,0} b_{t,0}}{\omega}\right\} \quad (44)$$

with $a_{t,i} = a_t, J_t = i$ and $b_{t,i} = b_t, J_t = i$, $i = 0, 1$. The conditional posterior is a Bernoulli with $P(0) = P_c(0)/(P_c(0) + P_c(1))$.
It is very easy to sample from the ξ^Y sequence. After some calculation it is possible to obtain

$$p(H_t^y|J_t = 1, rest) \propto N\left(\frac{\sigma_v^2(y_t - \mu)^2 - \rho\sigma_v b_t + (\mu_y + \rho_j H_t^v)\omega_t/\sigma_y^2)\sigma_y^2}{\sigma_v^2\sigma_y^2 + \omega_t}, \frac{\sigma_y^2\omega_t}{\sigma_v^2\sigma_y^2 + \omega_t}\right) \quad (45)$$

Analogous algebra leads to

$$p(H_t^V|J_t = 1, rest) \propto N\left(\frac{f_t - \rho\sigma_v(y_t - \mu - H_t^y) - \omega_t/\mu_v + \rho_j(H_t^Y - \mu_y)\omega_t/\sigma_y^2}{\sigma_v^2\sigma_y^2 + \omega_t}, \frac{\sigma_y^2\omega_t}{\sigma_y^2 + \rho_j^2\omega_t}\right) \quad (46)$$

where $f_t = V_t - V_{t-1} - k(\theta - V_{t-1})$. If $J_t = 0$ then the posterior conditional simplify to the law of $H_t^i, i = y, v$.
Parameters involved in the diffusion's drifts are easy to sample too. It is trivial to check that μ, k and $k\theta$ are proportional to Normal distributions. The prior used are respectively $N(0, 25)$, $N(0, 1)$ and $N(0, 1)$. θ is obtained as a ratio of the previous two extraction.

Correlation among Brownian errors is explained by the parameter ρ. We assume a prior Uniform in $(-1, 1)$. The conditional posterior is then

$$p(\rho \mid rest) \propto \frac{1}{(\sqrt{1-\rho^2})^n} \exp\Big\{ -\frac{1}{2} \sum_{t=1}^{n} \Big(\frac{\sigma_v^2 a_t^2 + b_t^2 - 2\rho\sigma_v a_t b_t}{\omega_t} \Big) \Big\} \mathbb{I}_{[-1,1]}(\rho) \quad (47)$$

The variance of the error term of the volatility process is σ_v^2. The prior we chose is $U(0, 1)$.

$$p(\sigma_v^2 \mid rest) \propto \frac{1}{\sigma_v^n} \exp\Big\{ -\frac{1}{2} \Big(\sigma_v^2 \sum_{t=1}^{n} \frac{a_t^2}{\omega_t} + \sum_{t=1}^{n} \frac{b_t^2}{\omega_t} - 2\rho\sigma_v \sum_{t=1}^{n} \frac{b_t a_t}{\omega_t} \Big) \Big\} \mathbb{I}_{[0,1]}(\sigma_v) \quad (48)$$

The conditional posterior both for ρ and σ_v^2 seems too complicate to handle directly. For this reason I decided to use ARMS to simulate from them.
The intensity λ is sampled from

$$p(\lambda \mid rest) \sim Beta\Big(2 + \sum_{t=1}^{n} J_t, 40 + n - \sum_{t=1}^{n} J_t \Big) \quad (49)$$

For μ_v, given that the prior is an $IG(\alpha_0, \beta_0)$, $\alpha_0 = 20, \beta_0 = 10$, after some calculation it is possible to show that $p(\mu_v | rest) \propto IG(n + \alpha_0, z + \beta_0)$ with $z = \sum H^v$.
The parameter ρ_j represent the correlation between jump sizes. The prior imposed is $N(0, 4)$. It is easy to show that its conditional distribution is proportional to a Normal distribution $p(\rho_j | rest) \propto N(c/(b + 1/4), 1/(4b + 1))$ in which $e_t = H_t^y - \mu_y$, $b = (\sum_{t=1}^{n} {H_t^v}^2)/\sigma_y^2$ and $c = (\sum_{t=1}^{n} e_t H_t^v)/\sigma_y^2$.
The variance of the conditional distribution of the return jump size is σ_y^2. Imposing an Inverse Gamma prior $IG(\alpha_0 = 5, \beta_0 = 20)$ leads to a conjugate posterior. The conditional posterior is then an Inverse Gamma $IG(\alpha + n/2, \beta + z/2)$, and where $z = \sum_{t=1}^{n} \left(H_t^y - \mu_y - \rho_j H_t^v \right)^2$.
Finally μ_y is Normal with mean $\sum m_t (n/\sigma_y^2 + 0.01)^{-1}$ and variance $(n/\sigma_y^2 + 0.01)^{-1}$ with $m_t = H_t^y - \rho_j H_t^v$, provided that the prior is $N(0, 100)$

B Results

Table 1: Indexes, daily series

Price Index	Symbol	Period
Financial Times Stock Exchange 100	FTSE	4/7/1988- 4/7/2003
Standard & Poor's 500 Composite	S&P500	4/7/1988- 4/7/2003
Dow Jones Composite 65 Stock Ave.	DJC	4/7/1988- 4/7/2003

Table 2: Descriptive statistics for annualized Index returns

	FTSE	S&P 500	DJC
Mean	5.0856	8.4266	7.5844
Volatility	16.6671	16.3609	15.0988
Maximum	5.9025	5.5732	5.3455
Minimum	-5.8853	-7.1127	-8.1497
Skewness	-0.1069	-0.1577	-0.3319
Kurtosis	3.0248	4.1393	5.2012
Jarque-Bera	1494.345	2801.362	4469.79
P-value	(0.0000)	(0.0000)	(0.0000)
Observations	3914	3914	3914

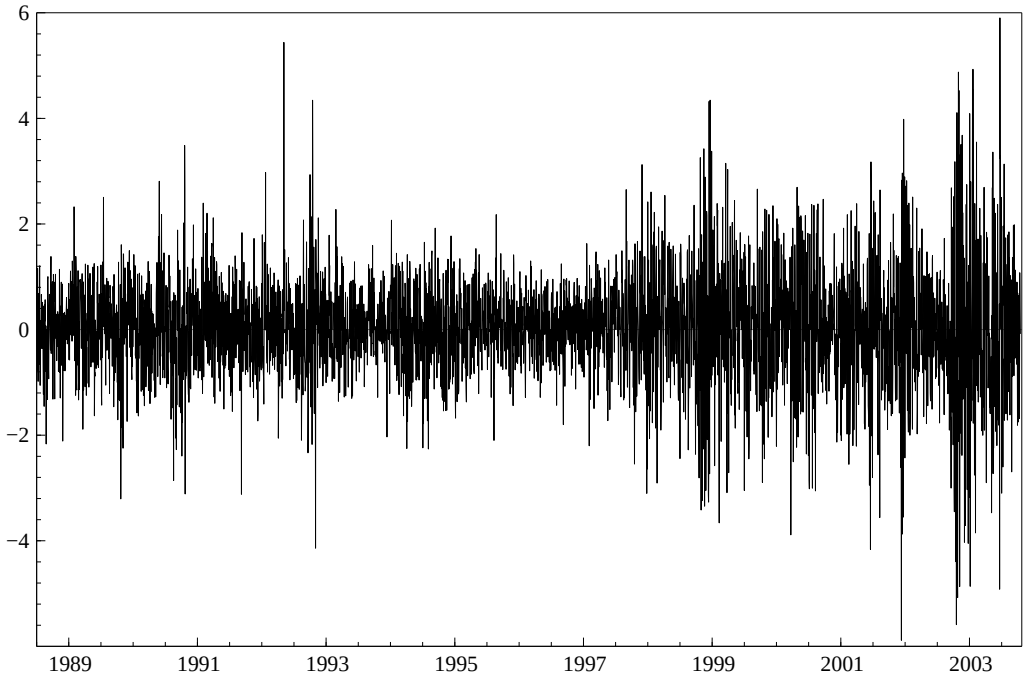

Figure 1: FTSE Returns - July 5,1988 - July 4,2003

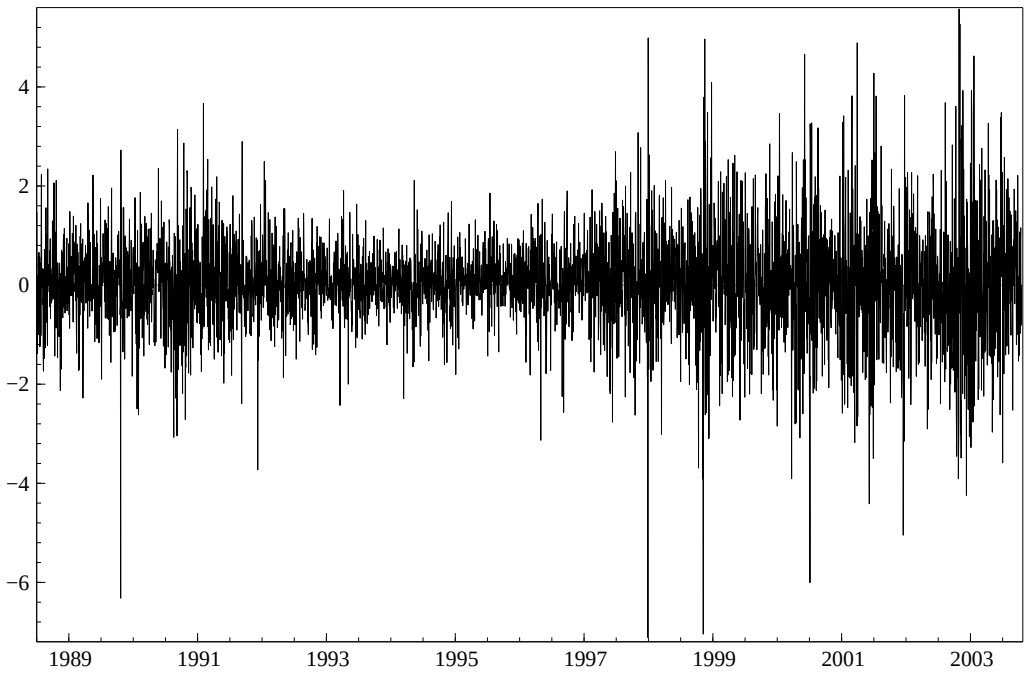

Figure 2: S&P 500 Returns - July 5,1988 - July 4,2003

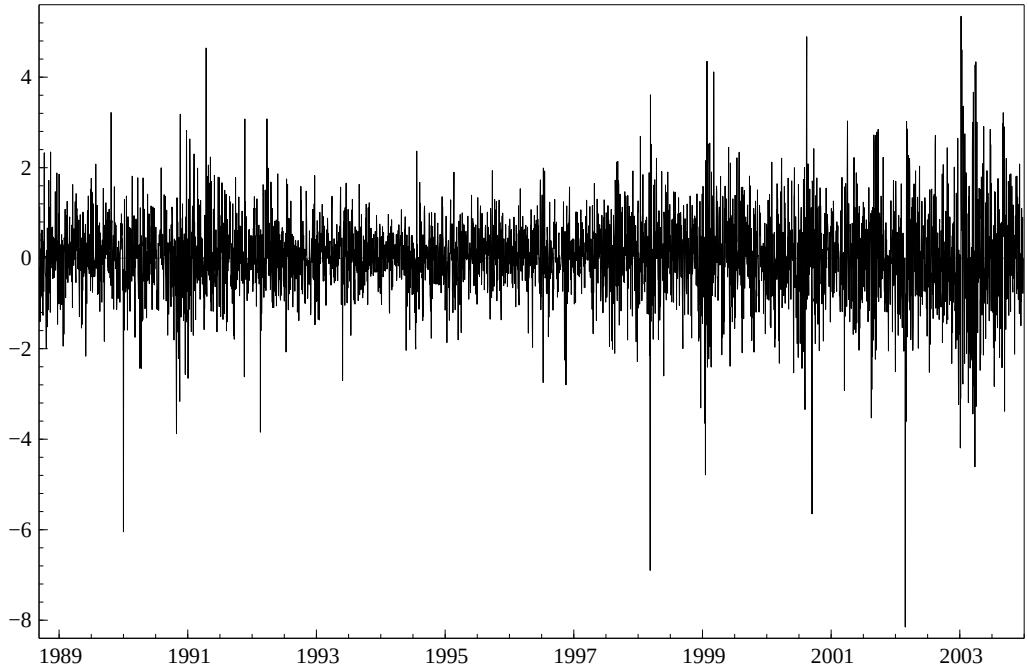

Figure 3: Dow Jones Composite 65 Average Stock Returns - July 5,1988 - July 4,2003

Table 3: *Simulated Data* - The first column show results obtained with *delayed rejection* on 3 stages. The second one the results based on a *Random Walk chain*

	$True$	$D.R.$	$R.W.Chain$
μ	0.05	0.0504	0.0524
		(0.00105)	(0.00264)
		[0.016,0.085]	[0.014,0.089]
κ	0.03	0.0340	0.0385
		(0.00065)	(0.00164)
		[0.023,0.046]	[0.024,0.052]
θ	0.5	0.4742	0.4940
		(0.01013)	(0.00940)
		[0.309,0.682]	[0.340,0.706]
σ_v	0.1	0.1160	0.1183
		(0.00088)	(0.00131)
		[0.103,0.133]	[0.106,0.132]
ρ	-0.5	-0.4208	-0.4491
		(0.0194)	(0.02633)
		[-0.57,-0.22]	[-0.61,-0.28]
λ	0.008	0.0148	0.0159
		(0.00038)	(0.00094)
		[0.008,0.023]	[0.008,0.026]
ρ_j	-0.4	-0.5925	0.9228
		(0.24078)	(0.19715)
		[-1.97,0.870]	[-0.56,2.814]
μ_v	1.0	0.7112	0.6006
		(0.02424)	(0.02439)
		[0.489,1.007]	[0.417,0.901]
μ_y	-2.0	-1.2891	-2.5695
		(0.29979)	(0.2637)
		[-3.02,0.776]	[-4.79,-0.68]
σ_y	3.5	2.9313	2.8996
		(0.09519)	(0.05200)
		[2.199,3.866]	[2.171,3.793]

Figure 4: *Simulated data* - Estimated rejection rate for the algorithm used to infer the model. Dotted line: Independence chain (medium line); Dashed line: Random walk chain (upper line); Solid line: Delayed Rejection (lower line)

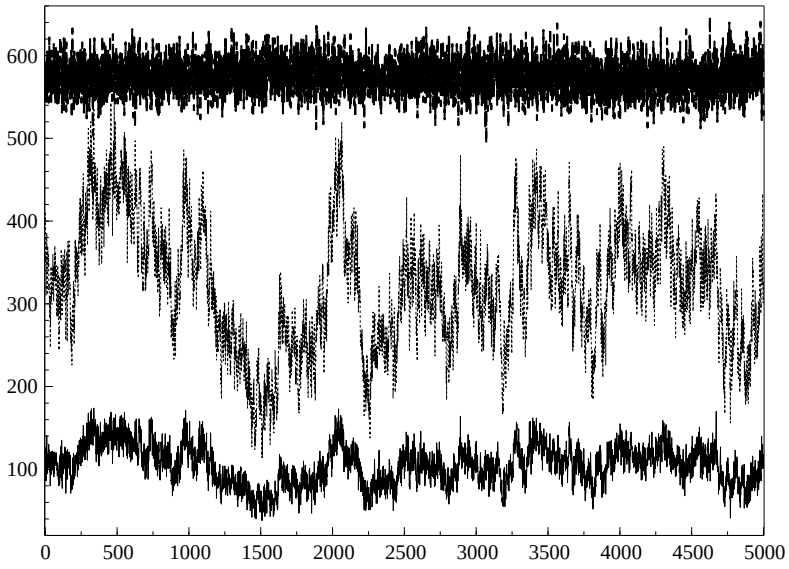

Figure 5: *Simulated data* - True (a) and estimated volatility processes (b).

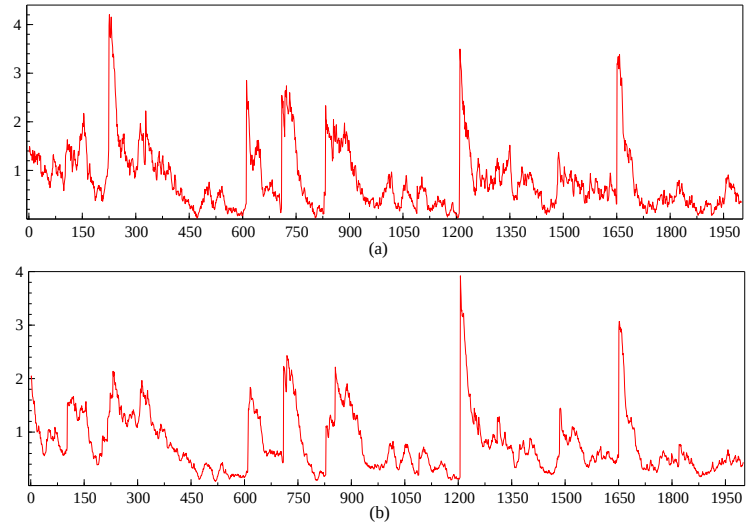

Figure 6: *Simulated data* - True (a) and estimated probability jump processes (b).

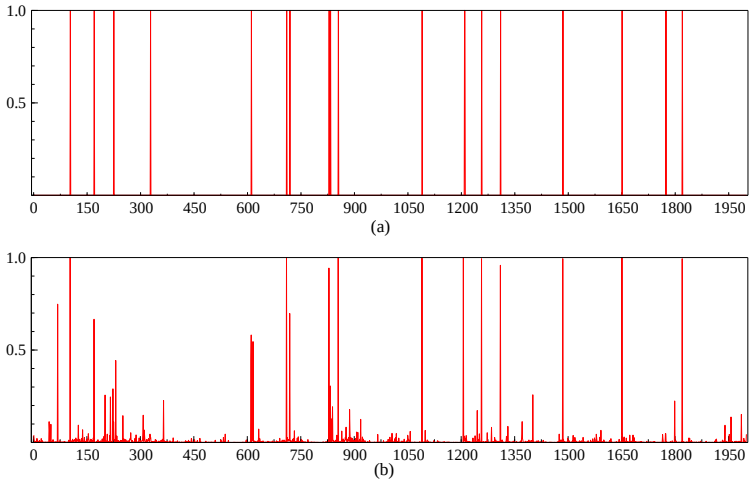

Table 4: *FTSE* - First column show results for the Heston's model. The second one the results for the Bates's model and on the third and fourth the Contemporaneous Jump Stochastic Volatility estimated with a random walk chain and with a delayed Rejection method respectively.

	SV	SVJ	$SVCJRW$	$SVCJ$
μ	0.37804	0.34913	0.35424	0.35776
	(0.00075)	(0.00126)	(0.00124)	(0.00085)
	[0.345,0.410]	[0.311,0.385]	[0.322,0.386]	[0.323,0.390]
κ	0.02470	0.02334	0.02357	0.03360
	(0.00099)	(0.00062)	(0.00084)	(0.00135)
	[0.003,0.060]	[0.004,0.051]	[0.005,0.038]	[0.009,0.063]
θ	1.13738	1.03846	0.85647	0.78465
	(0.02047)	(0.01515)	(0.02280)	(0.01182)
	[0.671,2.564]	[0.620,0.176]	[0.497,1.716]	[0.516,1.296]
σ_v	0.13123	0.12432	0.11115	0.12175
	(0.00185)	(0.00111)	(0.000802)	(0.00137)
	[0.102,0.196]	[0.102,0.176]	[0.103,0.131]	[0.103,0.161]
ρ	-0.07195	-0.07927	-0.14487	-0.10701
	(0.00580)	(0.00643)	(0.01076)	(0.00785)
	[-0.251,0.093]	[-0.259,0.087]	[-0.318,0.025]	[-0.299,0.076]
λ		0.02124	0.01041	0.01141
		(0.00084)	(0.00045)	(0.00044)
		[0.006,0.044]	[0.003,0.025]	[0.004,0.021]
ρ_j			-0.92304	-0.96041
			(0.19650)	(0.14603)
			[-3.964,1.592]	[-3.248,0.830]
μ_v			0.47698	0.55692
			(0.00887)	(0.01292)
			[0.328,0.698]	[0.379,0.815]
μ_y		1.20930	2.27814	2.27270
		(0.05283)	(0.12626)	(0.10749)
		[0.403,2.369]	[0.722,4.132]	[0.667,3.931]
σ_y		1.81939	1.95446	1.95544
		(0.02700)	(0.03273)	(0.03721)
		[1.369,2.494]	[1.353,2.679]	[1.351,2.657]

Table 5: *S&P 500* - First column show results for the Heston's model. The second one the results for the Bates's model and on the third and fourth the Contemporaneous Jump Stochastic Volatility estimated with a random walk chain and with a delayed Rejection method respectively.

	SV	SVJ	$SVCJRW$	$SVCJ$
μ	0.33817	0.27325	0.30755	0.30580
	(0.00066)	(0.00149)	(0.00123)	(0.00099)
	[0.308,0.369]	[0.235,0.308]	[0.274,0.338]	[0.273,0.338]
κ	0.01158	0.01193	0.01833	0.02603
	(0.00055)	(0.00053)	(0.00098)	(0.00127)
	[0.001,0.036]	[0.001,0.032]	[0.002,0.035]	[0.003,0.052]
θ	1.91813	1.43636	0.75231	0.70708
	(0.05952)	(0.05879)	(0.02342)	(0.03275)
	[0.743,6.136]	[0.594,4.876]	[0.354,1.876]	[0.383,1.690]
σ_v	0.12726	0.11381	0.11609	0.12408
	(0.00129)	(0.00091)	(0.00083)	(0.00139)
	[0.104,0.176]	[0.101,0.151]	[0.106,0.133]	[0.106,0.154]
ρ	-0.24906	-0.29625	-0.35679	-0.30640
	(0.00511)	(0.00849)	(0.01179)	(0.00959)
	[-0.392,-0.080]	[-0.478,-0.098]	[-0.501,-0.155]	[-0.497,-0.106]
λ		0.06230	0.02256	0.02467
		(0.00191)	(0.00064)	(0.00077)
		[0.033,0.102]	[0.012,0.036]	[0.011,0.041]
ρ_j			-1.73113	-2.33007
			(0.25020)	(0.31992)
			[-5.300,2.126]	[-6.914,1.852]
μ_v			0.37591	0.37536
			(0.00712)	(0.00649)
			[0.263,0.520]	[0.258,0.517]
μ_y		0.62210	1.26642	1.48429
		(0.01181)	(0.08863)	(0.11648)
		[0.256,1.039]	[-0.161,2.598]	[-0.046,3.361]
σ_y		1.84464	2.39534	2.26498
		(0.01996)	(0.03295)	(0.04678)
		[1.473,2.305]	[1.808,3.070]	[1.464,3.046]

Table 6: *DOW JONES* - First column show results for the Heston's model. The second one the results for the Bates's model and on the third and fourth the Contemporaneous Jump Stochastic Volatility estimated with a random walk chain and with a delayed Rejection method respectively.

	SV	SVJ	$SVCJRW$	$SVCJ$
μ	0.32813	0.30037	0.30865	0.30971
	(0.00055)	(0.00086)	(0.00116)	(0.00063)
	[0.299,0.357]	[0.268,0.330]	[0.277,0.339]	[0.279,0.339]
κ	0.03149	0.02630	0.03458	0.04248
	(0.00089)	(0.00073)	(0.00124)	(0.00134)
	[0.009,0.063]	[0.006,0.058]	[0.014,0.056]	[0.018,0.069]
θ	0.88656	0.83137	0.61888	0.60933
	(0.00829)	(0.01027)	(0.00797)	(0.00702)
	[0.530,1.504]	[0.516,1.599]	[0.412,0.962]	[379,0.885]
σ_v	0.14216	0.12136	0.11647	0.12578
	(0.00153)	(0.00114)	(0.00119)	(0.00114)
	[0.111,0.192]	[0.102,0.159]	[0.104,0.138]	[0.105,0.154]
ρ	-0.15691	-0.19619	-0.25083	-0.20458
	(0.00474)	(0.00781)	(0.00915)	(0.00658)
	[-0.303,-0.015]	[-0.389,-0.018]	[-0.400,-0.084]	[-0.371,-0.043]
λ		0.02512	0.01414	0.01437
		(0.00068)	(0.00048)	(0.00054)
		[0.011,0.045]	[0.006,0.024]	[0.006,0.026]
ρ_j			-2.27543	-2.64816
			(0.22512)	(0.28933)
			[-5.356,0.714]	[-6.365,1.429]
μ_v			0.46714	0.49273
			(0.00986)	(0.00959)
			[0.313,0.673]	[0.330,0.711]
μ_y		0.42276	1.27282	1.52653
		(0.02168)	(0.12008)	(0.16268)
		[-0.219,1.023]	[-0.387,3.012]	[-0.769,3.701]
σ_y		2.26954	2.46479	2.36237
		(0.02554)	(0.05279)	(0.06451)
		[1.732,2.928]	[1.637,3.333]	[1.337,3.574]

Figure 7: *FTSE* - Autocorrelation functions for the averaged volatility according to different sampling schemes.

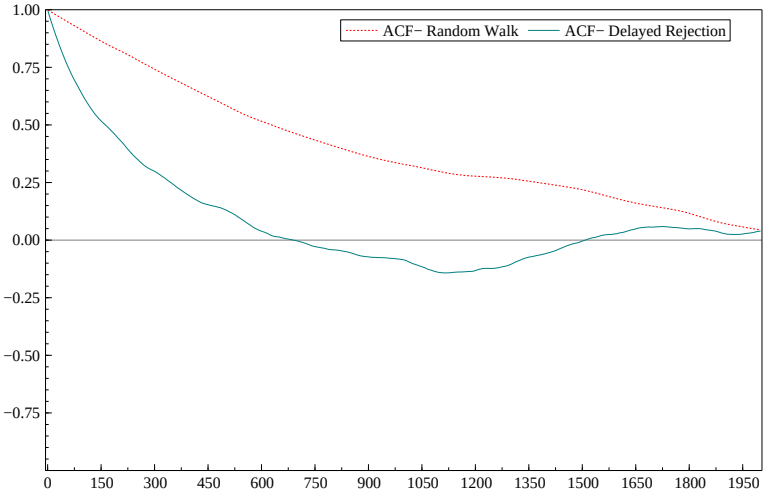

Figure 8: *S&P 500* - Autocorrelation functions for the averaged volatility according to different sampling schemes.

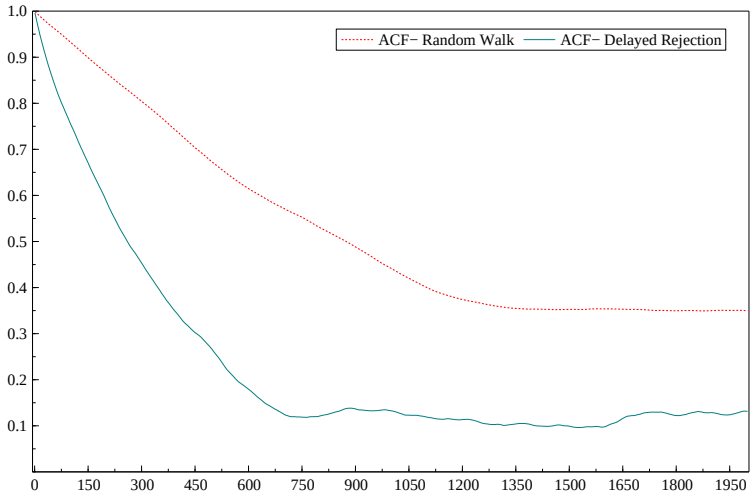

Figure 9: *DOW JONES* - Autocorrelation functions for the averaged volatility according to different sampling schemes.

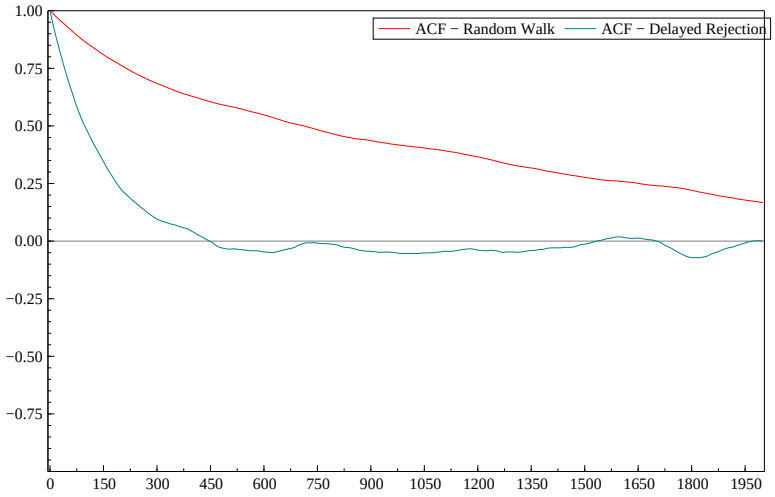

Figure 10: *FTSE* - Estimated volatility processes: (a) SV model; (b) SVJ model; (c) SVCJ model.

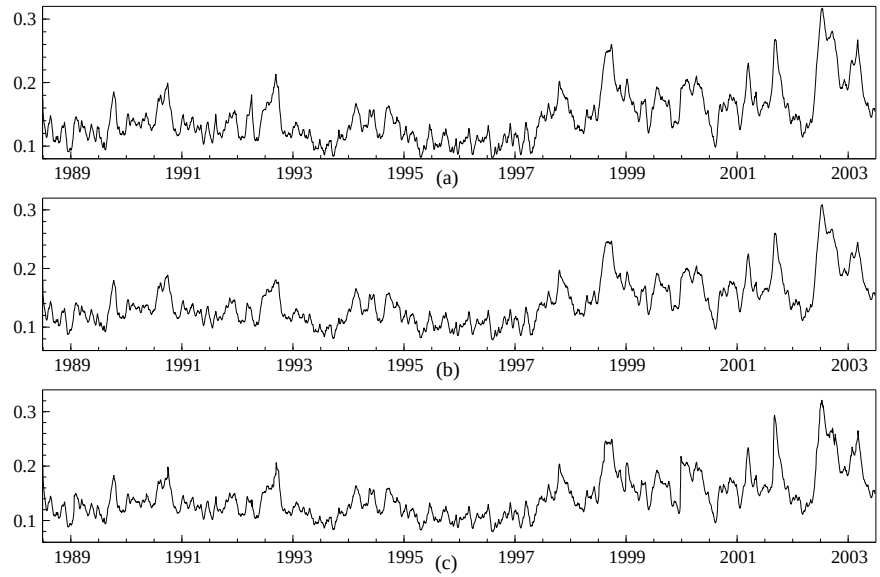

Figure 11: *FTSE* - Estimated probability of jump: (a) SVJ model; (b) SVCJ model.

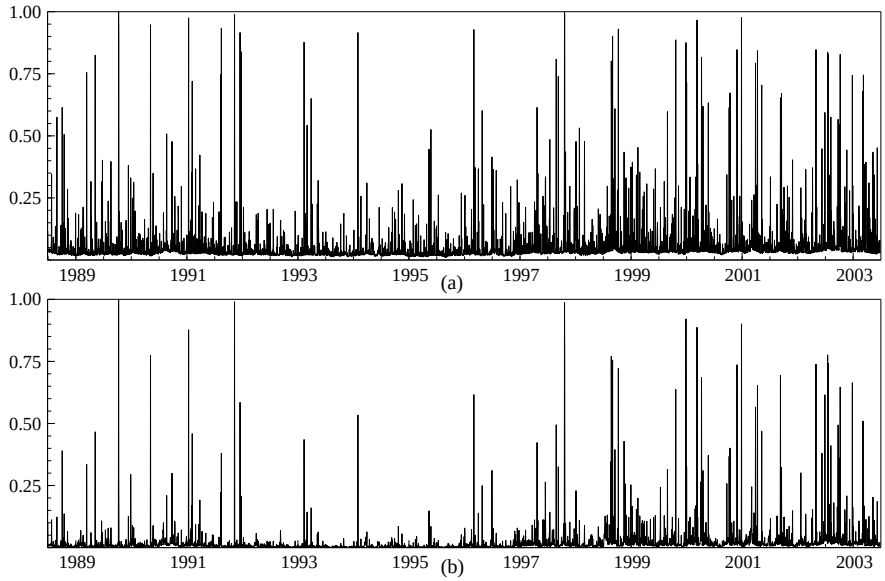

Figure 12: *S&P 500* - Estimated volatility processes: (a) SV model; (b) SVJ model; (c) SVCJ model.

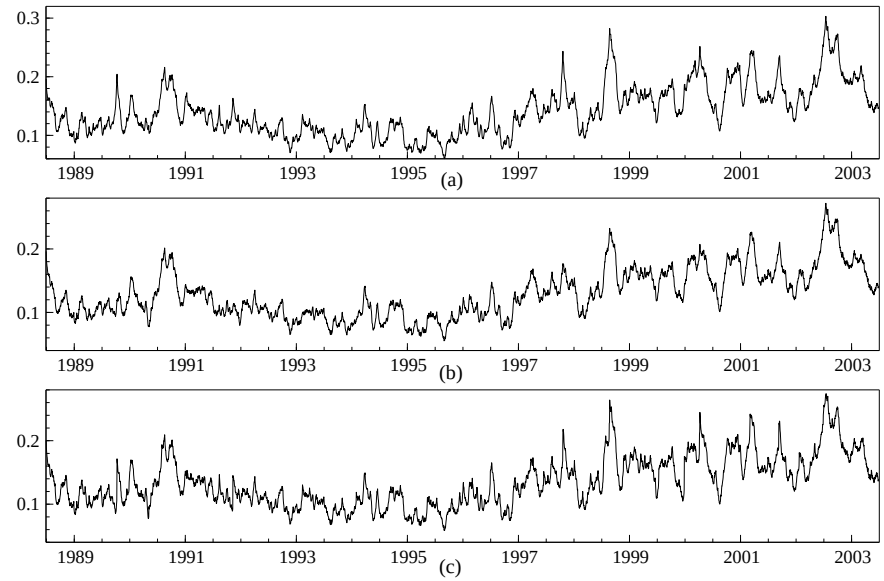

Figure 13: *S&P 500* - Estimated probability of jump: (a) SVJ model; (b) SVCJ model.

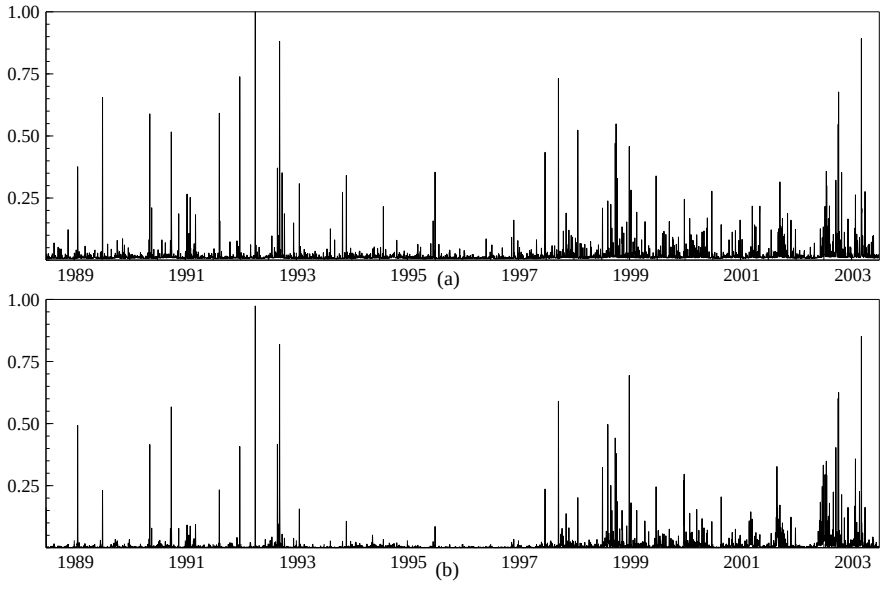

Figure 14: *Dow Jones* - Estimated volatility processes: (a) SV model; (b) SVJ model; (c) SVCJ model.

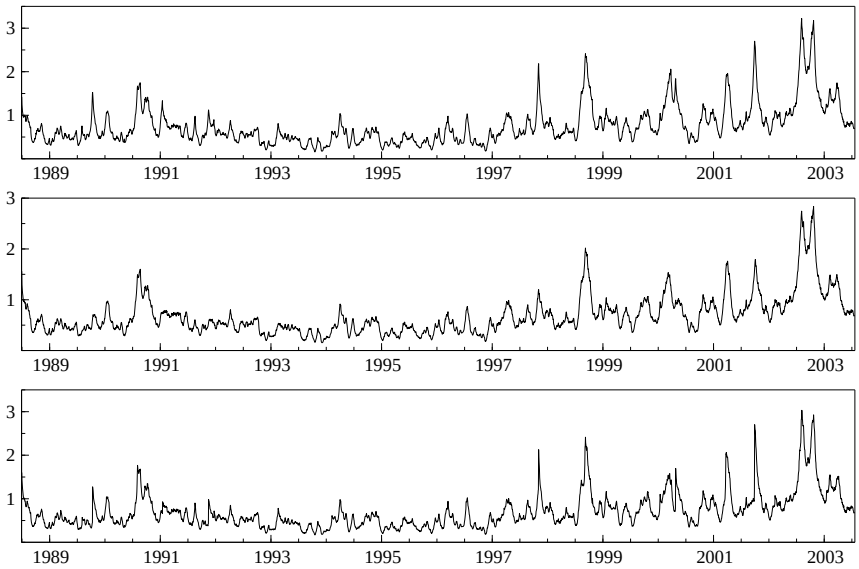

Figure 15: *Dow Jones* - Estimated probability of jump: (a) SVJ model; (b) SVCJ model.

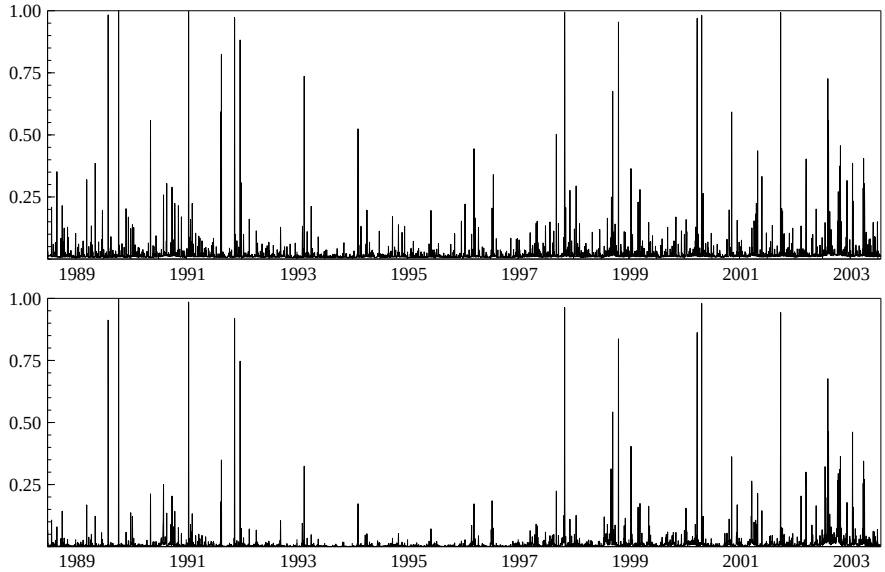

Table 7: *FTSE* - log-Bayes Factor. Entry (i, j) indicates the Bayes factor in favour of model j versus model i.

	SVCJ	SVJ	SV
SVCJ	1		
SVJ	4.8912	1	
SV	17.0307	12.1359	1
log-lik.	-5293.6	-5299.9	-5318.8
Marg. lik.	-5316.7	-5321.6	-5333.7

Table 8: *S&P500* - log-Bayes Factor. Entry (i, j) indicates the Bayes factor in favour of model j versus model i.

	SVCJ	SVJ	SV
SVCJ	1		
SVJ	1.5665	1	
SV	54.3685	52.80192	1
log-lik.	-5124.8	-5127.2	-5186.0
Marg.lik.	-5149.0	-5150.5	-5203.3

Table 9: *DJC* - log-Bayes Factor. Entry (i, j) indicates the Bayes factor in favor of model j versus model i.

	SVCJ	SVJ	SV
SVCJ	1		
SVJ	33.3291	1	
SV	36.8007	3.4716	1
log-lik	-5472.4	-5509.6	-5516.3
Marg. lik.	-5497.1	-5530.4	-5533.9

C Conclusion

The use of simulation techniques seems to provide a reliable and accurate toolbox to handle even continuous time stochastic processes. The use of Markov Chain Monte Carlo techniques together with sequential Monte Carlo methods allows to infer many stochastic volatility models and to provide a battery of diagnostic tools. The big advantage of the MCMC algorithms is that they allow to simplify the solution of many unfeasible models with standard methods, by dividing the problem into many simpler ones. This is the case of the models analyzed in this thesis, where the likelihood function is not known in closed form because of the presence of non observable components.

The use of an adaptive method sensibly increases the efficiency of the estimates for the latent processes, at least in the affine jump diffusion specification. An important advantage of the Delayed Rejection method is that it maintains the properties and the tractability of the plain Metropolis-Hastings method, and at the same time preserves the algorithm from a bad behaviour due to an imprecise choice of the proposal distributions. On the other side the introduction of further Metropolis-Hastings steps slow down the run of the software. For this reasons some care has to be taken when the algorithm is planned.

Anyway, the use of simulation techniques seems to me very useful in financial econometrics. A possible extension of the results obtained in this thesis can include the study of models for option pricing based on more general stochastic dynamics. Many different alternatives based on Lévy processes are growing in importance. At this purpose, the use of Monte Carlo simulations seems to be appropriate to do inference for derivatives and to efficiently estimate the risk-premium parameters.

Acknowledgements

Financial support from MIUR, Grant 2002135473, is gratefully acknowledged.
We thank Bjørn Eraker, Fabio Rigat, Silvano Bordignon, Nunzio Cappuccio, Diego Lubian, Alessandro Platania and all the participants at the Financial Econometric Lunch Group at Duke University for helpful comments. All errors are my own.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers to wp@stat.unipd.it.

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy