*Journal of Econometrics and Statistics*
Vol. 5, Issue 1, 2025, pp. 89-106

# Particle Filters and Adaptive Metropolis-Hastings Sampling Applied to Volatility Models

Iago Carvalho Cunha[a] and Ralph dos Santos Silva[b]

[a] *Instituto Nacional de Metrologia, Qualidade e Tecnologia, Brazil*
[b] *Instituto de Matemática, Universidade Federal de Rio de Janeiro, Brazil*

**ABSTRACT**
Markov Chain Monte Carlo methods are widely used in Bayesian statistical inference to sample from the posterior distribution from a target distribution. However, for non-Gaussian and non-linear state space models, one can find difficulties in calculating the exact likelihood. To overcome problems in calculating the likelihood function, it is possible to use approximations made by particle filter methods. Furthermore, an adaptive Metropolis-Hastings algorithm may be applied since its proposal distribution is updated with previous draws from the posterior distribution. In this way, this paper discusses the applicability of adaptive Metropolis-Hastings (AMH) algorithms with random walk or independent proposals combined with estimated likelihoods through particle filters. We also propose a few model comparison criteria that can be easily integrated to the AMH. Moreover, we estimate non-linear and non-Gaussian volatility models for three time series of real index returns.

**KEYWORDS**
Diminishing adaptation, sequential Monte Carlo methods, state space model

## 1. Introduction

Nowadays is almost impossible to analyze economic data without measuring volatility, often being considered more important than any other measure in a time series of stock prices. However, the volatility is an unobservable measure and it has often a property called the leverage effect phenomenon (possibly a negative correlation between return and volatility). Finally, it is important to keep in mind the volatility clustering property.

In this paper, we fit historical series of daily log-returns of stock market indexes with the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model [5] with noise and the stochastic volatility (SV) model [12] and a few variants, including the version with leverage effect. The unknowns of the models are estimated through the Bayesian approach [2].

Markov Chain Monte Carlo (MCMC) simulation methods are widely applied to sample from a joint probability distribution. These methods are generally used in Bayesian inference where the posterior distribution of the unknown parameters is

CONTACT Author[b]. Email: ralph@im.ufrj.br

**Article History**
Received : 17 November 2024; Revised : 21 December 2024; Accepted : 27 December 2024; Published : 08 January 2025

**To cite this paper**

Iago Carvalho Cunha & Ralph dos Santos Silva (2025). Particle Filters and Adaptive Metropolis-Hastings Sampling Applied to Volatility Models. *Journal of Econometrics and Statistics.* 5(1), 89-106.

often difficult to calculate exactly. However, it is common to come across problems when calculating the likelihood, commonly for non-linear non-Gaussian models, or in situations where the choice of powerful proposals is not particularly easy.

When the likelihood does not have a closed form, we can approximate it using, for example, the standard particle filter (or sequential importance resampling - SIR) of Gordon et al. [8] and the auxiliary particle filter of Pitt and Shephard [14]. Moreover, Andrieu et al. [1] proved that MCMC methods still converge to the correct posterior distribution even if the simulated likelihood via SIR or ASIR is used.

To work around problems when choosing effective proposal distributions to use on the MCMC method, we can apply a few adaptive Metropolis-Hastings (AMH) sampling techniques. In such methods, the parameters of the proposal distribution are tuned by using previous draws and the difference between these successive parameters of the proposal converges to zero (diminishing adaptation). Important theoretical and practical contributions to diminishing adaptation sampling were made by Haario et al. [9], Haario et al. [10], and Roberts and Rosenthal [17] through the adaptive random walk Metropolis sampling (ARWMS), and Giordani and Kohn [7] by the independent Metropolis-Hastings sampling (AIMHS) with a proposal distribution based on a mixture of normals.

Thus, we combine the two AMH schemes to either SIR or ASIR in order to estimate the unknowns of the GARCH with noise and SV models. Our applications include a small simulation study to verify if the combined algorithms recover the true values of the parameters, and the estimation the parameters for the models applied to three important stock market indexes. The results for real data are compared using marginal likelihoods and a few likelihood-based information criteria.

The outline of this paper is organized as follows. Short backgrounds are given in Sections 2, 3 and 4 about state space models and particle filters, adaptive Metropolis-Hastings algorithms, and model selection, respectively. The GARCH model with noise and the SV models are presented in Section 5. The applications are in Section 6. Section 7 concludes.

## 2. State Space Models and Particle Filters

A state space model can be represented by an observation equation given by $f(y_t|x_t;\theta)$, $t = 1,\ldots,n$, and a system equation given by $f(x_t|x_{t-1};\theta)$, $t = 2,\ldots,n$, where $\theta$ and $f(.)$ represent a parameter vector and general probability (density) functions, respectively. Note that, $y_{1:n} = (y_1,\ldots,y_n)$ denote the history of measurements and $x_{1:n} = (x_1,\ldots,x_n)$ the history of states up to time $n$. The initial state $x_1$ distribution is given by $f(x_1|\theta)$.

The main problem for a state space model is to evaluate the following integrals from these equations:

$$f(x_t|y_{1:t-1};\theta) = \int f(x_t|x_{t-1};\theta)f(x_{t-1}|y_{1:t-1};\theta)dx_{t-1}, \tag{1}$$

which is used to update the posterior distribution at time $t$, that is,

$$f(x_t|y_{1:t};\theta) = \frac{f(y_t|x_t;\theta)f(x_t|y_{1:t-1};\theta)}{f(y_t|y_{1:t-1};\theta)}, \tag{2}$$

and

$$f(y_t|y_{1:t-1};\theta) = \int f(y_t|x_t;\theta)f(x_t|y_{1:t-1};\theta)dx_t. \quad (3)$$

In principle, the Equations (1)–(3) allow us, for a given $\theta$, to obtain the likelihood function of the observations $y_{1:n}$,

$$f(y_{1:n}|\theta) = f(y_1|\theta)\prod_{t=2}^{n} f(y_t|y_{1:t-1};\theta). \quad (4)$$

Except in a few cases such as linear Gaussian models, the integrals given in Equations (1)–(3) are in general hard to solve. To work around this problem, the SIR and the ASIR can be used to approximate these distributions, in particular the likelihood function.

### 2.1. *Standard Particle Filter*

The standard particle filter, also known as the sampling importance resampling (SIR) method, was proposed by Gordon et al. [8]. Suppose that we have a sample $x_{t-1}^{(\ell)}$, $\ell = 1,\dots,L$ with probabilities $\pi_{t-1}^{(\ell)}$ from $f(x_{t-1}|y_{1:t-1};\theta)$. It is easy to notice that the simplest values of $\pi_{t-1}^{(\ell)}$ are $1/L$. An approximation to Equation (1) is given by:

$$f(x_t|y_{1:t-1};\theta) \approx \sum_{\ell=1}^{L} f(x_t|x_{t-1}^{(\ell)};\theta)\pi_{t-1}^{(\ell)}. \quad (5)$$

Therefore, $f(x_t|y_{1:t-1};\theta)$ can be viewed as a mixture density with $L$ components where $f(x_t|x_{t-1}^{(\ell)};\theta)$ represents the system equation conditional at each particle $x_{t-1}^{(\ell)}$. That would give us a sample $\widetilde{x}_t^{(\ell)}$, $\ell = 1,\dots,L$, from the density $f(x_t|y_{1:t-1};\theta)$. Now, we can update the posterior distribution using Equation (2). We obtain a sample $\widetilde{x}_t^{(\ell)}$, $\ell = 1,\dots,L$ from $f(x_t|y_{1:t};\theta)$ by assigning a probability of

$$\widetilde{\pi}_t^{(\ell)} = \frac{f(y_t|\widetilde{x}_t^{(\ell)};\theta)\pi_{t-1}^{(\ell)}}{\sum_{j=1}^{L} f(y_t|\widetilde{x}_t^{(j)};\theta)\pi_{t-1}^{(j)}} \quad (6)$$

to $\widetilde{x}_t^{(\ell)}$. Thus, we have a sample $\widetilde{x}_t^{(\ell)}$, $\ell = 1,\dots,L$ with probabilities $\widetilde{\pi}_t^{(\ell)}$ from $f(x_t|y_{1:t};\theta)$. Note that from Equations (1) and (2) the predictive function can be approximated by

$$f_s(y_t|y_{1:t-1};\theta) \approx \sum_{j=1}^{L} f(y_t|x_t^{(j)};\theta)\pi_{t-1}^{(j)}, \quad (7)$$

which is a component of the likelihood function (an unbiased estimator of the likelihood as shown in Pitt et al. [15]). Finally, we resample $L$ values (with replacement) from the particles $\widetilde{x}_t^{(\ell)}$ with weights $\widetilde{\pi}_t^{(\ell)}$ to obtain a sample from $f(x_t|y_{1:t};\theta)$, then restarts the procedure for time $t+1$.

### 2.2. *Auxiliary Particle Filter*

The auxiliary particle filter (ASIR) of Pitt and Shephard [14] can be seen as a generalization of the SIR method and the main idea is to sample from a higher dimension joint density with the aid of an auxiliary variable. We note that from Equations (2) and (5),

$$f(x_t|y_{1:t};\theta) \approx \sum_{\ell=1}^{L} f(y_t|x_t;\theta) f(x_t|x_{t-1}^{(\ell)};\theta)\pi_{t-1}^{(\ell)}. \tag{8}$$

Introducing an auxiliary variable $\ell$ which can be viewed as an index to the mixture in Equation (8), we are able to adapt the particle filter in a more efficient way. The density we wish to approximate becomes:

$$f(x_t,\ell|y_{1:t};\theta) \propto f(y_t|x_t;\theta) f(x_t|x_{t-1}^{(\ell)};\theta)\pi_{t-1}^{(\ell)}, \quad \text{for} \quad \ell = 1,\dots,L, \tag{9}$$

such that

$$f(\ell|y_{1:t};\theta) \approx \frac{1}{f(y_t|y_{1:t-1};\theta)} \int f(y_t|x_t;\theta) f(x_t|x_{t-1}^{(\ell)};\theta) dx_t \pi_{t-1}^{(\ell)}$$

where

$$f(y_t|y_{1:t-1};\theta) \approx \sum_{\ell=1}^{L} \int f(y_t|x_t;\theta) f(x_t|x_{t-1}^{(\ell)};\theta) dx_t \pi_{t-1}^{(\ell)}. \tag{10}$$

Now, if we are able to sample from $f(x_t,\ell|y_{1:t};\theta)$, then we can discard the sampled values of $\ell$ and get back to our filtering density in Equation (8). The next step is to sample from $f(x_t,\ell|y_{1:t};\theta)$ using sampling importance resampling algorithm. That is, we make $K$ proposals $(x_t^{(k)},\ell^{(k)}), k = 1,\dots,K$ from some proposal density $g(x_t,\ell|y_{1:t};\theta)$ and compute the weights

$$\tilde{\pi}_t^{(k)} = \frac{1}{f_a(y_t|y_{1:t-1};\theta)} \times \frac{f(y_t|x_t^{(k)};\theta) f(x_t^{(k)}|x_{t-1}^{(\ell^{(k)})};\theta)\pi_{t-1}^{\ell^{(k)}}}{g(x_t^{(k)},\ell^{(k)}|y_{1:t};\theta)}. \tag{11}$$

From Equation (10), the predictive function can be approximated by:

$$f(y_t|y_{1:t-1};\theta) \approx f_a(y_t|y_{1:t-1};\theta) = \sum_{k=1}^{K} \frac{f(y_t|x_t^{(k)};\theta) f(x_t^{(k)}|x_{t-1}^{(\ell^{(k)})};\theta)\pi_{t-1}^{\ell^{(k)}}}{g(x_t^{(k)},\ell^{(k)}|y_{1:t};\theta)}, \tag{12}$$

which in turn can be used to normalize the weights in Equation (11). Usually, $K$ and $L$ are equal. When Equations (4) and (12) are combined, they produce an unbiased estimator of the likelihood [15]. Finally, we resample $L$ values from the above sample to obtain a sample from $f(x_t|y_{1:t};\theta)$ corresponding to particles $x_t^{(\ell)}$ with weights $\pi_t^{(\ell)} = 1/L$. This gives us the approximation in Equation (8) which then restarts the procedure for time $t+1$.

The choice of the proposal density $g(\cdot)$ is left completely to the researcher, however there are a few particular cases for $g(\cdot)$ that receive specific nomenclatures. This is the case of the generic auxiliary particle filter and the fully adapted particle filter briefly described below.

*Generic Auxiliary Particle Filter*

Assume that $z_t^{(\ell)}$ is some point estimate (e.g. mean or median) of the distribution of $x_t|x_{t-1}^{(\ell)}$. Then, if we approximate $g(\cdot)$ by

$$g(x_t, \ell|y_{1:t}; \theta) \propto f(y_t|z_t^{(\ell)}; \theta) f(x_t|x_{t-1}^{(\ell)}; \theta) \pi_{t-1}^{(\ell)}, \quad \text{for} \quad \ell = 1, \ldots, L,$$

we have what the authors call the generic auxiliary particle filter.

*Fully Adapted Particle Filter*

Suppose again that we have particles $x_{t-1}^{(\ell)}$ with attached probabilities $\pi_{t-1}^{(\ell)}$. If we are able to rewrite $f(y_t|x_t; \theta) f(x_t|x_{t-1}^{(\ell)}; \theta) \pi_{t-1}^{(\ell)}$ as the product $g(x_t|\ell, y_{1:t}; \theta) g(\ell|y_{1:t}; \theta)$ where $g(x_t|\ell, y_{1:t}; \theta)$ has a known closed-form (probability density function) easy to sample from, then the particle filter is fully adapted and, as a consequence, the weights $\pi_{t-1}^{(\ell)}$ have the same value for all $\ell = 1, \ldots, L$.

## 3. Adaptive Metropolis-Hastings

The Metropolis-Hastings (MH) algorithm [11] is a Markov Chain Monte Carlo based method employed to generate random samples from a probability distribution. Suppose the chain is in the iteration state $\theta_{m-1}$ and a value $\theta_m^p$ is generated from a proposed auxiliary distribution $q_m(\theta|\theta_{m-1})$. The new value $\theta_m^p$ is accepted with probability:

$$\alpha(\theta_{m-1}, \theta_m^p) = \min\left\{1, \frac{f(\theta_m^p)}{f(\theta_{m-1})} \frac{q_m(\theta_{m-1}|\theta_m^p)}{q_m(\theta_m^p|\theta_{m-1})}\right\},$$

and take $\theta_m = \theta_{m-1}$ otherwise (see, for example, Tierney [19]). Note that $f(\cdot)$ is the distribution of interest and, in a Bayesian context, $f(\cdot)$ can be the posterior density.

In adaptive sampling the parameters of the proposal density $q_m(\theta|\theta_{m-1})$ of the MH algorithm are estimated from the iterates $\theta_1, \ldots, \theta_{m-1}$. Under appropriate regularity conditions (diminishing adaptation) the sequence of iterates $\theta_m, m \geqslant 1$ converges to draws from the target distribution [7, 10, 17]. Next, we briefly explain the two proposals used in this paper.

### 3.1. *Adaptive Random Walk Metropolis*

The adaptive random walk Metropolis sampling (ARWMS) algorithm [17] can be divided into two phases, the first one takes place until iteration $m_0$, defined by the researcher to start the algorithm, and the second one from iteration $m_0$ to $M$. For the first phase of the algorithm, the proposed distribution is given by $q_m(\theta|\theta_{m-1}) = N(\theta_{m-1}; (0.1)^2 I_d/d)$, where $N(\mu; \Sigma)$ is a multivariate $d$-dimensional normal density function with mean $\mu$ and covariance matrix $\Sigma$. $I_d$ is a $d$-dimensional identity matrix.

And for the second phase, when $m > m_0$, the proposed distribution is given by:

$$q_m(\theta|\theta_{m-1}) = \beta N(\theta_{m-1}; (0.1)^2 I_d/d) + (1-\beta)N(\theta_{m-1}; (2.38^2)\Sigma_m/d), \quad (13)$$

where $\beta$ is a small positive constant and, in this paper, equals to 0.05. And $\Sigma_m$ represents the covariance matrix estimated, iteratively, through the $m-1$ iterations. The part with less variation (covariance matrix equal to $(0.1)^2 I_d/d$) avoids the algorithm getting stuck at problematic values and second part (with the covariance matrix equal to $(2.38^2)\Sigma_n/d$) is optimal in a multi-dimensional context [16].

### 3.2. *Adaptive Independent Metropolis Hastings*

The adaptive independent Metropolis-Hastings sampling (AIMHS) method [7] can also be divided into two phases. For both phases, the proposed density is given by a mixture of four terms according to the equation below:

$$q_m(\theta|\theta_{m-1}) = \sum_{j=1}^{4} \beta_j q_{jm}(\theta|\lambda_{jm}), \quad (14)$$

with $\beta_j \geqslant 0$, for $j = 1, \ldots, 4$ and $\sum_{j=1}^{4} \beta_j = 1$, where $\lambda_{jm}$ holds all parameters of the density $q_{jm}(\theta|\lambda_{jm})$.

In the first phase, $q_{1m}(\theta|\lambda_{1m})$ is an initial proposal (via Laplace approximation of the posterior density or by other methods) and $q_{2m}(\theta|\lambda_{2m})$ is a heavy tailed version of the former. The $q_{3m}(\theta|\lambda_{3m})$ carries the adaptive part of the proposal, being an estimate of the target density calculated through a normal mixing using k-harmonic means clustering (each update is done after running a certain amount of iterations - block scheme). And finally, $q_{4m}(\theta|\lambda_{4m})$ is a version of $q_{3m}(\theta|\lambda_{3m})$ with heavy tails. However, the first phase begins with $\beta_3$ and $\beta_4$ being equal to zero until a sufficient number of iterations is reached to obtain $q_{3m}(\theta|\lambda_{3m})$ and, consequently, $q_{4m}(\theta|\lambda_{4m})$. We start this phase with $\beta_1 = 0.8$ and $\beta_2 = 0.2$, then we use $\beta_1 = 0.15$, $\beta_2 = 0.05$, $\beta_3 = 0.7$ and $\beta_4 = 0.1$.

And finally, in the second phase, $q_{1m}(\theta|\lambda_{1m})$ is defined as the last form assumed by $q_{3m}(\theta|\lambda_{3m})$ in the first phase. The densities $q_{2m}(\theta|\lambda_{2m})$ and $q_{4m}(\theta|\lambda_{4m})$ are constructed in the same manner as the first phase and $q_{3m}(\theta|\lambda_{3m})$ is maintained until the next update (by the block scheme).

### 3.3. *Adaptive Sampling with Simulated Likelihood*

The posterior distribution of $\theta$ is given by $f(\theta|y_{1:n}) \propto f(y_{1:n}|\theta)f(\theta)$ and this holds for the cases where the likelihood function $f(y_{1:n}|\theta)$ is calculated exactly (with $f(\theta)$ being the prior distribution). Nonetheless, Andrieu et al. [1] showed that Markov Chain Monte Carlo samplers still converge to the correct posterior density even when an unbiased estimator of likelihood, $\widehat{f}(y_{1:n}|\theta)$, is applied, such as those in Equations (7) and (12) given by particle filters with finite number of particles.

The simulated likelihood via particle filter algorithms may be seen as $f(y_{1:n}|\theta, u)$, where $u$ is a set of auxiliary variables that are not function of $\theta$ such that $f(y_t|y_{1:t-1}; \theta; u)$ is equal to $f_s(y_t|y_{1:t-1}; \theta)$ or $f_a(y_t|y_{1:t-1}; \theta)$ obtained from Equations (7) and (12), respectively. Now, let $f(y_t|y_{1:t-1}; \theta; u)$ obtained from the particle filter be the estimate of $f(y_t|y_{1:t-1};\theta)$. Then, $\widehat{f}(y_{1:n}|\theta) = f(y_1|\theta;u)\prod_{t=2}^{n} f(y_t|y_{1:t-1};\theta;u)$ is the unbiased estimate of the likelihood [15] given by $f(y_{1:n}|\theta)$.

## 4. Model Selection

### 4.1. *Estimating the Marginal Likelihood*

Marginal likelihoods are often used to compare models and can be seen as the probability of the data given the model type. Thus the higher its value, the more adjusted is the model to the data set. Following the previous notation, the marginal likelihood is given by

$$f(y_{1:n}) = \int f(y_{1:n}|\theta)f(\theta)d\theta. \tag{15}$$

Suppose that $h(\theta)$ is an approximation to $f(\theta|y_{1:n})$ which can be evaluated explicitly. Bridge sampling [13] estimates the marginal likelihood as follows. Let

$$t(\theta) = \left(\frac{f(y_{1:n}|\theta)f(\theta)}{U} + h(\theta)\right)^{-1},$$

where $U$ is a positive constant. Let

$$B = \int t(\theta)h(\theta)f(\theta|y_{1:n})d\theta.$$

Then,

$$B = \frac{B_1}{f(y_{1:n})} \text{ where } B_1 = \int t(\theta)h(\theta)f(y_{1:n}|\theta)f(\theta)d\theta.$$

Suppose the sequence of iterates $\{\theta^{(j)}, j = 1,\dots,M\}$ is generated from the posterior density $f(\theta|y_{1:n})$ and a second sequence of iterates $\{\tilde{\theta}^{(k)}, k = 1,\dots,K\}$ is generated from $h(\theta)$. Then,

$$\widehat{B} = \frac{1}{M}\sum_{j=1}^{M} t(\theta^{(j)})q(\theta^{(j)}),\quad \widehat{B}_1 = \frac{1}{K}\sum_{k=1}^{K} t(\tilde{\theta}^{(k)})f(y|\tilde{\theta}^{(k)})f(\tilde{\theta}^{(k)}) \text{ and } \widehat{f}_{BS}(y_{1:n}) = \frac{\widehat{B}_1}{\widehat{B}}$$

are estimates of $B$ and $B_1$, respectively, while $\widehat{f}_{BS}(y)$ is the bridge sampling estimator of the marginal likelihood $f(y_{1:n})$.

We take $h(\theta)$ from the adaptive independent Metropolis-Hastings (AIMHS) with mixture of normals proposal. Although $U$ can be any positive constant, it is more efficient if $U$ is a reasonable estimate of $f(y_{1:n})$. One way to do so is to take $\widehat{U} = f(y_{1:n}|\theta^*)f(\theta^*)/h(\theta^*)$, where $\theta^*$ is the posterior mean of $\theta$ obtained from the posterior simulation.

An alternative method to estimate of the marginal likelihood $f(y_{1:n})$ is to use importance sampling [6] based on the proposal distribution $h(\theta)$ as before (as the proposal from the AIMHS). That is,

$$\widehat{f}_{IS}(y_{1:n}) = \frac{1}{K}\sum_{k=1}^{K}\frac{f(y_{1:n}|\theta^{(k)})f(\theta^{(k)})}{h(\theta^{(k)})}.$$

Since the proposal distribution of the AIMHS has at least one heavy tailed component, the importance sampling ratios are likely to be bounded and well-behaved.

### 4.2. *Likelihood Based Information Criteria*

In the adaptive Metropolis-Hastings sampling, the log likelihood function is always evaluated as a component of the posterior distribution. In that case, each draw $\theta^{(j)}$ from the posterior distribution produces also the corresponding log-likelihood value, $\log f(y_{1:n}|\theta^{(j)})$. That can in turn be used to compute several likelihood based information criteria.

Let $f(y_{1:n}|\theta_\ell, \mathcal{M}_\ell)$ be the likelihood for model $\mathcal{M}_\ell$ and define the deviance as $D(\theta_\ell) = -2\log f(y_{1:n}|\theta_\ell, \mathcal{M}_\ell)$. The deviance information criterion (DIC) is defined as

$$DIC(\mathcal{M}_\ell) = 2E[D(\theta_\ell)|y_{1:n}, \mathcal{M}_\ell] - D(E[\theta_\ell|y_{1:n}, \mathcal{M}_y]). \tag{16}$$

The draws from $\theta_\ell^{(j)}$ and $\log f(y_{1:n}|\theta_\ell^{(j)}, \mathcal{M}_\ell)$, $j = 1, \ldots, M$, can be used to approximate $E[D(\theta_\ell)|y_{1:n}, \mathcal{M}_\ell]$ and $E[\theta_\ell|y_{1:n}, \mathcal{M}_\ell]$ by $M^{-1}\sum_{j=1}^{M} D(\theta_\ell^{(j)})$ and $M^{-1}\sum_{j=1}^{M}\theta_\ell^{(j)}$, respectively. Finally, approximations to DIC can be easily derived. The estimation $\hat{d}_\ell$ of the number of model parameters on the DIC is given by $\hat{d}_\ell = E[D(\theta_\ell)|y_{1:n}, \mathcal{M}_\ell] - D(E[\theta_\ell|y_{1:n}, \mathcal{M}_y])$. Thus, DIC may be rewritten as $DIC = D(E[\theta_\ell|y_{1:n}, \mathcal{M}_y]) + 2\hat{d}_\ell$.

Similarly, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), their expected versions, EAIC and EBIC, can also be defined. Note that given a set of models for a given data set, the lowest value of a criterion indicates the best model (but different criteria may not lead to different models). For more details on these information criteria, see [18] and references therein.

## 5. Modelling Volatility

Most financial studies focus on the analysis of returns series rather than the use of asset prices series. The reason we use a series of returns has two factors, the returns information serves the interests of investors and has more interesting statistical properties than the price series. Thus, let $P_t$ be the price of an asset at time $t$, the log-return at time $t$ is given by: $y_t = \log(P_t) - \log(P_{t-1})$, which is used in our applications.

### 5.1. *Generalized Autoregressive Conditionally Heteroscedastic Model with Noise*

A generalized autoregressive conditionally heteroscedastic (GARCH) model [5] is used to model the variance of a time series using values of the past squared means of the observations and past variances. The observation and system equations of the

GARCH(1,1) model with noise are given by:

$$
\begin{aligned}
y_t &= x_t + \epsilon_t, \text{ where } \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \\
x_{t+1} &= \omega_t, \qquad \text{where } \quad \omega_t \sim \mathcal{N}(0, \tau_t^2(x_t)) \\
\tau_{t+1}^2 &= \beta_0 + \beta_1 x_t^2 + \beta_2 \tau_t^2,
\end{aligned}
$$

where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean $\mu$ and variance $\sigma$, respectively. This model has the following restrictions on the parameters: $\sigma^2 > 0$, $\beta_j > 0$ for $j = 0, 1, 2$ and $\beta_1 + \beta_2 < 1$ (stationary condition). Thus, we assume the following prior distribution: $\sigma^2 \sim \mathcal{HN}(c_1^2)$, $\beta_0 \sim \mathcal{HN}(c_2^2)$, $(\beta_1, \beta_2) \sim \mathcal{U}(\{$over the triangle defined by (0,0), (0,1) and (1,0)$\})$, and $x_0 \sim \mathcal{N}(0, \tau_0^2)$ with $\tau_0^2 = \beta_0/(1 - \beta_1 - \beta_2)$, where $\mathcal{HN}(c_i^2)$, $i = 1, 2$, is a half-normal distribution with location parameter set to 0 and $c_i^2$, $i = 1, 2$, as the scale parameter. Here, $\mathcal{U}$ denotes a continuous uniform distribution.

### 5.2. *Stochastic Volatility Model and Its Variants*

The observation and system equations of the stochastic volatility (SV) model is given by [12]:

$$
\begin{aligned}
y_t &= e^{x_t/2} \epsilon_t \\
x_{t+1} &= \xi_{t+1} + \omega_t, \quad \text{where} \quad \xi_{t+1} = \alpha + \phi(x_t - \alpha)
\end{aligned}
$$

and $\epsilon_t$ is the observation error with mean 0 and variance 1.

First, we consider that $\epsilon_t$ and $\omega_t$ are independent, with $\omega_t \sim \mathcal{N}(0, \tau^2)$ and $\epsilon_t$ can be distributed as a standard normal distribution, $\mathcal{N}(0, 1)$, a standard skew normal distribution [3], denoted as $\mathcal{SN}(\lambda, 0, 1)$, a $t$ distribution with 3 degrees of freedom ($t(3)$) or a skew $t$ distribution [4] also with 3 degrees of freedom, denoted as $\mathcal{S}t(\lambda, 3)$, where $\lambda$ is a parameter of skewness. Additionally, $\epsilon_t$ and $\omega_t$ may have a bivariate normal distribution with correlation $\rho$ and this model is known as a stochastic volatility model with leverage effect.

To complete our model specification, we assume the following prior distribution: $\tau^2 \sim \mathcal{IG}(a_1, b_1)$, $\phi \sim Beta(a_2, b_2)$, $\alpha \sim \mathcal{N}(a_3, b_3^2)$, $\lambda \sim \mathcal{N}(a_4, b_4^2)$, $x_0 \sim \mathcal{N}(\alpha, \tau^2/(1 + \phi^2))$, where $\mathcal{IG}(a_1, b_1)$ is the inverse-gamma distribution with $a_1$ and $b_1$ as shape and scale parameters; and $Beta(a_2, b_2)$ is the beta distribution with $a_2$ and $b_2$ as shape parameters.

## 6. Applications

In this section we carry out a simulation study that consist of estimating the parameters of five simulated series for GARCH(1,1) model with noise, stochastic volatility model with Gaussian noise and stochastic volatility model with leverage. Each simulated series has 1.000 observations. Next, we model daily log-return data of three stock market indexes - namely BOVESPA, NASDAQ and S&P500 - from January 2012 to March 2016 resulting in a time series with more than one thousand observations each.

Note that it is possible to estimate the likelihood of the GARCH with noise model by applying both particle filters (SIR and ASIR). In addition, for this model, it is possible to apply the fully adapted particle filter (see A). However, it is not possible to use a fully adapted particle filter for the SV model, so we tried to apply the generic version of ASIR, but the estimates, when found, were not as expected and the computational time spent extremely high. Therefore, we have chosen to apply only SIR method for the SV models (see B for the SV model with leverage). In addition, to determine the number of particles of the filters, we use the methodology proposed by Pitt et al. [15]. In our applications, we set $c_1 = c_2 = 10$, $a_1 = b_1 = a_2 = b_2 = 1$, $a_3 = a_4 = 0$ and $b_3 = b_4 = 10^6$ as parameters for the prior distributions.

The estimation strategy used in all datasets and models is to first create an initial estimate for the parameters and covariance matrix through the ARWMS, then use them as initial values in the AIMHS. Therefore, all results shown below refer only for the last part when we use AIMHS.

### 6.1. *Simulation*

The main objective of this simulation study is to verify if the algorithm is actually estimating the true parameters. For this, we generate 5 time series of the intended models and observe if the estimation is being done correctly. The parameters used to generate the data of GARCH(1,1) model with noise were $\sigma^2 = 0.00009$, $\beta_0 = 0.000002$, $\beta_1 = 0.15$ and $\beta_2 = 0.84$. For the stochastic volatility model with Gaussian noise were $\tau^2 = 0.20$, $\alpha = -9.6$ and $\phi = 0.84$. And for the stochastic volatility model with leverage were $\tau^2 = 0.11$, $\alpha = -11$, $\phi = 0.98$ and $\rho = -0.7$. To obtain the results of this section, we ran all AMH algorithms with 50.000 iterations with the first half of them being discarded for the calculation of the final estimates.

In Tables 1, 2, 3 and 4 we can observe the posterior mean, median, standard deviation and credibility interval of 95% ($CI_{0.025}$ and $CI_{0.975}$ are, respectively, the lower limit and upper limit of the interval) for the parameters of each model.

**Table 1.** Posterior mean, median, standard deviation and credibility interval for the parameters of GARCH(1,1) model with noise using SIR filter.

| Replica | Parameters | Posterior estimations | | | | |
|---|---|---|---|---|---|---|
| | | Mean | Std. dev. | $CI_{0.025}$ | Median | $CI_{0.975}$ |
| 1 | $\sigma^2$ | 0.0000630 | 0.0000258 | 0.0000076 | 0.0000637 | 0.0001097 |
| | $\beta_0$ | 0.0000037 | 0.0000020 | 0.0000009 | 0.0000033 | 0.0000090 |
| | $\beta_1$ | 0.1463449 | 0.0453058 | 0.0790899 | 0.1408561 | 0.2472860 |
| | $\beta_2$ | 0.8290534 | 0.0459327 | 0.7252073 | 0.8343345 | 0.8992350 |
| 2 | $\sigma^2$ | 0.0000883 | 0.0000253 | 0.0000319 | 0.0000910 | 0.0001317 |
| | $\beta_0$ | 0.0000061 | 0.0000053 | 0.0000010 | 0.0000046 | 0.0000220 |
| | $\beta_1$ | 0.2257102 | 0.1003751 | 0.0829156 | 0.2067775 | 0.4762093 |
| | $\beta_2$ | 0.7226548 | 0.1125832 | 0.4477203 | 0.7420469 | 0.8822469 |
| 3 | $\sigma^2$ | 0.0000529 | 0.0000269 | 0.0000059 | 0.0000524 | 0.0001051 |
| | $\beta_0$ | 0.0000037 | 0.0000023 | 0.0000008 | 0.0000033 | 0.0000100 |
| | $\beta_1$ | 0.0905114 | 0.0390999 | 0.0337985 | 0.0833210 | 0.1868716 |
| | $\beta_2$ | 0.8807635 | 0.0440161 | 0.7734476 | 0.8890788 | 0.9410084 |
| 4 | $\sigma^2$ | 0.0000546 | 0.0000325 | 0.0000045 | 0.0000549 | 0.0001182 |
| | $\beta_0$ | 0.0000049 | 0.0000025 | 0.0000009 | 0.0000045 | 0.0000093 |
| | $\beta_1$ | 0.1269757 | 0.0345311 | 0.0762891 | 0.1201576 | 0.2073463 |
| | $\beta_2$ | 0.8506429 | 0.0345098 | 0.7679393 | 0.8602096 | 0.9070856 |
| 5 | $\sigma^2$ | 0.0000718 | 0.0000213 | 0.0000256 | 0.0000741 | 0.0001085 |
| | $\beta_0$ | 0.0000032 | 0.0000021 | 0.0000008 | 0.0000027 | 0.0000088 |
| | $\beta_1$ | 0.1865094 | 0.0585892 | 0.0933376 | 0.1799826 | 0.3186721 |
| | $\beta_2$ | 0.8007600 | 0.0591066 | 0.6671012 | 0.8084501 | 0.8941634 |

In order to obtain the estimates of the GARCH(1,1) model with noise parameters using SIR filter (Table 1), 3.000 particles were used for the preliminary part (ARWMS) and 2.000 in the final part (AIMHS). Notice that for all replicas, all credibility intervals contain the true parameter values, which indicates satisfactory behaviour of our approach.

However, to obtain the estimates of the GARCH(1,1) model with noise parameters using ASIR filter, 50 particles were used for both the preliminary and final parts. Note that this algorithm is much more efficient than the one using SIR filter, because it uses much fewer particles (saves computational time) and obtains results as satisfactory as the previous one, as can be verified in Table 2.

**Table 2.** Posterior mean, median, standard deviation and credibility interval for the parameters of GARCH(1,1) model with noise using ASIR filter.

| Replica | Parameters | Posterior estimations | | | | |
|---|---|---|---|---|---|---|
| | | Mean | Std. dev. | $CI_{0.025}$ | Median | $CI_{0.975}$ |
| 1 | $\sigma^2$ | 0.0000622 | 0.0000253 | 0.0000107 | 0.0000633 | 0.0001090 |
| | $\beta_0$ | 0.0000038 | 0.0000020 | 0.0000010 | 0.0000034 | 0.0000086 |
| | $\beta_1$ | 0.1473738 | 0.0469848 | 0.0755105 | 0.1413896 | 0.2562844 |
| | $\beta_2$ | 0.8276800 | 0.0473448 | 0.7176016 | 0.8336273 | 0.9008290 |
| 2 | $\sigma^2$ | 0.0000904 | 0.0000251 | 0.0000344 | 0.0000935 | 0.0001340 |
| | $\beta_0$ | 0.0000057 | 0.0000051 | 0.0000010 | 0.0000042 | 0.0000192 |
| | $\beta_1$ | 0.2363900 | 0.1041273 | 0.0817055 | 0.2210820 | 0.4786691 |
| | $\beta_2$ | 0.7152533 | 0.1143224 | 0.4405373 | 0.7351936 | 0.8782488 |
| 3 | $\sigma^2$ | 0.0000514 | 0.0000279 | 0.0000034 | 0.0000505 | 0.0001068 |
| | $\beta_0$ | 0.0000038 | 0.0000024 | 0.0000007 | 0.0000032 | 0.0000099 |
| | $\beta_1$ | 0.0903956 | 0.0400999 | 0.0364844 | 0.0820313 | 0.1903948 |
| | $\beta_2$ | 0.8812510 | 0.0437155 | 0.7677189 | 0.8886806 | 0.9442178 |
| 4 | $\sigma^2$ | 0.0000585 | 0.0000297 | 0.0000050 | 0.0000585 | 0.0001154 |
| | $\beta_0$ | 0.0000045 | 0.0000026 | 0.0000009 | 0.0000041 | 0.0000108 |
| | $\beta_1$ | 0.1307535 | 0.0370271 | 0.0707828 | 0.1264888 | 0.2162167 |
| | $\beta_2$ | 0.8482669 | 0.0385817 | 0.7639958 | 0.8517576 | 0.9141380 |
| 5 | $\sigma^2$ | 0.0000718 | 0.0000209 | 0.0000262 | 0.0000734 | 0.0001082 |
| | $\beta_0$ | 0.0000032 | 0.0000021 | 0.0000008 | 0.0000027 | 0.0000085 |
| | $\beta_1$ | 0.1895439 | 0.0600698 | 0.0929562 | 0.1821909 | 0.3298645 |
| | $\beta_2$ | 0.7979995 | 0.0612526 | 0.6584258 | 0.8055708 | 0.8956539 |

Table 3 shows the results of the estimations made for the parameters of the SV model with Gaussian noise using SIR filter. For this model, we use 350 particles in both algorithms (first ARWMS, then AIMHS). Notice that all credibility intervals also contain the true values of the parameters. Finally, Table 4 refers to results on the SV model with leverage using SIR filter estimations where 1.750 particles for both ARWMS and AIMHS.

This simulation study is not exhaustive and does not include all models used in the real data application. However, it can be seen that the combined AMH with SIR or ASIR algorithms are able to recover the true values of the parameters and it is expected that the same behaviour is found in all other models.

**Table 3.** Posterior mean, median, standard deviation and credibility interval for the parameters of SV model with Gaussian noise using SIR filter.

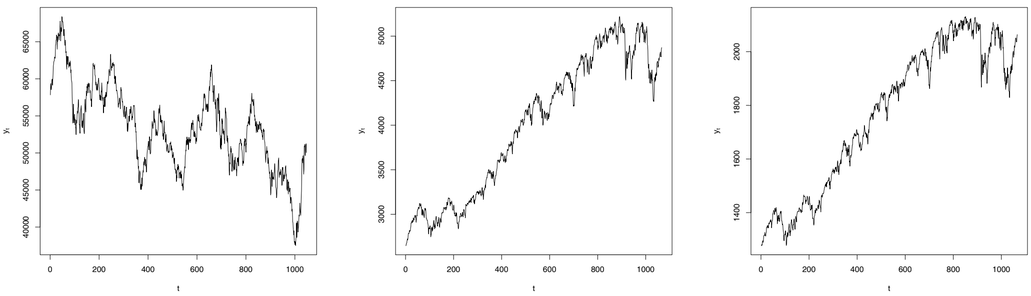
| Replica | Parameters | Posterior estimations | | | | |
|---|---|---|---|---|---|---|
| | | Mean | Std. dev. | $CI_{0.025}$ | Median | $CI_{0.975}$ |
| 1 | $\tau^2$ | 0.2209872 | 0.0537652 | 0.1344007 | 0.2140449 | 0.3433991 |
| | $\phi$ | 0.8233984 | 0.0444285 | 0.7214496 | 0.8283471 | 0.8959334 |
| | $\alpha$ | -9.6640155 | 0.1045723 | -9.8683794 | -9.6628164 | -9.4591694 |
| 2 | $\tau^2$ | 0.2307925 | 0.0617954 | 0.1360358 | 0.2217310 | 0.3726260 |
| | $\phi$ | 0.8234385 | 0.0460480 | 0.7156061 | 0.8289964 | 0.8975777 |
| | $\alpha$ | -9.4129225 | 0.1029042 | -9.6095164 | -9.4162568 | -9.2053341 |
| 3 | $\tau^2$ | 0.2571744 | 0.0593681 | 0.1613926 | 0.2504576 | 0.3912914 |
| | $\phi$ | 0.8386231 | 0.0371395 | 0.7578607 | 0.8417097 | 0.9017371 |
| | $\alpha$ | -9.5249343 | 0.1145634 | -9.7559453 | -9.5251857 | -9.2956183 |
| 4 | $\tau^2$ | 0.2448653 | 0.0675171 | 0.1391726 | 0.2355385 | 0.4004809 |
| | $\phi$ | 0.7908226 | 0.0511464 | 0.6745926 | 0.7955630 | 0.8759804 |
| | $\alpha$ | -9.7202447 | 0.0934316 | -9.9040991 | -9.7201533 | -9.5385561 |
| 5 | $\tau^2$ | 0.2616024 | 0.0633620 | 0.1616784 | 0.2547169 | 0.4088322 |
| | $\phi$ | 0.7973314 | 0.0450955 | 0.6948007 | 0.8012077 | 0.8710793 |
| | $\alpha$ | -9.7114517 | 0.0956238 | -9.8984855 | -9.7100075 | -9.5280789 |

**Table 4.** Posterior mean, median, standard deviation and credibility interval for the parameters of SV model with leverage using SIR filter.

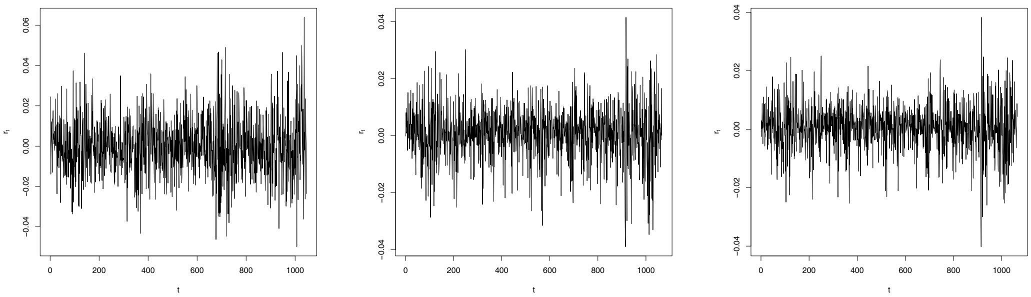
| Replica | Parameters | Posterior estimations | | | | |
|---|---|---|---|---|---|---|
| | | Mean | Std. dev. | $CI_{0.025}$ | Median | $CI_{0.975}$ |
| 1 | $\tau^2$ | 0.1300850 | 0.0208358 | 0.0940193 | 0.1280196 | 0.1747818 |
| | $\phi$ | 0.9852177 | 0.0061611 | 0.9722945 | 0.9857468 | 0.9956883 |
| | $\alpha$ | -10.850556 | 1.0132163 | -13.471119 | -10.649624 | -9.4863125 |
| | $\rho$ | -0.7031984 | 0.0619436 | -0.8091877 | -0.7079091 | -0.5718210 |
| 2 | $\tau^2$ | 0.1167414 | 0.0187372 | 0.0850963 | 0.1149657 | 0.1584491 |
| | $\phi$ | 0.9820628 | 0.0059646 | 0.9697027 | 0.9823247 | 0.9928902 |
| | $\alpha$ | -10.814652 | 0.5032284 | -11.944318 | -10.776913 | -9.9164817 |
| | $\rho$ | -0.7130920 | 0.0630058 | -0.8197065 | -0.7176833 | -0.5789504 |
| 3 | $\tau^2$ | 0.1403466 | 0.0216788 | 0.1036933 | 0.1376562 | 0.1888396 |
| | $\phi$ | 0.9813912 | 0.0056285 | 0.9697771 | 0.9815278 | 0.9927614 |
| | $\alpha$ | -10.071973 | 0.6470769 | -11.125004 | -10.131074 | -8.6076082 |
| | $\rho$ | -0.6913779 | 0.0585517 | -0.7884205 | -0.6965849 | -0.5628501 |
| 4 | $\tau^2$ | 0.1404871 | 0.0204211 | 0.1057169 | 0.1382987 | 0.1849773 |
| | $\phi$ | 0.9823382 | 0.0047879 | 0.9726977 | 0.9824653 | 0.9913064 |
| | $\alpha$ | -10.921940 | 0.4967491 | -11.980299 | -10.870996 | -10.062043 |
| | $\rho$ | -0.7975224 | 0.0453794 | -0.8753869 | -0.8017720 | -0.6978566 |
| 5 | $\tau^2$ | 0.0928752 | 0.0162011 | 0.0666495 | 0.0915056 | 0.1293698 |
| | $\phi$ | 0.9679708 | 0.0096341 | 0.9468060 | 0.9687180 | 0.9846366 |
| | $\alpha$ | -10.848615 | 0.2664703 | -11.364805 | -10.853792 | -10.320197 |
| | $\rho$ | -0.6733129 | 0.0858376 | -0.8204434 | -0.6786633 | -0.4914737 |

### 6.2. *Real data*

As said before, we model daily log-return data of BOVESPA, NASDAQ and S&P500 indexes from January 2012 to March 2016. Figure 1 shows the prices and log returns for all indexes. For all series, $t = 200$ corresponds to October 2012, $t = 400$ to July 2013, $t = 600$ to May 2014, $t = 800$ to February 2012 and $t = 1000$ to December 2015. As expected, the log returns are around zero with increasing volatilities when prices tend to decrease.



(a) BOVESPA prices ($P_t$). (b) NASDAQ prices ($P_t$.) (c) S&P500 prices ($P_t$).



(d) BOVESPA returns ($y_t$). (e) NASDAQ returns ($y_t$). (f) S&P500 returns ($y_t$).

**Figure 1.** BOVESPA, NASDAQ and SP500 price and log-return series.

Table 5 presents a few descriptive statistics of all log return series. They have means and medians about zero and positive excess kurtoses. However, NASDAQ and S&P500 have negative skewness whilst BOVESPA has positive skewness. Thus, all log return series present evidence of non-normality.

**Table 5.** Descriptive statistics of the indexes return series.

| Statistics | Indexes | | |
|---|---|---|---|
| | BOVESPA | NASDAQ | S&P500 |
| Mean | -0.00014 | 0.00057 | 0.00045 |
| Standard Deviation | 0.01481 | 0.00958 | 0.00830 |
| Median | -0.00096 | 0.00088 | 0.00049 |
| Skewness | 0.24131 | -0.32704 | -0.24698 |
| Kurtosis | 0.75280 | 1.44709 | 1.66029 |

Next, we apply a few volatility models to these data sets using AMH algorithms combined with particle filters methods. In order to perform the model comparisons by means of likelihood-based information criteria and marginal likelihoods, we initially estimated the parameters for our data sets (BOVESPA, NASDAQ and S&P500 return series) using GARCH and SV models presented in Sections 5.1 and 5.2, respectively. We ran all AMH algorithms with 200.000 iterations with the first half of them being discarded for the calculation of final estimates. In addition, we first generate an initial estimate of the parameters and covariance matrix using the ARWMS method to use them as initial values in AIMHS method.

Furthermore, we are interested in the results given by the output of the AIMHS since this algorithm allows computations of all model selection criteria in Section 4. Remember that the lowest value of a likelihood-based information criterion indicates the best model whilst largest value of a marginal likelihood, such as those given by $\widehat{f}_{BS}(y)$ and $\widehat{f}_{IS}(y)$, results also the best model. The values in bold refer to the model most adjusted to the data according to the respective selection criteria.

To obtain the results shown in Table 6, we used the following quantity of particles in all filters: 2.000 for the GARCH(1,1) model with noise and SIR, 50 for the same model with ASIR, 200 for the SV($\mathcal{N}$) model, 50 for the SV($t_3$) model, 230 for the SV($S\mathcal{N}$) model, 45 for the SV($St_3$) model and 350 for the SV with leverage model. Note that all model selection criteria indicate the GARCH model with noise as the most adjusted to BOVESPA return series - either with SIR or ASIR.

**Table 6.** Model comparisons by means of likelihood-based information criteria and marginal likelihoods for BOVESPA series.

| Model | PF | *AIC* | *BIC* | *EAIC* | *EBIC* | *DIC* | $\widehat{f}_{BS}(y)$ | $\widehat{f}_{IS}(y)$ |
|---|---|---|---|---|---|---|---|---|
| GARCH | SIR | **-5998** | **-5978** | **-6001** | **-5981** | -6012 | $2.2 \times 10^{-230}$ | $2.1 \times 10^{-230}$ |
| | ASIR | -5990 | -5970 | -6001 | -5981 | **-6020** | $\mathbf{2.3 \times 10^{-230}}$ | $\mathbf{2.2 \times 10^{-230}}$ |
| SV($\mathcal{N}$) | SIR | -5975 | -5960 | -5977 | -5962 | -5985 | $1.9 \times 10^{-232}$ | $1.9 \times 10^{-232}$ |
| SV($t_3$) | | -5910 | -5895 | -5908 | -5893 | -5911 | $5.5 \times 10^{-246}$ | $5.7 \times 10^{-246}$ |
| SV($S\mathcal{N}$) | | -5974 | -5955 | -5974 | -5954 | -5982 | $3.6 \times 10^{-235}$ | $3.5 \times 10^{-235}$ |
| SV($St_3$) | | -5897 | -5877 | -5901 | -5881 | -5912 | $2.2 \times 10^{-249}$ | $2.3 \times 10^{-249}$ |
| SV lev | | -5989 | -5970 | -5987 | -5967 | -5992 | $1.9 \times 10^{-231}$ | $2.2 \times 10^{-231}$ |

Therefore, the posterior mean, median, standard deviation and credibility interval of 95% of the GARCH(1,1) model with noise are given in Table 7.

**Table 7.** Posterior mean, median, standard deviation and credibility interval for the parameters of GARCH(1,1) model with noise using ASIR applied to BOVESPA series.

| Parameters | Summary of the posterior distribution | | | | |
|---|---|---|---|---|---|
| | Mean | Std. dev. | $CI_{0.025}$ | Median | $CI_{0.975}$ |
| $\sigma^2$ | 0.0000913 | 0.0000259 | 0.0000294 | 0.0000942 | 0.0001342 |
| $\beta_0$ | 0.0000020 | 0.0000015 | 0.0000005 | 0.0000016 | 0.0000060 |
| $\beta_1$ | 0.1512039 | 0.0517068 | 0.0676217 | 0.1448260 | 0.2735384 |
| $\beta_2$ | 0.8408180 | 0.0512353 | 0.7203291 | 0.8471988 | 0.9208007 |

For the models applied to NASDAQ return series, we used the following quantity of particles in all filters: 2.000 for the GARCH(1,1) model with noise and SIR, 75 for the same model with ASIR, 350 for the SV($\mathcal{N}$) model, 200 for the SV($t_3$) model, 400 for the SV($S\mathcal{N}$) model, 200 for the SV($St_3$) model and 400 for the SV with leverage model. Model comparison results are shown in Table 8.

In addition, with the exception of DIC that indicated GARCH with ASIR as the model most adjusted to the data, all criteria pointed to the SV model with $\epsilon_t$ having a skew-normal distribution. Now, when considered only marginal likelihood criterion for model selection, stochastic volatility model with skew Gaussian noise is considered

**Table 8.** Model comparisons by means of likelihood-based information criteria and marginal likelihoods for NASDAQ series.

| Model | PF | *AIC* | *BIC* | *EAIC* | *EBIC* | *DIC* | $\widehat{f}_{BS}(y)$ | $\widehat{f}_{IS}(y)$ |
|---|---|---|---|---|---|---|---|---|
| GARCH | SIR | -6942 | -6922 | -6941 | -6921 | -6948 | $1.3 \times 10^{-25}$ | $1.3 \times 10^{-25}$ |
| | ASIR | -6912 | -6891 | -6940 | -6921 | **-6978** | $1.4 \times 10^{-25}$ | $1.3 \times 10^{-25}$ |
| SV($\mathcal{N}$) | SIR | -6955 | -6940 | -6953 | -6938 | -6957 | $1.6 \times 10^{-19}$ | $1.6 \times 10^{-19}$ |
| SV($t_3$) | | -6906 | -6891 | -6907 | -6893 | -6915 | $8.6 \times 10^{-30}$ | $8.4 \times 10^{-30}$ |
| SV($S\mathcal{N}$) | | **-6975** | **-6955** | **-6971** | **-6951** | -6976 | $\mathbf{7.1 \times 10^{-18}}$ | $\mathbf{6.7 \times 10^{-18}}$ |
| SV($St_3$) | | -6924 | -6905 | -6920 | -6900 | -6923 | $3.6 \times 10^{-29}$ | $4.1 \times 10^{-29}$ |
| SV lev | | -6954 | -6934 | -6957 | -6937 | -6968 | $9.3 \times 10^{-20}$ | $8.8 \times 10^{-30}$ |

the best one applied to NASDAQ data. Hence, the posterior mean, median, standard deviation and credibility interval of 95% of this model are given in Table 9.

**Table 9.** Posterior mean, median, standard deviation and credibility interval for the parameters of stochastic volatility model with skew Gaussian noise using SIR applied to NASDAQ series.

| Parameters | Summary of the posterior distribution | | | | |
|---|---|---|---|---|---|
| | Mean | Std. dev. | $CI_{0.025}$ | Median | $CI_{0.975}$ |
| $\tau^2$ | 0.2032857 | 0.0590969 | 0.1139529 | 0.1945681 | 0.3426287 |
| $\phi$ | 0.8365621 | 0.0441244 | 0.7357273 | 0.8422046 | 0.9071204 |
| $\alpha$ | -9.5973375 | 0.1023758 | -9.7981681 | -9.5970571 | -9.3955176 |
| $\lambda$ | 0.0014686 | 0.0003089 | 0.0008629 | 0.0014690 | 0.0020713 |

Finally, for the models applied to S&P500 return series, we used the following quantity of particles in all filters: 4.000 for the GARCH(1,1) model with noise and SIR, 50 for the same model with ASIR, 250 for the SV($\mathcal{N}$) model, 200 for the SV($t_3$) model, 400 for the SV($S\mathcal{N}$) model, 200 for the SV($St_3$) model and 1.750 for the SV with leverage model. Results in Table 10 indicates that, among the models applied to S&P500 return series, all selection criteria point to the SV model with leverage as the model most adjusted to the data, except for the DIC that, as for NASDAQ data, indicated GARCH with ASIR.

**Table 10.** Model comparisons by means of likelihood-based information criteria and marginal likelihoods for S&P500 series.

| Model | PF | *AIC* | *BIC* | *EAIC* | *EBIC* | *DIC* | $\widehat{f}_{BS}(y)$ | $\widehat{f}_{IS}(y)$ |
|---|---|---|---|---|---|---|---|---|
| GARCH | SIR | -7275 | -7255 | -7273 | -7253 | -7279 | $6.0 \times 10^{-171}$ | $6.0 \times 10^{-171}$ |
| | ASIR | -7184 | -7164 | -7273 | -7253 | **-7370** | $6.5 \times 10^{-171}$ | $6.4 \times 10^{-171}$ |
| SV($\mathcal{N}$) | SIR | -7298 | -7283 | -7295 | -7280 | -7297 | $1.1 \times 10^{-162}$ | $1.1 \times 10^{-162}$ |
| SV($t_3$) | | -7247 | -7232 | -7245 | -7230 | -7249 | $6.5 \times 10^{-174}$ | $6.2 \times 10^{-174}$ |
| SV($S\mathcal{N}$) | | -7309 | -7289 | -7309 | -7289 | -7317 | $5.4 \times 10^{-162}$ | $5.2 \times 10^{-162}$ |
| SV($St_3$) | | -7251 | -7231 | -7250 | -7231 | -7257 | $7.0 \times 10^{-175}$ | $8.2 \times 10^{-175}$ |
| SV lev | | **-7319** | **-7300** | **-7323** | **-7303** | -7335 | $\mathbf{5.0 \times 10^{-158}}$ | $\mathbf{5.6 \times 10^{-158}}$ |

Again, the marginal likelihood criterion indicates the stochastic volatility model with leverage as the best one applied to S&P500 data. Therefore, the posterior mean, median, standard deviation and credibility interval of 95% of this model are given in Table 11.

It is worth mentioning that applications were performed in three real series because they had different behaviours, which was confirmed by the results of the model comparisons, which pointed to different models in each of the series.

**Table 11.** Posterior mean, median, standard deviation and credibility interval for the parameters of stochastic volatility model with leverage using SIR applied to S&P500 series.

| Parameters | Summary of the posterior distribution | | | | |
|---|---|---|---|---|---|
| | Mean | Std. dev. | $CI_{0.025}$ | Median | $CI_{0.975}$ |
| $\tau^2$ | 0.1082816 | 0.0204612 | 0.074450 | 0.1063194 | 0.1550263 |
| $\phi$ | 0.9828918 | 0.0155592 | 0.9423327 | 0.9874117 | 0.9990852 |
| $\alpha$ | -11.096292 | 1.8050948 | -16.746751 | -10.532022 | -9.3502158 |
| $\rho$ | -0.6686287 | 0.0956189 | -0.8154062 | -0.6827698 | -0.4396470 |

## 7. Concluding remarks

This paper deals with modelling volatility through GARCH(1,1) model with noise and a few stochastic volatility models. These models are in the class of non-linear or non-Gaussian state space models. In order to infer on the static parameters and the state vector, we have proposed to work with particle filters and adaptive Metropolis-Hastings algorithms. The particle filters are suitable for obtaining the filtering distributions as well as to obtain an unbiased estimate of the likelihood. The latter is coupled into an adaptive Metropolis-Hastings scheme to sample from the posterior of the static parameters. The proposed method used in this paper is a powerful tool since it allows inference in a large class of models, such as change the prior distributions, without much effort in implementing the MCMC or to worry about proposal distributions and how to choose the hyperparameters. On the other hand, due to generality of our proposed approach, the resulting algorithm may be slow compared to other known methods. In any case, theoretical properties guarantees that the algorithm really draws a sample from the correct posterior distribution.

Moreover, we have also applied the mentioned models above to simulated series and three log-returns data sets - namely BOVESPA, NASDAQ and S&P500. In our applications and methodology, we computed likelihood-based information criteria and marginal likelihoods to do model comparisons. From the Bayesian perspective and in our algorithms, all these measures for model comparisons are easily obtained in the adaptive independent Metropolis-Hastings sampling, which is another advantage of our approach.

For future work, there is a need of more detailed versions of particle filters in order to reduce the variability of the likelihood estimator, thus improving on the convergence and other properties of the adaptive Metropolis-Hastings sampling. In addition, the methodology may also be applied to other class of state space models, including multivariate ones.

## References


[1] Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72, 269–342. doi: 10.1111/j.1467-9868.2009.00736.x.

[2] Arroyo, A.H., Silva, R.S., Migon, H.S., 2023. Bayesian quantile stochastic frontier models. Journal of Econometrics and Statistics 3, 157–184. doi:.

[3] Azzalini, A., 1985. A class of distributions wich includes the normal ones. Scandinavian Journal of Statistics 12, 171–178.

[4] Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbations of symmetry with emphasis on a multivariate skew $t$ distribtuion. Journal of Royal Statistical Society,

Series B 65, 367–389. doi: 10.1111/1467-9868.00391.

[5] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327. doi: 10.1016/0304-4076(86)90063-1.

[6] Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339. doi: 10.2307/1913710.

[7] Giordani, P., Kohn, R., 2010. Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals. Journal of Computational and Graphical Statistics 19, 243–259.

[8] Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings F – Radar and Signal Processing 140, 107–113.

[9] Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. Computational Statistics 14, 375–395. doi: 10.1007/s001800050022.

[10] Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7, 223–242.

[11] Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their application. Biometrika 57, 97–109. doi: 10.1093/biomet/57.1.97.

[12] Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business & Economic Statistics 12, 371–389. doi: 10.2307/1392199.

[13] Meng, X.L., Wong, W.H., 1996. Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. Statistica Sinica 6, 831–860.

[14] Pitt, M.K., Shephard, N., 1999. Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association 94, 590–599. doi: 10.1080/01621459.1999.10474153.

[15] Pitt, M.K., Silva, R.S., Giordani, P., Kohn, R., 2012. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. Journal of Econometrics 171, 134–151. doi: 10.1016/j.jeconom.2012.06.004.

[16] Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. Annals of Applied Probability 7, 110–120. doi: 10.1214/aoap/1034625254.

[17] Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. Journal of Computational and Graphical Statistics 18, 349–367. doi: 10.1198/jcgs.2009.06134.

[18] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statististical Society: Series B 64, 583–639. doi: 10.1111/1467-9868.00353.

[19] Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762. doi: 10.1214/aos/1176325750.

## Appendix A. Fully Adapted Particle Filter for the GARCH(1,1) with Noise

First, consider the GARCH(1,1) model with noise given in Section 5.1. Moreover, we omit the dependence of $\tau_t^2$ on $x_{t-1}$ for while. It follows that

$$\begin{aligned} -2\log(f(y_t|x_t)f(x_t|x_{t-1}^{(\ell)})) &= \kappa + \log\sigma^2 + \log\tau_t^{2(\ell)} + \frac{(y_t - x_t)^2}{\sigma^2} + \frac{x_t^2}{\tau_t^{2(\ell)}} \\ &= \log\tau_t^{2*} + \frac{1}{\tau_t^{2*}}\left(x_t - \delta_t^{(\ell)}\right)^2 + \Delta_t^{(\ell)}, \end{aligned}$$

where $\kappa$ is a constant,

$$\tau_t^{2*} = \left(\frac{1}{\sigma^2} + \frac{1}{\tau_t^{2(\ell)}}\right)^{-1} \quad \text{and} \quad \delta_t^{(\ell)} = \frac{y_t \tau_t^{2*}}{\sigma^2}$$

while

$$\Delta_t^{(\ell)} = \kappa + \log \sigma^2 + \log \tau_t^{2(\ell)} - \log \tau_t^{2*} + \frac{y_t^2}{\sigma^2} - \frac{[\delta_t^{(\ell)}]^2}{\tau_t^{2*}}.$$

Hence,

$$g(\ell | y_{1:t}) \propto \exp\left(-\frac{\Delta_t^{(\ell)}}{2}\right) \pi_{t-1}^{(\ell)} \quad \text{and} \quad g(x_t | \ell, y_{1:t}) \sim N(\delta_t^{(\ell)}, \tau_t^{2*}).$$

## Appendix B. Standard Particle Filter for the SV with Leverage

Here, consider the stochastic volatility model with leverage given in Section 5.2. It follows that

$$f(y_t, x_t | x_{t-1}) = \frac{1}{2\pi e^{x_t/2} \tau \sqrt{1-\rho^2}} \exp(-u_t/2),$$

where

$$\begin{aligned} u_t &= \frac{1}{1-\rho^2}\left[\frac{y_t^2}{e^{x_t}} + \frac{(x_t - \xi_t)^2}{\tau^2} - 2\rho \frac{y_t}{e^{x_t/2}} \frac{(x_t - \xi_t)}{\tau}\right] \\ &= \frac{1}{1-\rho^2}\left[\frac{1}{e^{x_t}}\left(y_t - \rho \frac{(x_t - \xi_t)}{\tau} e^{x_t/2}\right)^2\right] + \frac{(x_t - \xi_t)^2}{\tau^2} - \kappa(x_t), \end{aligned}$$

where $\kappa(x_t) = \rho^2 e^{x_t} (x_t - \xi_t)^2 / \tau^2$. Taking into consideration that

$$f(y_t, x_t | x_{t-1}) = f(y_t | x_t, x_{t-1}) f(x_t | x_{t-1}),$$

thus

$$y_t | x_t, x_{t-1} \sim \mathcal{N}\left(\rho \frac{(x_t - \xi_t)}{\tau} e^{x_t/2}, e^{x_t}(1-\rho^2)\right) \quad \text{and} \quad x_t | x_{t-1} \sim \mathcal{N}(\xi_t, \tau^2).$$

Note that $\kappa(x_t)$ is treated as a constant of $f(y_t | x_t, x_{t-1})$.