

Estimation of Stochastic Volatility Models Using Optimized Filtering Algorithms

Saba Infante
Yachay Tech

Cesar Luna
Universidad de Carabobo

Luis Sánchez
Universidad Técnica
de Manabí

Aracelis Hernández
Universidad de Carabobo

Abstract

In this paper, we describe and implement two recursive filtering algorithms, the optimized particle filter, and the Viterbi algorithm, which allow the joint estimation of states and parameters of continuous-time stochastic volatility models, such as the Cox Ingersoll Ross and Heston model. In practice, good parameter estimates are required so that the models are able to generate accurate forecasts. To achieve the objectives the proposed algorithms were implemented using daily empirical data from the time series of the *S&P500* returns of the stock exchange index. The proposed methodology facilitates computational calculations of the marginal likelihood of states and allows the reconstruction of unknown states in a suitable way, and reliable estimation of the parameters. To measure the quality of estimation of the algorithms, we used the square root of the mean square error and relative deviation standard as measures of goodness of fit. The estimated errors are insignificant for the analyzed data and the two models considered. We also calculated the execution times of the algorithms, demonstrating that the Viterbi algorithm has less execution time than the optimized particle filter.

Keywords: stochastic volatility models, optimized particle filter, Viterbi algorithm.

1. Introduction

The Black-Scholes model described by [Black and Scholes \(1973\)](#) is an equation derived from the financial mathematics used to determine the prices of certain financial assets. The assumptions of this model are given in terms of an ideal scenario where it is assumed that it is possible to conduct continuous trading, the markets are perfect, the interest rate is risk-free and fixed and the price of the underlying asset behaves as a random variable that is modeled as a stochastic process. However, it has been shown by empirical studies that these assumptions are not realistic because the model does not explain the true impact of change on financial markets, such as changes in volatility. Currently, there are more sophisticated models that incorporate volatility as a random variable related to risk factors. Two most representative models are the model of Cox Ingersoll Ross (CIR) (see [Cox, Ingersoll, and Ross \(1985\)](#)) and the model of Heston (see [Heston \(1993\)](#)).

The problem with these models lies in the estimation of the parameters. When the maximum likelihood estimation method is used, there is a first drawback related to obtaining a solution

in closed form for the transition density when the solution of the transition density between prices and volatility is not known in closed form, see [Ait-Sahalia \(2002\)](#) for detail.

A second problem arises when the time series of prices are only partially observed, and volatility is unknown when integrating the joint density to obtain the marginal density, the process results in the handling of analytically intractable integrals, which need to be solved by numerical or technical Monte Carlo methods.

In recent years some solutions have been proposed to solve the above problems. For example, [Ait-Sahalia and Kimmel \(2007\)](#) and [Ait-Sahalia and Kimmel \(2010\)](#), proposed closed-form approximations to the log-likelihood function. MCMC methods in [Eraker \(2001\)](#), [Eraker \(2004\)](#), [Eraker, Johannes, and N. \(2003\)](#), [Jacquier, Polson, and P. \(2004\)](#) have been used for such purposes. In [Bates \(1996\)](#), [Johannes, Polson, and Stroud \(2009\)](#) and [Christoffersen, Jacobs, and Mimouni \(2010\)](#), proposed computational algorithms based on filter theory. [Hurn, Lindsay, and McClelland \(2005\)](#) described a maximum likelihood method for estimating the parameters of the Heston volatility model using market index data and the prices of the options inscribed in that index using a particle filter algorithm and the study shows the efficiency of the filter on simulated and real data taken from the *S&P 500* index. [Kleppe, Jun, and Skaug \(2009\)](#) implemented a method for maximum likelihood estimation in stochastic volatility models, which does not require observations of either the price option or the volatility. In order to integrate the volatility of the latent process from the joint density of profitability and volatility, the use of a technique of modified importance sampling for the approximation of the continuous time model using the Euler-Maruyama scheme was proposed. [Kleppe, Jun, and Skaug \(2014\)](#) develop a maximum likelihood method to estimate partially observed diffusion models based on data sampled at discrete times. The method combines two estimation techniques, the first, proposed by [Ait-Sahalia \(2008\)](#), and which was used to obtain an exact approximation of the joint transition probability density of latent states and observed measurements. The second was proposed in [Richard and Zhang \(2007\)](#) and used an importance sampling technique to integrate the latent states and obtain an approximation of the likelihood function. [Javaheri, Lautier, and Galli \(2003\)](#) introduced the use of filtering algorithms in the world of the quantitative finances. In particular, the Kalman filter, the extended Kalman filter, the unscented Kalman filter, the nonlinear Kushner's filter and the particle filter are implemented for stochastic volatility models and structure models of commodity price terms. [Spall \(2003\)](#) proposed a dual estimation algorithm based on Markov chain Monte Carlo to estimate states and parameters in a state space model. Recently, in [Javaheri \(2015\)](#) established a methodology to estimate parameters in volatility models of time series of market price. In particular, problems related to stochastic volatility models, statistical inference techniques, filtering algorithm and optimization are considered, the consistency of the parameters is demonstrated, and recommends the use of observations of the markets of better quality options.

The main contribution of this article is based on the use of the filtering theory for the joint estimation of states and parameters ([Haykin \(2001\)](#), [Liu and West \(2001\)](#)) in the stochastic volatility models of CIR and Heston. In particular, the optimized particle filter algorithm proposed by [Yang and Xing \(2011\)](#) was implemented in this context. Assuming the known parameters, the Viterbi algorithm, [Viterbi \(1967\)](#), was implemented to estimate the maximum a posteriori states. For both models, we obtain the estimated likelihood and show how the optimized algorithms work in an application with real data. In summary, the proposed in this article:

- Two stochastic volatility models are formulated, these models are defined on terms of stochastic processes of continuous times with continuous space state.
- The dynamic system in continuous time is approximate by a discrete system, i.e., the derivatives in continuous time are approximated by difference equations in discrete time, which can be expressed in the form of space state model.
- Once that the models are in the form space state, it is necessary to estimate the states and parameters for which two algorithms are proposed.

The rest of article is summarized as follows: in section 2, Cox Ingersoll Ross model is defined, this model describes the evolution of interest rates (i.e specifies that the instantaneous interest rate follows a stochastic differential equation) and can be used in the valuation of interest rate derivatives; the section 3 contains the Heston model, this is a mathematical model that describing the evolution of the volatility of an underlying asset, assumes that the price of the asset is determined by a stochastic process; in section 4, optimized particle filter is using for generate new observations into sampling process and also optimizes it, through optimized particle filter, particles are moved towards regions where they have larger values of posterior density function; in section 5 the Viterbi algorithm are developed, this algorithm is a maximum a posteriori (MAP) estimation method that rely on a particle cloud representation of the filtering distribution which evolves through time using importance sampling and resampling ideas. The MAP estimation is then performed using a classical dynamic programming technique applied to the discretised version of the state space model; in section 6 the results obtained for two different models are shown, in section 7 contains a final discussion and conclusions and lastly, section 8 contains the acknowledgements.

2. Cox Ingersoll Ross model

The Cox Ingersoll Ross (CIR) model, describes the evolution of the interest rate by the following stochastic differential equation

$$dr_t = \beta(\mu - r_t)dt + \sigma\sqrt{r_t}dB_t \quad (1)$$

where r_t is the interest rate process, B_t is a Brownian motion (which models the random risk factor of the market), β , μ and σ are parameters. The parameter β corresponds to the speed of adjustment, μ to the mean and σ to volatility. The drift factor $\beta(\mu - r_t)$ is exactly the same as in the Vasicek model. It ensures mean reversion of the interest rate towards the long run value μ , with speed of adjustment governed by the strictly positive parameter β . The CIR model ensures that the process has mean reversion and avoids the possibility of negative interest rates. A process satisfying the equation (1) is called the CIR process, the transition density of the CIR model is not Gaussian.

Stochastic volatility models are generally constructed based on mean reversion of volatility parameter, which reflect through observations that periods of low volatility tend to be followed by a reversion to a more moderate long-term level.

In order to solve analytically the equation given in (1), we need to introduce the chain rule for stochastic differentials, called Itô's Lemma, which is defined as follows:

Theorem 1 (Itô's Lemma). *Let B_t be a Brownian motion and x_t be an Itô drift-diffusion process which satisfies the stochastic differential equation $dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dB_t$. Suppose that $y_t = u(x_t, t)$, $u : \mathbf{R} \times [0, T] \rightarrow \mathbf{R}$, is continuous, $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x}$, $\frac{\partial^2 u}{\partial x^2}$, exist and are continuous then*

$$\begin{aligned} dy_t &= \frac{\partial u}{\partial t}(x_t, t)dt + \frac{\partial u}{\partial x}(x_t, t)dx_t + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(x_t, t)(dx_t)^2 \\ &= \left(\frac{\partial u}{\partial t}(x_t, t) + \frac{\partial u}{\partial x}(x_t, t)\mu(x_t, t) + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(x_t, t)\sigma(x_t, t)^2 \right) dt + \frac{\partial u}{\partial x}(x_t, t)\sigma(x_t, t)dB_t \end{aligned}$$

The term $(dx_t)^2$ is interpreted using the following identities:

- $dt dt = 0$.
- $dt dB_t = dB_t dt = 0$
- $dB_t dB_t = dt$.

Now, using the transformation $v_t = \ln(r_t) - \ln(\sigma^2)$ and applying Itô's Lemma in CIR model, where

$$\mu(x_t, t) = \beta(\mu - r_t) \quad , \quad \sigma(x_t, t) = \sigma\sqrt{r_t} \quad (2)$$

$$\frac{\partial v_t}{\partial r_t} = \frac{1}{r_t}, \quad \frac{\partial^2 v_t}{\partial r_t^2} = -\frac{1}{r_t^2}, \quad \text{and} \quad \frac{\partial v_t}{\partial t} = 0 \quad (3)$$

then

$$dv_t = \left(\frac{\partial v_t}{\partial t} + \frac{\partial v_t}{\partial r_t} \mu(x_t, t) + \frac{1}{2} \frac{\partial^2 v_t}{\partial r_t^2} \sigma^2(x_t, t) \right) dt + \frac{\partial v_t}{\partial r_t} \sigma(x_t, t) dB_t \quad (4)$$

we have

$$dv_t = \frac{1}{r_t} \left\{ \beta(\mu - r_t) - \frac{1}{2} \frac{\sigma^2 r_t}{r_t^2} \right\} dt + \frac{\sigma \sqrt{r_t}}{r_t} dB_t \quad (5)$$

$$= \frac{1}{r_t} \left\{ \beta(\mu - r_t) - \frac{1}{2} \frac{\sigma^2}{r_t} \right\} dt + \frac{\sigma}{\sqrt{r_t}} dB_t \quad (6)$$

$$= \left\{ \beta \left(\frac{\mu}{r_t} - 1 \right) - \frac{\sigma^2}{2r_t^2} \right\} dt + \exp\left(-\frac{v_t}{2}\right) dB_t \quad (7)$$

where

$$v_t = \ln(r_t) - \ln(\sigma^2) = \ln\left(\frac{r_t}{\sigma^2}\right) \quad (8)$$

taking exponential and solving, we obtain

$$\frac{\sigma}{\sqrt{r_t}} = \exp\left(-\frac{v_t}{2}\right) \quad (9)$$

In order to illustrate the methodology of estimation of parameters and states proposed in this article, we use the reparametrized version of the CIR model in continuous time, which is discussed in [Chib, Pitt, and Shephard \(2006\)](#) and [Poyiadjis, Doucet, and Singh \(2011\)](#), where the volatility follows a square root process given by

$$v_{t+1} = \mu + v_t + \beta \exp(-v_t) + \exp\left(-\frac{v_t}{2}\right) \zeta_{t+1} \quad (10)$$

and

$$y_t = \sigma_\zeta \exp\left(-\frac{v_t}{2}\right) \xi_t \quad (11)$$

where ζ_{t+1} and ξ_t are two identically distributed independent random variables, $\zeta_{t+1} \sim N(0, \sigma_\zeta^2)$ and $\xi_t \sim N(0, \sigma_\xi^2)$. We are interested in estimating the states $v_{0:t} = (v_0, \dots, v_t)$ and parameters $\Theta = (\mu, \beta, \sigma_\zeta)$ of system placed in (10), and (11), given the observation $y_{1:t} = (y_1, \dots, y_t)$.

3. The Heston model

The Heston model (1993) is a mathematical model used to describe the behavior of a bivariate stochastic process between stock prices s_t and its variance v_t . It initially emerged as a generalization of the Black and Scholes option pricing model, but assuming that volatility is a stochastic process. The model is governed by the following system of stochastic differential equations

$$ds_t = \mu s_t dt + \sqrt{v_t} s_t dB_{1,t} \quad (12)$$

and

$$d(\sqrt{v_t}) = -\beta \sqrt{v_t} dt + \delta dB_{2,t} \quad (13)$$

If $y_t = (\sqrt{v_t})^2$ and Itô's Lemma is applied, then

$$dy_t = 2\sqrt{v_t} d(\sqrt{v_t}) + \frac{1}{2} [2(d(\sqrt{v_t}))^2] \quad (14)$$

$$dy_t = 2\sqrt{v_t} (-\beta \sqrt{v_t} dt + \delta dB_{2,t}) + (-\beta \sqrt{v_t} dt + \delta dB_{2,t})^2 \quad (15)$$

developing, we obtain

$$dy_t = (\delta^2 - 2\beta v_t)dt + 2\delta\sqrt{v_t}dB_{2,t} = 2\beta\left(\frac{\delta^2}{2\beta} - v_t\right)dt + 2\delta\sqrt{v_t}dB_{2,t} \quad (16)$$

If we make $k = 2\beta$; $\theta = \frac{\delta^2}{2\beta}$ and $\sigma = 2\delta$, we get:

$$dy_t = dv_t = k(\theta - v_t)dt + \sigma\sqrt{v_t}dB_{2,t} \quad (17)$$

to obtain this last result, the following properties were used $dt dt = 0$, $dt dB_t = dB_t dt = 0$ and $dB_t dB_t = dt$.

Then the given system in (12) and (18) becomes

$$ds_t = \mu s_t dt + \sqrt{v_t} s_t dB_{1,t} \quad (18)$$

$$dv_t = k(\theta - v_t)dt + \sigma\sqrt{v_t}dB_{2,t} \quad (19)$$

and

$$dB_{1,t} dB_{2,t} = \rho dt \quad (20)$$

where the system variables are defined as follows:

- s_t : is the asset price,
- v_t : is the volatility of the asset,
- μ : is the expected return on the asset,
- θ : is the long-term price change,
- k : is the rate at which volatility tends towards its long-term average,
- σ : is the volatility of $\sqrt{v_t}$,
- ρ : is the correlation of Brownian motions,
- $dt = t_k - t_{k-1}$: is a small increase over time,
- $dB_{1,t}$: is a standard one dimensional Brownian motion,
- $dB_{2,t}$: is a standard one dimensional Brownian motion.

The equations (12) and (13) is a system of stochastic differential equations, where the solutions represent a set of random variables x_t and v_t indexed by real numbers $t > 0$ is called a continuous-time stochastic process. Each instance, or realization of the stochastic process is a choice from the random variables x_t and v_t for each t , and is therefore a function of t . The equations (12) and (13) are, by definition, an integral equation (integrating from t to $t + dt$)

$$\int_t^{t+dt} ds_t = \int_t^{t+dt} \mu s_t dt + \int_t^{t+dt} \sqrt{v_t} s_t dB_{1,t} \quad (21)$$

and

$$\int_t^{t+dt} dv_t = \int_t^{t+dt} k(\theta - v_t)dt + \int_t^{t+dt} \sigma\sqrt{v_t}dB_{2,t} \quad (22)$$

where the meaning of the last integral in (21) and (22), are called an Ito integral.

The simplest effective computational method for the approximation of ordinary differential equations is Euler's method. The Euler-Maruyama method is the analogue of the Euler method for ordinary differential equations. To develop an approximate solution on the interval $[c; d]$, assign a grid of points

$$c = \tau_0 < \tau_1 < \dots < \tau_n = d \quad (23)$$

approximate x values

$$\omega_0 < \omega_1 < \dots < \omega_n \quad (24)$$

will be determined at the respective t points, given the SDE initial value problem.

The simplest way to discretize the process in the equations (21) and (22) is to use Euler discretization, this discretization is necessary to find the solution of the equations (18) and (19). In the following, is shown the process of discretization of s_t , v_t and $\ln(s_t)$.

- Discretization of the process s_t

The stochastic differential equation given in (12) and (13) in integral form can be expressed as follows

$$s_{t+dt} = s_t + \int_t^{t+dt} \mu s_u du + \int_t^{t+dt} \sqrt{v_u} s_u dB_{1,u} \quad (25)$$

Using a discretization of Euler we obtain that

$$\int_t^{t+dt} \mu s_u du \approx \mu s_t dt \quad (26)$$

and

$$\int_t^{t+dt} \sqrt{v_u} s_u dB_{1,u} = \sqrt{v_t} s_t (B_{t+dt} - B_t) = \sqrt{v_t} s_t z_u \quad ; \quad z_u \sim N(0, 1) \quad (27)$$

so the stochastic differential equation of the stock price in its discretized version of Euler is given by

$$s_{t+dt} = s_t + \mu s_t dt + \sqrt{v_t} s_t z_u \quad (28)$$

- Discretization of the process v_t

With a relation similar to the discretization of the stochastic difference equation of s_t , a version for v_t is obtained, which is written as follows

$$v_{t+dt} = v_t + \int_t^{t+dt} k(\theta - v_u) du + \int_t^{t+dt} \sigma \sqrt{v_u} dB_{2,u} \quad (29)$$

Using a discretization of Euler we obtain that

$$\int_t^{t+dt} k(\theta - v_u) du \approx k(\theta - v_t) dt \quad (30)$$

and

$$\int_t^{t+dt} \sigma \sqrt{v_u} dB_{2,u} \approx \sigma \sqrt{v_t} (B_{t+dt} - B_t) = \sigma \sqrt{v_t} z_s \quad ; \quad z_s \sim N(0, 1)$$

where $\text{Corr}(z_u, z_s) = \rho$. Finally, the equation of the volatility in the discrete version of Euler is given by

$$v_{t+dt} = v_t + k(\theta - v_t) dt + \sigma \sqrt{v_t} dt z_s \quad (31)$$

To avoid obtaining a negative variance, v_t is replaced by $v_t^+ = \max(0, v_t)$, or $v_t^+ = |v_t|$.

- Discretization of the process $\ln(s_t)$

If we now consider $y_t = \ln(s_t)$, using Itô's Lemma, we obtain

$$dy_t = d\ln(s_t) = \left(\mu - \frac{1}{2}v_t\right) dt + \sqrt{v_t} dB_{1,t}$$

Integrating we obtain

$$\ln(s_{t+dt}) = \ln(s_t) + \int_0^t \left(\mu - \frac{1}{2}v_u\right) du + \int_0^t \sqrt{v_u} dB_{1,u}$$

$$\begin{aligned}
 &= \ln(s_t) + \left(\mu - \frac{1}{2}v_t\right) dt + \sqrt{v_t}(B_{1,t+dt} - B_{1,t}) \\
 &= \ln(s_t) + \left(\mu - \frac{1}{2}v_t\right) dt + \sqrt{v_t}dz_u
 \end{aligned} \tag{32}$$

then Euler's discretization for $\ln(s_t)$ is given by

$$\ln(s_{t+dt}) = \ln(s_t) + \left(\mu - \frac{1}{2}v_t\right) dt + \sqrt{v_t}dz_u \tag{33}$$

or equivalently

$$s_{t+dt} = s_t \exp \left[\left(\mu - \frac{1}{2}v_t\right) dt + \sqrt{v_t}dz_u \right] \tag{34}$$

- The model in the state space

The Heston's stochastic volatility model can be written as state space models, these models are useful for describing data in many different areas, such as financial time series, environmental data, and clinical trial, among other applications; and are used to estimate latent state processes (unknown processes), based on measurements from the observation process. Also, allows to write equations (18) and (19) in discretized form in time, facilitating the computational implementation of the estimation algorithms, this representation as follows

$$v_{t+1} = v_t + \kappa(\theta - v_t) + \sigma\sqrt{v_t}z_s \tag{35}$$

$$y_{t+1} = y_t + \left(\mu - \frac{1}{2}v_t\right) + \sqrt{v_t}z_u \tag{36}$$

where $dt = 1$ and $y_{t+1} = \ln s_{t+1}$ are assumed. The equation given in (35) is the state equation and the equation given in (36) is the observation equation. A way to generate z_u and z_s with correlation ρ , consists of generating identically distributed independent normal random variables $z_1 \sim N(0, 1)$ and $z_2 \sim N(0, 1)$, such that $z_u = z_1$ and $z_s = \rho z_1 + \sqrt{1 - \rho^2}z_2$. The equations given in (35) and (36) can be rewritten as

$$v_{t+1} = v_t + \kappa(\theta - v_t) + \sqrt{v_t}\sigma(\rho z_1 + \sqrt{1 - \rho^2}z_2) \tag{37}$$

$$y_{t+1} = y_t + \left(\mu - \frac{1}{2}v_t\right) + \sqrt{v_t}z_1 \tag{38}$$

The model parameters are $\Theta = (\kappa, \theta, \mu, \sigma, \rho)$. The joint distribution of $(v_{t+1}, y_{t+1})^T$ is given by

$$\begin{pmatrix} v_{t+1} \\ y_{t+1} \end{pmatrix} \sim N(\mu_t; \Sigma_t) \tag{39}$$

where

$$\mu_t = \begin{pmatrix} \mathbf{E}(v_{t+1}) \\ \mathbf{E}(y_{t+1}) \end{pmatrix} \tag{40}$$

and

$$\Sigma_t = \begin{pmatrix} \mathbf{Var}(v_{t+1}) & \mathbf{Cov}(v_{t+1}, y_{t+1}) \\ \mathbf{Cov}(y_{t+1}, v_{t+1}) & \mathbf{Var}(y_{t+1}) \end{pmatrix} \tag{41}$$

The bivariate transition density of a correlated normal distribution is given by

$$p(v_{t+1}, y_{t+1}) = \frac{1}{2\pi\sqrt{\mathbf{Var}(v_{t+1})}\sqrt{\mathbf{Var}(y_{t+1})}\sqrt{(1 - \rho^2)}} \exp \left\{ -\frac{1}{2(1 - \rho^2)}Q(v_{t+1}, y_{t+1}) \right\} \tag{42}$$

Where $Q(v_{t+1}, y_{t+1}) = q$

$$q = \left[\left(\frac{v_{t+1} - \mathbf{E}(v_{t+1})}{\sqrt{\mathbf{Var}(v_{t+1})}} \right)^2 - 2\rho \left(\frac{v_{t+1} - \mathbf{E}(v_{t+1})}{\sqrt{\mathbf{Var}(v_{t+1})}} \right) \left(\frac{y_{t+1} - \mathbf{E}(y_{t+1})}{\sqrt{\mathbf{Var}(y_{t+1})}} \right) + \left(\frac{y_{t+1} - \mathbf{E}(y_{t+1})}{\sqrt{\mathbf{Var}(y_{t+1})}} \right)^2 \right]$$

Note that

$$\mathbf{E}(v_{t+1}) = v_t + \kappa(\theta - v_t) \quad ; \quad \mathbf{Var}(v_{t+1}) = \sigma^2 v_t \quad (43)$$

$$\mathbf{E}(y_{t+1}) = y_t + \left(\mu - \frac{1}{2}v_t\right) \quad ; \quad \mathbf{Var}(y_{t+1}) = v_t \quad (44)$$

and

$$\mathbf{Cov}(v_{t+1}, y_{t+1}) = \sigma\sqrt{v_t} \left[y_t + \left(\mu - \frac{3}{2}v_t\right) \right] \quad (45)$$

Therefore, marginal distributions $p(v_{t+1}|v_t)$ and $p(y_{t+1}|y_t)$ are given by

$$p(v_{t+1}|v_t) \sim N(v_t + \kappa(\theta - v_t), \sigma^2 v_t) \quad (46)$$

$$p(y_{t+1}|y_t) \sim N\left(y_t + \left(\mu - \frac{1}{2}v_t\right), v_t\right) \quad (47)$$

The conditional distribution $p(y_{t+1}|v_{t+1})$, is given by

$$p(y_{t+1}|v_{t+1}) \sim N\left(y_t + \left(\mu - \frac{1}{2}v_t\right) + \frac{\rho}{\sigma}(v_{t+1} - (v_t + \kappa(\theta - v_t))), v_t(1 - \rho^2)\right) \quad (48)$$

Finally, the joint transition density can be rewritten as

$$p(v_{t+1}, y_{t+1}) = \frac{1}{2\pi\sigma v_t \sqrt{(1 - \rho^2)}} \exp\left\{-\frac{1}{2(1 - \rho^2)}Q(v_{t+1}, y_{t+1})\right\} \quad (49)$$

where

$$Q(v_{t+1}, y_{t+1}) = \left[\frac{v_{t+1} - (v_t + \kappa(\theta - v_t))}{\sigma\sqrt{v_t}} \right]^2 - 2\rho \left[\frac{v_{t+1} - (v_t + \kappa(\theta - v_t))}{\sigma\sqrt{v_t}} \right] \left[\frac{y_{t+1} - (y_t + (\mu - \frac{1}{2}v_t))}{\sqrt{v_t}} \right] + \left[\frac{y_{t+1} - (y_t + (\mu - \frac{1}{2}v_t))}{\sqrt{v_t}} \right]^2$$

One way to eliminate the correlation between the errors associated with the state and observation equations in the state space model is to subtract on both sides of the equation (37), a multiple of the quantity $y_t + (\mu - \frac{1}{2}v_t) + \sqrt{v_t}z_1 - y_{t+1} = 0$. This procedure would eliminate the correlation (Javaheri (2015)); that is to say

$$v_{t+1} = v_t + \kappa(\theta - v_t) + \sqrt{v_t}\sigma(\rho z_1 + \sqrt{1 - \rho^2}z_2) - \rho\sigma \left[y_t + \left(\mu - \frac{1}{2}v_t\right) + \sqrt{v_t}z_1 - y_{t+1} \right]$$

then, the state space model with uncorrelated noises is as follows

$$v_{t+1} = v_t + \kappa(\theta - v_t) + \sqrt{v_t}\sigma(\rho z_1 + \sqrt{1 - \rho^2}z_2) - \rho\sigma \left[y_t + \left(\mu - \frac{1}{2}v_t\right) + \sqrt{v_t}z_1 - y_{t+1} \right]$$

$$y_{t+1} = y_t + \left(\mu - \frac{1}{2}v_t\right) + \sqrt{v_t}z_1$$

or equivalently:

$$v_{t+1} = v_t + \kappa(\theta - v_t) - \rho\sigma \left[y_t + \left(\mu - \frac{1}{2}v_t\right) - y_{t+1} \right] + \sqrt{v_t}\sigma\sqrt{1 - \rho^2}z_2$$

$$y_{t+1} = y_t + \left(\mu - \frac{1}{2}v_t \right) + \sqrt{v_t}z_1$$

where z_1 and z_2 are independent and identically distributed standard Gaussian random variables. Using this system of equations the calculations are faster and the filters are implemented more efficiently.

4. Optimized particle filter (FPO)

In the previous sections, two models of stochastic volatility, the CIR model and the Heston model, were proposed, then a discretization of both was performed, and then represented in the state space form. Now, it is proposed to define the optimized particulate filter to be used to estimate the states, solutions of the SDE, and parameters of the proposed models.

Consider a dynamic system

$$v_t = f(v_{t-1}, \phi) + \epsilon_t \quad (50)$$

$$y_t = h(v_t, \phi) + \zeta_t \quad (51)$$

where v_t is unobserved state of the system, y_t are observations at time $t \in \{0, \dots, T\}$, $\phi \in \mathbf{R}^m$ is parameters vector, $f(\cdot|\cdot)$ is a known density given by the evolution of the states, $h(\cdot|\cdot)$ is also a known density by the observations, which may be linear or non-linear, ϵ_t and ζ_t are estimation errors of the state and observation equations which can be Gaussian or non Gaussian. It is assumed that $\{v_t\}$ is a Markov process generated according to the evolution of the previous state, where the observation process $\{y_t\}$ is conditional independent of the process $\{v_t\}$.

When ϕ is known the inference falls on the posterior distribution $p(v_{0:t}|y_{1:t})$, where the joint distribution is given by

$$p(v_{0:t}, y_{1:t}) = f(v_0) \prod_{i=0}^t f(v_{i+1}|v_i) \prod_{i=1}^t h(y_i|v_i) \quad (52)$$

$f(v_0)$ is an initial state, $v_{0:t} = (v_0, \dots, v_t)$ and $y_{1:t} = (y_1, \dots, y_t)$, denote the states and observations from a time 0 to t . If ϕ is unknown parameter, an a priori distribution $p(\phi)$ is assigned and the Bayesian inference falls on the estimation of the posterior distribution of parameters and states, given by

$$p(v_{0:t}, \phi_t | y_{1:t}) \propto p(v_{0:t}, y_{1:t} | \phi_t) p(\phi_t) \quad (53)$$

In particular, we are interested in estimating the filtering distribution $p(v_t|y_{1:t})$, the predictive distribution $p(v_{t+1}|y_{1:t})$, a posteriori mean $\mu_{t|t} = \mathbf{E}(v_t|y_{1:t})$ and a posteriori covariance $\Sigma_{t|t} = \mathbf{E}[(v_t - m_{t|t})(v_t - m_{t|t})^T]$, which are calculated as follows:

- Prediction of the filtering distribution:

$$p(v_k|y_{1:t-1}) = \int p(v_t|v_{t-1})p(v_{t-1}|y_{1:t-1})dv_{t-1} \quad (54)$$

- Update of the filtering distribution:

$$p(v_t|y_{1:t}) = \frac{p(y_t|v_t)p(v_t|y_{1:t-1})}{\int p(y_t|v_t)p(v_t|y_{1:t-1})dv_t} \quad (55)$$

- The predictive distribution:

$$p(x_{k+1}|\mathbf{y}_{1:k}) = \int p(x_{k+1}|x_k)p(x_k|\mathbf{y}_{1:k-1})dx_k \quad (56)$$

- The a posteriori mean:

$$\mu_{k|k} = \int x_k p(x_k | \mathbf{y}_{1:k}) dx_k \quad (57)$$

- The a posteriori covariance:

$$\Sigma_{k|k} = \int (x_k - m_{k|k})(x_k - m_{k|k})^T p(x_k | \mathbf{y}_{1:k}) dx_k \quad (58)$$

In models with nonlinear structures and with non-Gaussian errors, a posteriori distributions $p(v_{0:t} | y_{1:t})$ or $p(\phi, v_{0:t} | y_{1:t})$ does not have closed form expression, so making inference is complicated; in practice, it is difficult to calculate the integrals given in equations (54) and (55). It is therefore necessary to use approximation Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo methods (SMC). The MCMC and SMC provide efficient computational tools for inference in state space models (See [Andrieu, Doucet, and Holenstein \(2010\)](#)).

In this work we propose a method that allows to simultaneously estimate the states and parameters of the posterior distribution, using a sequence of discrete particles generated by a particle filter algorithm. The approximation is given by

$$p(v_{0:t}, \phi_t | y_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta\{v_t^{(i)}, \phi_t^{(i)}\} \quad (59)$$

where $\{v_t^{(i)}, \phi_t^{(i)}\}_{i=1}^N$ are the particles of states and parameters and δ is the Dirac delta, which represents the generated distribution of the N particles. Taking into account the approximation given in (59), the problem now focuses on how to extract the samples sequentially from the posterior distribution. This step is complicated because the propagation of states depends on the parameters, and vice versa. To solve this situation, to use a density of importance based on the sampling technique of importance is suggested ([Carvalho, Johannes, Lopez, and Nicholas \(2010\)](#)).

The proposed methodology combines the maximum likelihood method with an optimized stochastic particle filter algorithm. The particle filter algorithm works properly for dynamic models with Gaussian and non-Gaussian distributions, with linear and non-linear structures. On the other hand, the classical method of estimation by maximum likelihood implies the optimization of the parameters estimated in such way that the observed data are most likely to be chosen and the likelihood function is defined as

$$L(\phi) = f(v_0) \prod_{i=0}^t f(v_{i+1} | v_i, \phi) \prod_{i=1}^t h(y_{i+1} | v_{i+1}, \phi) \quad (60)$$

where: $v_0 \sim N(\mu_0, \Sigma_0)$, $f(v_{i+1} | v_i, \phi)$ is given by the equation (50), $h(y_{i+1} | v_{i+1}, \phi)$ is given by the equation (51), $\phi = (\beta, \mu, \sigma)$, denotes the parameters of the CIR model and $\phi = (\kappa, \theta, \mu, \sigma, \rho)$ are the Heston model parameters. In practice, one uses the log-likelihood which is numerically better behaved and satisfies

$$l(\phi) = \ln L(\phi) = \ln f(v_0) + \sum_{i=0}^t \ln f(v_{i+1} | v_i, \phi) + \sum_{i=1}^t \ln h(y_{i+1} | v_{i+1}, \phi) \quad (61)$$

The maximum likelihood method is most commonly done by taking the partial derivative of the equation given in (61) for each parameter and setting it equal to zero and the system depending on the parameters is solved.

On the other hand, if we consider the joint density between observations $y_{1:t}$, states $v_{0:t}$ and parameters ϕ , we have

$$p(v_{0:t}, y_{1:t}, \phi) = p(y_{1:t} | v_{0:t}, \phi) p(v_{0:t} | \phi) p(\phi) \quad (62)$$

Furthermore;

$$p(v_{0:t}, y_t, y_{1:t-1}, \phi) = p(y_t | y_{1:t-1}, v_{0:t}, \phi) p(y_{1:t-1}, v_{0:t}, \phi) = p(y_1, v_{0:t}, \phi) \prod_{i=2}^t p(y_i | y_{1:i-1}, v_{0:t}, \phi)$$

where:

$$\begin{aligned} p(y_t | y_{1:t-1}, v_{0:t}, \phi) &= \int p(y_t | v_t, \phi) p(v_t | y_{1:t-1}, \phi) dv_t = \int p(y_t | v_t, \phi) p(v_t | v_{t-1}, \phi) p(v_{t-1} | y_{1:t-1}, \phi) dv_{t-1} \\ &= \int p(y_t | v_{t-1}, \phi) p(v_{t-1} | y_{1:t-1}, \phi) dv_{t-1} \end{aligned} \quad (63)$$

and

$$p(y_1, \phi) = \int p(y_1 | v_0, \phi) p(v_0) dv_0 \quad (64)$$

The equation given in (63) represents the likelihood of the data at time t . Taking logarithm in the equation (63), we get

$$l(p(v_{0:t}, y_{1:t}, \phi)) = \ln[p(y_1, v_{0:t}, \phi)] + \sum_{i=2}^t \ln[p(y_i | y_{1:i-1}, v_{0:t}, \phi)] \quad (65)$$

To simplify the computation, the cost function is chosen as the predicted likelihood at time t , that is:

$$C(\phi) = p(y_t | y_{1:t-1}, \phi) = \int p(y_t | v_t, \phi) p(v_t | y_{1:t-1}, \phi) dv_t \quad (66)$$

However, except in a few simple cases, it is impossible to compute the optimal filter and the likelihood in closed form, the numerical approximation methods are required; this is

$$C(\phi) = p(y_t | y_{1:t-1}, \phi) \approx \frac{1}{N} \sum_{i=1}^N p(y_t | v_t^{(i)}, \phi) \quad (67)$$

The problem of maximizing the cost function is equivalent to finding the zeros of the gradient $\nabla C(\phi)$. A recursive procedure to estimate the parameter vector ϕ such that $\nabla C(\phi) = 0$ proceeds as follows

$$\phi_{t+1} = \phi_t + \gamma_t \hat{\nabla} C(\phi_t) \quad (68)$$

where $\hat{\nabla} C(\phi)$ is the estimation of gradient at the point ϕ_{t-1} and $\{\gamma_t > 0\}$ denotes a sequence of decreasing step-size. One selects a step-size sequence satisfying $\gamma_t \rightarrow 0$, $\sum_{t=1}^{\infty} \gamma_t = \infty$. Under appropriate conditions, the iteration defined in (68) will converge to the true value of ϕ in some stochastic sense. The essential part of the equation (68) is how to obtain the gradient estimate, however, in most cases, it is impossible to compute the closed-form gradient and we must resort to the numerical approximation.

On the other hand, the particulate filters (Gordon, Salmond, and Smith (1993), Doucet, De Freitas, and Gordon (2001)), are based on importance sampling where $\{v_t^{(i)}\}_{i=1}^N$ is simulated sequentially from some importance density function $q(v_t | y_{1:t})$ and all particles v_t are selected according to a weight of importance ω_t , defined by

$$\omega_t = \frac{p(v_t | y_{1:t})}{q(v_t | y_{1:t})} \quad (69)$$

To develop the particle filter algorithm in detail consider that $\{v_t^{(i)}, \omega_t^{(i)}\}_{i=1}^N$ is a random sample that characterizes the filtered density function $p(v_t | y_{1:t})$, where $\{v_t^{(i)}\}_{i=1}^N$ is a set of particles obtained with associated weights $\{\omega_t^{(i)}\}_{i=1}^N$. Then a posteriori distribution also called

filtered distribution at time t can be approximated by an empirical distribution formed by the points of mass or particles, as follows

$$p_N(v_t|y_{1:t}) \approx \sum_{i=1}^N \omega_t^{(i)} \delta(v_t - v_t^{(i)}) \quad ; \quad \sum_{i=1}^N \omega_t^{(i)} = 1 \quad (70)$$

where $\delta(\cdot)$ is the Dirac delta function and the weights are updated recursively by

$$\omega_t^{(i)} = \frac{p(y_t|v_t^{(i)})p(v_t^{(i)}|v_{t-1}^{(i)})}{q(v_t^{(i)}|v_{t-1}^{(i)}, y_t)} \omega_{t-1}^{(i)} \quad (71)$$

If in particular, we consider:

$$q(v_t^{(i)}|v_{t-1}^{(i)}, y_t) = p(v_t^{(i)}|v_{t-1}^{(i)}) \quad (72)$$

then we have the modified weights given by

$$\tilde{\omega}_t^{(i)} = \tilde{\omega}_{t-1}^{(i)} p(y_t|v_t^{(i)}) \quad (73)$$

That is, the particles are taken directly from the a priori distribution at time t . The filtered density $p_N(v_t|y_{1:t})$, can be approximated by

$$p_N(v_t|y_{1:t}) \approx \sum_{i=1}^N \tilde{\omega}_t^{(i)} \delta(v_t - v_t^{(i)}) \quad ; \quad \sum_{i=1}^N \tilde{\omega}_t^{(i)} = 1 \quad (74)$$

Crisan and Doucet (2002) proved that when $N \rightarrow \infty$ the equation given in (74) approaches the true a posteriori distribution $p(v_t|y_{1:t})$.

In particular, optimized particle filter algorithm proposed to be used in this work, using a set of particles and weights $\{v_{t-1}^{(i)}, \omega_{t-1}^{(i)}\}$, which approximate the filtered distribution $p(v_{t-1}|y_{1:t-1})$ in at time $t-1$ is initialized, and given the parameters estimated at a time $t-1$, $\phi_{t-1} = (\beta_{t-1}, \mu_{t-1}, \sigma_{t-1})$ (CIR Model) or $\phi_{t-1} = (\kappa_{t-1}, \theta_{t-1}, \mu_{t-1}, \sigma_{t-1}, \rho_{t-1})$ (Heston model). The cost function can be approximated by

$$\hat{C}(\phi_{t-1}) = \hat{p}(y_t|y_{1:t-1}, \phi_{t-1}) = \sum_{i=1}^N \omega_{t-1}^{(i)} p(y_t|v_t^{(i)}, \phi_{t-1}) \quad (75)$$

where the particles

$$v_t^{(i)}|v_{t-1}^{(i)} \sim q(v_t|v_{t-1}^{(i)}, \phi_{t-1}) = p(v_t|v_{t-1}^{(i)}, \phi_{t-1}) \quad (76)$$

are obtained using a one-step ahead state evolution prediction.

The stochastic optimization technique apply only in the cases where a closed form solution is not available and the approximation method introduced in Spall (1987), Spall (1988) can be used. It is assumed that $C(\phi)$ is a differentiable function in ϕ and the minimum point of ϕ corresponding to a zero point of the gradient. The gradient estimate

$$\hat{\nabla}C(\phi_t) = \left(\hat{\nabla}C_1(\phi_t), \dots, \hat{\nabla}C_m(\phi_t) \right) \quad (77)$$

is given by

$$\hat{\nabla}C_j(\phi_t) = \frac{\hat{C}(\phi_t + c_t \Delta_t) - \hat{C}(\phi_t - c_t \Delta_t)}{2c_t \Delta_{t,j}} \quad (78)$$

c_t denotes a sequence of positive scalars such that $c_t \rightarrow 0$ and $\Delta_t = (\Delta_{t,1}, \dots, \Delta_{t,m})$ is a m -dimensional random perturbation vector. Careful selection of algorithm parameters

(a, c, A, α, r) is required to ensure the convergence. In [Yang and Xing \(2011\)](#) the following values are indicated

$$\gamma_t = \frac{a}{(A + k + 1)^\alpha} \quad ; \quad c_t = \frac{c}{(t + 1)^r} \quad ; \quad \alpha = 0.602 \quad ; \quad r = 0.101 \quad (79)$$

The value of c is chosen empirically and it is approximately equal to the standard deviation of the measurement noise in $C(\phi)$. The values of a, A can be chosen together to ensure effective practical performance. Each components of Δ_t is usually generated from a Bernoulli 1 distribution with probability of $\frac{1}{2}$ for each of 1 independently.

In cases where the gradient has more than one zero point, then the algorithm may only converge to a local minimum. In [Spall \(1988\)](#), some modifications in the algorithm to allow it to search for the global solution among multiple local solutions are given.

We present here how to incorporate maximum-likelihood within the particle filtering framework. The algorithm proceeds as follows:

Algorithm: Optimized Particle Filter

- **Step 1.** Initialization: For $i = 1, \dots, N$, sample $\phi_0^{(i)} \sim p(\phi_0)$ and $v_0^{(i)} \sim p(v_0)$ assign initial important weights as $\omega_0^{(i)} = \frac{1}{N}$ and the initial estimate of states and parameters is:

$$v_0 = \sum_{i=1}^N \tilde{\omega}_0^{(i)} v_0^{(i)} \quad ; \quad \phi_0 = \sum_{i=1}^N \omega_0^{(i)} \phi_0^{(i)} \quad (80)$$

- **Step 2.** Propagation of states and parameters: suppose that states, parameters and weights are known at time $t - 1$:

$$\{v_{t-1}^{(i)}, \phi_{t-1}^{(i)}, \omega_{t-1}^{(i)}\} \quad ; \quad i = 1, \dots, N \quad (81)$$

We proceed to propagate the states and parameters at time t , as follows:

$$\phi_t^{(i)} | v_{t-1}^{(i)}, \phi_{t-1}^{(i)} \sim p(\phi_t | v_{t-1}^{(i)}, \phi_{t-1}^{(i)}) \quad (82)$$

$$v_t^{(i)} | v_{t-1}^{(i)}, \phi_t^{(i)} \sim p(v_t | v_{t-1}^{(i)}, \phi_t^{(i)}) \quad (83)$$

To carry out this procedure a step of propagation of the algorithm in [Liu and West \(2001\)](#) can be implemented. In this case, the parameters can be propagated, as follows

$$\phi_t^{(i)} = \nu \phi_{t-1}^{(i)} + (1 - \nu) m_t^{(i)} + \epsilon_t^{(i)} \quad ; \quad \epsilon_t^{(i)} \sim N(0, h^2 V_t) \quad (84)$$

where

$$m_t^{(i)} = \nu \phi_{t-1}^{(i)} + (1 - \nu) \bar{\phi}_{t-1} \quad ; \quad \bar{\phi}_{t-1} = \frac{1}{N} \sum_{j=1}^N \phi_{t-1}^{(j)} \quad (85)$$

$$V_t = \frac{1}{N} \sum_{j=1}^N \left(\phi_{t-1}^{(j)} - \bar{\phi}_{t-1} \right) \left(\phi_{t-1}^{(j)} - \bar{\phi}_{t-1} \right)^T \quad (86)$$

The constants $\nu = \sqrt{1 - h^2}$ and h measure the contraction and the degree of dispersion of the mixture respectively. The states propagate from a density of importance

$$q(v_t | v_{t-1}^{(i)}, \phi_t^{(i)}) = p(v_t | v_{t-1}^{(i)}, \phi_t^{(i)}) \quad (87)$$

- **Step 3.** Cost function evaluation: for each parameter particle $\phi_t^{(i)}$, generate a m -dimensional simultaneous perturbation vector $\Delta_t^{(i)}$. Compute the perturbed parameter particle $(\phi_t^{(i)} + c_t \Delta_t^{(i)})$ and $(\phi_t^{(i)} - c_t \Delta_t^{(i)})$. For $i = 1, \dots, N$

- Sample:

$$\tilde{v}_+^{(i)} \sim p(v_t | v_{t-1}^{(i)}, \phi_t^{(i)} + c_t \Delta_t^{(i)}) \quad (88)$$

- Compute the likelihood $\tilde{v}_+^{(i)}$:

$$p(y_t | \tilde{v}_+^{(i)}, \phi_t^{(i)} + c_t \Delta_t^{(i)}) \quad (89)$$

- Sample:

$$\tilde{v}_-^{(i)} \sim p(v_t | v_{t-1}^{(i)}, \phi_t^{(i)} - c_t \Delta_t^{(i)}) \quad (90)$$

- Compute the likelihood $\tilde{v}_-^{(i)}$:

$$p(y_t | \tilde{v}_-^{(i)}, \phi_t^{(i)} - c_t \Delta_t^{(i)}) \quad (91)$$

- Evaluate cost function

$$\hat{C}(\phi_t^{(i)} + c_t \Delta_t^{(i)} | y_{1:t}) = p(y_t | \tilde{v}_+^{(i)}, \phi_t^{(i)} + c_t \Delta_t^{(i)}) \quad (92)$$

and

$$\hat{C}(\phi_t^{(i)} - c_t \Delta_t^{(i)} | y_{1:t}) = p(y_t | \tilde{v}_-^{(i)}, \phi_t^{(i)} - c_t \Delta_t^{(i)}) \quad (93)$$

- **Step 4.** Gradient approximation: for each parameter particle, the corresponding gradient:

$$\hat{\nabla} C(\phi_t^{(i)}) = (\hat{\nabla} C_1(\phi_t^{(i)}), \dots, \hat{\nabla} C_m(\phi_t^{(i)})) \quad (94)$$

where:

$$\hat{\nabla} C_j(\phi_t^{(i)}) = \frac{\hat{C}(\phi_t^{(i)} + c_t \Delta_t^{(i)}) - \hat{C}(\phi_t^{(i)} - c_t \Delta_t^{(i)})}{2c_t \Delta_{t,j}^{(i)}} \quad (95)$$

and $\Delta_{t,j}^{(i)}$ denotes the j -th component of $\Delta_t^{(i)}$.

- **Step 5.** Parameter update: for each $i = 1, \dots, N$, compute:

$$\phi_{t+1}^{(i)} = \phi_t^{(i)} + \gamma_t \hat{\nabla} C(\phi_t^{(i)}) \quad (96)$$

- **Step 6.** Re-sampling: given the samples of particles $\{v_t^{(i)}, \phi_t^{(i)}\}$, compute the normalized importance weights at time t :

$$\hat{\omega}_t^{(i)} = \hat{\omega}_{t-1}^{(i)} p(y_t | v_t^{(i)}, \phi_t^{(i)}) \quad (97)$$

The important weights are normalized:

$$\tilde{\omega}_t^{(i)} = \frac{\hat{\omega}_t^{(i)}}{\sum_{j=1}^N \hat{\omega}_t^{(j)}} \quad , \quad \sum_{i=1}^N \tilde{\omega}_t^{(i)} = 1 \quad (98)$$

Those particles $\tilde{\omega}_t^{(i)}$ with low weights are discarded and only the particles with high weights are selected. Re-assign importance weights $\tilde{\omega}_t^{(i)} = \frac{1}{N}$.

- **Step 7.** Output: the obtained weighted particles $\{v_t^{(i)}, \phi_t^{(i)}, \tilde{\omega}_t^{(i)}\}$, for $i = 1, \dots, N$ compute:

- Filtered density of states:

$$p_N(v_t | v_{0:t-1}, y_{1:t}, \phi) = \sum_{i=1}^N \tilde{\omega}_t^{(i)} \delta(v_t - v_t^{(i)}) \quad (99)$$

- The estimate mean of states:

$$\mathbf{E}(v_t|v_{0:t-1}, y_{1:t}, \phi) = \bar{v}_t = \sum_{i=1}^N \tilde{\omega}_t^{(i)} v_t^{(i)} \quad (100)$$

- The estimate mean of parameters:

$$\mathbf{E}(\phi_t|v_{0:t}, y_{1:t}, \phi_{1:t-1}) = \bar{\phi}_t = \sum_{i=1}^N \tilde{\omega}_t^{(i)} \phi_t^{(i)} \quad (101)$$

- The estimate variance of states:

$$\tilde{\Sigma}_t^v = \sum_{i=1}^N \tilde{\omega}_t^{(i)} (v_t^{(i)} - \bar{v}_t)(v_t^{(i)} - \bar{v}_t)^T \quad (102)$$

- The estimate variance of parameters:

$$\tilde{\Sigma}_t^\phi = \sum_{i=1}^N \tilde{\omega}_t^{(i)} (\phi_t^{(i)} - \bar{\phi}_t)(\phi_t^{(i)} - \bar{\phi}_t)^T \quad (103)$$

- The estimate likelihood:

$$l_t = \sum_{j=1}^{N_{simul}} \tilde{\omega}_j^{(i)} \quad (104)$$

$t = t + 1$ and return to step 2.

It is known that the likelihood in the k -th step is given by the equation (63), so the total likelihood can be defined as $L_{1:N} = \prod_{i=1}^N l_i$, so the log of the likelihood can be maximized by

$$\ln(L_{1:N}) = \sum_{i=1}^N \ln(l_i) \quad (105)$$

Now l_t can be written, such as:

$$l_t = \int p(y_t|v_t, \phi_{1:t}) \frac{p(v_t|y_{1:t-1}, \phi_{1:t})}{q(v_t|v_{t-1}, y_{1:t}, \phi_{1:t})} q(v_t|v_{t-1}, y_{1:t}, \phi_{1:t}) dv_t \quad (106)$$

Since by construction the $v_t^{(i)}$ is generated from $q(v_t|v_{t-1}^{(i)}, \phi_{t-1}^{(i)})$ and taking into account the readjustment of the weights ω_t^i up to a constant $\frac{1}{N_{simul}}$ during the resampling step, the likelihood can be approximated by

$$l_t = \sum_{j=1}^{N_{simul}} \tilde{\omega}_j^i \quad (107)$$

which will give us an approximate interpretation of the likelihood as the total weights.

5. Viterbi algorithm

In addition, it is proposed to implement the Viterbi algorithm with the objective of comparing these results with those obtained by the optimized particle filter. The Viterbi algorithm is a maximum a posteriori (MAP) estimation method that uses the a posteriori density maximization, whose structure is not available analytically and is approximated by a Bayesian filter. In the literature it is noted that the maximum weights are taken as the MAP estimator (Zhou, Chellappa, and Moghaddam (2004), Candy (2007), Godsill, Doucet, and West (2001), Saha, Boers, Driessen, Mandal, and Bagchi (2009), Boers and Driessen (2007), among others).

In this article, we intend to estimate the unknown states of the stochastic CIR and Heston models, assuming the known parameters through the MAP estimation. The method consists of sample particles $\{v_{0:t}^{(i)}\}_{i=1}^N$ according to the methodology used to approximate the given filtered density in (99) and then selecting

$$\hat{v}_{0:t}^{MAP}(t) = \arg \max_{v_{0:t}^{(i)}} p(v_{0:t}^{(i)} | y_{1:t}) \quad (108)$$

If the support of $q(\cdot)$ includes the support of $p(v_{0:t}^{(i)} | y_{1:t})$ then the estimator converges asymptotically when $N \rightarrow \infty$ to $\hat{v}_{0:t}^{MAP}(t)$ (Godsill *et al.* (2001)). The choice of $q(\cdot)$ has a great influence on the performance of the algorithm.

Sampling directly from $p(v_{0:t} | y_{1:t})$ is usually impossible, but an approximation can be used by the particle filter, then the following approximation can be obtained

$$\hat{v}_{0:t}^{MAP}(t) = \arg \max_{v_{0:t} \in \{v_{0:t}^{(i)}\}_{i=1}^N} p(v_{0:t} | y_{1:t}) \quad (109)$$

This is an easy method to implement, but has the degenerative problem that affects particle filters, the quality of the estimators decrease as t is increased. To solve this problem, an alternative is to use a dynamic programming algorithm. The method assumes that the filtered distribution given in the equation (99) has been calculated and stored for each time t . Then the approximation $\hat{v}_{0:t}^{MAP}(t)$ is obtained by

$$\hat{v}_{0:t}^{MAP}(t) = \arg \max_{v_{0:t} \in \bigotimes_{k=1}^t \{v_{0:t}^{(i)}\}_{i=1}^N} p(v_{0:t} | y_{1:t}) \quad (110)$$

To evaluate the function given in the equation (110) to gross force would imply an exhaustive search of all possible trajectories of the model given by the equation (50). However, the maximizing function is given by the equation (61), which is additive. Then the maximizing function is

$$\hat{v}_{0:t}^{MAP}(t) = \arg \max_{v_{0:t}} \left[\ln f(v_0) + \sum_{i=0}^t \ln f(v_{i+1} | v_i) + \sum_{i=1}^t h(y_{i+1} | v_{i+1}) \right] \quad (111)$$

This property allows the use of the dynamic programming technique called Viterbi algorithm (Viterbi (1967)), which allows estimation of $\hat{v}_{0:t}^{MAP}(t)$ as follows

$$\hat{v}_{0:t}^{MAP}(t) = \arg \max_{v_{0:t} \in \bigotimes_{k=1}^t \{v_{0:t}^{(i)}\}_{i=1}^N} \left[\ln f(v_0) + \sum_{i=0}^t \ln f(v_{i+1} | v_i) + \sum_{i=1}^t h(y_{i+1} | v_{i+1}) \right] \quad (112)$$

The Viterbi algorithm is a widely used technique for estimation in discrete state space models of hidden Markov processes as given by equations (50) and (51), they are widely used for recognition voice and decoding codes in information theory (Godsill *et al.* (2001)), among other applications. The algorithm proceeds as follows:

Algorithm: Viterbi

- **Step 1.** for $1 \leq i \leq N$, do:

$$\varphi_1(i) = \ln f(v_0^{(i)}) + \ln \left(f(v_1^{(i)} | v_0^{(i)}) \right) + \ln \left(h(y_1 | v_0^{(i)}) \right) \quad (113)$$

- **Step 2.** for $2 \leq k \leq t$ and $2 \leq i \leq N$, do:

$$\varphi_k(i) = \ln \left\{ h(y_k | v_k^{(i)}) + \max_i \left[\varphi_{k-1}(i) + \ln f(v_0^{(i)}) + \ln \left(f(v_k^{(i)} | v_{k-1}^{(i)}) \right) \right] \right\}$$

and

$$\psi_k(i) = \arg \max \left[\varphi_{k-1}(i) + \ln f(v_0^{(i)}) + \ln \left(f(v_k^{(i)} | v_{k-1}^{(i)}) \right) \right]$$

- **Step 3.** Do:

$$i_t = \arg \max_i [\varphi_t(i)] \quad (114)$$

and

$$\hat{v}_t^{MAP}(t) = v_t^{(i_t)} \quad (115)$$

Finally, the sequence of maximized a posteriori states is obtained

$$\hat{v}_{0:t}^{MAP}(t) = (\hat{v}_0^{MAP}(t), \dots, \hat{v}_t^{MAP}(t))^T \quad (116)$$

6. Results

In order to show the proposed methodology, the joint parameter and state values of the CIR and Heston stochastic models are estimated using 2610 daily data from the time series of the *S&P500* index. The time series analyzed correspond to the period from 29/09/2006 up to 30/09/2016 and it is available at the following address: <http://us.spindices.com/indices/equity/sp-500>. The algorithms was programmed in the Octave GNU programming environment, in an Intel CPU Core i7 3.6GHz with 16GB RAM running 64Bit Windows. To analyze the data series, the following transformation was performed

$$R_t = \frac{s_t - s_{t-1}}{s_{t-1}} \quad (117)$$

where: R_t denotes the net return on an investment of a *S&P500* asset at the end of the month between the time $t - 1$ and the time t ; s_t denotes the *S&P500* index at the end of the month in time t and s_{t-1} denotes the *S&P500* index at the end of the month at time $t - 1$. In the Figure 1 a graph is displayed where the daily *S&P500* is observed for the raw data and for the transformed data in the figure 2, observing a rather volatile behavior of the time series.

To initialize the FPO algorithm in the CIR model, a priori the following values were chosen

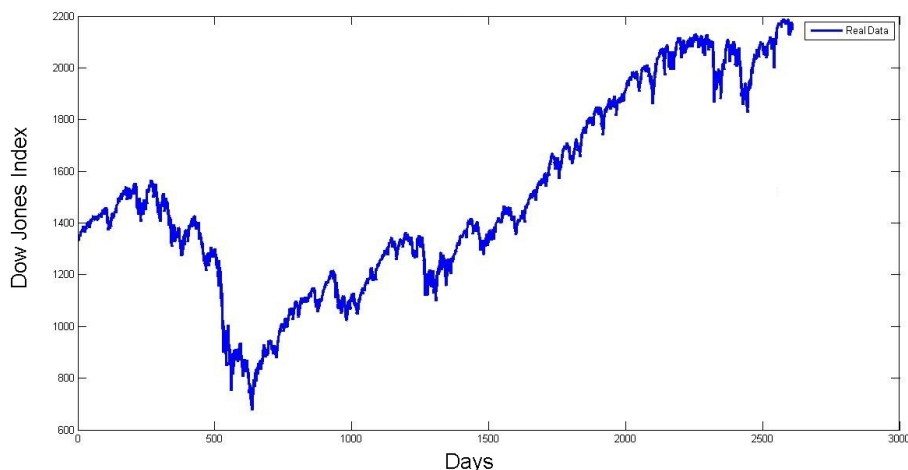


Figure 1: Representation of raw data of the Dow Jones Index.

$c = 0.1$, $a = 1$, $h = 0.00001$, $\gamma_t = 0.0001$, $\Delta_{i,j}^{(i)} = 0.0001$, $h = 0.00001$, $\alpha = 0.602$ and $A = 253.8$. In the Heston model the following values were chosen $c = 0.1$, $a = 1$, $\gamma_t = 0.00001$, $\Delta_{i,j}^{(i)} = 0.0001$, $h = 0.00001$, $A = 1021.2$ and $\alpha = 0.602$.

In the Table 1 the estimates of the parameters of the CIR model are shown. In the Figure 3 and 4, the states estimated by the FPO (red color) and the Viterbi algorithm (red color) and the observed data (blue color) are shown. For the CIR model, much variability in the

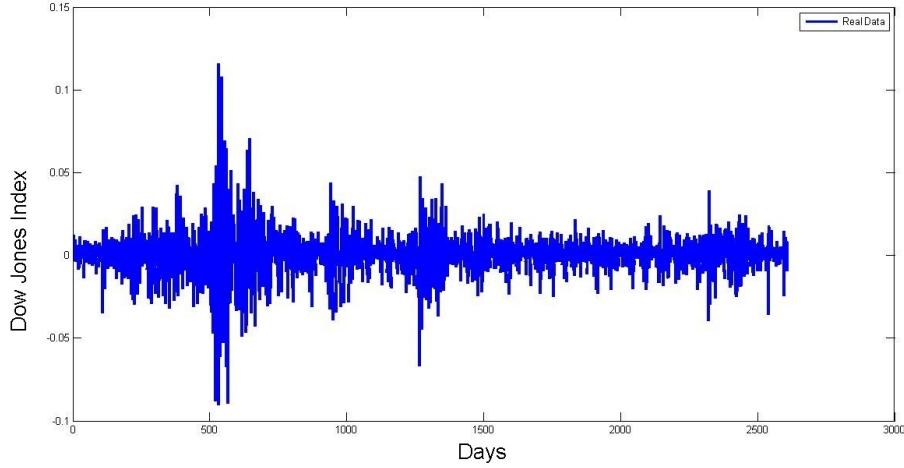


Figure 2: Representation of transformed data of the Dow Jones Index.

approximate states by the Viterbi algorithm can be observed, whereas the states estimated by the FPO fit quite well to the real data. Since the Viterbi algorithm only estimates the states, the parameters required by the CIR model are those estimated by the FPO algorithm.

In the Table 2 the estimations of the parameters of the Heston model are shown. In the

Table 1: Parameters estimated by the CIR model using the FPO.

Parameters	FPO Algorithm
$\hat{\mu}$	-0.00339
$\hat{\beta}$	0.001
$\hat{\sigma}_{\zeta}$	0.1

Figures 5 and 6 the states estimated by the FPO (red color) and by the Viterbi algorithm (red color), together with the observed data (blue color) are shown. As with the CIR model, states estimated by the Heston model show erratic behavior when calculated by the FPO and Viterbi algorithms with respect to the actual data. In the Table 3 the execution times of

Table 2: Estimated parameters of the Heston model using the FPO.

Parameters	FPO Algorithm
$\hat{\mu}$	-0.00339
$\hat{\sigma}$	100
$\hat{\kappa}$	3.3065
$\hat{\theta}$	0.0215
$\hat{\rho}$	-0.9108

the algorithms used are shown, speed and efficiency in the calculations for a high number of data in the cases analyzed in this work are observed.

The Root Mean Squared Error (RMSE) was also calculated as a measure of goodness of fit for the comparison between the estimated states \hat{v}_i and the observed true values v_i . The formula provides a quantitative measurement for the comparison between two models, the smaller the RMSE value, the closer the estimated values to the observed values will be. The

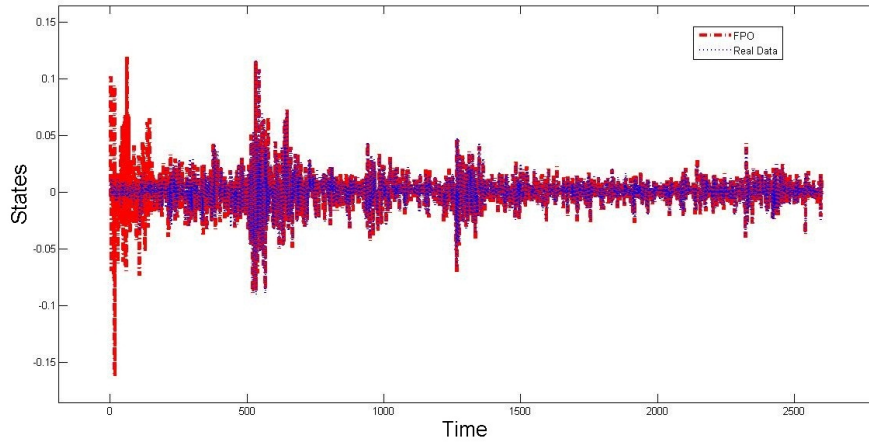


Figure 3: Real and estimated state by the FPO algorithm, CIR model.

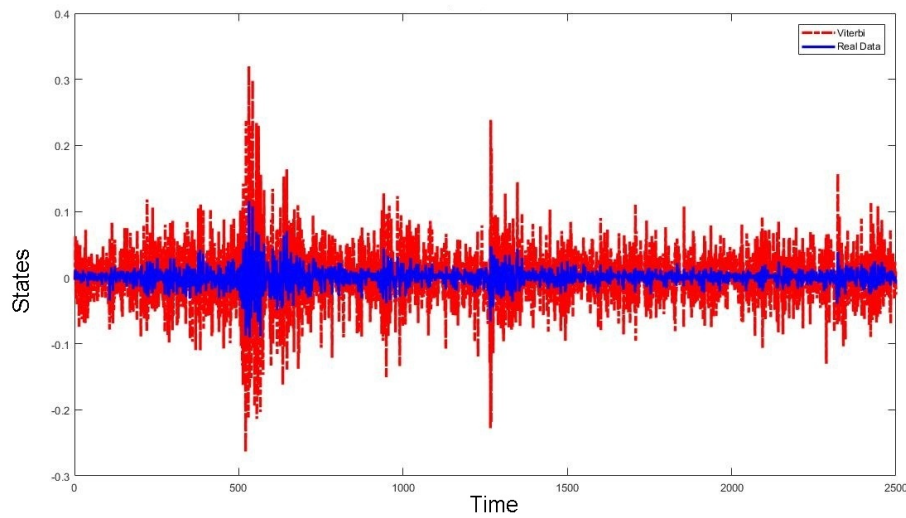


Figure 4: Real and estimated state by the Viterbi algorithm, CIR model.

RMSE is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{v}_i - v_i)^2} \quad (118)$$

A measure of relative dispersion of the data, which takes into account its magnitude, is given by the relative standard deviation (RSD), where it is a measure of the relative dispersion of a data set estimated \hat{v}_i , which is obtained by dividing the standard deviation of the set by its mean arithmetic and is usually expressed in percentage terms.

$$RSD = \frac{s}{\bar{v}_i} 100\% \quad (119)$$

In the Table 4, we show the errors estimated by the two optimized filtering algorithms proposed for the two models considered, observing little variability between the real states and the estimated states. In addition, the relative standard deviation is observed showing low variability in the estimated states.

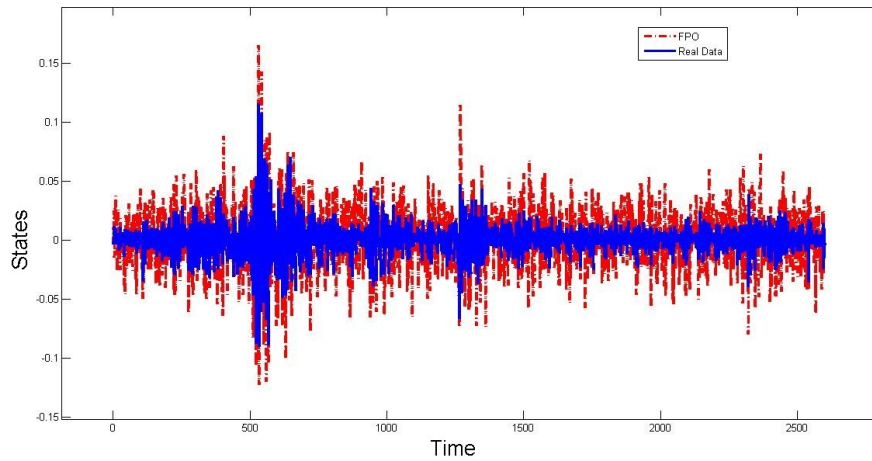


Figure 5: Real and estimated state by the FPO algorithm, Heston model.

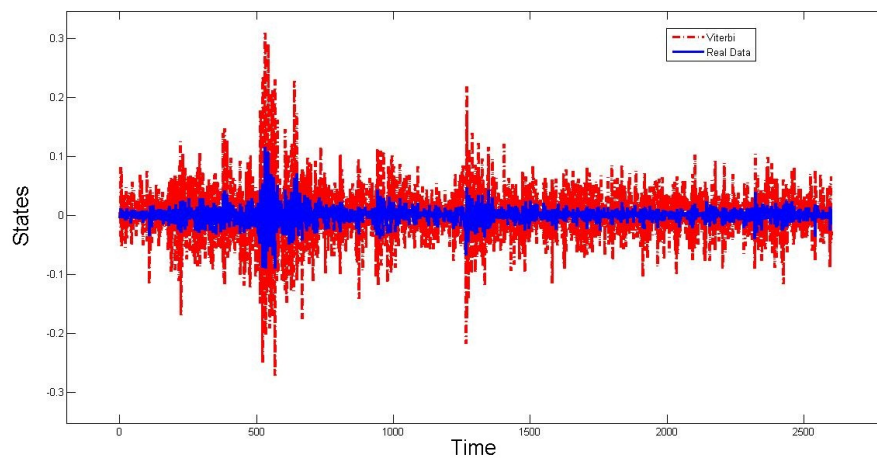


Figure 6: Real and estimated state by the Viterbi algorithm, Heston model.

7. Discussion and conclusions

There are economic models used to determine the prices of certain financial assets. However, it has been shown that the assumptions are not realistic because the model does not explain the true impact of the change in financial markets, such as changes in volatility. Volatility is what helps us to know the status of investors, that is, the feeling of levels of complacency or confidence as well as fear and extreme panic, these two opposite levels are very interesting when it comes to making a decision about a certain value. Currently, financial models are estimated with the maximum likelihood estimation, presenting a drawback related to obtaining the solution in closed form when the density of transition between prices and volatility is not known in a closed manner, also when prices are partially and partially observed. Volatility is unknown, the process in the maximum likelihood estimation results in the management of analytically intractable integrals. In this sense, this research presents a Bayesian estimation methodology that estimates the states and market parameters of the *S&P500* index considering volatile models and incomplete observations on line with low execution time and immediate response of the algorithms, observing a good estimate of the real data, describing the market of the *S&P500* index when the market is presented with confidence or complacency because the volatility is very low and the market rises slowly without scarcely scaring,

Table 3: Elapsed time of the algorithms for the CIR and Heston models.

Algorithm	Model 1	Elapsed time (second)	Model 2	Elapsed time (second)
FPO	CIR	0.589036	Heston	1.165637
Viterbi	CIR	0.059803	Heston	0.148185

Table 4: RSME and RSD estimated for the CIR and Heston models.

Algorithm	Model 1	RSME	Model 2	RSME
FPO	CIR	0.0462	Heston	0.0427
Viterbi	CIR	0.0926	Heston	0.0415
Algorithm	Model 1	RSD	Model 2	RSD
FPO	CIR	30.66	Heston	23.20
Viterbi	CIR	21.16	Heston	36.21

so investors feel safe and this feeling is the prelude to the falls, also, describing the market in the points presented with fear or panic because the volatility is very high and the market falls sharply scare the great part of the investors, therefore the investors feel insecure and this feeling is the prelude to the rises.

In this article, two algorithms for estimation of states and parameters in stochastic volatility models were proposed. The methodology used is a useful and elegant tool that serves to approximate in a recursive way the parameters and the filtered distribution of the states not observed from a process observed with errors using weighted random samples. The FPO algorithm is based on a particle filter structure and a gradient stochastic approximation algorithm is used to optimize the cost function. The estimation of states and parameters is performed simultaneously. On the other hand, the Viterbi algorithm uses the properties of the hidden Markov chains in conjunction with dynamic programming techniques to obtain an optimal sequence of states that maximizes the joint posterior distribution of all states. To estimate the parameters and to reconstruct the unknown states in the two models, both algorithms were implemented, observing little variability with respect to the real states. To measure the relative success of the estimation algorithms, the RMSE and RSD was calculated showing that the estimates produced by the two models have small errors. Finally, we compared the execution times of the algorithms and showed that the Viterbi algorithm has shorter execution time than the FPO.

Some of the works under investigation currently related to this methodology, we can point out the study of the stochastic behavior of the prices of a commodity taking into account the reversal of the average in terms of futures contract prices and real prices. The implementation of algorithms type Metropolis Hastings, Monte Carlo Sequential, and Particle Independent Metropolis Hastings to make simultaneous estimation of the parameters and states solutions in models of stochastic differential equations in situations where the diffusion process is partially observed. Finally it is developing a mixed effects model defined through a stochastic differential equation, these models arise in many fields of research such as: clinical trials, growth studies in agriculture, dispersion processes of epidemiological diseases, series of financial time, population dynamics (including predatory dam systems), and intracellular processes, among many other applications.

8. Acknowledgments

We are infinitely grateful to the Editor Matthias Templ, for their valuable suggestions and contributions to improving this manuscript.

References

- Ait-Sahalia Y (2002). “Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach.” *Econometrica*, **70**, 223–262.
- Ait-Sahalia Y (2008). “Closed-Form Likelihood Expansions for Multivariate Diffusions.” *Annals of Statistics*, **36**(2), 906–937.
- Ait-Sahalia Y, Kimmel R (2007). “Maximum Likelihood Estimation of Stochastic Volatility Models.” *Journal of Financial Economics*, **83**, 507–551.
- Ait-Sahalia Y, Kimmel R (2010). “Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions.” *Journal of Financial Economics*, **98**(1), 113–144.
- Andrieu C, Doucet A, Holenstein R (2010). “Particle Markov Chain Monte Carlo Methods.” *Journal of the Royal Statistical Society B*, **72**(3), 269–342.
- Bates D (1996). “Jump and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options.” *Review of Financial Studies*, **9**, 69–107.
- Black F, Scholes M (1973). “The Pricing of Options and Corporate Liabilities.” *Journal of Political Economy*, **81**(3), 637–654.
- Boers Y, Driessen JN (2007). “The Mixed Labelling Problem in Multitarget Particle Filter.” *Proceedings of the 10th International Conference on Information Fusion*.
- Candy JV (2007). “Bootstrap Particle Filtering.” *IEEE Signal Processing Magazine*, **24**(34), 73–85.
- Carvalho C, Johannes M, Lopez H, Nicholas P (2010). “Particle Learning and Smoothing.” *R News*, **25**(1), 88–106.
- Chib S, Pitt MK, Shephard N (2006). “Likelihood Based Inference for Diffusion Driven State Space Models.” *Working paper. Oxford: Nuffield College*.
- Christoffersen P, Jacobs P, Mimouni K (2010). “Volatility Dynamics for the S and P500: Evidence From Realized Volatility, Daily Returns and Option Prices.” *Review of Financial Studies*, **23**, 3141–3189.
- Cox JC, Ingersoll JA, Ross SA (1985). “A Theory of the Term Structure of Interest Rates.” *Econometrica*, **53**, 385–407.
- Crisan D, Doucet A (2002). “A Survey of Convergence Results on Particle Filtering Methods for Practitioners.” *IEEE Transactions on Signal Processing*, **50**(3), 736–746.
- Doucet A, De Freitas J, Gordon N (2001). *Sequential Monte Carlo in Practice*. Springer Verlag, New York.
- Eraker B (2001). “MCMC Analysis of Diffusion Models with Application to Finance.” *Journal of Business and Economic Statistics*, **19**, 177–191.
- Eraker B (2004). “Do Stock Prices and Volatility Jump. Evidence From Spot and Option Prices.” *Journal of Finance*, **59**, 1367–1403.
- Eraker B, Johannes M, N P (2003). “The Impact of Jumps in Equity Index Volatility and Returns.” *Journal of Finance*, **58**, 1269–1300.
- Godsill S, Doucet A, West M (2001). “Maximum a Posteriori Sequence Estimation Using Monte Carlo Particle Filters.” *Annals of the Institute of Statistical Mathematics*, **53**, 82–9.

- Gordon N, Salmond D, Smith AF (1993). “Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation.” *Radar and Signal Processing, IEE Proceedings F*, **140**(2), 107–113.
- Haykin S (2001). *Kalman Filtering and Neural Networks*. John Wiley & Sons, New York.
- Heston S (1993). “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options.” *The Review of Financial Studies*, **6**, 327–343.
- Hurn AS, Lindsay KA, McClelland AJ (2005). “Estimating the Parameters of Stochastic Volatility Models using Option Price Data.” *School of Economics and Finance, Queensland University of Technology. Working Paper*, **87**.
- Jacquier E, Polson N, P R (2004). “Bayesian Analysis of Stochastic Volatility Models with Leverage Effect and Fat Tails.” *Journal of Econometrics*, **1**, 185–212.
- Javaheri A (2015). “Inside Volatility Filtering.” *Published by John Wiley and Sons, Inc., Hoboken, New Jersey*.
- Javaheri A, Lautier D, Galli A (2003). “Filtering in Finance.” *Wilmott*, **5**.
- Johannes MS, Polson NG, Stroud JR (2009). “Optimal Filtering of Jump Diffusions: Extracting Latent States From Asset Prices.” *Review of Financial Studies*, **22**, 2759–2799.
- Kleppe TS, Jun JY, Skaug H (2009). “Stimulated Maximum Likelihood Estimation of Continuous Time Stochastic Volatility Models.” *Research Collection School of Economics*.
- Kleppe TS, Jun JY, Skaug H (2014). “Maximum Likelihood Estimation of Partially Observed Diffusion Models.” *Journal of Econometrics*.
- Liu J, West M (2001). “Combining Parameter and State Estimation in Simulation-Based Filtering.” *In Sequential Monte Carlo Methods in Practice*, **10**, 197–224.
- Poyiadjis G, Doucet A, Singh S (2011). “Particle Approximations of the Score and Observed Information Matrix in State Space Models with Application to Parameter Estimation.” *Biometrika*, **98**(1), 65–80.
- Richard JF, Zhang W (2007). “Efficient High-Dimensional Importance Sampling.” *Journal of Econometrics*, **2**, 1385–1411.
- Saha S, Boers Y, Driessen H, Mandal P, Bagchi A (2009). “Particle Based MAP State Estimation: A Comparison.” *12th International Conference on Information Fusion Seattle WA USA July 6-9*.
- Spall JC (1987). “A Stochastic Approximation Technique for Generating Maximum Likelihood Parameter Estimates.” *Proceedings of the American Control conference*, pp. 1161–1167.
- Spall JC (1988). “Implementation of the Simultaneous Perturbation Algorithm for Stochastic Optimization.” *IEEE Transactions on Aerospace and Electronic Systems*, **34**(3), 817–823.
- Spall JC (2003). “Estimation via Markov Chain Monte Carlo.” *IEEE Control System Magazine*, **2**, 34–45.
- Viterbi AJ (1967). “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm.” *IEEE Transactions on Information Theory IT-13*, pp. 260–269.
- Yang X, Xing K (2011). “Joint State and Parameter Estimation in Particle Filtering and Stochastic Optimization.” *Stochastic Optimization*.

Zhou SK, Chellappa R, Moghaddam B (2004). “Visual Tracking and Recognition Using Appearance Adaptive Models in Particle Filters.” *IEEE Transaction on Image Processing*, **11**, 1491–1506.

Affiliation:

Saba Infante
School of Mathematical Sciences and Information Technology
Yachay Tech University
Ecuador
E-mail: sinfante@yachaytech.edu.ec \ sinfante64@gmail.com

Cesar Luna
Departamento de Matemática
Facultad de Ciencias y Tecnología
Venezuela
E-mail: cluna@uc.edu.ve

Luis Sánchez
Departamento de Matemática y Estadística
Universidad Técnica de Manabí
Ecuador
E-mail: lasanchez@utm.edu.ec \ uccursos@gmail.com

Aracelis Hernández
Departamento de Matemática
Facultad de Ciencias y Tecnología
Venezuela
E-mail: arhernan@uc.edu.ve \ arhernan65@gmail.com