Expert Systems With Applications 54 (2016) 228–240

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Early online detection of high volatility clusters using Particle Filters

Karel Mundnich[a], Marcos E. Orchard[a,*]

Electrical Engineering Department, Universidad de Chile, Av. Tupper 2007, Santiago 8370451, Chile

ARTICLE INFO

Keywords:
Bayesian inference
Risk-sensitive particle filters
Stochastic volatility estimation
Event detection

ABSTRACT

This work presents a novel online early detector of high-volatility clusters based on uGARCH models (a variation of the GARCH model), risk-sensitive particle-filtering-based estimators, and hypothesis testing procedures. The proposed detector utilizes Risk-Sensitive Particle Filters (RSPF) to generate an estimate of the volatility probability density function (PDF) that offers better resolution in the areas of the state-space that are associated with the incipient appearance of high-volatility clusters. This is achieved using the Generalized Pareto Distribution for the generation of particles. Risk-sensitive estimates are used by a detector that evaluates changes between prior and posterior probability densities via asymmetric hypothesis tests, allowing early detection of sudden volatility increments (typically associated with early stages of high-volatility clusters). Performance of the proposed approach is compared to other implementations based on the classic Particle Filter, in terms of its capability to track regions of the state-space associated to a greater financial risk. The proposed volatility cluster detection scheme is tested and validated using both simulated and actual IBM's daily stock market data.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Volatility of returns is a well-studied variable in Finance, mainly because its relevance in pricing and risk management. Since the work of Mandelbrot in the 1960's, it has been widely accepted that volatility presents itself in temporal clusters, where large price variations are followed by large variations (Cont, 2005; Mandelbrot, 1963). Multiple researchers have tried to model the complex behavior of volatility, being the GARCH model (Bollerslev, 1986) the first to capture these temporal cluster properties. The wide recognition for the GARCH models has given rise to a whole family of structures, in which stochastic variations have been lately introduced.

On the one hand, from an engineering perspective, early online detection of high-volatility clusters in a stochastic environment poses an interesting problem, since detection algorithms must be designed to monitor a latent (non-observable) state; simultaneously tracking disturbances introduced by other non-measurable variables that are always present in complex systems (such as stock markets). In fact, from the standpoint of state-space modeling for financial time series, volatility is a non-observable state, while continuously compounded returns can be associated with daily measurements. Given that inference on financial volatility is necessary to detect high risk events, the challenge is then to propose detection frameworks based on accurate and precise estimates of this non-observable state.

On the other hand, in Finance, the words "early online detection" have reached unsuspected relevance in a world where is now possible to use information from high-volatility cluster detectors (which could have been originally focused on very specific and critical markets) for the implementation of online predictive strategies at a global-scale. Consider, for example, the implementation of intelligent expert systems that could recommend optimal corrective actions for Latin American markets based on online anomaly detectors analyzing Asian markets during the early morning. Indeed, the development of tools for online early detection of high-volatility clusters (such as the one proposed in this article) generates appropriate conditions for the implementation of novel online schemes for optimal decision-making in Finance; a task where the whole community working on expert and intelligent systems may contribute in the near future.

These fundamental questions have motivated in recent years substantial research with focus in the detection of structural breaks (or model parameter changes) in financial variables, with the purpose of understanding market shocks or anomalies (Chan & Koop, 2014; Chen, Gerlach, & Liu, 2011; He & Maheu, 2010; Rapach & Strauss, 2008; Ross, 2013). Given the complexities involved in modeling volatility, several approaches have been proposed, including new models such as the structural break GARCH (SB-GARCH). For these models that include stochastic volatility and

* Corresponding author. Tel.: +56 229784215; fax: +56 226720162.
E-mail addresses: kmundnic@ing.uchile.cl (K. Mundnich), morchard@ing.uchile.cl (M.E. Orchard).

http://dx.doi.org/10.1016/j.eswa.2016.01.052
0957-4174/© 2016 Elsevier Ltd. All rights reserved.

breaks, the most common approach for the estimation of volatility has been sequential Monte Carlo methods (a.k.a. Particle Filters) (Arulampalam, Maskell, Gordon, & Clapp, 2002; Doucet, Godsill, & Andrieu, 2000), because of its good performance, flexibility, and the possibility to estimate model parameters online (Liu and West, 2001, chap. 10). Further efforts have been spent in the study of jumps (or discontinuities) of returns and volatility (Andersen, Tim, & Diebold, 2007; Eraker, Johannes, & Polson, 2003; Laurent, Lecourt, & Palm, 2014; Lee & Mykland, 2008), although these approaches impose restrictions for the quality of data that may be difficult to address. In fact, most of the available tests for detection of jumps include high-frequency, intra-day information of the studied variables or high liquidity of the assets (Laurent et al., 2014), or offline tests.

In this regard, we propose a novel online early detector of high-volatility clusters based on unobserved GARCH models (Tobar, 2010) (uGARCH, a variation of the GARCH model), Risk-Sensitive Particle-Filters (RSPF) estimators (Thrun, Langford, & Verma, 2002), and hypothesis testing procedures. The proposed detector utilizes risk-sensitive particle filters to generate an estimate of the volatility probability density function (PDF) that offers better resolution in the areas of the state-space that are associated with the incipient appearance of high-volatility clusters. Risk-sensitive estimates are used by a detector that evaluates changes between prior and posterior probability densities via asymmetric hypothesis tests, allowing early detection of sudden volatility increments (typically associated with early stages of high-volatility clusters). This algorithm is tested in simulated data (where volatility is known), as well as IBM's stock market data, where volatility has to be estimated (since ground truth cannot be measured). To the best of our knowledge, this is the first attempt in financial econometrics to perform online detection of events by contrasting the information that is present in priors and posterior probability densities estimates in Bayesian estimation frameworks.

The main contributions of this article are:

- Implementation and assessment of a novel method for the generation of volatility estimators, based on RSPF, that provides better resolution in the areas of the state-space associated with the appearance of high-volatility clusters.
- Implementation and assessment of early detection schemes for high-volatility clusters based on the comparison between prior and posterior particle-filtering-based estimates.
- A throughout performance comparison between RSPF and classic sequential Monte Carlo methods in terms of their effectiveness when used in early detection of high-volatility clusters.

The structure of this article is as follows. Section 2 presents a literature review on the use of Bayesian frameworks for Financial volatility estimation. Section 3 presents the proposed method for early detection of high-volatility clusters. Section 4 presents performance measures to be used in the assessment of obtained results, provides a sensitivity analysis for the proposed method using simulated data (where the ground truth value of the unmeasured state is known), and finalizes with a throughout performance analysis for the proposed method based on actual IBM stock data. Section 5 presents a few interesting general remarks, while Section 6 shows the main conclusions related to this research.

2. A Bayesian framework for volatility estimation

Monte Carlo (MC) and *Markov Chain Monte Carlo* (MCMC) methods have been widely used to approximate integrals and probability density functions (Tobar, 2010). Nevertheless, their use in Bayesian inference is not direct, since this problem involves a sequence of time-variant probability density functions; while MCMC assumes that the objective density is time-invariant. This prompted the development of a sequential version of Monte Carlo integration, one that is able to use measurements to improve recursive estimation.

The first section of this section introduces the uGARCH model, a stochastic volatility model based on the well-known GARCH(1,1) model (Bollerslev, 1986). Then, the tracking problem is presented in Section 2.2, providing insight about the problems encountered in a Bayesian filtering framework. Also, Monte Carlo integration and the *importance sampling* method are presented. This opens the possibility to explore the Particle Filter and the Risk Sensitive Particle Filter, which may be employed in a stochastic volatility estimation framework. Finally, Section 2.3 explains the need for online parameter estimation.

2.1. The uGARCH model

The uGARCH model can be seen as a state-space structure that allows the implementation of a Bayesian framework for the purpose of volatility estimation.

The uGARCH model (Tobar, 2010) assumes that the dynamics of volatility are not driven by the observed process $u_k = r_k - \mu_{k|k-1}$. Instead, they are driven by a non-observable process u'_k which has the same distribution as u_k. The uGARCH model is defined as:

$$\sigma_k^2 = \omega + \alpha\sigma_{k-1}^2\eta_k^2 + \beta\sigma_{k-1}^2, \tag{1}$$

$$r_k = \mu + \sigma_k\epsilon_k, \tag{2}$$

where r_k is a process of returns, σ_k is the stochastic volatility, $\mu \in \mathbb{R}^+$, $\omega \in \mathbb{R}^+$, and α, $\beta > 0$ are parameters, with $\alpha + \beta < 1$. Furthermore, $\epsilon_k \sim \mathcal{N}(0,1)$ and $\eta_k \sim \mathcal{N}(0, \sigma_\eta^2)$ are *i.i.d.*[1] processes for every time step k.

It is necessary to note from Eqs. (1) and (2) that the subscripts are not written conditionally: at time step k, σ_k^2 is not known without uncertainty, given Σ_{k-1}.

To completely define the model, it is necessary to present the state transition distribution $p(\sigma_k^2|\sigma_{k-1}^2)$ and the likelihood $p(r_k|\sigma_k^2)$:

$$p(\sigma_k^2|\sigma_{k-1}^2) = \frac{1}{\sqrt{2\pi\sigma_\eta^2\alpha\sigma_{k-1}^2\left(\sigma_k^2 - \omega + \beta\sigma_{k-1}^2\right)}} \cdot \exp\left[\frac{\sigma_k^2 - \omega + \beta\sigma_{k-1}^2}{2\sigma_\eta^2\alpha\sigma_k^2}\right], \quad \sigma_k^2 \geq \omega + \beta\sigma_{k-1}^2. \tag{3}$$

$$p(r_k|\sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left(-\frac{(r_k-\mu)^2}{2\sigma_k^2}\right). \tag{4}$$

For the complete derivation of Eq. (4), please refer to Mundnich (2013). The calculation and presentation of Eq. (4) makes the use of a generic Particle Filtering approach for volatility estimation in this model possible.

2.2. The Particle Filter

State-space models consider a transition equation that describes the prior distribution of a hidden Markov process $\{x_k; k \in \mathbb{N}\}$, called the state process, and an observation equation describing the likelihood of the observation $\{z_k; k \in \mathbb{N}\}$ (Doucet et al., 2000):

$$x_k = f(x_{k-1}, v_{k-1}), \tag{5}$$

$$z_k = h(x_k, w_k), \tag{6}$$

where $f(\cdot,\cdot)$ is a state-transition function with corresponding $\{v_{k-1}, k \in \mathbb{N}\}$ *i.i.d.* innovation process, and $h(\cdot,\cdot)$ is the observation

[1] Independent and identically distributed.

function with $\{w_k, k \in \mathbb{N}\}$ its corresponding *i.i.d.* noise process. In particular, the objective of tracking is to recursively estimate x_k from all available measurements $z_{1:k} = \{z_i;\ i = 1, \ldots, k\}$ up to time k.

Within a Bayesian estimation framework, all relevant information about $x_{0:k}$ given the observations $z_{0:k}$ can be obtained from the posterior distribution $p(x_{0:k}|z_{0:k})$. In many applications, nevertheless, it is sufficient to calculate the marginal conditional distribution $p(x_k|z_{0:k})$. In particular, the intention of the Bayesian approach is to recursively construct $p(x_k|z_{1:k})$, using Arulampalam et al. (2002):

$$p(x_k|z_{1:k}) = \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{p(z_k|z_{1:k-1})}, \tag{7}$$

$$= \frac{p(z_k|x_k)\int p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1}}{\int p(z_k|x_k)p(x_k|z_{1:k-1})dx_k}. \tag{8}$$

Eq. (8) forms the basis for the Bayesian optimal solution in the mean square error sense. In most cases, this expression is only conceptual, and cannot be determined analytically. In a restricted set of cases, the optimal solution may be found (Kalman, 1960). This is possible only if the noises v_k and w_k are additive and Gaussian and the functions $f(\cdot,\cdot)$ are $h(\cdot,\cdot)$ are linear.

Particle Filters are a class of algorithms developed to solve Eq. (8) through sequential Monte Carlo simulations when the integrals are intractable due to possible nonlinearities in the model involved or noise processes that do not possess standard distributions. Solving these integrals is achieved through the *Importance Sampling* principle. The key idea is to represent the required posterior density function by a set of random samples which serve as support points with associated weights and to compute estimates based on these samples and weights, this is:

$$p(x_k|z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^{(i)}\delta(x_k - x_k^{(i)}), \tag{9}$$

The former approximation may be obtained using an *importance density* $q(x_{0:k}|z_{1:k})$ to generate random samples $x_k^{(i)}$, where the weights $w_k^{(i)}$ are calculated using:

$$w_k^{(i)} \propto \frac{p(x_{0:k}^{(i)}|z_{1:k})}{q(x_{0:k}^{(i)}|z_{1:k})}, \tag{10}$$

$$\propto w_{k-1}^{(i)} \frac{p(z_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, z_k)}. \tag{11}$$

This algorithm is generally called *Sampling Importance Sampling* (SIS), and denotes the simplest form to solve Eq. (8).

The position of the particles and consequent performance of the filter is greatly determined by the importance density $q(x_k|x_{k-1}^{(i)}, z_k)$ from which particles are drawn. The structure of the Particle Filter algorithm and importance densities usually employed do not regard the problem of high risk and low-likelihood event tracking. In the case where unlikely events may conduce to great loss or high costs, it is natural extend the Particle Filter algorithm to track these low probability states.

The Risk Sensitive Particle Filter is proposed as an extension of the 'Classic' Particle Filter, where the particles are generated from an importance density that is the product of the combination of the posterior density function and a risk functional.

Risk Sensitive Particle Filters generate samples that are distributed according to Thrun et al. (2002):

$$q(x_k|x_{k-1}^{(i)}, z_k) = \gamma_k r(x_k)p(x_k|z_{1:k}), \tag{12}$$

where

$$\gamma_k = \frac{1}{\int r(x_k)p(x_k|z_{1:k})} \tag{13}$$

is a normalizing constant that ensures that the importance density is indeed a probability density function. Hence, the position of samples are generated according to the likelihood of a certain state event $x_k^{(i)}$ and its risk $r(x_k^{(i)})$.

Considering the former approach, the Classic Particle Filter needs a simple modification. First, the initial set of particles $x_0^{(i)}$ is generated from $\gamma_0 r(x_0)p(x_0)$, and Eq. (11) is updated to

$$w_l^{(i)} = \frac{r(x_k^{(i)})p(z_k|x_k^{(i)})}{r(x_{k-1}^{(i)})}. \tag{14}$$

In this work, the authors propose an importance density $q(x_k|x_{k-1}^{(i)}, z_k)$, for which they assume that a risk functional $r(x_k^{(i)})$ exists.

2.3. Online parameter estimation with Particle Filters

In the context of state estimation, it is sometimes necessary to handle an online estimation scheme for a model parameter vector. Although parameters α and β have been presented as fixed in the uGARCH model, this is not necessarily adequate, given possible structural breaks in the data. The stock market suffers from variations and regime shifts, and these variations may be considered as parameter changes through time. This is true not only for time series derived from the stock market, but for a diverse range of applications where state tracking is intended.

To understand the problems of parameter estimation outside a Bayesian context, let θ be a vector parameter. The maximum likelihood estimate of the vector parameter θ is obtained by maximizing the log-likelihood function (Kitagawa & Sato, 2001, chap. 9):

$$l(\theta) = \log[L(\theta)] = \sum_{k=1}^{\kappa} \log[p(z_k|z_{1:k-1}, \theta)], \tag{15}$$

where

$$p(z_k|z_{1:k-1}, \theta) = \int p(z_k|x_k, \theta)p(x_k|z_{1:k-1}, \theta)dx_k \tag{16}$$

needs to be approximated through Monte Carlo.

The maximization of Eq. (15) for the estimation of θ is not always direct, and approximations over Eq. (16) make this method impractical, due to the high computational costs involved if parameter estimation is intended for every time step. Thus, a different perspective is necessary to approach the online parameter estimation problem. This idea is attacked through the artificial evolution of parameters.

The first ideas about introducing random disturbances to particles were proposed by Gordon, Salmond, and Smith (1993), and are currently widely used in financial econometrics. In their work, the authors propose to introduce random disturbances to the positions of particles (called *roughening penalties*) in order to combat degeneracy. This idea has been extended in order to estimate online a vector of fixed model parameters, which is referred to as *artificial evolution* (Liu & West, 2001, chap. 10). Artificial evolution of parameters is a simple and powerful idea, nevertheless, it requires careful handling because of the inherent model information loss given by the consideration of time-varying parameters that are fixed.

Instead of estimating the vector parameter θ through maximum likelihood, the Bayesian framework may be introduced to estimate θ online. This is achieved by augmenting the state vector x_k with unknown parameters θ as:

$$x_k' = \begin{bmatrix} \theta_k \\ x_k \end{bmatrix}, \tag{17}$$

where $\theta_k = \theta$ implies the consideration of an extended model where parameters are time-varying. Then, an independent,

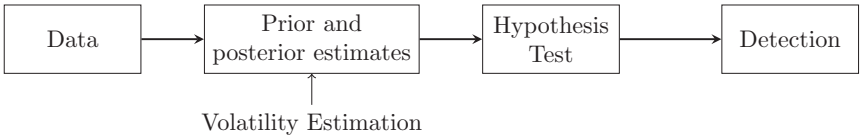


Fig. 1. Detection flow chart. Data is served as an input for the PF-based estimation, which produces prior and posterior estimates that are given as the input for the hypothesis test, which results in the early detection of high volatility clusters.

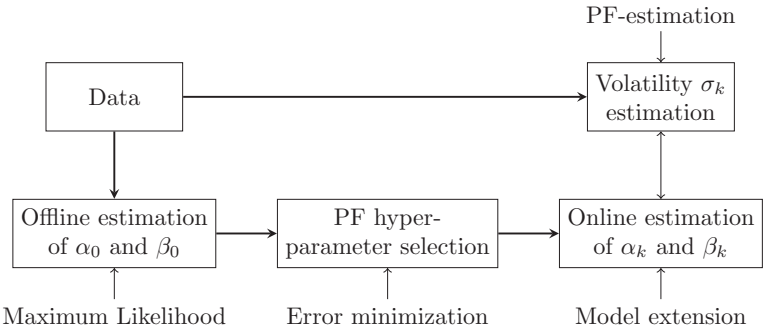


Fig. 2. Estimation flow chart. In an initialization step, PF hyper-parameter selection is performed through error minimization. Then, offline parameter estimation is performed using maximum likelihood. Finally, the PF estimates in parallel the model parameters α_k and β_k, as well as the stochastic volatility σ_k.

zero-mean normal increment is added to the parameter at each time step (Liu & West, 2001, chap. 10):

$$\theta_k = \theta_k + \zeta_k, \tag{18}$$

$$\zeta_k \sim \mathcal{N}(0, W_k), \tag{19}$$

where W_k is a variance matrix and θ_k and ζ_k are conditionally independent given Σ_k. The key motivation is that the artificial evolution of parameters gives new values for each iteration, and thus, weight assignment in Particle Filters considers the likelihood of the state and parameter values.

3. Detection of high volatility clusters using PF-based estimation methods

This chapter describes the implementation details followed to create an online high volatility cluster detection scheme. In the context of Bayesian estimation in state-space models, volatility arises as a non-observable state. Therefore, simulated stock market data is used to correctly implement, analyze and assess the proposed Bayesian filtering framework.

Our approach proposes the use of the GARCH(1,1) volatility model to create a volatility time series – after defining the value of some model parameters – and consequently generate a returns series for the given volatility at every time step. This is useful to measure the effectiveness of estimation frameworks and detection schemes.

The detection scheme presented in this section is founded upon a Bayesian estimation framework, which is based in Particle Filter-based estimation (Fig. 1). Therefore, it is mandatory to comprehend the details of the estimation process in order to understand the construction of the detection scheme. These details include the online hyper-parameter estimation and volatility estimation in the uGARCH model, and the construction of prior and posterior estimates (Arulampalam et al., 2002). To introduce these concepts, Section 3.1 first describes the data used in the development of this work. Then details about offline hyper-parameter estimation are given. This offline estimation is used as the input for online hyper-parameter estimation, which is performed in parallel to the volatility estimation (Fig. 2).

3.1. Data

The assessment of Bayesian estimation frameworks and detection schemes requires data sets where observations and the state are known for every instant in a given period. This allows the evaluation of filtering schemes and consequent comparison of the implemented techniques. Given that the volatility of a returns series is not observable, it is mandatory to generate data sets where the algorithms can be tested and fine-tuned. This section provides details about artificially generated data used during this work, and presents the acquisition and post-processing necessary to apply the proposed algorithms in stock market data.

3.1.1. Simulated data

The simulated data has been generated using a GARCH(1,1) model, where model parameters α and β are chosen in such a way that $\alpha + \beta$ is a value close to 1. In observed financial time series, it is not possible to ensure that the values of the model parameters α and β are fixed for a given time window. For this reason, volatility time series are created using time-dependent parameters over the studied time span. In particular, time series of 500 steps have been generated, with a parameter change in the step 250 (see Table 1). This change resembles a regime shift (or structural break) in the market (Tobar, 2010).

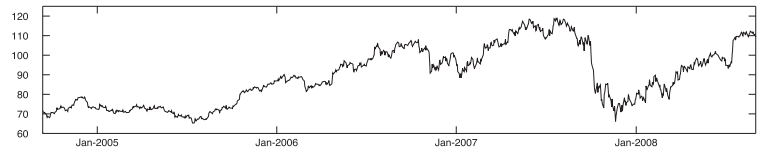
The model used for data generation is:

$$\sigma^2_{k|k-1} = \omega + \alpha u^2_{k-1} + \beta \sigma^2_{k-1|k-2}, \tag{20}$$

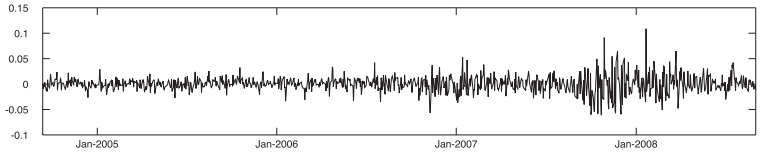
$$r_k = \mu + u_k, \tag{21}$$

Table 1
GARCH(1,1) model parameters for each data set. The arrow (→) indicates a change in the parameter value at time step 250. Note that parameters μ and ω are constant for each set.

Parameter	GARCH1	GARCH3
μ	9×10^{-4}	9×10^{-4}
ω	10×10^{-6}	10×10^{-6}
α	0.20 → 0.10	0.20 → 0.12
β	0.60 → 0.85	0.60 → 0.80



(a) Adjusted closing price.



(b) Returns.

Fig. 3. IBM's adjusted closing stock prices and corresponding daily returns from September 12th, 2005 to September 1st, 2009.

where $\mu_{k|k-1} = \mu$ and ω are considered constant parameters in the studied time span, and $\sigma_0^2 = 0.5 \times 10^{-4}$.

The implemented method for data generation creates data sets in which there are large volatility variations and small volatility variations. These are necessary for the correct assessment of the proposed detection algorithm, as well as providing the necessary environment to assess the robustness of the method against model variations.

3.1.2. Stock market data

A section of IBM daily stock prices is used to apply the developed algorithm for early high volatility cluster detection. The data is extracted from (Yahoo!, 2013), with information between January 1st, 1962 and May 17th, 2013 for a total of 12,933 data points. The data considered for filtering is the *adjusted closing price*, which is commonly used for analysis of historical data. Data in which the proposed algorithm is applied is shown in Fig. 3, which corresponds to 1000 data points between September 12th, 2005 and September 1st, 2009. This data set includes the dramatic market fall occurred in October 2008.

3.2. Implementation issues related to Bayesian filtering

3.2.1. Estimation of model parameters α and β

Model parameters α and β have a high impact on volatility series. These parameters have the power to drive the variation speed of a volatility series and to control the average of the series over time. Hence, it is of great importance to have good estimates of both model parameters to adequately estimate volatility.

In financial time series, it is impossible to know if model parameters α and β are fixed for a given time window in a data set. Therefore, it is necessary to estimate these model parameters online. This work includes two stages of model parameter estimation: estimation through maximum likelihood in a training set and online estimation in test data points.

Estimation of model parameter initial conditions through maximum likelihood. Model parameter estimation has been performed in both simulated data sets and stock market data through maximum likelihood. This is plausible due to the similar structure in both the GARCH(1,1) model and the uGARCH model. In particular, this task has been accomplished using the `garchfit` function available in the Financial Toolbox of MATLAB®.

In the simulated data sets, model parameter estimation is performed using the first 150 steps for each returns time series. For IBM's stock prices, model parameter estimation is performed over the first 200 time steps of the returns time series. These estimations are used as the initial conditions for online model parameter estimation.

Online model parameter estimation. Section 2.3 describes the reasons for using online model parameter estimation in a Particle Filtering scheme. In this work, parameters α and β of the uGARCH model are allocated into an extended state vector,

$$(x_k)' = \begin{bmatrix} \alpha_k \\ \beta_k \\ \sigma_k^2 \end{bmatrix}, \tag{22}$$

where α_k and β_k are parameters considered to be time-variant, and are called pseudo-particles.

Maximum likelihood estimates α_0 and β_0 are used to compute the initial conditions α_1 and β_1, which include a random perturbation for every particle (i):

$$\alpha_1^{(i)} = \mathcal{N}(\alpha_0, 0.1 \cdot \alpha_0), \tag{23}$$

$$\beta_1^{(i)} = \mathcal{N}(\beta_0, 0.1 \cdot \beta_0). \tag{24}$$

The initial conditions are used to drive the noise variance of the parameters in the extended state vector (see Eq. (17)):

$$\alpha_k^{(i)} \sim \mathcal{N}(\alpha_{k-1}^{(i)}, \alpha_1^{(i)} \sigma_{\alpha,\beta}^2) \tag{25}$$

$$\beta_k^{(i)} \sim \mathcal{N}(\beta_{k-1}^{(i)}, \beta_1^{(i)} \sigma_{\alpha,\beta}^2) \tag{26}$$

There are two major drawbacks with this method:

- $\mathbb{P}(\alpha_k^{(i)} < 0) > 0$ and $\mathbb{P}(\beta_k^{(i)} < 0) > 0 \quad \forall i, k$,
- $\mathbb{P}(\alpha_k^{(i)} + \beta_k^{(i)} > 1) > 0 \quad \forall i, k$,

both of which are not permitted in the uGARCH model. In particular, they have been addressed in the following way:

- For each $\alpha_k^{(i)} < 0$, let $\alpha_k^{(i)} = 10^{-5}$. Similarly, for each $\beta_k^{(i)} < 0$, let $\beta_k^{(i)} = 10^{-5}$.
- The Particle Filter self-regulates from the cases where $\alpha_k^{(i)} + \beta_k^{(i)} > 1$, provided that these cases have very low likelihood, which translate into very low values of corresponding weights. Hence, no saturation condition has been used for the upper bound.

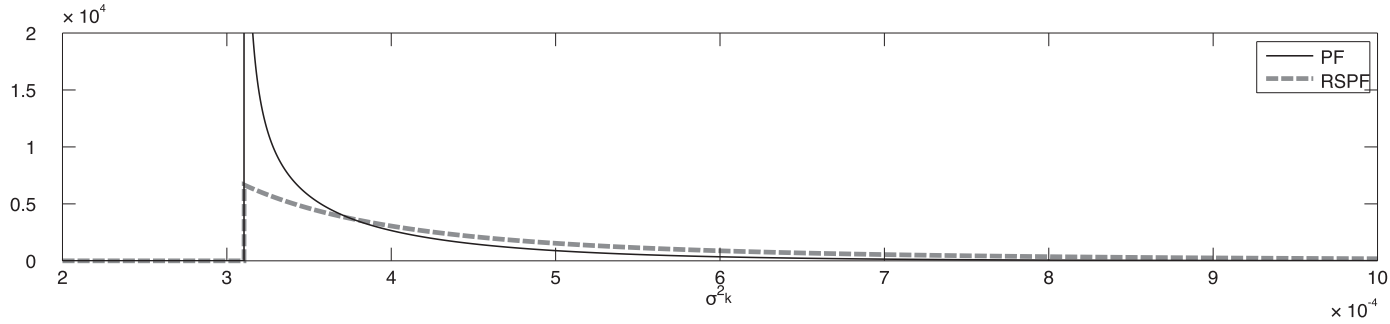


Fig. 4. Comparison between $q_{PF}(\sigma_k^2|\sigma_{k-1}^{2(i)})$ and $q_{RSPF}(\sigma_k^2|\sigma_{k-1}^{2(i)})$. In this example, $\sigma_{k-1}^{2(i)} = 5 \times 10^{-4}, \alpha = 0.2, \beta = 0.6, \omega = 1.0468 \times 10^{-5}$, and $\sigma_\eta^2 = 0.7$. The RSPF obtains particles from a fat-tailed distribution to ensure a higher resolution in risk-sensitive areas.

3.2.2. Particle Filters

Volatility estimation in both simulated data and stock market data has been performed using two different Particle Filtering schemes, including classic and risk sensitive approaches. The estimation is performed using 100 particles. Due to the inherent randomness of the filtering processes, these are repeated 10 times. Also, both filters are implemented with a resampling stage, where residual resampling is used. Next, details about each particular filter are presented.

Classic Particle Filter. The Classic Particle Filter (PF) for volatility estimation in the uGARCH model has been implemented using an importance density equal to

$$q_{PF}(x_k|x_{k-1}^{(i)}, z_k) = p(\sigma_k^2|\sigma_{k-1}^{2(i)}). \tag{27}$$

A closed expression for $p(\sigma_k^2|\sigma_{k-1}^{2(i)})$ has been given in Eq. (4). Thus, samples are generated according to:

$$\sigma_k^{(i)} \sim p(\sigma_k^2|\sigma_{k-1}^{2(i)}), \tag{28}$$

which leads to the following weight update equation:

$$w_k^{(i)} = w_{k-1}^{2(i)} p(r_k|\sigma_k^{2(i)}). \tag{29}$$

Risk Sensitive Particle Filter. In the search for an importance density function that could be used to propose a risk sensitive approach towards volatility estimation, it was necessary to find a distribution with very specific characteristics. First, the probability density function needs a localization parameter that lets both the Classic Particle Filter's (PF) and Risk Sensitive Particle Filter's (RSPF) importance density have the same support. Second, the RSPF's importance density should have a fatter tail than the PF's density. The proposed RSPF uses the Generalized Pareto Distribution as the importance density function, which is commonly used to model the tails of other distributions; since it is able to model exponential, polynomial and even finite tails.

The Generalized Pareto Distribution is defined as follows (Embrechts, Kluppelberg, & Mikosch, 1997; Mathworks, 2013):

$$f_{GPD}(x|k,\sigma,\theta) = \begin{cases} \frac{1}{\sigma}\left[1+k\frac{x-\theta}{\sigma}\right]^{-1-\frac{1}{k}} & \text{if } \begin{cases} k>0, & \text{for } \theta < x \\ k<0, & \text{for } \theta < x < \sigma/k \end{cases} \\ \frac{1}{\sigma}\exp\left[-\frac{x-\theta}{\sigma}\right] & \text{if } \quad k=0, \text{ for } \theta < x \end{cases} \tag{30}$$

This distribution has two special cases, where it is reduced to other distributions:

- If $k = 0$ and $\theta = 0$, the generalized Pareto distribution is equivalent to the exponential distribution.
- If $k > 0$ and $\theta = \sigma/k$, the generalized Pareto distribution is equivalent to the Pareto distribution.

The probability density function of the GPD has three parameters. These can be interpreted as follows:

- k: Shape parameter,
- σ: Scale parameter,
- θ: Threshold (location) parameter.

These parameters cannot take any value if one wants to ensure the convergence of the first and second moments of the GDP, since

$$\mathbb{E}[X] = \theta + \frac{\sigma}{1+k}, \text{ for } k < 1,$$
$$\mathbb{V}ar[X] = \frac{\sigma^2}{(1-k)^2(1-2k)}, \text{ for } k < 1/2, \tag{31}$$

Considering that the variance is defined for $k < 1/2$, the parameters of the probability density function of the GPD have been used in the following way to utilize it as the importance density of the RSPF:

$$k = 0.49, \tag{32}$$

$$\sigma = 0.3\sigma_{k-1}^{2(i)}, \tag{33}$$

$$\theta = \omega + \beta_{k-1}^{(i)}\sigma_{k-1}^{2(i)}, \tag{34}$$

where $\beta_k^{(i)}$ is the $(i)^{\text{th}}$ pseudo-particle for the online estimation of the uGARCH parameter β. Parameter k has been fixed in the aforementioned value to reproduce the shape of $p(\sigma_k^2|\sigma_{k-1}^2)$ (see Eq. (4)). Parameter σ gives the scale to $f_{GPD}(x|k, \sigma, \theta)$. Given that $\max\{f_{GPD}(x|k,\sigma,\theta)\} = 1/\sigma$, using a scaled previous-step particle, a desired fat tail with similar shape to $p(\sigma_k^2|\sigma_{k-1}^2)$ is obtained. Parameter θ sets the location of the density of the GPD and is set to be equivalent to $\omega + \beta_{k-1}^{(i)}$ (see Eq. (4)), this is, the support of $f_{GPD}(x|k, \sigma, \theta)$ is set to be equivalent to the support of $p(\sigma_k^2|\sigma_{k-1}^2)$. Hence, the importance density function employed is

$$q_{RSPF}(x_k|x_{k-1}^{(i)}, z_k) = f_{GPD}(\sigma_k^2|0.49; 0.3\sigma_{k-1}^{2(i)}; \omega + \beta_{k-1}^{(i)}\sigma_{k-1}^{2(i)}). \tag{35}$$

Particles are drawn from

$$\sigma_k^{2(i)} \sim f_{GPD}(\sigma_k^2|0.49; 0.3\sigma_{k-1}^{2(i)}; \omega + \beta_{k-1}^{(i)}\sigma_{k-1}^{2(i)}), \tag{36}$$

and the weight update equation is

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(r_k|\sigma_k^{2(i)})p(\sigma_k^{(i)}|\sigma_{k-1}^{(i)})}{f_{GPD}(\sigma_k^{2(i)}|0.49; 0.3\sigma_{k-1}^{2(i)}; \omega + \beta_{k-1}^{(i)}\sigma_{k-1}^{2(i)})}. \tag{37}$$

A visual comparison of $q_{PF}(x_k|x_{k-1}^{(i)}, z_k)$ and $q_{RSPF}(x_k|x_{k-1}^{(i)}, z_k)$ is shown in Fig. 4. Notice that both importance densities are defined over the same support, and $q_{RSPF}(x_k|x_{k-1}^{(i)}, z_k)$ has a fatter tail than $q_{PF}(x_k|x_{k-1}^{(i)}, z_k)$. Hence, the design conditions for the RSPF's importance density are met.

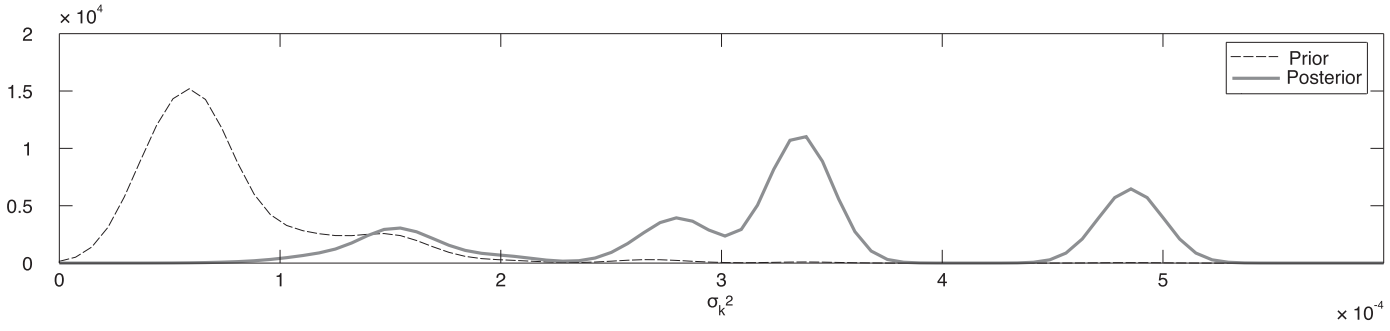


Fig. 5. Examples of prior and posterior densities of the RSPF in a volatility filtering process.

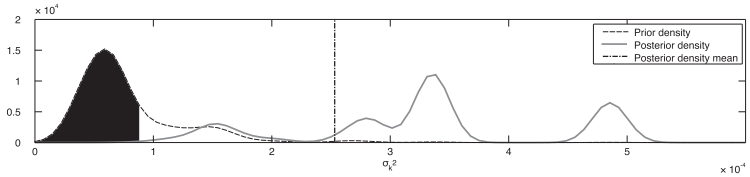


Fig. 6. Hypothesis test example. The filled area under the smoothed prior density represents the 70% confidence interval for the smoothed prior density. Here, the null hypothesis is accepted.

3.3. Detection using a hypothesis test

The Particle Filter, such as any other Bayesian filtering framework, predicts through model dynamics (Eq. (5)), and updates the estimation with the new measurement. Therefore, for every time step, the Particle Filter produces a prior estimate and a posterior estimate:

$$\text{Prior:} \quad \hat{p}(x_k|z_{1:k-1}) = \sum_{i=1}^{N_s} w_{k-1}^{(i)} \delta(x_k - x_k^{(i)}), \tag{38}$$

$$\text{Posterior:} \quad \hat{p}(x_k|z_{1:k}) = \sum_{i=1}^{N_s} w_k^{(i)} \delta(x_k - x_k^{(i)}), \tag{39}$$

Differences between prior and posterior densities may be considerable if model dynamics diverge from actual measurements. This is the case when unlikely events such as unexpected market falls or machine ruptures occur. Fig. 5 shows the vast differences that may occur between prior and posterior density estimates. A detection scheme through hypothesis test exploits these differences to design rapid change detectors in the estimated state.

To accept or reject the null hypothesis $\mathcal{H}_0$, the implemented test considers the 70% confidence interval of the prior density, and contrasts it with the mean of the posterior density. The confidence interval is calculated using Parzen windows and a Normal kernel, whose bandwidth σ_{kernel} is obtained through Silverman's thumb rule (Principe, 2010). If the mean of the posterior density is greater than the 70% interval bound of the prior density, the null hypothesis is accepted. Fig. 6 shows an example of the designed hypothesis test, where an unlikely event occurs and the null hypothesis $\mathcal{H}_0$ is accepted.

4. Results obtained for the proposed detection strategy

This section describes the results obtained for volatility estimation using Particle Filters, and detection of high volatility clusters using information derived from the filtering process. During the training stage, hyper-parameter selection for the setup of the PF algorithm is achieved through a sensibility analysis and subsequent selection through the smallest associated estimation error (in percentage). These results are used to select the PF algorithm hyper-parameters to be utilized in the detection scheme, where estimation through PF is crucial. After the hyper-parameter selection stage, results for three different detection approaches are presented.

The performance measure introduced in this chapter may only be used in simulated data, where the true volatility is known. Hence, quantitative results showing performance measures results are presented for simulated data, and qualitative results are presented for returns series derived from IBM stock prices.

4.1. Performance measure: accuracy indicator

Section 2 describes the non-observability property of the state associated with financial volatility. The performance measure described in this section assume knowledge of the ground truth data, and as a consequence, results may be analyzed only in simulated data. The following sections consider $\hat{\sigma}_k^2$ as the estimated volatility and σ_k^2 as the *true* volatility (this is, the ground truth).

Accuracy of Particle Filters, including the Classic and Risk Sensitive approaches, is compared in terms of error (in percentage). The accuracy indicator is defined as follows:

$$i^{EX}(k) = \frac{|\hat{\sigma}_k^2 - \sigma_k^2|}{\sigma_k^2} \cdot 100. \tag{40}$$

Given that $i^{EX}(k)$ is defined for every time step k of the filtering process, one can obtain an average of $i^{EX}(k)$ over the filtering time window $T_i, \ldots, T_f$:

$$I^{EX} = \frac{1}{T_f - T_i} \sum_{k=T_i}^{T_f} i^{EX}(k) = \frac{1}{T_f - T_i} \sum_{k=T_i}^{T_f} \frac{|\hat{\sigma}_k^2 - \sigma_k^2|}{\sigma_k^2} \cdot 100. \tag{41}$$

Furthermore, given that I^{EX} is defined only for one filtering process, one can obtain an average of I^{EX} over the amount of realizations of the filtering process, which include 10 in this work:

$$\bar{I}^{EX} = \frac{1}{10} \sum_{n=1}^{10} I^{EX}(n). \tag{42}$$

In particular, index $\bar{I}^{EX}$ serves as the base to compare the error (in percentage) for each filtering process, for each set of hyper-parameters. Hence, the best set of parameters is obtained by observing the smaller index $\bar{I}^{EX}$.

Table 2
Parameter estimation using MATLAB®'s `garchfit` function from the Financial Toolbox.

Parameter	GARCH1	GARCH3
α	0.1477	0.2612
β	0.2768	0.6699
$\alpha+\beta$ (<1)	0.4245	0.9311

4.2. Volatility estimation through Particle Filters for an early high volatility cluster detection scheme

4.2.1. Fitting model parameters

The first step towards a filtering process starts with the estimation of model parameters that are used as the initial conditions in the extended GARCH model (refer to Section 3.2.1). As previously mentioned, for the uGARCH model this can be achieved by maximum likelihood, assuming that the model is in fact a GARCH model (Table 1).

Table 2 shows the results of model parameter estimation for the simulated data sets. The parameters are obtained using the `garchfit` function of MATLAB® over the training window of each data set. The true values for this time window are $\alpha = 0.2$ and $\beta = 0.6$ for every data set. From Table 2, GARCH1 is the data set that obtains the poorest parameter estimates from data contained within the training window.

This estimation has direct incidence over the filtering process, since these values are used as initial conditions for the extended uGARCH model, where the dynamics are non-observable. Initial conditions in non-observable systems are of great importance in the outcome of the a Bayesian filtering process. If the system is non-observable, the state may follow one of an infinite number of possible paths that match the current observations. Therefore, accurate initial conditions are necessary to achieve an unbiased estimate of the state.

4.2.2. Particle Filter hyper-parameter selection

This section presents the results of the sensibility analysis of hyper-parameters of the Classic Particle Filter. These parameters have been tested to find the combination that minimizes the estimation error $\bar{I}^{EX}$. The tested parameters include R_{th} (resampling threshold), $\sigma_{\alpha,\beta}$ (pseudo-particle standard deviation), and σ_η (process noise).

To find the hyper-parameter values that minimize the estimation error, the following hyper-parameter mesh is used:

- $R_{th} = \{0.5, 0.6, 0.7\}$,
- $\sigma_{\alpha,\beta} = \begin{Bmatrix} 0.0010, & 0.0015, & 0.0020, & 0.0025, & 0.0030 \\ 0.0035, & 0.0040, & 0.0050, & 0.0075, & 0.0100 \\ 0.0125, & 0.0150, & 0.0175, & 0.0200, & 0.0225 \\ 0.0250, & 0.0275, & 0.0300, & 0.0350, & 0.0400 \end{Bmatrix}$,
- $\sigma_\eta = \{0.5, 0.6, 0.7\}$.

It should be noted that the selected values that are employed to create the mesh for $\sigma_{\alpha,\beta}$ were placed at irregular intervals. Since there is a tendency to have better estimations with lower values of $\sigma_{\alpha,\beta}$, a better resolution has been given to the interval of smaller values.

Every set of hyper-parameters is used to run 10 times each filtering process over the complete time window $T = \{1, \ldots, 500\}$ of every set of simulated data (GARCH1 and GARCH3). The error is computed over the interval $T' = \{151, \ldots, 500\}$, which excludes the training interval.

Table 3a and b shows the percentage error for the mean of the 10 filtering routines for each set of parameters. Since results are 3-dimensional, the tables show the results for $\sigma_\eta = 0.7$, which is the noise process value that minimizes the error for every data set.

Table 3
Estimation error $\bar{I}^{EX}$ for different parameter values $\sigma_{\alpha,\beta}$ for the two data sets. These tables show the results for $\sigma_\eta = 0.7$.

R_{th}	0.0010	0.0015	0.0020	0.0025	0.0030	0.0035	0.0040	0.0050	0.0075	0.0100	0.0125	0.015	0.0175	0.02	0.0225	0.025	0.0275	0.03	0.035	0.04
Data set GARCH1. The minimum error is obtained for $\sigma_{\alpha,\beta} = 0.0350$, $R_{th} = 0.6$.																				
0.5	51.0611	51.2075	50.1367	50.1331	47.4300	48.0407	46.5103	44.3850	38.7528	30.2866	26.1898	23.0201	21.7539	21.6545	21.4255	20.5647	20.3163	19.9770	19.7330	19.9307
0.6	51.4037	50.8755	50.6492	49.4681	49.1941	47.3560	47.2149	45.0067	36.1592	31.3635	27.4839	25.0151	22.8430	21.6427	20.7188	21.0935	20.1355	19.7223	**19.4162**	20.1503
0.7	51.4245	50.8909	50.8915	49.6318	48.5846	47.6203	47.2209	45.2000	36.7997	30.1693	27.4139	24.1124	22.9076	21.5613	21.4261	20.0818	20.5648	20.6290	19.9582	20.0205
Data set GARCH3. The minimum error is obtained for $\sigma_{\alpha,\beta} = 0.0035$, $R_{th} = 0.7$.																				
R_{th}	0.0010	0.0015	0.0020	0.0025	0.0030	0.0035	0.0040	0.0050	0.0075	0.0100	0.0125	0.015	0.0175	0.02	0.0225	0.025	0.0275	0.03	0.035	0.04
0.5	19.3664	20.0746	17.8894	17.7761	18.0384	17.3640	16.3491	18.7394	19.4852	22.1763	24.1930	26.4535	27.9966	31.0447	32.1416	33.1691	35.9803	38.5001	38.3597	41.2628
0.6	18.1588	17.5985	17.4676	17.9119	17.2962	17.4517	18.1153	17.9320	19.3590	20.3719	24.5801	25.5974	29.5183	31.5754	32.1902	32.8693	37.5007	37.0170	39.5580	42.0956
0.7	19.7638	18.9000	18.2602	18.5192	17.4313	**16.1627**	17.2223	18.4436	18.9302	21.1907	24.3899	26.4036	27.6496	29.5650	35.0783	31.5716	37.9800	35.0379	39.5324	43.3656

Table 4
Summary of the sensibility analysis for the Classic Particle Filter.

Parameter	GARCH1	GARCH3	Mean
R_{th}	0.7	0.7	0.63
$\sigma_{\alpha,\beta}$	0.0350	0.0035	0.0141
σ_{η}	0.7	0.7	0.7
Minimum error $\bar{I}^{EX}$	19.4162	16.1627	22.9247
$\alpha+\beta$ (<1)	0.4245	0.9311	0.7609

The first thing to note in Table 3a and b is that the minimum percentage error for the filtering process on these data sets is bounded approximately between 16% and 19%. Second, it is very important to notice that errors are very similar for each one of the columns of the tables. This means that the resampling threshold R_{th} has a limited impact on the estimates when contrasted to the ground truth from an accuracy perspective. This is very important since one can simply employ an average of R_{th} over all data sets without losing estimation accuracy, or simply select the value that is most often the best value. Also, data shows that minimization occurs over a convex space which lets one assume that there is in fact a set of parameters which minimize the estimation error. These results also demonstrate that higher noises (this is, greater particle variability) do not translate into better estimates. In fact, there is a small subset of the parameter space where Bayesian filters such as the Particle Filter may work properly.

Table 4 shows the summary of selected hyper-parameters for each data set and its arithmetic mean, calculated using the information for every data set. Error values increase hugely towards the left side of the columns of Table 3a. This most probably occurs due to the poor estimation of initial conditions through maximum likelihood. Since the initial conditions are far from the ideal values, more variability is needed in the artificial evolution equations included within the Particle Filter algorithm in order to effectively learn and find the correct intervals where these parameters lie. On the other hand, error values increase hugely towards the right side of Table 3b. Initial conditions are very close to the ideal values, low noise variabilities are needed in order to find the correct intervals where these parameters lie.

The inherent non-observability issues of volatility imply that using an average value for $\sigma_{\alpha,\beta}$ over all the data sets where the filtering process is applied will result in poor estimations for the certain data sets.

To choose specific hyper-parameter values R_{th}, $\sigma_{\alpha,\beta}$ and σ_{η}, it is necessary to consider that hyper-parameters R_{th} and σ_{η} have a very small incidence in the estimation error given the parameter mesh. Thus, both of these parameters are set to 0.7. For parameter $\sigma_{\alpha,\beta}$, if one considers the arithmetic mean, results for the GARCH1 data set are far from optimum. Nevertheless, this is the proposed value used in the detection scheme. As a summary, the values considered for the proposed detection algorithms are the following:

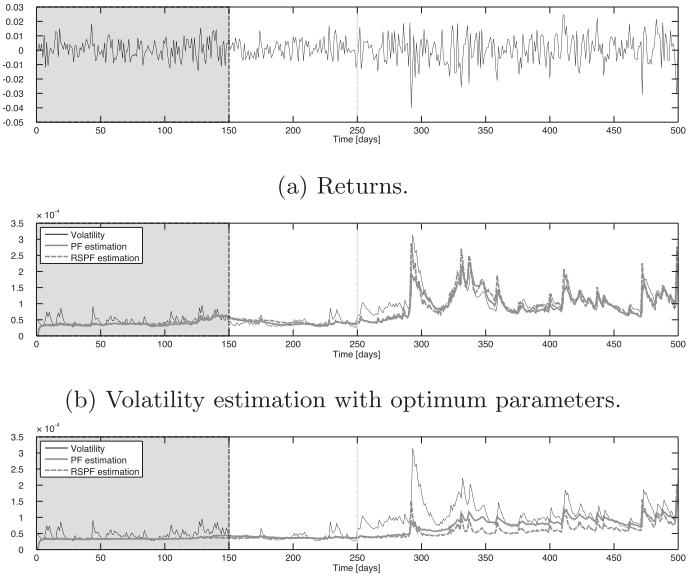
$$R_{th} = 0.6, \tag{43}$$

$$\sigma_{\alpha,\beta} = 0.0141, \tag{44}$$

$$\sigma_{\eta} = 0.7. \tag{45}$$

4.2.3. PF and RSPF-based volatility estimation results

Filtering results with optimum hyper-parameters for each data set. This section presents the results obtained for volatility estimation using the hyper-optimum parameters for the Classic Particle Filtering processes, described in Section 4.2.2. These hyper-parameters have also been applied and used in the RSPF. Results are shown in Figs. 7b and 8b.

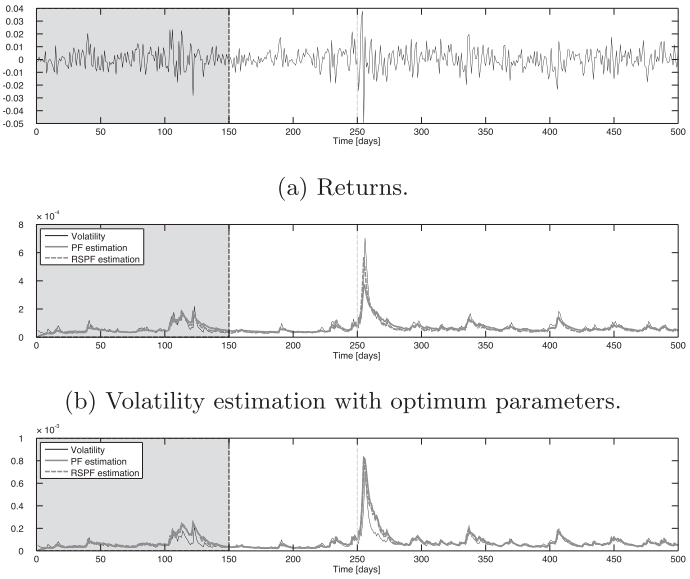


(a) Returns.

(b) Volatility estimation with optimum parameters.

(c) Volatility estimation with averaged hyper-parameters.

Fig. 7. Volatility estimation in GARCH1 data set. Thin line represents ground truth volatility, coarse line represents the PF estimation, and coarse dashed line represents the RSPF estimation.



(a) Returns.

(b) Volatility estimation with optimum parameters.

(c) Volatility estimation with averaged hyper-parameters.

Fig. 8. Volatility estimation in data set GARCH3. Thin line represents ground truth volatility, coarse line represents the PF estimation, and coarse dashed line represents the RSPF estimation.

Analysis of the estimation performance in each data set uncovers many interesting findings that need to be addressed. The comments about results are discussed separately for every data set.

- GARCH1 (Fig. 7): In this data set, there is a large change in the model parameters α and β for the simulated data at time step 250. Up to time step 250, both the PF and RSPF are only able to track the trend of the volatility curve, but there is no reaction to sudden changes. This behavior changes in time step 250, where there is a tendency towards capturing rapid volatility changes. The filter demonstrates the results of the learning process at time step 290, where a hefty volatility cluster

occurs. There is correct tracking of volatility shape with a very small estimation bias. This occurs between time steps 290 and 500, which corresponds to the end of the time window.

- GARCH3 (Fig. 8): There is excellent filtering performance throughout the time window. Accurate estimation, excellent shape tracking. For this data set, only about 100 time steps are necessary for the algorithm to learn and adapt.

The algorithms need at most 300 data points to learn and correctly adapt to the observed data. In general, there is good tracking of trends before this turning point, but if the algorithm is able to adapt properly, both tendency and shape are correctly tracked.

One important aspect from these results is that estimation performance depends vastly on the value $\sigma_{\alpha,\beta}$. Previous experiments demonstrate that low values of $\sigma_{\alpha,\beta}$ (this is, lower than the optimum) result in underestimation of financial volatility in both PF approaches, while higher values of $\sigma_{\alpha,\beta}$ are conducive to overestimation of volatility. From the purpose of tracking the shape of the envelope that characterizes the evolution of volatility in time, this is irrelevant, unless the filters lose the ability to track due to lack of particle variability. Given these results, it could be convenient to separate $\sigma_{\alpha,\beta}$ into σ_α and σ_β (this is, to consider separate sources of uncertainty on each pseudo-particle that extends the model). This is important for two reasons: First, is it necessary to understand that α multiplies the process noise σ_η in Eq. (1) and therefore, the process noise in the extended uGARCH model is the result of the multiplication of two random variables: $\alpha_k \sim \mathcal{N}(\alpha_{k-1}, \alpha_1\sigma_\alpha)$ and η_k^2. Second, variables α and β introduce different behaviors in the model, since the former is associated to innovations and the latter is associated to the memory of the model.

According to Rachev, Hsu, Bahasheva, and Fabozzi (2008), the value of $\alpha+\beta$ in the GARCH model is the "*process persistence parameter*, since it determines the speed of the mean-reversion of volatility to its long-term average. A higher value for $\alpha+\beta$ implies that the effect of the shocks of volatility, u_k^2, dies out slowly". In Table 4, the estimated value of $\alpha+\beta$ was included for each data set. Although there is no apparent relation between the filtering performance of the Particle Filters and the value of $\alpha+\beta$, there is in fact one relation that needs attention: The data set in which the estimated value of α was bigger, the filtering performance was more accurate and errors were systematically lower (Table 3b).

From a detection perspective, it is necessary to notice that the RSPF is usually more capable of tracking correctly sudden rises of volatility. In these cases, estimations of the RSPF are better than the PF estimations, since the latter tends to under estimate. This seems a natural result considering the construction of both PFs: The RSPF grants more resolution to high volatility areas, resulting in a better estimation of sudden volatility rises.

Continuing with the PF and RSPF comparison, the PF usually outperforms the RSPF in terms of estimation accuracy. Albeit sudden volatility changes from low to high values, the PF is less biased than the RSPF.

As a final comment, one should notice that the RSPF outperforms the Classic PF in terms of 1-step prediction in cases where volatility experiences sudden increments. This occurs due to the construction of the uGARCH model, in comparison to the GARCH model. Comparing both dynamics equations,

$$\text{GARCH: } \sigma_{k|k-1}^2 = \omega + \alpha u_{k-1}^2 + \beta\sigma_{k-1|k-2}^2,$$

$$\text{uGARCH: } \sigma_k^2 \quad = \omega + \alpha\sigma_{k-1}^2\eta_k^2 + \beta\sigma_{k-1}^2,$$

where $u_k = \sigma_{k|k-1}\epsilon_k$, the innovations process in the GARCH model depends on the value of the previous step of the returns process, while the volatility dynamics of the uGARCH model are time independent of the returns series. Since the simulated data was generated according to a GARCH model and cases associated with filtering through PF schemes is based on the uGARCH model, this 1-step prediction in sudden volatility rises is possible.

Filtering results with averaged hyper-parameters (as used in detection scheme). This section presents the results obtained for financial volatility estimation using the averaged hyper-parameters in the Classic Particle Filtering implementation (Eqs. (43)–(45)). The use of averaged hyper-parameter values is performed as an attempt to provide a more realistic solution to the problem of interest. This algorithm is intended to be used in stock market data and the hyper-parameters need to be estimated. These hyper-parameters have also been applied and used in the RSPF. Results are shown in Figs. 7c and 8c.

The previous section describes the phenomenon of over and underestimation related to the selected value of $\sigma_{\alpha,\beta}$. Given that the previous experiment showed results for the optimum value of this parameter, this situation was not apparent. Nonetheless, the new experiment makes this behavior palpable. A detailed analysis for each filtering process is given below.

- GARCH1 (Fig. 7c): The corresponding figure clearly shows underestimation of the state. Nevertheless, an interesting result is that shape tracking is extremely accurate, which is essential for the correct operation of the proposed detection algorithms.
- GARCH3 (Fig. 8c): Results for this data set are extremely interesting because of the ample robustness of the filtering performance to variations of the value $\sigma_{\alpha,\beta}$. Shape tracking and estimation accuracy are almost intact in contrast to the use of optimum parameters.

The anomalous behavior occurring in data set GARCH1 (Fig. 7c) may be explained again by the estimation through maximum likelihood of parameter α.

Comparing the performance of the PF and RSPF, there is again a clear response from the RSPF towards estimating correctly sudden changes in volatility from low to high values. This is correct for sudden changes, since the PF tends to be less biased in average. This behavior is extremely important for the detection scheme, since correct performance from the proposed detection techniques can be obtained even though the optimal parameters are not used in simulated or real data.

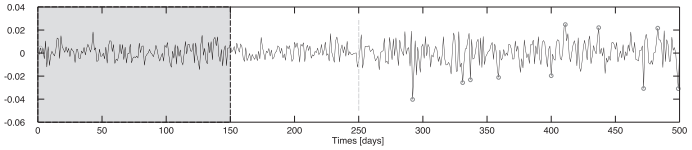
4.3. Early detection of high volatility clusters using a hypothesis test

This section presents results obtained from the proposed hypothesis test to capture early rises in volatility. Figs. 9 and 10 show these results. These figures contain 3 subfigures, which correspond to (a) returns series, (b) volatility series, RSPF prior and posterior estimation, and confidence interval, (c) detection points.

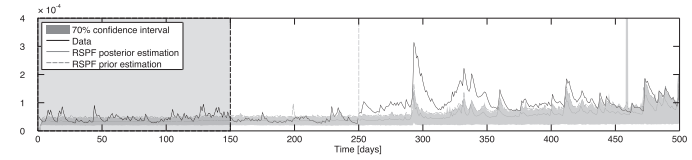
Figs. 9 to 10 show that the detector works correctly, since it is able to capture early rises of volatility which transform into high volatility clusters. The detector can also be interpreted as a local peak detector in the returns series, which is expected. Since the hypothesis test contrasts the dynamics of the model (prior) and the updated dynamics through the observations (posterior), it is clear that detections will occur mainly when local peaks of returns occur.

A detailed analysis of the results for each of the data sets is given ahead.

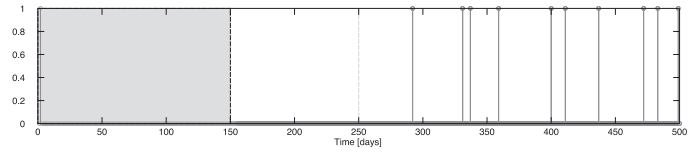
- GARCH1 (Fig. 9): All of the high volatility clusters are detected, except for the high volatility variation due to regime shift at time step 250. This regime shift introduces a notorious mean variation in volatility, which the test is not able to capture, since there are no vast variations in the returns series. High volatility sub-clusters around time step 350 are also detected.



(a) Returns.

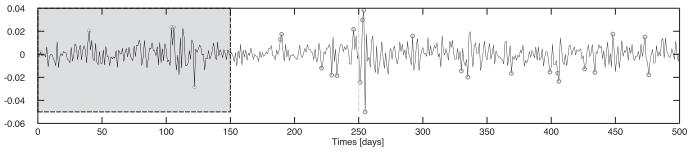


(b) Volatility estimation.

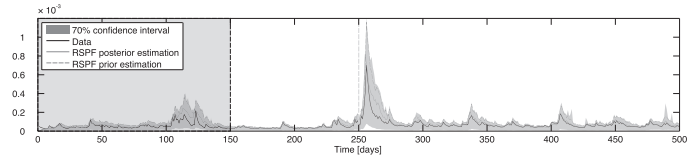


(c) Detections.

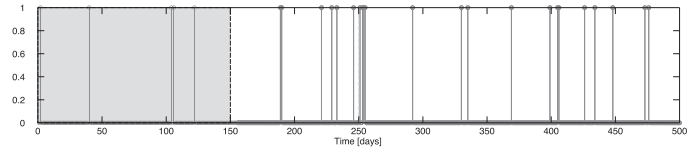
Fig. 9. Hypothesis test-based detection for data set GARCH1.



(a) Returns.



(b) Volatility estimation.



(c) Detections.

Fig. 10. Hypothesis test-based detection for data set GARCH3.

- GARCH3 (Fig. 10): All of the major sudden volatility rises are detected, except for the higher volatility episode starting at approximately time step 320. Continuing over this line of thought, the detector works as expected, although the performance measures do not correctly express the actual detector capacity.

The detection results obtained through a proposed hypothesis test show that the detector is very sensitive even to mild high volatility clusters, when the estimation framework works properly. This translates into the difficulty of measuring correctly the performance of the algorithm, since there is no possible definable hard limit between low and high volatility clusters. In fact, one can only use a diffuse definition.

Other results that need to be addressed correspond to the robustness of the algorithm and the employed value of $\sigma_{\alpha,\beta}$ in the estimation stage. This parameter has tremendous implications over the estimation performance, but not over detection performance. If shape is tracked correctly, the hypothesis test-based detector performs exceptionally well, even under extreme estimation biases.

Table 5
Parameter estimation of the GARCH(1,1) model through maximum likelihood for IBM's returns series between September 12th, 2005 and September 1th, 2009.

GARCH parameter	Value
ω	2.5690×10^{-6}
α	0.0647
β	0.9234
μ	6.9333×10^{-4}

4.4. Case study: early detection of high volatility clusters in IBM's stock data

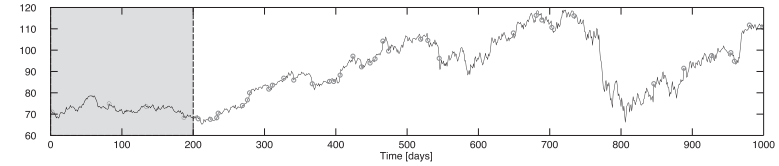
IBM stock price series are usually used as examples for the study of returns series and volatility series (Tsay, 2010). There are various episodes since the year 1962 which are interesting events to explore, including the market falls of 1987 and 2008. As mentioned in Sction 3.1.2, the data employed for this case study involves adjusted closing prices between September 12th, 2005 and September 1st, 2009.

The data observed here does not include the ground truth values for volatility, which means that volatility can only be estimated and therefore, there is no possibility to quantify the detector's performance. Analysis is solely based upon observation of the obtained results and qualitative interpretation of the data.

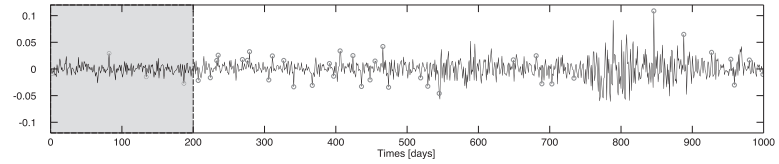
Table 5 displays the estimated parameters for the GARCH(1,1) model in the first 200 data points of the series, which serve as the training period. The parameters ω and μ are left fixed in the extended uGARCH model, while estimations of α and β are used as initial conditions for the online estimation of these parameters. The estimation exhibits a very low value for α, while β has a large value. Given that $\alpha+\beta=0.9881$ and that evidence shows that usually $\alpha+\beta$ is close to 1, one might assume that the estimation is good. Given that the value of α is small, the pseudo-particle standard deviation used is equal to $\sigma_{\alpha,\beta}=0.04$. Moreover, $R_{th}=0.7$ and $\sigma_{\eta}=0.7$.

Fig. 11 exhibits the obtained results from volatility estimation and early detection of high volatility clusters. In particular, details about the adjusted price series, returns series, volatility estimation, detections and the training window may be observed. Analysis of this Fig. 11c shows that volatility estimation of both the PF and the RSPF are extremely close, and the differences between most estimations occur, although mildly, in sudden volatility rises, where the RSPF has a faster reaction towards unlikely values. This is more visible at the beginning of bigger high volatility clusters, from time step 500 and onwards.

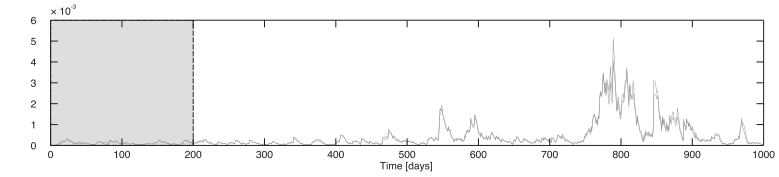
Estimations obtained from the RSPF are used as the base of the hypothesis test-based detector, which showed the best results in the previous sections. One may observe that most of the small high volatility clusters between time steps 200 and 500 are detected. There are some false positives and false negatives, but these are minor. In the time window that includes time steps 500–1000, all of the major volatility clusters are detected in a very early stage, including the high volatility cluster starting at time step 750, conducive to the big stock market drop of the year 2008. Moreover, in this time window, there are only 2 false positives, which occur after the last high volatility cluster. All of the other detections need to be considered true positives.



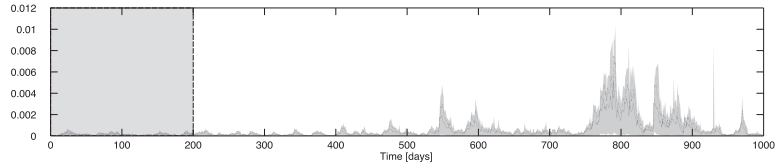
(a) Prices.



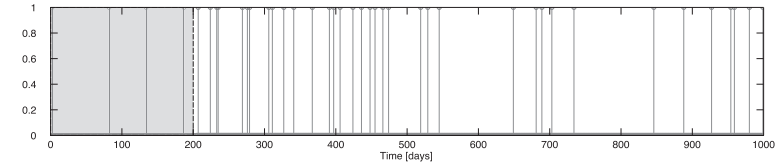
(b) Returns.



(c) Volatility estimation through PF and RSPF.



(d) Confidence interval, prior and posterior estimations for RSPF.



(e) Detections.

Fig. 11. Early hypothesis test-based detection of high volatility clusters in IBM's stock data.

5. Discussion

The results obtained from the PF and RSPF are aligned with the literature: they are suitable frameworks that may offer excellent estimation performance for stochastic volatility estimation. Nevertheless, non-observability issues may produce poor results, a problem that needs to be correctly addressed. Analysis of the parameter estimation of the GARCH model together with the sensibility analysis including noise values demonstrate that estimation performance is extremely dependent on four aspects:

1. Correct initial conditions for particle population.
2. Adequate characterization of process noise sources.
3. Correct initial conditions of pseudo-particles if the state-space model has been extended to include online parameter estimation.
4. Adequate characterization of process noise sources for pseudo-particle variability within artificial evolution-based approaches.

Inadequate values can lead to algorithms with inability to learn, or extremely biased estimates. Moreover, there is an important relationship between points 3 and 4: if estimates of the GARCH model (which in this case are used as the initial conditions) are too low or inaccurate, higher noise values for these pseudo-particles are needed to improve the learning capabilities of the PF algorithm. As a consequence, the parameter $\sigma_{\alpha,\beta}$ should be separated into σ_α and σ_β, this is, use a separate dispersion value for the noise process of each pseudo-particle which extends the model.

A performance comparison between the Classic Particle Filter and the proposed Risk Sensitive approach shows that the Risk Sensitive algorithm behaves better for purposes of tracking sudden volatility changes from low to high values. The greater particle resolution offered by the Risk Sensitive Particle Filter in areas of high volatility give this algorithm a very high performance in these cases.

This filtering approach, combined with the proposed detection technique based on the contrast of prior and posterior estimations of the Risk Sensitive Particle Filter through a hypothesis test proves that early detection of high volatility clusters is possible with a small error. Important aspects associated with the performance ensure that the detection is extremely robust to biased estimates, which are related to sub-optimal dispersion values of noise. In particular, if the Particle Filter does not lose the ability to learn and track the shape of volatility, the proposed hypothesis test-based detector excels in early detection of high volatility clusters.

6. Conclusions

This work presents and explores the use of Particle Filtering frameworks for the online detection of variations in financial returns that may conduce to high volatility clusters. Our approach uses a novel volatility estimator based on a Risk Sensitive Particle Filters (RSPF) that employs the Generalized Pareto Distribution to generate particles in areas associated to higher risk.

The methods proposed include the use of a simple stochastic variation of the GARCH model – the uGARCH model– in order to capture volatility variations of financial returns that may lead to high-volatility clusters. This model has been chosen in order to diminish the complexity of our method, while simultaneously helping to track disturbances introduced by other non-measurable factors (often found in complex systems such as the stock markets). This efforts result in a simple, but effective, detection scheme based on the comparison of prior and posterior PDF estimates through a hypothesis test. The proposed method proves (both with simulated and actual financial data) that early detection of high volatility clusters is possible with a small error using low-complexity models and risk-sensitive approaches in the detection framework.

Future work will focus on exploring connections with the problem of jumps detection in financial variables. Our approach offers a framework that is independent from the stochastic volatility model structure; thus representing a plausible option for online jumps detection in financial econometrics.

Acknowledgements

This work has been partially supported by FONDECYT Chile Grant No. 1140774 and Advanced Center for Electrical and Electronic Engineering, Basal Project FB0008.

References

Andersen, T. G., Tim, & Diebold, F. X. (2007). Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics, 89*, 701–720.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing, 50*, 174–188.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, 307–327.

Chan, J. C., & Koop, G. (2014). Modelling breaks and clusters in the steady states of macroeconomic variables. *Computational Statistics and Data Analysis, 76*, 186–193.

Chen, C. W., Gerlach, R., & Liu, F.-C. (2011). Detection of structural breaks in a time-varying heteroskedastic regression model. *Journal of Statistical Planning and Inference, 141*, 3367–3381.

Cont, R. (2005). Volatility clustering in financial markets: empirical facts and agent–based models. *Long memory in economics*. Springer.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential monte carlo sampling methods for Bayesian filtering. *Statistics and Computing, 10*, 197–208.

Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). Modelling extremal events for insurance and finance. *Stochastic modelling and applied probability*: vol. 33 (1st ed.). Springer-Verlag.

Eraker, B., Johannes, M., & Polson, N. (2003). The impact of jumps in volatility and returns. *The Journal of Finance, 58*(3), 1269–1300.

Gordon, N., Salmond, D., & Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F: Radar and Signal Processing, 140*, 107–113.

He, Z., & Maheu, J. M. (2010). Real time detection of structural breaks in GARCH models. *Computational Statistics and Data Analysis, 54*, 2628–2640.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering, 82*, 35–45.

Kitagawa, G., & Sato, S. (2001). Monte carlo smoothing and self-organizing state-space model. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice in statistics for engineering and information science* (pp. 177–195). Springer-Verlag.

Laurent, S., Lecourt, C., & Palm, F. C. (2014). Testing for jumps in conditionally gaussian ARMA-GARCH models, a robust approach. *Computational Statistics and Data Analysis*. doi:10.1016/j.csda.2014.05.015. (in press).

Lee, S. S., & Mykland, P. A. (2008). Jumps in financial markets: a new nonparametric test and jump dynamics. *The Review of Financial Studies, 21*(6), 2535–2563.

Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential monte carlo methods in practice in statistics for engineering and information science* (pp. 197–223). Springer-Verlag.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business, 36*, 394–419.

Mathworks (2013). Generalized pareto distribution. retrieved from http://www.mathworks.com/help/stats/generalized-pareto-distribution.html.

Mundnich, K. (2013). Early detection of high volatility clusters using particle filters. Electrical Engineering thesis. Department of Electrical Engineering, University of Chile. retrieved from http://tesis.uchile.cl/handle/2250/115486.

Principe, J. C. (2010). Information theoretic learning: Renyi's entropy and kernel perspectives. *Information Science and statistics* (1st ed.). Springer.

Rachev, S., Hsu, J., Bahasheva, B., & Fabozzi, F. (2008). Bayesian methods in finance. *The Frank J. Fabozzi Series* (1st ed.). John Wiley & Sons, Inc..

Rapach, D. E., & Strauss, J. K. (2008). Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics, 23*, 65–90.

Ross, G. J. (2013). Modelling financial volatility in the presence of abrupt changes. *Physica A, 392*, 350–360.

Thrun, S., Langford, J., & Verma, V. (2002). Risk sensitive particle filters*: vol. 14. Advances in Neural Information Processing Systems* (pp. 961–968). MIT Press.

Tobar, F. A. (2010). *Inferencia de volatilidad de retornos financieros usando filtro de partículas*. Master's thesis Universidad de Chile.

Tsay, R. (2010). Analysis of financial time series. *Wiley Series in Probability and Statistics* (3rd ed.). Chicago: John Wiley & Sons, Inc.

Yahoo! (2013). International Business Machines Corporation – NYSE. retrieved from http://finance.yahoo.com/q?s=IBM.