UFRJ

Universidade Federal do Rio de Janeiro

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

Iago Carvalho Cunha

2018

UFRJ

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

Iago Carvalho Cunha

Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

Orientador: Ralph dos Santos Silva

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

Iago Carvalho Cunha

Orientador: Ralph dos Santos Silva

Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

Prof. Ralph dos Santos Silva
IM-UFRJ

Prof. Carlos Antônio Abanto Valle
IM-UFRJ

Prof. Rodrigo dos Santos Targino
EMAp/FGV

CIP - Catalogação na Publicação

Cunha, Iago Carvalho
CC972p Particle filters and adaptive Metropolis Hastings sampling applied to volatility models / Iago Carvalho Cunha. -- Rio de Janeiro, 2018.
54 f.

Orientador: Ralph dos Santos Silva.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto de Matemática, Programa de Pós-Graduação em Estatística, 2018.

1. Análise Bayesiana. 2. MCMC. 3. Filtro de Partículas. 4. Amostragem Adaptativa. I. Silva, Ralph dos Santos, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

ABSTRACT

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

Iago Carvalho Cunha

Advisor: Ralph dos Santos Silva

Abstract da Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da UFRJ como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

Markov Chain Monte Carlo methods are widely used in Bayesian statistical inference to sample the posterior distribution from a target distribution. However, especially for non-Gaussian and non-linear state space models, one can find difficulties in calculating the exact likelihood, thus the proposal distribution of a Metropolis-Hastings algorithm. To overcome problems in calculating the likelihood function, it is possible to use approximations made by particle filter methods (also known as sequential Monte Carlo methods). Furthermore, an adaptive Metropolis-Hastings algorithm may be applied since its proposal distribution is updated with previous draws from the posterior distribution. In this way, this thesis discusses the applicability of adaptive Metropolis-Hastings algorithms with random walk or independent proposal combined with estimated likelihood functions through particle filters. For this, we will estimate non-linear and non-Gaussian volatility models for three series of real index returns.

Key-words: Adaptive MCMC; Dynamic models; Sequential Monte Carlo methods.

RESUMO

Filtro de Partículas e Metropolis-Hastings Adaptativo Aplicados a Modelos de Volatilidade

Iago Carvalho Cunha

Orientador: Ralph dos Santos Silva

Resumo da Dissertação de Mestrado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática da UFRJ como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências Estatísticas.

Métodos de Markov Chain Monte Carlo são bastante utilizados na inferência estatística bayesiana para amostrar da distribuição a posteriori a partir de uma distribuição proposta. Contudo, principalmente para modelos de espaço de estado não lineares e não Gaussianos, pode-se encontrar dificuldades no cálculo da função de verossimilhança exata ou definição da distribuição proposta do algoritmo. Para contornar os problemas no cálculo da função de verossimilhança é possível utilizar aproximações feitas por métodos de filtro de partículas (também conhecidos como métodos sequenciais de Monte Carlo). E, além disso, algoritmos de amostragem adaptativa de Metropolis-Hastings podem ser particularmente úteis no ajuste da distribuição proposta. Neste tipo de amostragem, a distribuição proposta é atualizada a medida que o algoritmo progride. Desta maneira, essa dissertação discute o uso de algoritmos de amostragem adaptativa através de Metropolis-Hastings com propostas passeio aleatório ou independente, e função de verossimilhança estimada através de filtro de partículas. Para isso, iremos estimar modelos de volatilidade não-lineares e ou não-Gaussianos para três séries de retorno de índices reais.

Palavras-chave: MCMC adaptativo, Modelos dinâmicos; Métodos sequenciais de Monte Carlo.

Para

Minha mãe, Maria Celia Miranda de Carvalho

AGRADECIMENTOS

Em especial, ao meu orientador Ralph dos Santos Silva pela compreensão;
Aos meus familiares que me apoiaram; e
A todos os professores do DME/IM/UFRJ.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Nowadays is almost impossible to analyze economic data, especially time series data, without considering volatility, often being considered more important than any other measure in a time series of stock prices, for example. According to Liu et al. (1999, pp. 1390), "The volatility of stock price changes is a measure of how much the market is liable to fluctuate", which leads us to believe that, in order to become a good investor, it is necessary to seriously worry about market fluctuations.

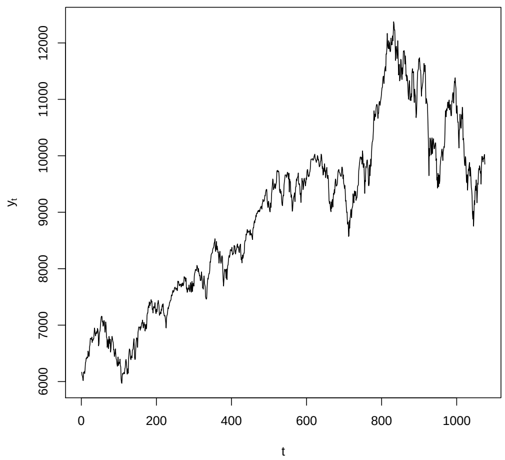
Moreover, another problem with volatility is that it is an unobservable measure, and therefore we need a way of estimating it for both historical series and forecasts, depending on the interest. Still according to Liu et al. (1999, pp. 1391), "The term volatility represents a generic measure of the magnitude of market fluctuations." In the literature, there are several ways of calculating it quantitatively. Note that, given its theoretical definition, volatility resembles the statistical definition of variance.

It is interesting to note that volatility has a property called the leverage effect phenomenom, first seen in Black (1976) and stated as the tendency of an asset's returns to be negatively correlated with its volatility. That is, positive changes are accompanied by small variations in the market or vice versa.

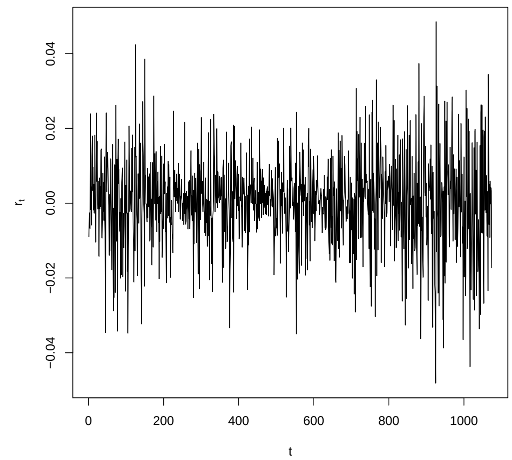
And, finally, about volatility it is important to keep in mind the volatility clustering property. Mandelbrot (1963) was the first to study this effect and, according to him, consists of saying that "...large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes..." Several empirical studies were carried out in the 1990s to verify this effect. Granger & Ding (1995), for example, found that a series of absolute returns elevated to the n–th power has strong autocorrelation.

In Figure 1.1, we can see the daily price series over time t (y_t), the series of daily log-returns over time t (r_t), the autocorrelation for the log-returns series over time t and for the squared log-returns series over time t (r_t^2), belonging to an index of a Stock Exchange market collected from January 2012 to March 2016.

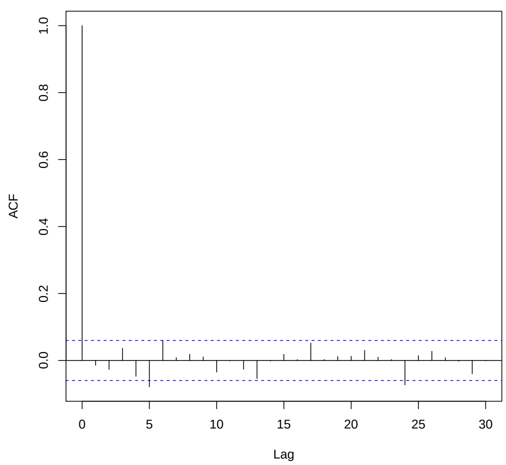
Notice that we can empirically perceive the effect of volatility clustering as established by



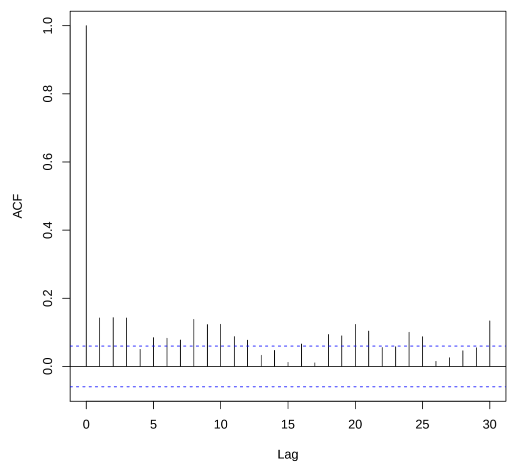
(a) Price series (y_t)



(b) Log-return series (r_t)



(c) Autocorrelation of r_t series



(d) Autocorrelation of r_t^2 series

Figure 1.1: Price and log-return series on the top and the autocorrelation for the log-returns and for the squared log-returns series on the bottom. In Figures 1.1a and 1.1b, $t = 200$ corresponds to october 2012, $t = 400$ to july 2013, $t = 2 = 600$ to may 2014, $t = 800$ to february 2012 and $t = 1000$ to december 2015.

Granger & Ding (1995). There is no significant autocorrelation between the log-returns r_t (Figure 1.1c), but it is possible to see that it appears when we work with squared log-returns

1.1d. This gives a glimpse that models that work with estimation of volatility are suitable for these types of data.

Therefore, we will use models that estimate volatility to perform the desired analyzes. In this way, we will work with historical series of stock market indexes, more specifically with the series of daily log-returns of these indexes. Besides that, it is important to remind that it is not the focus of this dissertation to estimate volatility in any way. More details on log-returns and the advantages of using it can be seen in Section 2.5.1

Among the most important volatility models for economic studies, we can cite the Autoregressive Conditional Heteroscedasticity (ARCH) model developed by Engle (1982), which relates the error variance (volatility) to an autoregressive (AR) model. The ARCH model was so important to the literature of volatility models that the author won the 2003 Nobel Prize for economics for it. Afterwards, Bollerslev (1986) developed the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, which consists of relating the error variance to an autoregressive-moving-average (ARMA) model. However, in this thesis, we will use the GARCH model with noise which is a more flexible version of the basic model. See Section 2.5.2 for more details about ARCH and GARCH models.

In addition, we can also mention the stochastic volatility (SV) models suggested in Jacquier et al. (1995), which uses probability distributions to model volatility. According to Jacquier et al. (1995, p. 4), such models "...allows both the conditional mean and variance to be driven by separate stochastic processes." In more general cases, SV models are also able to capture the leverage effect through stochastic volatility with leverage (SV-lev) models. More details on these models can be found in Section 2.5.3.

In this thesis we are interested in making only investigations on the unknowns of the models from the Bayesian point of view, and we will use only state space models (SSM) to analyze the data. Therefore, we need a way to estimate the posterior distribution of the parameters.

Markov Chain Monte Carlo (MCMC) simulation methods are widely applied in statistical inference to sample from a probability distribution. These methods are generally used in Bayesian inference where the posterior distribution of the unknown parameters is often difficult to calculate exactly. In this approach, the parameters are generated from a proposal distribution, or several such proposal distributions, with the generated proposals accepted or rejected using the Metropolis-Hastings (MH) method (Hastings, 1970). However, it is common to come across problems when calculating the likelihood, commonly for non-linear non-Gaussian models, or in situations where the choice of powerful proposals is not particularly easy. More details about the Metropolis-Hastings method and other basic concepts can be found in Chapter 2.

When the likelihood does not have a closed form we can approximate it using, for example, Particle Filters, also known as Sequential Monte Carlo methods. Two particle filtering methods are the standard particle filter, also known as Sequential Importance Resampling (SIR) or bootstrap filter developed by Gordon et al. (1993) and the auxiliary particle filter (ASIR) proposed by Pitt & Shephard (1999). Moreover, Andrieu et al. (2010) proved that MCMC methods still converge to the correct posterior distribution even if the simulated likelihood via SIR or ASIR is used. The particle filters SIR and ASIR will be detailed in Chapter 2, Section 2.2.

To work around problems when choosing effective proposal distributions to use on the MCMC method, we can apply some adaptive sampling (AS) techniques. In such methods, the parameters of the proposal distribution are tuned by using previous draws and the difference between these successive parameters of the proposal converges to zero (diminishing adaptation). Important theoretical and practical contributions to diminishing adaptation sampling were made by Haario et al. (1999), Haario et al. (2001), Roberts & Rosenthal (2007), Roberts & Rosenthal (2009) and Giordani & Kohn (2010).

The methods proposed by Haario et al. (1999), improved by Haario et al. (2001) with further contributions by Roberts & Rosenthal (2009) are based on the classical Metropolis algorithm (Metropolis et al., 1953) and called adaptive random walk Metropolis sampling (ARWMS) method. Furthermore, Giordani & Kohn (2010) suggested a method based on the independent Metropolis-Hastings algorithm with a proposal distribution based on a mixture of normals. Such methods will be detailed in Chapter 2, Section 2.3.

Chapter 3 contains applications using a generalized autoregressive conditionally heteroscedastic (GARCH) model with noise and several stochastic volatility (SV) models. Here, we use both simulated and real data. First we perform a simulation study to verify if the algorithms recover the true values of the parameters, and then estimate the parameters of real data. The results for real data will be compared using marginal likelihoods and a few likelihood based information criteria, which are explained in Section 2.4. The data used was collected from four stock market indexes around the world.

1.1 Objective

The main objective of this thesis is to discuss the applicability, in volatility models, of adaptive sampling algorithms with the likelihood function simulated by particle filters.

In this way, firstly, to verify the consistency of the algorithms, we will perform a simulation study that consists of estimating the parameters of 5 simulated series from a GARCH with noise model, stochastic volatility with Gaussian noise model and stochastic volatility with leverage model. For all of them the adaptive independent Metropolis-Hastings sampling (AIMHS) suggested by Giordani & Kohn (2010) will be used. Besides that, GARCH with noise likelihood will be estimated using both SIR and ASIR filters, while the others only with SIR.

Then, we will estimate a GARCH with noise model and several SV models again via AIMHS using real daily data from three stock market indexes. Furthermore, the likelihoods will be simulated with both SIR and ASIR methods for GARCH model and with SIR for SV models. It is important to note that to initialize the AIMHS algorithm in all applications an initial estimate was made for all the parameters of the model using the ARWMS method. We will work on the followings indexes: BOVESPA, from Brazil, NASDAQ and S&P500, from United States collected from January 2012 to March 2016.

Next, all estimated models using real data will be compared using various comparison measures, such as the marginal likelihood estimated by a bridge sampling estimator and by an importance sampling approach. Beyond them, several likelihood based information criteria, such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), their expected versions, EAIC and EBIC, and the deviance information criterion (DIC), will also be calculated to compare the models.

Chapter 2 gives a brief description of the methodology and the models. More specifically, Section 2.2 contains details about state space models and particle filters and Section 2.3 discusses about adaptive Metropolis-Hastings methods. Chapter 3 contains results of estimated models for both simulated and real data. Finally, in Chapter 4, conclusions will be drawn and further works described.

Chapter 2

METHODOLOGY AND MODELS

2.1 Statistical Inference

In the present day it is relatively simple to obtain data from various sources, as well as surveys, design of experiments, and so on. It is common in several areas of knowledge that the researcher has an interest in extrapolating the analysis made from such data to the entire population of interest if the analysis is being done from a sample. The extrapolation of these results is called statistical inference.

To understand the fundamentals of inference, it is important to know the difference between population and sample. A population of interest is the set of individuals, in general, with a common characteristic that we are interested in studying. However, it is not always possible to collect or analyse all the individuals of a population, in this way, we collect a sample, which consists of a portion of the individuals of the population. It is important to emphasize that individuals of interest will not always be people or living beings. They can be, for example, a price index of a stock market.

Moreover, statistical inference consists of first identifying a statistical model of the process that generates the data and then deducing the propositions from the model. In this way, the inference uses statistical models that have unknown components, also known as parameters. Because they are unknown, such parameters have an associated uncertainty and, thus, it is the investigator's role to try to reduce this uncertainty. Thus, it is clear that parameter estimation can be treated as a decision problem.

In decision theory, the concepts of loss function, decision rule and risk of a decision rule are fundamental. Thus, imagine the parameters space Θ, the space of possible results of an experiment Ω, and the space of possible actions A.

In this way, the loss function can be defined as $L(\theta, a)$ for all $(\theta, a) \in \Theta \times A$ and assumes values in the set of real numbers, where θ represents a specific parameter belonging to Θ and a a particular action of A. That is, $L(\theta, a)$ represents the loss associated when taking action a if θ occurs. Since the loss function is almost never known, it is normal to use the expected loss in decision making.

In addition, a $\delta : \Omega \to A$ decision rule is a function that represents the action to be taken if $\mathbf{y}$ is an observed value of a sample of the population $\mathbf{Y}$. It is important to note that, from the classical point of view, the decision to be made is based only on the information obtained from the sample collected and, on the other hand, from the Bayesian point of view, the decision is also based in a distribution *a priori* of the parameters, in other words, on the beliefs of the researcher or previous informations of the parameters that are unknown.

And finally, from a Bayesian point of view, the risk of a decision rule is the expected value of the posterior loss and will be denoted by $R(\delta)$. If it is possible to find a decision rule δ^* that minimizes the risk for any δ, then δ^* is called the optimal decision rule, also known as Bayes rule. In other words, δ^* is optimal if $R(\delta^*) < R(\delta)$, $\forall \delta$.

Note that the classical or frequentist approach and the Bayesian approach are the two main approaches on statistical inference. In this thesis we will be interested only in the Bayesian approach. More details on statistical inference from the classical point of view can be seen in Casella & Berger (2002).

2.1.1 Bayesian Estimation

Bayesian estimation is based on Bayes' theorem, initially proposed by Thomas Bayes in the early eighteenth century, and later presented by Pierre-Simon Lapace in the year 1812 in its modern formulation. Note that in Bayesian inference the parameters are assumed to be random variables, so it makes sense to say that the parameters have associated distributions.

Under this approach, when we want to estimate a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d$, we first need to calculate the distribution *a posteriori* of $\boldsymbol{\theta}$ given the information obtained from a sample $\mathbf{y}$ of the population $\mathbf{Y}$ of interest, also known as $p(\boldsymbol{\theta}|\mathbf{y})$.

The posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ can be obtained through a direct application of Bayes' theorem, which is:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \tag{2.1}$$

where $p(\mathbf{y}|\boldsymbol{\theta})$ is the information obtained from the sample $\mathbf{y}$, also knows as the likelihood function, $p(\boldsymbol{\theta})$ is the prior distribution and $p(\mathbf{y})$ is a normalization constant that does not depend on $\boldsymbol{\theta}$. $p(\mathbf{y})$ can be calculated as:

$$p(\mathbf{y}) = \int_{\mathbf{\Theta}} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.2}$$

In other words, we can say that the posterior distribution is proportional to the likelihood function multiplied by the prior distribution, that is,

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{2.3}$$

Now, recalling from Section 2.1, if we want to find an optimal point estimator for $\boldsymbol{\theta}$, we need to minimize the risk of the expected value of a posterior loss function of interest. According to Migon et al. (2015), the most common loss functions are the quadratic loss function, absolute loss function, and 0-1 loss function.

In this case, the risk of a decision rule is defined by

$$R(\delta) = E_{\boldsymbol{\theta}|\mathbf{y}}\left[L(\boldsymbol{\theta},\delta)\right] = \int_{\mathbf{\Theta}} L(\boldsymbol{\theta},\delta\mathbf{y})p(\boldsymbol{\theta}|(y))d\boldsymbol{\theta}. \tag{2.4}$$

The quadratic loss, absolute loss, and 0-1 loss functions are defined, respectively, by $L(\boldsymbol{\theta},\delta) = (\boldsymbol{\theta}-\delta)^2$, $L(\boldsymbol{\theta},\delta) = |\boldsymbol{\theta}-\delta|$ and $L(\boldsymbol{\theta},\delta) = \lim_{\epsilon\to\infty} I_{|\boldsymbol{\theta}-\delta|}([\epsilon,\infty))$. And the Bayes estimator (that is, the optimal decision rule) associated to each one are, respectively, the posterior expected value ($E(\boldsymbol{\theta}|\mathbf{y})$), the posterior median ($median(\boldsymbol{\theta}|\mathbf{y})$) and the posterior mode ($mode(\boldsymbol{\theta}|\mathbf{y})$).

However, point estimates also generate a great loss of the information contained in the sample collected. In this way, if we are interested in finding a set of values in which there is a high probability of finding the parameter of interest, we can use the concept of credibility or Bayesian regions. That is, $\mathbf{C} \in \mathbf{\Theta}$ is a region with $100(1-\alpha)\%$ credibility if $P(\boldsymbol{\theta} \in \mathbf{C}|\mathbf{y}) \geq 1-\alpha$. In this case, $1-\alpha$ is also known as the credibility level.

Note that with the above definition we can find an infinity different regions $\mathbf{C}$, but if we are interested in the one with the smallest possible size, then we must find the one with the highest posterior density associated with $\boldsymbol{\theta}$, or the region of highest posterior density (HPD). Thus, the region $\mathbf{C}$ of HPD for $\boldsymbol{\theta}$ with credibility level $100(1-\alpha)\%$ is the region given by $C = \boldsymbol{\theta} \in \mathbf{\Theta} : p(\boldsymbol{\theta}|\mathbf{y}) \geq k(\alpha)$, where $k(\alpha)$ is the largest constant such that $P(\boldsymbol{\theta} \in \mathbf{C}|\mathbf{y}) \geq 1-\alpha$. See Migon et al. (2015, pp. 137-139) for more details.

2.2 State Space Models and Particle Filtering

A state space model, in general, can be represented by an observation equation given by $p(y_t|x_t;\boldsymbol{\theta})$ and a system equation (also known as model equation) given by $p(x_t|x_{t-1};\boldsymbol{\theta})$,

$t = 1, \ldots, T$, where $\boldsymbol{\theta}$ and $p(.)$ represent a parameter vector and general probability (density) functions, respectively. There may even exist nonlinearities between the observation and system equations. Note that, $y_{1:T} = (y_1, \ldots, y_T)$ denote the history of measurements and $x_{1:T} = (x_1, \ldots, x_T)$ the history of states up to time T. The initial state x_0 distribution is given by $p(x_0|\boldsymbol{\theta})$.

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

$$p(x_t|y_{1:t-1};\boldsymbol{\theta}) = \int p(x_t|x_{t-1};\boldsymbol{\theta})p(x_{t-1}|y_{1:t-1};\boldsymbol{\theta})dx_{t-1} \tag{2.5}$$

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

$$p(x_t|y_{1:t};\boldsymbol{\theta}) = \frac{p(y_t|x_t;\boldsymbol{\theta})p(x_t|y_{1:t-1};\boldsymbol{\theta})}{p(y_t|y_{1:t-1};\boldsymbol{\theta})}, \quad \text{and} \tag{2.6}$$

$$p(y_t|y_{1:t-1};\boldsymbol{\theta}) = \int p(y_t|x_t;\boldsymbol{\theta})p(x_t|y_{1:t-1};\boldsymbol{\theta})dx_t. \tag{2.7}$$

The equations (2.5)–(2.7) allow, in principle, the filtering for a given $\boldsymbol{\theta}$ and to obtain the likelihood function of the observations $\boldsymbol{y} = y_{1:T}$,

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{t=1}^{T} p(y_t|y_{1:t-1};\boldsymbol{\theta}) \ , \tag{2.8}$$

with $y_{1:0}$ representing a null observation.

If the likelihood function (2.8) is available, then maximum likelihood or Bayesian methods through the Metropolis-Hastings algorithm may be employed to make inference of the vector of parameters $\boldsymbol{\theta}$. When both the observation and the system equations are normal and linear, the likelihood function may be calculated analytically using the Kalman filter (West & Harrison, 1997, pp. 103-104).

However, according to West & Harrison (1997, pp. 506-507), when the observation and the system equations are nonlinear and/or non-Gaussian, the integrals given in Equations (2.5)–(2.7) are in general difficult to solve. To work around this problem, the SIR (Gordon et al., 1993) and the ASIR (Pitt & Shephard, 1999) are methods proposed to approximate all distributions in those equations, specially the likelihood function.

2.2.1 Standard Particle Filtering

The Standard Particle Filtering, also known as the Sampling Importance Resampling (SIR) method or the bootstrap filter, was proposed by Gordon et al. (1993) for "...implementing recursive Bayesian filters" when one needs a method for nonlinear or non-Gaussian models, that is, "...it may be applied to any state transition or measurement model." In this section, the SIR method will be described.

Suppose that we have a sample $x_{t-1}^{(\ell)}$, $\ell = 1, \ldots, L$ with probabilities $\pi_{t-1}^{(\ell)}$ from $p(x_{t-1}|y_{1:t-1};\boldsymbol{\theta})$. It is easy to notice that the simplest values of $\pi_{t-1}^{(\ell)}$ are $1/L$. An approximation to equation (2.5) is given by:

$$p(x_t|y_{1:t-1};\boldsymbol{\theta}) \approx \sum_{\ell=1}^{L} p(x_t|x_{t-1}^{(\ell)};\boldsymbol{\theta})\pi_{t-1}^{(\ell)}. \tag{2.9}$$

Therefore, $p(x_t|y_{1:t-1};\boldsymbol{\theta})$ can be viewed as a mixture density with L components. So we just have to pass the sample $x_{t-1}^{(\ell)}$ in the system equation, that is, $p(x_t|x_{t-1}^{(\ell)};\boldsymbol{\theta})$. That would give us a sample $\widetilde{x}_t^{(\ell)}$, $\ell = 1, \ldots, L$ from the density $p(x_t|y_{1:t-1};\boldsymbol{\theta})$.

Now we can update the posterior distribution using Equation (2.6). We obtain a sample $\widetilde{x}_t^{(\ell)}$, $\ell = 1, \ldots, L$ from $p(x_t|y_{1:t};\boldsymbol{\theta})$ by assigning a probability of

$$\widetilde{\pi}_t^{(\ell)} = \frac{p(y_t|\widetilde{x}_t^{(\ell)};\boldsymbol{\theta})\pi_{t-1}^{(\ell)}}{\sum_{j=1}^{L} p(y_t|\widetilde{x}_t^{(j)};\boldsymbol{\theta})\pi_{t-1}^{(j)}} \tag{2.10}$$

to $\widetilde{x}_t^{(\ell)}$. Thus,we have a sample $\widetilde{x}_t^{(\ell)}$, $\ell = 1, \ldots, L$ with probabilities $\widetilde{\pi}_t^{(\ell)}$ from $p(x_t|y_{1:t};\boldsymbol{\theta})$.

Note that from Equations (2.5) and (2.6) the predictive function can be approximated by

$$p_s(y_t|y_{1:t-1};\boldsymbol{\theta}) \approx \sum_{j=1}^{L} p(y_t|x_t^{(j)};\boldsymbol{\theta})\pi_{t-1}^{(j)}, \tag{2.11}$$

which is the denominator in Equation (2.10). And this is also a component of the likelihood function (Equation 2.8). When we combine Equations 2.8 and 2.11, it is possible to produce an unbiased estimator of the likelihood (Pitt et al., 2012).

We finally resample L values from the particles $\widetilde{x}_t^{(\ell)}$ with weights $\widetilde{\pi}_t^{(\ell)}$ to obtain a sample from $p(x_t|y_{1:t};\boldsymbol{\theta})$ which then restarts the procedure for time $t+1$.

2.2.2 Auxiliary Particle Filtering

Pitt & Shephard (1999) stated that "A fundamental problem with existing particle filters is that their mixture structure means that it is difficult to adapt the SIR, rejection, or MCMC sampling methods without greatly slowing the running of the filter." And to solve this problem they proposed the auxiliary particle filter (ASIR) which can be seen as a generalization of the SIR method showed in the previous section. The idea of the ASIR method is to sample from a higher dimension joint density with the aid of an auxiliary variable. Here we review the ASIR method.

We note that from equations (2.6) and (2.9),

$$p(x_t|y_{1:t};\boldsymbol{\theta}) \propto \sum_{\ell=1}^{L} p(y_t|x_t;\boldsymbol{\theta})p(x_t|x_{t-1}^{(\ell)};\boldsymbol{\theta})\pi_{t-1}^{(\ell)}. \tag{2.12}$$

Introducing an auxiliary variable ℓ which can be viewed as an index to the mixture in Equation (2.12), we will be able to adapt the particle filter in a more efficient way (Pitt & Shephard, 1999, pp. 592). This auxiliary variable is used only to help in simulation. Thus, the density we wish to approximate becomes:

$$p(x_t,\ell|y_{1:t};\boldsymbol{\theta}) \propto p(y_t|x_t;\boldsymbol{\theta})p(x_t|x_{t-1}^{(\ell)};\boldsymbol{\theta})\pi_{t-1}^{(\ell)}, \quad \text{for} \quad \ell = 1,\ldots,L, \tag{2.13}$$

such that

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

where

$$p(y_t|y_{1:t-1};\boldsymbol{\theta}) = \sum_{\ell=1}^{L}\int p(y_t|x_t;\boldsymbol{\theta})p(x_t|x_{t-1}^{(\ell)};\boldsymbol{\theta})dx_t\pi_{t-1}^{(\ell)}. \tag{2.14}$$

Now, if we are able to sample from $p(x_t,\ell|y_{1:t};\boldsymbol{\theta})$, then we can discard the sampled values of ℓ and get back to our filtering density (2.12). In most cases, we cannot do these integrals exactly.

Then, the next step is to sample from $p(x_t,\ell|y_{1:t};\boldsymbol{\theta})$ using sampling importance resampling algorithm. That is, we make K proposals $(x_t^{(k)},\ell^{(k)}), k = 1,\ldots,K$ from some proposal density $g(x_t,\ell|y_{1:t};\boldsymbol{\theta})$ and compute the weights

$$\widetilde{\pi}_t^{(k)} = \frac{1}{p_a(y_t|y_{1:t-1};\boldsymbol{\theta})} \times \frac{p(y_t|x_t^{(k)};\boldsymbol{\theta})p(x_t^{(k)}|x_{t-1}^{(\ell^{(k)})};\boldsymbol{\theta})\pi_{t-1}^{\ell^{(k)}}}{g(x_t^{(k)},\ell^{(k)}|y_{1:t};\boldsymbol{\theta})}. \tag{2.15}$$

Note from Equation (2.14) that the predictive function can be approximated by:

$$p(y_t|y_{1:t-1};\boldsymbol{\theta}) \approx p_a(y_t|y_{1:t-1};\boldsymbol{\theta}) \triangleq \sum_{k=1}^{K}\frac{p(y_t|x_t^{(k)};\boldsymbol{\theta})p(x_t^{(k)}|x_{t-1}^{(\ell^{(k)})};\boldsymbol{\theta})\pi_{t-1}^{\ell^{(k)}}}{g(x_t^{(k)},\ell^{(k)}|y_{1:t};\boldsymbol{\theta})}, \tag{2.16}$$

which in turn can be used to normalize the weights in Equation (2.15). Usually K and L are equal. When we combine Equations 2.8 and 2.16, it is possible to produce an unbiased estimator of the likelihood (Pitt et al., 2012).

We finally resample L values from the above sample to obtain a sample from $p(x_t|y_{1:t};\boldsymbol{\theta})$ corresponding to particles $x_t^{(\ell)}$ with weights $\pi_t^{(\ell)} \triangleq 1/L$. This gives us the approximation in equation (2.12) which then restarts the procedure for time $t+1$.

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

Generic Auxiliary Particle Filtering

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

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

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

Fully Adapted Particle Filtering

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

2.3 Adaptive Metropolis-Hastings

The Metropolis-Hastings (MH) algorithm (Hastings, 1970) is a Markov Chain Monte Carlo based method employed to generate random samples from a probability distribution. The idea is to generate a value from an auxiliary distribution and accept it with a given probability. This correction mechanism ensures the convergence of the chain to the distribution of interest (see, for example, Tierney (1994)).

Suppose the chain is in the iteration state $\boldsymbol{\theta}_{n-1}$ and a value $\boldsymbol{\theta}_n^p$ is generated from a proposed auxiliary distribution $q_n(\boldsymbol{\theta}|\boldsymbol{\theta}_{n-1})$. Note that the proposed distribution may depend on the current state of the chain. The new value $\boldsymbol{\theta}_n^p$ will be accepted with probability:

$$\alpha(\boldsymbol{\theta}_{n-1},\boldsymbol{\theta}_n^p) = \min\left\{1, \frac{p(\boldsymbol{\theta}_n^p)}{p(\boldsymbol{\theta}_{n-1})}\frac{q_n(\boldsymbol{\theta}_{n-1}|\boldsymbol{\theta}_n^p)}{q_n(\boldsymbol{\theta}_n^p|\boldsymbol{\theta}_{n-1})}\right\} ,$$

and take $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1}$ otherwise. Note that $p(\cdot)$ is the distribution of interest and, in a Bayesian context, $p(\cdot)$ can be the posterior density (see Section 2.1.1) in cases where it is difficult to generate its observations computationally, for example.

On the context of this thesis, adaptive sampling provides a way to initiate the MH algorithm with a proposal distribution that will have its parameters updated as the chains of the posterior densities are generated.

In other words, the parameters of the proposal density $q_n(\boldsymbol{\theta}|\boldsymbol{\theta}_{n-1})$ of the MH algorithm are estimated from the iterates $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n-1}$. Under appropriate regularity conditions the sequence of iterates $\boldsymbol{\theta}_n, n \geq 1$ converges to draws from the target distribution. See Haario et al. (2001), Roberts & Rosenthal (2007), Roberts & Rosenthal (2009), Giordani & Kohn (2010) and Silva et al. (2010) for more details.

Now, the problem lies in choosing which probability density function will be used as proposal. In the literature we can find some options. Amongst them the adaptive random walk Metropolis proposal of Roberts & Rosenthal (2009) and the adaptive independent Metropolis-Hastings density proposal by Giordani & Kohn (2010).

2.3.1 Adaptive Random Walk Metropolis

Here we will briefly describe the adaptive random walk Metropolis sampling (ARWMS) algorithm proposed by Roberts & Rosenthal (2009) which was based on the proposal of Haario et al. (2001). The proposed algorithm can be divided into two phases, the first one will take place until iteration n_0, defined by the researcher to start the algorithm, and the second one from iteration n_0 to n. In the applications of this thesis, all n_0 will be equal to 500.

For the first phase of the algorithm, the proposed distribution will be given by $q_n(\boldsymbol{\theta}|\boldsymbol{\theta}_{n-1}) = N(\boldsymbol{\theta}_{n-1}; (0.1)^2 I_d/d)$, where $N(\boldsymbol{\mu}; \boldsymbol{\Sigma})$ is a multivariate d-dimensional normal density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. I_d is a d-dimensional identity matrix.

And for the second phase, when $n > n_0$, the proposed distribution will be given by:

$$q_n(\boldsymbol{\theta}|\boldsymbol{\theta}_{n-1}) = (1-\beta)N(\boldsymbol{\theta}_{n-1}; (2.38^2)\boldsymbol{\Sigma}_n/d) + (\beta)N(\boldsymbol{\theta}_{n-1}; (0.1)^2 I_d/d), \tag{2.17}$$

where β is a small positive constant and, in this thesis, equals to 0.05. And $\boldsymbol{\Sigma}_n$ represents the covariance matrix estimated through the $n-1$ iterations.

The motivation for choosing these structures for $q_n(\boldsymbol{\theta}|\boldsymbol{\theta}_{n-1})$ can be seen in Roberts & Rosenthal (2009) and Silva et al. (2010) with more details. In general terms, the part with less variation (covariance matrix equal to $(0.1)^2 I_d/d$) is "...to avoid the algorithm getting stuck at problematic values." (Roberts & Rosenthal, 2009, pp. 351). The second part, with covariance matrix equal to $(2.38^2)\boldsymbol{\Sigma}_n/d$ is optimal in a multi-dimensional context, according to Roberts et al. (1997) and Roberts & Rosenthal (2001).

2.3.2 Adaptive Independent Metropolis Hastings

The adaptive independent Metropolis-Hastings sampling (AIMHS) method, proposed by Giordani & Kohn (2010) and discussed in Silva et al. (2010), for adaptive sampling can also be divided into two phases. For both phases, the proposed density is given by a mixture of four terms according to the equation below:

$$q_n(\boldsymbol{\theta}|\boldsymbol{\theta}_{n-1}) = \sum_{j=1}^{4} \beta_j q_j(\boldsymbol{\theta}|\boldsymbol{\lambda}_{jn}), \qquad \beta_j \geq 0, \quad \text{for} \quad j = 1, \ldots, 4 \quad \text{and} \quad \sum_{j=1}^{4} \beta_j = 1, \quad (2.18)$$

where $\boldsymbol{\lambda}_{jn}$ holds all parameters of the density $q_j(\boldsymbol{\theta}|\boldsymbol{\lambda}_{jn})$.

In the first phase, $q_1(\boldsymbol{\theta}|\boldsymbol{\lambda}_{1n})$ is given, if available, by a Laplace approximation of the posterior density (our density of intereset), otherwise is given by a Gaussian density constructed from several iterations of a method of adaptive random walk Metropolis, which will be use in all applications of this thesis. And $q_2(\boldsymbol{\theta}|\boldsymbol{\lambda}_{2n})$ has the same components of $q_1(\boldsymbol{\theta}|\boldsymbol{\lambda}_{1n})$, but with heavier tails, and particularly in this dissertation, its covariance matrices will be five times those of $q_1(\boldsymbol{\theta}|\boldsymbol{\lambda}_{1n})$.

Now, still in the first phase, $q_3(\boldsymbol{\theta}|\boldsymbol{\lambda}_{3n})$ carries the adaptive part of the proposal, being an estimate of the target density calculated through a method suggested by Giordani & Kohn (2010) of a normal mixing using k-harmonic means clustering. And finally, $q_4(\boldsymbol{\theta}|\boldsymbol{\lambda}_{4n})$ is a version of $q_3(\boldsymbol{\theta}|\boldsymbol{\lambda}_{3n})$ with heavy tails, more specifically its covariance component has ten times that of $q_3(\boldsymbol{\theta}|\boldsymbol{\lambda}_{3n})$.

However, this phase begins with β_3 and β_4 being equal to zero until a sufficient number of iterations is reached to obtain $q_3(\boldsymbol{\theta}|\boldsymbol{\lambda}_{3n})$ and, consequently, $q_4(\boldsymbol{\theta}|\boldsymbol{\lambda}_{4n})$. As suggested by Silva et al. (2010), we will start this phase with $\beta_1 = 0.8$ and $\beta_2 = 0.2$, then we will use $\beta_1 = 0.15$, $\beta_2 = 0.05$ $\beta_3 = 0.7$ and $\beta_4 = 0.1$.

And finally, in the second phase, $q_1(\boldsymbol{\theta}|\boldsymbol{\lambda}_{1n})$ is defined as the last form assumed by $q_3(\boldsymbol{\theta}|\boldsymbol{\lambda}_{3n})$ in the first phase. The densities $q_2(\boldsymbol{\theta}|\boldsymbol{\lambda}_{2n})$ and $q_4(\boldsymbol{\theta}|\boldsymbol{\lambda}_{4n})$ are constructed in the same manner as the first phase and $q_3(\boldsymbol{\theta}|\boldsymbol{\lambda}_{3n})$ is maintained (only in the first iteration of the second phase).

Note that, as seen in the ARWM method, we also have parts of the proposal density structure with less variation (lighter tails) and others with greater variation. This type of structure helps to explore the entire sample space of the posterior density and, at the same time, prevent the algorithm from being stuck.

It is worthwhile to point out that it is computationally expensive to update $q_3(\boldsymbol{\theta}|\boldsymbol{\lambda}_{3n})$ and, consequently, $q_4(\boldsymbol{\theta}|\boldsymbol{\lambda}_{4n})$ on every iteration. To minimize this problem, updates are done through a block scheme. That in turn enables us to draw candidate values from the proposal and to evaluate both the proposal and the target densities in blocks using parallel computing. After each sampling block is completed, we then proceed with the usual Metropolis-Hastings method and update the proposal distribution before a new sampling block is started. For more details on the update scheme, see Silva et al. (2010).

2.3.3 Adaptive Sampling with Simulated Likelihood

We know from the beginning of Section 2.3 that $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ and this proportionality relation holds for the cases where the likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$ may be calculated exactly. Nonetheless, Andrieu et al. (2010) showed that Markov Chain Monte Carlo samplers still converge to the correct posterior density when we use, instead of exactly likelihood functions, the simulated ones (Equations (2.11) and (2.16)) obtained by particle filters with finite number of particles.

The simulated likelihood via particle filter algorithms may be seen as the density of $\boldsymbol{y}|\boldsymbol{\theta}, u$, where u is a set of auxiliary variables that are not function of $\boldsymbol{\theta}$ such that $p(y_t|y_{1:t-1};\boldsymbol{\theta};u)$ is equal to $p_s(y_t|y_{1:t-1};\boldsymbol{\theta})$ or $p_a(y_t|y_{1:t-1};\boldsymbol{\theta})$ obtained from Equations (2.11) and (2.16), respectively. The unbiasedness of the simulated likelihood may be verified (See Pitt et al. (2012) and Del Moral (2004) for more details on this property).

Now, additionally, let $p(y_t|y_{1:t-1};\boldsymbol{\theta};u)$ obtained from the particle filter be the estimate of $p(y_t|y_{1:t-1};\boldsymbol{\theta})$. Then $\widehat{p}(\boldsymbol{y}|\boldsymbol{\theta};u) = \prod_{t=1}^{T} p(y_t|y_{1:t-1};\boldsymbol{\theta};u)$ is the unbiased estimate of the likelihood given by $p(\boldsymbol{y}|\boldsymbol{\theta})$.

2.4 Model Selection

2.4.1 Estimating the Marginal Likelihood

Marginal likelihoods are often used to compare two or more models and can be seen as the probability of the data given the model type. Thus the higher its value, the more adjusted is the model to the data set.

For a given model, let $\boldsymbol{\theta}$ be the vector of model parameters, $p(\boldsymbol{y}|\boldsymbol{\theta})$ the likelihood of the observations $\boldsymbol{y}$ and $p(\boldsymbol{\theta})$ the prior for $\boldsymbol{\theta}$. The marginal likelihood is defined as

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.19}$$

Suppose that $q(\boldsymbol{\theta})$ is an approximation to $p(\boldsymbol{\theta}|y)$ which can be evaluated explicitly. Bridge sampling (Meng & Wong, 1996) estimates the marginal likelihood as follows. Let

$$t(\boldsymbol{\theta}) = \left(\frac{p(y|\boldsymbol{\theta})p(\boldsymbol{\theta})}{U} + q(\boldsymbol{\theta})\right)^{-1},$$

where U is a positive constant. Let

$$\begin{aligned} A &= \int t(\boldsymbol{\theta})q(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}\ . \qquad \text{Then,} \\ A = \frac{A_1}{p(\boldsymbol{y})} \qquad \text{where} \qquad A_1 &= \int t(\boldsymbol{\theta})q(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}\ . \end{aligned} \tag{2.20}$$

Suppose the sequence of iterates $\{\boldsymbol{\theta}^{(j)}, j = 1, \ldots, M\}$ is generated from the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$ and a second sequence of iterates $\{\tilde{\boldsymbol{\theta}}^{(k)}, k = 1, \ldots, M\}$ is generated from $q(\boldsymbol{\theta})$. Then

$$\widehat{A} \;=\; \frac{1}{M}\sum_{j=1}^{M} t(\boldsymbol{\theta}^{(j)})q(\boldsymbol{\theta}^{(j)}), \quad \widehat{A}_1 = \frac{1}{M}\sum_{k=1}^{K} t(\tilde{\boldsymbol{\theta}}^{(k)})p(y|\tilde{\boldsymbol{\theta}}^{(k)})p(\tilde{\boldsymbol{\theta}}^{(k)}) \quad \text{and} \quad \widehat{p}_{BS}(y) = \frac{\widehat{A}_1}{\widehat{A}}$$

are estimates of A and A_1, respectively, while $\widehat{p}_{BS}(y)$ is the bridge sampling estimator of the marginal likelihood $p(\boldsymbol{y})$.

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

An alternative method to estimate of the marginal likelihood $p(\boldsymbol{y})$ is to use importance sampling based on the proposal distribution $q(\boldsymbol{\theta})$ (Geweke, 1989; Chen & Shao, 1997). That is,

$$\widehat{p}_{IS}(\boldsymbol{y}) = \frac{1}{K}\sum_{k=1}^{K}\frac{p(\boldsymbol{y}|\boldsymbol{\theta}^{(k)})p(\boldsymbol{\theta}^{(k)})}{q(\boldsymbol{\theta}^{(k)})}.$$

Since the proposal distributions have at least one heavy tailed component, the importance sampling ratios are likely to be bounded and well-behaved, as it is in several examples in the literature. It is noteworthy that all calculations were optimized for better computational performance.

2.4.2 Likelihood Based Information Criteria

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

Let $p(\boldsymbol{y}|\boldsymbol{\theta}_\ell, \mathcal{M}_\ell)$ be the likelihood function for model $\mathcal{M}_\ell$ and $D(\Psi_\ell) = -2\log p(\boldsymbol{y}|\boldsymbol{\theta}_\ell, \mathcal{M}_\ell)$. The Akaike information criterion (AIC), the Bayesian information criterion (BIC), their expected versions, EAIC and EBIC, and the deviance information criterion (DIC) are defined as (Spiegelhalter et al., 2002)

$$AIC(\mathcal{M}_\ell) = D(E[\boldsymbol{\theta}_\ell|\boldsymbol{y}, \mathcal{M}_\ell]) + 2d_\ell, \tag{2.21a}$$

$$BIC(\mathcal{M}_\ell) = D(E[\boldsymbol{\theta}_\ell|\boldsymbol{y}, \mathcal{M}_\ell]) + \log(n)d_\ell, \tag{2.21b}$$

$$EAIC(\mathcal{M}_\ell) = E[D(\boldsymbol{\theta}_\ell)|\boldsymbol{y}, \mathcal{M}_\ell] + 2d_\ell, \tag{2.21c}$$

$$EBIC(\mathcal{M}_\ell) = E[D(\boldsymbol{\theta}_\ell)|\boldsymbol{y}, \mathcal{M}_\ell] + \log(n)d_\ell, \quad \text{and} \tag{2.21d}$$

$$DIC(\mathcal{M}_\ell) = 2E[D(\boldsymbol{\theta}_\ell)|\boldsymbol{y}, \mathcal{M}_\ell] - D(E[\theta_\ell|\boldsymbol{y}, \mathcal{M}_y]), \tag{2.21e}$$

respectively, with d_ℓ representing the number of parameters of the $\mathcal{M}_\ell$ model and n the number of observations in the data.

The draws from $\boldsymbol{\theta}_\ell^{(j)}$ and $\log p(\boldsymbol{y}|\boldsymbol{\theta}_\ell^{(j)})$, $j = 1, \ldots, M$, can be used to approximate to $E[D(\boldsymbol{\theta}_\ell)|\boldsymbol{y}, \mathcal{M}_\ell]$ and $E[\boldsymbol{\theta}_\ell|\boldsymbol{y}, \mathcal{M}_\ell]$ by $M^{-1}\sum_{j=1}^M D(\boldsymbol{\theta}_\ell^{(j)})$ and $M^{-1}\sum_{j=1}^M \boldsymbol{\theta}_\ell^{(j)}$, respectively. Finally, approximations to DIC, AIC, BIC, EAIC and EBIC can be easily derived.

Using AIC, it is accepted that among the evaluated models none is considered what really describes the relation between the dependent variable and the explanatory variables, or the "true model" and then, it is tried to choose the model that minimizes the divergence between the real unknown model and the adjusted one. On the other hand, with BIC it is implicit that there is a model that describes the relationship between the variables involved and the criterion tries to maximize the probability of choosing the true model. And, finally, DIC is a generalization of AIC, but estimates the number of model parameters instead of counting them.

The estimation $\hat{d}_\ell$ of the number of model parameters on the DIC criteria is given by $\hat{d}_\ell = E[D(\boldsymbol{\theta}_\ell)|\boldsymbol{y}, \mathcal{M}_\ell] - D(E[\theta_\ell|\boldsymbol{y}, \mathcal{M}_y])$. Thus, DIC may be rewriten as $DIC = D(E[\theta_\ell|\boldsymbol{y}, \mathcal{M}_y]) + 2\hat{d}_\ell$.

Note that given a set of models for a given data set, the lower the value found for all the criteria presented in this section the better the model. Therefore, we consider the best adjusted model for these criteria, the one with the lowest AIC, BIC, EAIC, EBIC and DIC values. For more details on these information criteria, see Spiegelhalter et al. (2002), including the discussions and references therein.

2.5 Modelling Volatility

2.5.1 Log-returns

Most financial studies focus on the analysis of returns series rather than the use of asset prices series. The reason we use a series of returns has two factors, the returns information serves the interests of investors and has more interesting statistical properties than the price series.

Thus, let y_t be the price of an asset at time t, the return R_t between the times $t-1$ and t is given by:

$$R_t = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1 \Rightarrow 1 + R_t = \frac{y_t}{y_{t-1}} \tag{2.22}$$

And the log-return at time t is given by:

$$r_t = \ln(1 + R_t) = \ln\left(\frac{y_t}{y_{t-1}}\right) = \ln(y_t) - \ln(y_{t-1}), \tag{2.23}$$

Which will be used in our applications. For more details about returns and log-returns, see Appendix A.

2.5.2 Generalized Autoregressive Conditionally Heteroscedastic Model with Noise

A generalized autoregressive conditionally heteroscedastic (GARCH) model (Engle, 1982; Bollerslev, 1986) is used to model the variance of a time series using values of the past squared means of the observations and past variances. The observation and system equations of the GARCH(1,1) model with noise is given by:

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

where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with μ and σ as location and scale parameters, respectively.

This model has the following restrictions on the parameters: $\sigma^2 > 0$, $\beta_j > 0$ for $j = 0, 1, 2$ and $\beta_1 + \beta_2 < 1$ (stationary condition). Thus, we assume the following prior distribution:

$$
\begin{aligned}
\sigma^2 &\sim \mathcal{HN}(a_1^2) \\
\beta_0 &\sim \mathcal{HN}(a_2^2) \\
(\beta_1, \beta_2) &\sim \mathcal{U}(\beta_1,\, \beta_2),\ \beta_1 > 0,\ \beta_2 > 0,\ \beta_1 + \beta_2 < 1 \\
\tau_0^2 &= \beta_0/(1 - \beta_1 - \beta_2) \\
x_0 &\sim \mathcal{N}(0, \tau_0^2),
\end{aligned}
$$

where $\mathcal{HN}(a_n^2)$, $i = 1, 2$, is a half-normal distribution with location parameter set to 0 and a_n, $i = 1, 2$, as the scale parameter. Remember that in this distribution the location parameter is different from the mean; and $\mathcal{U}(\beta_1,\, \beta_2)$ is a bivariate continuous uniform distribution.

Several algebraic developments are required to perform the fully adapted particle filter for this model and further details on this matter can be found in the Appendix B.

2.5.3 Stochastic Volatility Model and Its Variants

The observation and system equations of the stochastic volatility (SV) model is given by (Jacquier et al., 1994):

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

and ϵ_t is the observation error with mean 0 and variance 1.

First we consider that ϵ_t and ω_t are independent, with ω_t distributed as a normal distribution with mean $zero$ and variance τ^2 ($\mathcal{N}(0, \tau^2)$) and ϵ_t can be distributed, on this thesis, as a standard normal distribution, $\mathcal{N}(0, 1)$, a standard skew normal distribution, denoted as $\mathcal{SN}(\lambda, 0, 1)$, a

t distribution with 3 degrees of freedom $(t(3))$ or a skew t distribution also with 3 degrees of freedom, denoted as $\mathcal{S}t(\lambda, 3)$, where λ is a parameter of skewness.

But ϵ_t and ω_t may have a structure of correlation, then its distribution is given by a bivariate normal as follows

$$\begin{pmatrix} \epsilon_t \\ \omega_t \end{pmatrix} \Bigg| \rho, \tau \sim \mathcal{N}\left[0, \begin{pmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{pmatrix}\right].$$

and this particular case of the model is known as a stochastic volatility model with leverage. Notes on the explicit form of the observation and system equations of this case can be found in Appendix C.

The SV model, in general, has the following restrictions on the parameters: $\tau^2 > 0$ and $0 < \phi < 1$. Thus, we assume the following prior distribution:

$$\begin{aligned} \tau^2 &\sim \mathcal{IG}(a_1, b_1) \\ \phi &\sim Beta(a_2, b_2) \\ \alpha &\sim \mathcal{N}(a_3, b_3^2) \\ \lambda &\sim \mathcal{N}(a_4, b_4^2) \\ x_0 &\sim \mathcal{N}(\alpha, \tau^2/(1+\phi^2)). \end{aligned}$$

where $\mathcal{IG}(a_1, b_1)$ is the inverse-gamma distribution with a_1 and b_1 as shape and scale parameters; and $Beta(a_2, b_2)$ is the beta distribution with a_2 and b_2 as shape parameters.

Chapter 3

APPLICATIONS

First, in Section 3.1, we will carry out a simulation study that will consist of estimating the parameters of five simulated series for GARCH(1,1) model with noise, stochastic volatility model with Gaussian noise and stochastic volatility model with leverage. Each simulated series will have 1.000 observations.

Next, in Section 3.2, we will estimate the parameters of some models for daily log-return data of three stock market indexes worldwide from January 2012 to March 2016 resulting in a time series with more than one thousand observations each. Thus, from now on, read log-return every time we say return.

The stock market indexes colected were the stock market index of about 50 stocks that are traded on the São Paulo, Mercantile & Futures Exchange, known as BOVESPA index or simply IBOVESPA; the stock market index of the common stocks and similar securities listed on the NASDAQ stock market, known as NASDAQ Composite; and the stock market index based on the market capitalizations of 500 large companies having common stock listed on the New York Stock Exchange or the NASDAQ stock market, known as Standard and Poor's 500 or only S&P500.

Note that for GARCH model we are using it is possible to estimate the likelihood by applying both particle filters presented in Sections 2.2.1 and 2.2.2 (SIR and ASIR). In addition, for this model, it is possible to apply the fully adapted auxiliary particle filter, as shown in Appendix B.

However, for SV models it is not possible to use a fully adapted particle filter, so we tried to apply the generic version of ASIR, but the estimates, when found, were bad and the computational time spent extremely high. Therefore, we have chosen to apply only SIR method for these models.

In adittion, to determine the number of particles of the filters, we use the methodology proposed by Pitt et al. (2012) which states that "...the computing time is minimized by choosing N [the number of partciles] so that the standard deviation of the log-likelihood, at a posterior central value, is around 0.92." Moreover, for all applications using ASIR, the number of particles for both steps (K and L) will be equal.

In our applications, we set $a_1 = a_2 = 10$ as parameters for GARCH with noise models prior distributions (Section 2.5.2) and $a_1 = b_1 = a_2 = b_2 = 1$, $a_3 = a_4 = 0$ and $b_3 = b_4 = 1.0 \times 10^6$ as parameters for SV models prior distributions (Section 2.5.3). That is, all prior distributions used are non-informative.

The estimation strategy used in all datasets and models was to first create an initial estimate for the parameters and covariance matrix through the adaptive random walk Metropolis sampling (ARWMS) method (see Section 2.3.1 for the particular strategy used to this sampling technique), and then use them as initial values in the adaptive independent Metropolis-Hastings sampling (AIMHS) algorithms (see Section 2.3.2 for the particular strategy used to this sampling technique). Therefore, all results shown below in this Chapter refer only for the last part, that is when we use AIMHS.

It is worth highlighting that the analyzes were done on different computers and therefore the computational time spent will not be mentioned for any of the algorithms used.

3.1 Simulation

The main objective of this simulation study is to verify if the algorithm is actually estimating the true parameters. For this, we will generate 5 time series of the intended models and observe if the estimation is being done correctly.

The parameters used to generate the data of GARCH(1,1) model with noise were $\sigma^2 = 0.00009$, $\beta_0 = 0.000002$, $\beta_1 = 0.15$ and $\beta_2 = 0.84$. For the stochastic volatility model with Gaussian noise were $\tau^2 = 0.20$, $\alpha = -9.6$ and $\phi = 0.84$. And for the stochastic volatility model with leverage were $\tau^2 = 0.11$, $\alpha = -11$, $\phi = 0.98$ and $\rho = -0.7$.

To obtain the results of this section, we run all adaptive MCMC algorithms with 50.000 iterations with the first half of them being discarded for the calculation of the final estimates. In addition, as explained at the beginning of this chapter, we first generate an initial estimate of the parameters and covariance matrix using the ARWMS method to use them as initial values in AIMHS method.

In Tables 3.1, 3.2, 3.3 and 3.4 we can observe the posterior mean, median, standard deviation and credibility interval of 95% (CI$_{0.025}$ and CI$_{0.975}$ are, respectively, the lower limit and upper limit of the interval) for the parameters of each model.

In order to obtain the estimates of the GARCH(1,1) model with noise parameters using SIR filter (Table 3.1), 3.000 particles were used for the preliminary part (ARWMS method) and 2.000 in the final part. Notice that for all replicas, all credibility intervals contain the true parameters values, which indicates satisfactory behaviour of the estimators.

Table 3.1: Posterior mean, median, standard deviation and credibility interval for the parameters of GARCH(1,1) model with noise using SIR filter

Replica	Parameters	Posterior estimations				
		Mean	Std. dev.	$CI_{0.025}$	Median	$CI_{0.975}$
1	σ^2	0.0000630	0.0000258	0.0000076	0.0000637	0.0001097
	β_0	0.0000037	0.0000020	0.0000009	0.0000033	0.0000090
	β_1	0.1463449	0.0453058	0.0790899	0.1408561	0.2472860
	β_2	0.8290534	0.0459327	0.7252073	0.8343345	0.8992350
2	σ^2	0.0000883	0.0000253	0.0000319	0.0000910	0.0001317
	β_0	0.0000061	0.0000053	0.0000010	0.0000046	0.0000220
	β_1	0.2257102	0.1003751	0.0829156	0.2067775	0.4762093
	β_2	0.7226548	0.1125832	0.4477203	0.7420469	0.8822469
3	σ^2	0.0000529	0.0000269	0.0000059	0.0000524	0.0001051
	β_0	0.0000037	0.0000023	0.0000008	0.0000033	0.0000100
	β_1	0.0905114	0.0390999	0.0337985	0.0833210	0.1868716
	β_2	0.8807635	0.0440161	0.7734476	0.8890788	0.9410084
4	σ^2	0.0000546	0.0000325	0.0000045	0.0000549	0.0001182
	β_0	0.0000049	0.0000025	0.0000009	0.0000045	0.0000093
	β_1	0.1269757	0.0345311	0.0762891	0.1201576	0.2073463
	β_2	0.8506429	0.0345098	0.7679393	0.8602096	0.9070856
5	σ^2	0.0000718	0.0000213	0.0000256	0.0000741	0.0001085
	β_0	0.0000032	0.0000021	0.0000008	0.0000027	0.0000088
	β_1	0.1865094	0.0585892	0.0933376	0.1799826	0.3186721
	β_2	0.8007600	0.0591066	0.6671012	0.8084501	0.8941634

However, to obtain the estimates of the GARCH(1,1) model with noise parameters using ASIR filter, 50 particles were used for both the preliminary and final parts. Note that this algorithm is much more efficient than the one using SIR filter, because it uses much fewer particles (saves computational time) and obtains results as satisfactory as the previous one, as can be verified in Table 3.2.

Now, in Table 3.3, we can see the results of the estimations made for the parameters of the

Table 3.2: Posterior mean, median, standard deviation and credibility interval for the parameters of GARCH(1,1) model with noise using ASIR filter

Replica	Parameters	Posterior estimations				
		Mean	Std. dev.	$CI_{0.025}$	Median	$CI_{0.975}$
1	σ^2	0.0000622	0.0000253	0.0000107	0.0000633	0.0001090
	β_0	0.0000038	0.0000020	0.0000010	0.0000034	0.0000086
	β_1	0.1473738	0.0469848	0.0755105	0.1413896	0.2562844
	β_2	0.8276800	0.0473448	0.7176016	0.8336273	0.9008290
2	σ^2	0.0000904	0.0000251	0.0000344	0.0000935	0.0001340
	β_0	0.0000057	0.0000051	0.0000010	0.0000042	0.0000192
	β_1	0.2363900	0.1041273	0.0817055	0.2210820	0.4786691
	β_2	0.7152533	0.1143224	0.4405373	0.7351936	0.8782488
3	σ^2	0.0000514	0.0000279	0.0000034	0.0000505	0.0001068
	β_0	0.0000038	0.0000024	0.0000007	0.0000032	0.0000099
	β_1	0.0903956	0.0400999	0.0364844	0.0820313	0.1903948
	β_2	0.8812510	0.0437155	0.7677189	0.8886806	0.9442178
4	σ^2	0.0000585	0.0000297	0.0000050	0.0000585	0.0001154
	β_0	0.0000045	0.0000026	0.0000009	0.0000041	0.0000108
	β_1	0.1307535	0.0370271	0.0707828	0.1264888	0.2162167
	β_2	0.8482669	0.0385817	0.7639958	0.8517576	0.9141380
5	σ^2	0.0000718	0.0000209	0.0000262	0.0000734	0.0001082
	β_0	0.0000032	0.0000021	0.0000008	0.0000027	0.0000085
	β_1	0.1895439	0.0600698	0.0929562	0.1821909	0.3298645
	β_2	0.7979995	0.0612526	0.6584258	0.8055708	0.8956539

SV model with Gaussian noise using SIR filter. For this model, we use 350 particles in both algorithms. Notice that all credibility intervals also contain the true values of the parameters.

Finally, the results presented in Table 3.4 refer to the SV model with leverage using SIR filter estimations. Here, we use 1.750 particles for both ARWMS and AIMHS algorithms. Again, all credibility intervals contain the true values of the parameters.

This simulation study is not exhaustive for all the models that will be used for real data, however, in general, it can be seen that the algorithms are able to recover the true values of the

Table 3.3: Posterior mean, median, standard deviation and credibility interval for the parameters of SV model with Gaussian noise using SIR filter

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

parameters and it is expected that the same behaviour will be found in all other models.

3.2 Real data

3.2.1 Describing Data

In this section we make a descriptive analysis of the data used in the models. It is worth emphasizing that descriptive statistics do not intend to generalize their results to the population, that is, to make any kind of inference, but only to describe and summarize the data in a way that facilitates their understanding.

First, we want to have a brief notion of how the series behaves. As might be expected (see Section 2.5.1), all means are almost equal to zero, although, unlike IBOVESPA, NASDAQ and S&P500 have shown a positive return in the long run (mean greater than zero). In addition,

Table 3.4: Posterior mean, median, standard deviation and credibility interval for the parameters of SV model with leverage using SIR filter

Replica	Parameters	Posterior estimations				
		Mean	Std. dev.	$CI_{0.025}$	Median	$CI_{0.975}$
1	τ^2	0.1300850	0.0208358	0.0940193	0.1280196	0.1747818
	ϕ	0.9852177	0.0061611	0.9722945	0.9857468	0.9956883
	α	-10.8505564	1.0132163	-13.4711188	-10.6496241	-9.4863125
	ρ	-0.7031984	0.0619436	-0.8091877	-0.7079091	-0.5718210
2	τ^2	0.1167414	0.0187372	0.0850963	0.1149657	0.1584491
	ϕ	0.9820628	0.0059646	0.9697027	0.9823247	0.9928902
	α	-10.8146515	0.5032284	-11.9443184	-10.7769132	-9.9164817
	ρ	-0.7130920	0.0630058	-0.8197065	-0.7176833	-0.5789504
3	τ^2	0.1403466	0.0216788	0.1036933	0.1376562	0.1888396
	ϕ	0.9813912	0.0056285	0.9697771	0.9815278	0.9927614
	α	-10.0719730	0.6470769	-11.1250038	-10.1310744	-8.6076082
	ρ	-0.6913779	0.0585517	-0.7884205	-0.6965849	-0.5628501
4	τ^2	0.1404871	0.0204211	0.1057169	0.1382987	0.1849773
	ϕ	0.9823382	0.0047879	0.9726977	0.9824653	0.9913064
	α	-10.9219404	0.4967491	-11.9802993	-10.8709957	-10.0620430
	ρ	-0.7975224	0.0453794	-0.8753869	-0.8017720	-0.6978566
5	τ^2	0.0928752	0.0162011	0.0666495	0.0915056	0.1293698
	ϕ	0.9679708	0.0096341	0.9468060	0.9687180	0.9846366
	α	-10.8486145	0.2664703	-11.3648046	-10.8537918	-10.3201966
	ρ	-0.6733129	0.0858376	-0.8204434	-0.6786633	-0.4914737

all the means indexes return are approximately equal to their respective medians, which could indicate lack of asymmetry in the data. However IBOVESPA has a slightly positive skewness unlike the other two. It is worth noticing that the skewness measurement for the Gaussian distribution is equal to zero, that is, the further away from this value, the greater the asymmetry of the distribution (both positive and negative, depending on the signal).

Moreover, by looking only at the standard deviation, we could erroneously conclude that all series have an approximately equally concentration around the mean, but by the coefficient of variation (standard deviation divided by the mean), we find that the log-returns from IBOVESPA have higher variance than the others, i.e., heavier tails than NASDAQ Composite and S&P500 return series. This fact was confirmed by the kurtosis, which indicated almost double the value of IBOVESPA for the other ones. The kurtosis measurement we are using here is also equal to zero for the Gaussian distribution, so, despite the comparisons, all the series presented here have a leptokurtic distribution and, therefore, presents evidence of non-normality.

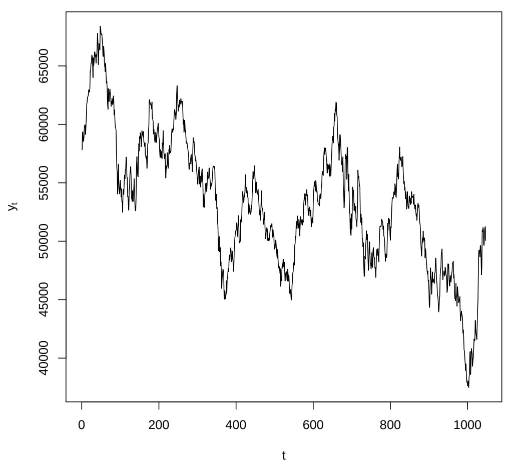
Table 3.5: Descriptive statistics of the indexes return series.

Statistics	Indexes		
	IBOVESPA	NASDAQ	S&P500
Mean	-0.00014	0.00057	0.00045
Standard Deviation	0.01481	0.00958	0.00830
Coefficient of variation	-107.3296	16.76162	18.40781
Median	-0.00096	0.00088	0.00049
Skewness	0.24131	-0.32704	-0.24698
Kurtosis	0.75280	1.44709	1.66029

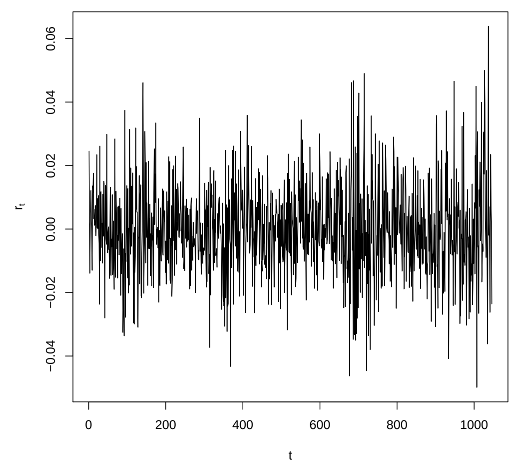
It is now interesting to perform a graphical analysis of the data. This type of analysis strengthens the information obtained in the descriptive analysis, as well as providing new information that can be of great help in investigating the data. The graph of the price series, for example, helps to verify the trend of the data, on the other hand through the graph of the return series it is possible to empirically observe the stationarity of r_t (see Section 2.5.1).

Furthermore, returns and squared returns autocorrelation graphs would help to identify whether it is really justifiable to apply models for volatility in these series. Absence of correlation for r_t indicates that the series does not have autocorrelation in the mean and, therefore, it would not be necessary to model this parameter. The presence of autocorrelation in r_t^2 would indicate volatility in the series, confirming the applicability of volatility models.

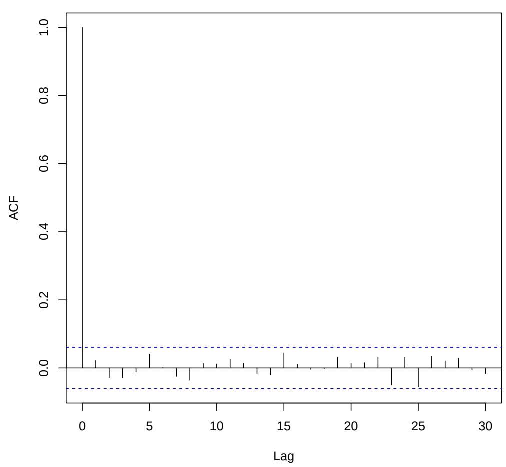
Analyzing Figure 3.1, which is the same present in Chapter 1, we can see, through Graph 3.1a, that IBOVESPA price series has several times of downturn followed by recovery periods, but that is not enough to recover the price in the long run. This instability seems to be reflected in the return series (Figure 3.1b), notice that there are moments of greater variation followed by others of lesser and vice-versa.



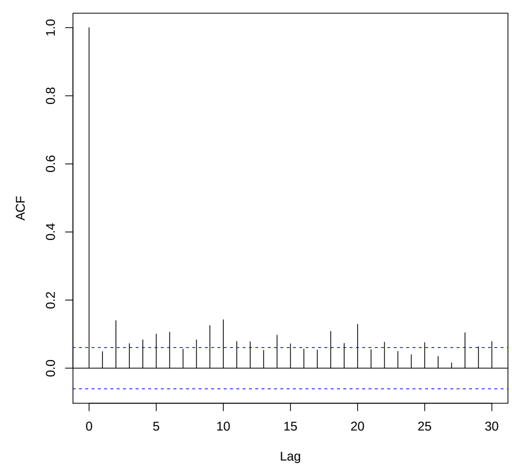
(a) Price series (y_t).



(b) Return series (r_t).



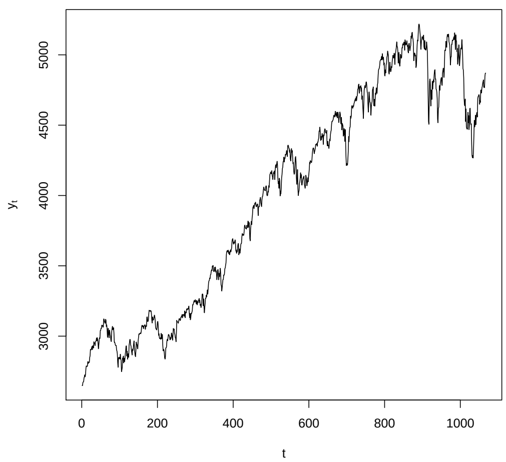
(c) Autocorrelation of r_t series.



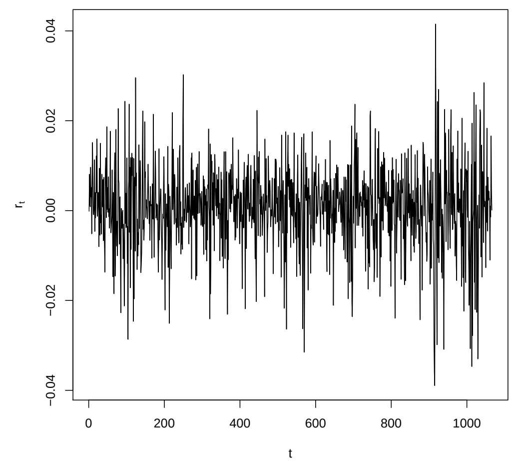
(d) Autocorrelation of r_t^2 series.

Figure 3.1: IBOVESPA price and log-return series on the top and the autocorrelation for the log-returns and for the squared log-returns series on the bottom. In Figures **??** and **??**, $t = 200$ corresponds to october 2012, $t = 400$ to july 2013, $t = 2 = 600$ to may 2014, $t = 800$ to february 2012 and $t = 1000$ to december 2015.

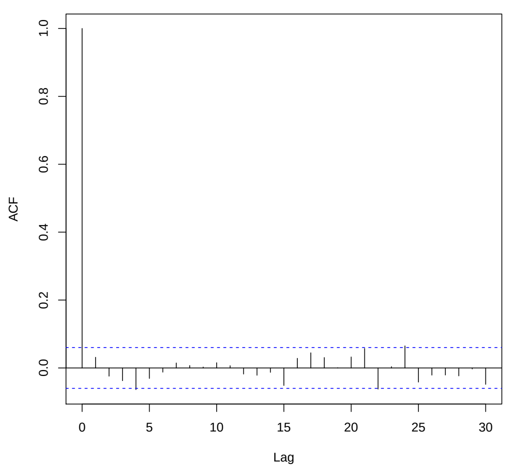
Moreover, it should be noted that returns do not show any significant autocorrelation (Figure 3.1c) whereas the squared returns have several points of autocorrelation (Figure 3.1d). Significant autocorrelation in r_t^2 shows us the volatility clustering effect for this series and this indicates that volatility models are suitable for these data.



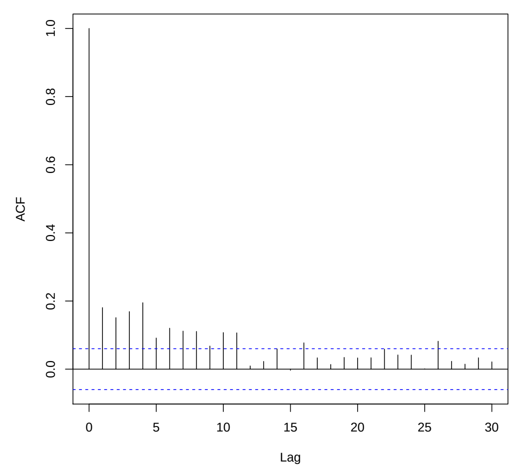
(a) Price series (y_t.)



(b) Return series (r_t).



(c) Autocorrelation of r_t serie.



(d) Autocorrelation of r_t^2 series.

Figure 3.2: NASDAQ price and log-return series on the top and the autocorrelation for the log-returns and for the squared log-returns series on the bottom. In Figures 3.2a and 3.2b, $t = 200$ corresponds to october 2012, $t = 400$ to july 2013, $t = 2 = 600$ to may 2014, $t = 800$ to february 2012 and $t = 1000$ to december 2015.

The price series of NASDAQ and S&P500 indexes differ from IBOVESPA price series mainly because of the long-term trend, as indicated by the means in the descriptive analysis (Table 3.5). Graphs 3.2a and 3.3a clearly show a positive trend, unlike that shown by IBOVESPA price series. In addition, also going in the opposite direction to that presented by IBOVESPA price series, NASDAQ and S&P500 price series appear to have more stability over time.

However, similar conclusions to those obtained through IBOVESPA returns can be taken from the graphs presented for the NASDAQ and S&P500 return series. That is, despite of apparently having less volatility, NASDAQ and S&P500 return series also seem to have moments of greater variation followed by others of lesser and vice-versa (see Figures 3.2b and 3.3b), which appears to be a clue to the leverage effect phenomenom.

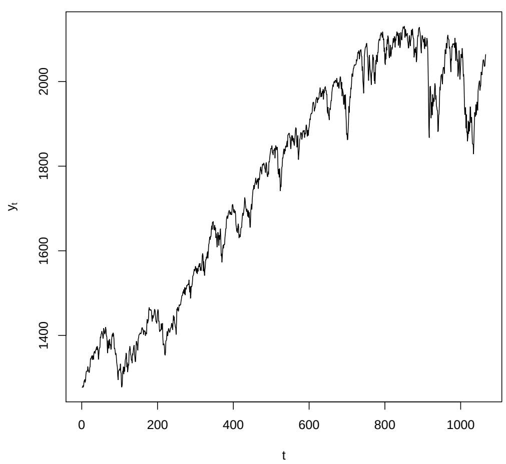
Besides, the autocorrelation of the return and squared return series for NASDAQ and S&P500 indexes also present a similar result to those presented for the IBOVESPA return series (see Figures 3.2c, 3.2d, 3.3c and 3.3d). That is, they do not show autocorrelation in r_t, but they have significant autocorrelation in r_t^2, which indicates the presence of volatility clustering, property of this type of series (Granger & Ding, 1995). And, in addition, this property indicates that models of volatility are suitable for these series.

Next, we apply a few volatility models to these data sets using adaptive Metropolis-Hastings algorithms combined with particle filters methods. We also do model comparisons by means of likelihood-based information criteria and marginal likelihoods.

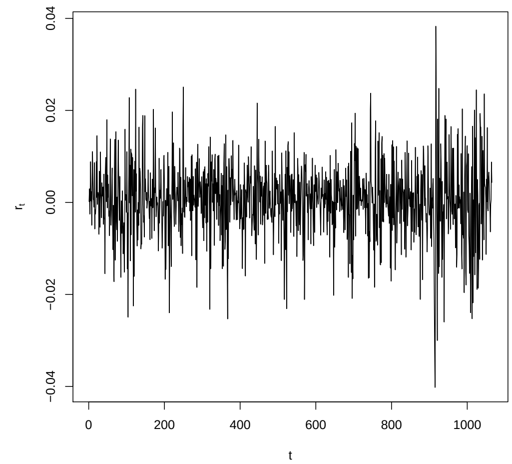
3.2.2 Model Comparisons

In order to perform the model comparisons, we initially estimated the parameters for our data sets (IBOVESP, NASDAQ and S&P500 return series) using GARCH and SV models presented in Sections 2.5.2 and 2.5.3. We run all adaptive MCMC algorithms with 200.000 iterations with the first half of them being discarded for the calculation of final estimates. In addition, as explained at the beginning of this chapter, we first generate an initial estimate of the parameters and covariance matrix using the ARWMS method to use them as initial values in AIMHS method.

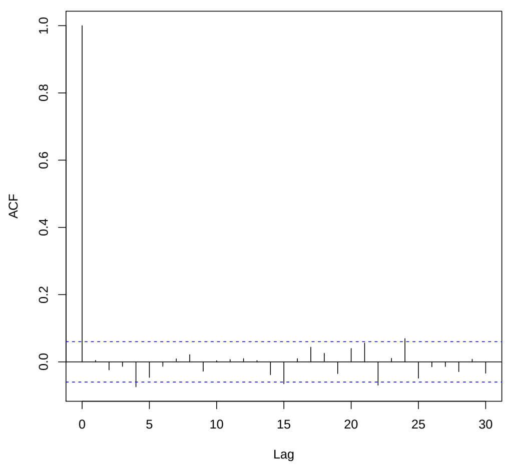
Furthermore, recall that we are interested in the results when we use AIMHS algorithm. Thus all model comparisons by means of likelihood-based information criteria and marginal likelihoods were calculated only when we use AIMHS. Therefore, the results presented in the following tables refer to this kind of adaptive sampling. The values in bold refer to the model most adjusted to the data according to the respective selection criteria. Moreover, the posterior estimations will



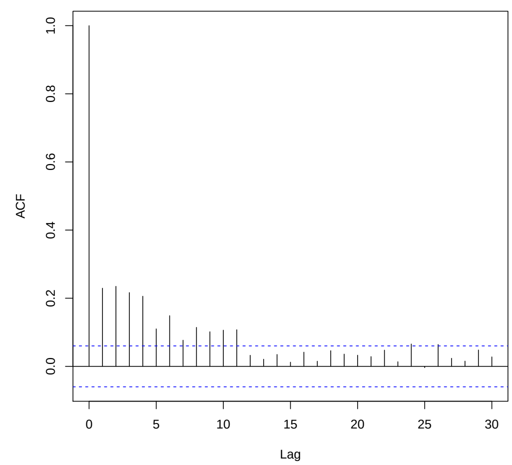
(a) Price series (y_t).



(b) Return series (r_t).



(c) Autocorrelation of r_t series.



(d) Autocorrelation of r_t^2 series.

Figure 3.3: S&P500 price and log-return series on the top and the autocorrelation for the log-returns and for the squared log-returns series on the bottom. In Figures 3.3a and 3.3b, $t = 200$ corresponds to october 2012, $t = 400$ to july 2013, $t = 2 = 600$ to may 2014, $t = 800$ to february 2012 and $t = 1000$ to december 2015.

be presented only for the more suitable models according to these comparisons criteria.

Remember that as seen in Sections 2.4 e 2.4.2, given a set of models for a given dataset, the lower the AIC, BIC , EAIC, EBIC and DIC, more adjusted to the data is the model. And, on the other hand, the greater the values of $\widehat{p}_{BS}(y)$ and $\widehat{p}_{IS}(y)$, the more to the data is the model. Thus, the values in bold refer to the model most adjusted to the data according to the respective selection criteria.

To obtain the results shown in Table 3.6, we used the following quantity of particles in all filters: 2.000 for the GARCH(1,1) model with noise and SIR filter, 50 for the same model with ASIR filter, 200 for the SV($\mathcal{N}$) model, 50 for the SV(t_3) model, 230 for the SV($S\mathcal{N}$) model, 45 for the SV(St_3) model and 350 for the SV with leverage model.

Note that a few selection criteria (AIC, BIC, EAIC and EBIC) indicated the GARCH model with SIR as the most adjusted to IBOVESPA return series, while DIC , $\widehat{p}_{BS}(y)$ and $\widehat{p}_{IS}(y)$ also indicated GARCH, but with ASIR.

Table 3.6: Model comparisons by means of likelihood-based information criteria and marginal likelihoods for IBOVESPA series

Model	PF	AIC	BIC	$EAIC$	$EBIC$	DIC	$\widehat{p}_{BS}(y)$	$\widehat{p}_{IS}(y)$
GARCH	SIR	**-5997.9**	**-5978.0**	**-6001.0**	**-5981.2**	-6012.1	2.2×10^{-230}	2.1×10^{-230}
	ASIR	-5989.5	-5969.6	-6000.5	-5980.6	**-6019.5**	$\mathbf{2.3 \times 10^{-230}}$	$\mathbf{2.2 \times 10^{-230}}$
SV($\mathcal{N}$)	SIR	-5974.6	-5959.7	-5976.9	-5962.0	-5985.2	1.9×10^{-232}	1.9×10^{-232}
SV(t_3)		-5909.7	-5894.8	-5907.5	-5892.6	-5911.3	5.5×10^{-246}	5.7×10^{-246}
SV($S\mathcal{N}$)		-5974.4	-5954.6	-5974.0	-5954.1	-5981.5	3.6×10^{-235}	3.5×10^{-235}
SV(St_3)		-5897.0	-5877.1	-5900.7	-5880.8	-5912.4	2.2×10^{-249}	2.3×10^{-249}
SV lev		-5989.4	-5969.6	-5986.5	-5966.7	-5991.6	1.9×10^{-231}	2.2×10^{-231}

Taking into account the marginal likelihoods criterias for model selection, GARCH(1,1) model with noise was considered the best one applied to IBOVESPA data. Therefore, the posterior mean, median, standard deviation and credibility interval of 95% of this model are given in Table 3.7.

For the models applied to NASDAQ return series (Table 3.8), we used the following quantity of particles in all filters: 2.000 for the GARCH(1,1) model with noise and SIR filter, 75 for the same model with ASIR filter, 350 for the SV($\mathcal{N}$) model, 200 for the SV(t_3) model, 400 for the SV($S\mathcal{N}$) model, 200 for the SV(St_3) model and 400 for the SV with leverage model.

Furthermore, with the exception of DIC that indicated GARCH with ASIR as the model most adjusted to the data, all criteria pointed to the SV model with the variation of ϵ_t having a skew-normal distribution (see Section 2.5.3).

Table 3.7: Posterior mean, median, standard deviation and credibility interval for the parameters of GARCH(1,1) model with noise using ASIR filter applied to IBOVESPA series

Parameters	Posterior estimations				
	Mean	Std. dev.	$\text{CI}_{0.025}$	Median	$\text{CI}_{0.975}$
σ^2	0.0000913	0.0000259	0.0000294	0.0000942	0.0001342
β_0	0.0000020	0.0000015	0.0000005	0.0000016	0.0000060
β_1	0.1512039	0.0517068	0.0676217	0.1448260	0.2735384
β_2	0.8408180	0.0512353	0.7203291	0.8471988	0.9208007

Table 3.8: Model comparisons by means of likelihood-based information criteria and marginal likelihoods for NASDAQ series

Model	PF	AIC	BIC	$EAIC$	$EBIC$	DIC	$\widehat{p}_{BS}(y)$	$\widehat{p}_{IS}(y)$
GARCH	SIR	-6942.0	-6922.2	-6940.9	-6921.0	-6947.7	1.3×10^{-25}	1.3×10^{-25}
	ASIR	-6911.9	-6891.3	-6940.4	-6920.5	**-6977.6**	1.4×10^{-25}	1.3×10^{-25}
SV($\mathcal{N}$)	SIR	-6954.9	-6940.0	-6952.9	-6938.0	-6956.9	1.6×10^{-19}	1.6×10^{-19}
SV(t_3)		-6906.2	-6891.3	-6907.4	-6892.5	-6914.7	8.6×10^{-30}	8.4×10^{-30}
SV($S\mathcal{N}$)		**-6974.5**	**-6954.6**	**-6971.1**	**-6951.2**	-6975.7	$\mathbf{7.1 \times 10^{-18}}$	$\mathbf{6.7 \times 10^{-18}}$
SV(St_3)		-6924.4	-6904.5	-6919.8	-6899.9	-6923.2	3.6×10^{-29}	4.1×10^{-29}
SV lev		-6954.0	-6934.1	-6956.9	-6937.0	-6967.8	9.3×10^{-20}	8.8×10^{-30}

As stated above, especially when considered only marginal likelihood criterias for model selection, stochastic volatility model with skew Gaussian noise was considered the best one applied to NASDAQ data. Therefore, the posterior mean, median, standard deviation and credibility interval of 95% of this model are given in Table 3.9.

Table 3.9: Posterior mean, median, standard deviation and credibility interval for the parameters of stochastic volatility model with skew Gaussian noise using SIR filter applied to NASDAQ series

Parameters	Posterior estimations				
	Mean	Std. dev.	$\text{CI}_{0.025}$	Median	$\text{CI}_{0.975}$
τ^2	0.2032857	0.0590969	0.1139529	0.1945681	0.3426287
ϕ	0.8365621	0.0441244	0.7357273	0.8422046	0.9071204
α	-9.5973375	0.1023758	-9.7981681	-9.5970571	-9.3955176
λ	0.0014686	0.0003089	0.0008629	0.0014690	0.0020713

Finally, for the models applied to S&P500 return series (Table 3.10), we used the following quantity of particles in all filters: 4.000 for the GARCH(1,1) model with noise and SIR filter, 50 for the same model with ASIR filter, 250 for the SV($\mathcal{N}$) model, 200 for the SV(t_3) model, 400 for the SV($S\mathcal{N}$) model, 200 for the SV(St_3) model and 1.750 for the SV with leverage model.

Besides that, among the models applied to S&P500 return series, all the selection criteria pointed to the SV model with leverage (see Section 2.5.3) as the model most adjusted to the data, except for the DIC that, as for NASDAQ data, indicated GARCH with ASIR.

Table 3.10: Model comparisons by means of likelihood-based information criteria and marginal likelihoods for S&P500 series

Model	PF	AIC	BIC	$EAIC$	$EBIC$	DIC	$\widehat{p}_{BS}(y)$	$\widehat{p}_{IS}(y)$
GARCH	SIR	-7274.6	-7254.8	-7272.7	-7252.8	-7278.8	6.0×10^{-171}	6.0×10^{-171}
	ASIR	-7183.6	-7163.8	-7273.0	-7253.1	**-7370.3**	6.5×10^{-171}	6.4×10^{-171}
SV($\mathcal{N}$)	SIR	-7297.9	-7283.0	-7294.6	-7279.7	-7297.3	1.1×10^{-162}	1.1×10^{-162}
SV(t_3)		-7246.5	-7231.5	-7244.6	-7229.7	-7248.8	6.5×10^{-174}	6.2×10^{-174}
SV($S\mathcal{N}$)		-7308.7	-7288.8	-7308.7	-7288.9	-7316.8	5.4×10^{-162}	5.2×10^{-162}
SV(St_3)		-7250.7	-7230.8	-7250.0	-7230.1	-7257.3	7.0×10^{-175}	8.2×10^{-175}
SV lev		**-7319.4**	**-7299.5**	**-7323.3**	**-7303.4**	-7335.2	$\mathbf{5.0 \times 10^{-158}}$	$\mathbf{5.6 \times 10^{-158}}$

Thus, especially when considered only marginal likelihood criterias for model selection, stochastic volatility model with leverage was considered the best one applied to S&P500 data. Therefore, the posterior mean, median, standard deviation and credibility interval of 95% of this model are given in Table 3.11.

Table 3.11: Posterior mean, median, standard deviation and credibility interval for the parameters of stochastic volatility model with leverage using SIR filter applied to S&P500 series

Parameters	Posterior estimations				
	Mean	Std. dev.	$\mathrm{CI}_{0.025}$	Median	$\mathrm{CI}_{0.975}$
τ^2	0.1082816	0.0204612	0.0744495	0.1063194	0.1550263
ϕ	0.9828918	0.0155592	0.9423327	0.9874117	0.9990852
α	-11.0962922	1.8050948	-16.7467513	-10.5320217	-9.3502158
ρ	-0.6686287	0.0956189	-0.8154062	-0.6827698	-0.4396470

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

Chapter 4

CONCLUDING REMARKS

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

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

For future work, more detailed versions of particle filters in order to reduce the variability of the likelihood estimator, thus improving on the convergence and other properties of the MCMC. The methodology can also be applied to other class of state space models, including multivariate ones.

REFERENCES

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 269–342.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, 177–181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Carmona, R. (2014). *Statistical analysis of financial data in R.* New York: Springer.

Casella, G., & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Duxbury Thomson Learning.

Chen, M. H., & Shao, Q. M. (1997). On Monte Carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics*, *25*(4), 1563–1594.

Del Moral, P. (2004). *Feynman-Kac formulae: Genealogical and interacting particle systems with applications*. New York: Springer.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, *57*(6), 1317–1339.

Giordani, P., & Kohn, R. (2010). Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals. *Journal of Computational and Graphical Statistics*, *19*(2), 243–259.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F – Radar and Signal Processing*, *140*(2), 107–113.

Granger, C. W. J., & Ding, Z. (1995). Some properties of absolute return: An alternative measure of risk. *Annales d'Économie et de Statistique*(40), 67–91.

Haario, H., Saksman, E., & Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, *14*(3), 375–395.

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, *7*(2), 223–242.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, *57*(1), 97–109.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *12*(4), 371–389.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1995). Models and priors for multivariate stochastic volatility. *CIRANO Working Papers*(95s–18).

Liu, Y., Gopikrishnan, P., Cizeau, P., Meyer, A., Peng, C. K., & Stanley, H. E. (1999). Statistical properties of the volatility of price fluctuations. *Physical Review*, *60*(2), 1390–1400.

Mandelbrot, H. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Meng, X. L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, *6*(4), 831–860.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Migon, H., Gamerman, D., & Louzada, F. (2015). *Statistical inference: An integrated approach* (2nd ed.). Boca Raton: CRC Press.

Pitt, M. K., & Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, *94*(446), 590–599.

Pitt, M. K., Silva, R. S., Giordani, P., & Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, *171*(2), 134–151.

Roberts, G. O., Gelman, A., & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, *7*(1), 110–120.

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, *16*, 351–367.

Roberts, G. O., & Rosenthal, J. S. (2007). Coupling and ergodicity of adaptive MCMC. *Journal of Applied Probability*, *44*(2), 458–475.

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, *18*(2), 349–367.

Silva, R., Kohn, R., Giordani, P., & Mun, X. (2010). A copula based approach to adaptive sampling. *ArXiv*.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statististical Society: Series B*, *64*(4), 583–639.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, *22*(4), 1701–1762.

West, L., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.). New York: Springer-Verlag.

Appendix A

LOG-RETURNS

The use of the return in its form given by Equation 2.22 (Section 2.5.1) already brings benefits, such as normalization of the data, that is, it is possible to easily compare several return series because they are in a comparable metric, which may not be true for price series with different values.

However, if we work with log-return series r_t (Equation 2.23), it is possible to achieve even better mathematical and statistical properties. One of the main advantages of using log-returns is when we are interested in calculating the compounding return. Consider an ordered sequence of n trades, the compounding return is given by $(1+R_1)(1+R_2),\cdots,(1+R_n) = \prod_i(1+R_i)$. However, $\ln(\prod_i(1+R_i)) = \sum_i \ln(1+R_i) = \ln(y_n) - \ln(y_0)$, which generates a great computational advantage.

In addition, the log-return series has properties of stationarity, that is, constant expected value equal to zero throughout the series; and ergodicity, that is, its statistical properties can be obtained from a sufficiently large random sample of the generating process. More details on the motivations of using log-returns series in financial studies can be seen in Carmona (2014). Therefore, in this thesis, we will only work with log-return series instead of price series.

Appendix B

FULLY ADAPTED PARTICLE FILTER FOR THE GARCH(1,1) PLUS NOISE

A GARCH(1,1) model is given by

$$\begin{aligned} y_t|x_t &\sim \mathcal{N}(x_t, \sigma^2) \\ x_{t+1}|x_t &\sim \mathcal{N}(0, \tau_t^2(x_t)) \\ \tau_{t+1}^2 &= \beta_0 + \beta_1 x_t^2 + \beta_2 \tau_t^2. \end{aligned}$$

We omit the dependence of τ_{t+1}^2 on x_t for while. Now, suppose we have particles $x_{t-1}^{(\ell)}$, $\ell = 1, \ldots, L$ with attached probabilities $\pi_{t-1}^{(\ell)}$

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

where

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

Hence,

$$-2\log(p(y_t|x_t)p(x_t|x_{t-1}^{(\ell)})) = \log\tau_t^{2*} + \frac{1}{\tau_t^{2*}}\left(x_t - \delta_t^{(\ell)}\right)^2 + \Delta_t^{(\ell)}$$

with

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

It follows that

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

Since this model is fully adapted, the $\pi_{t-1}^{(\ell)}$'s have equal weights for all $\ell = 1, \ldots, L$.

Appendix C

STOCHASTIC VOLATILITY WITH LEVERAGE

The observation and system equations of the SV model is given by

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

and

$$\left(\begin{matrix} \epsilon_t \\ \omega_t \end{matrix} \middle| \rho, \tau \right) \sim \mathcal{N}\left[0, \begin{pmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{pmatrix}\right].$$

Then,

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

where,

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

Taking into consideration that,

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

so,

$$\begin{aligned} y_t|x_t, x_{t-1} &\sim \mathcal{N}\left(\rho\frac{x_t-\xi_t}{\tau}e^{x_t/2}, e^{x_t}(1-\rho^2)\right) \\ x_t|x_{t-1} &\sim \mathcal{N}(\xi_t, \tau^2) \end{aligned}$$