Università Ca' Foscari Venezia

Dipartimento di Scienze Economiche

Dottorato di Ricerca in Economia, 1° Ciclo Nuova Serie

SIMULATION METHODS FOR NONLINEAR AND NON-GAUSSIAN MODELS IN FINANCE

Tesi di Dottorato di

Roberto Casarin

Coordinatore del Dottorato
Prof. Piero Gottardi

Tutore del Dottorando
Prof.ssa Monica Billio

Submitted version, January 2004

Table of Contents

Introduction

Assumptions concerning the probability distribution and the dynamic of asset returns are among the most important aspects in financial modelling, due to their implications on asset pricing and risk evaluation and management.
Since the early work due to Bachelier (1900), the dynamic of asset prices is assumed to be a Gaussian stochastic process. In order to guarantee the positivity of the price, successive studies slightly modify this hypothesis by assuming that the dynamic of the logarithm of the price is a Gaussian process. The use of the Gaussian hypothesis allows to obtain analytical solutions in many problems of asset pricing, asset portfolio management and risk evaluation.
Nevertheless, Gaussian models often fails to produce satisfactory results when applied to real data. Many contributions in the literature, see for example Fama (1965), Fama and Roll (1968), Mandelbrot (1963), Barnes and Downes (1973) and the more recent study due to Mandelbrot (1990), evidenced how observed asset returns exhibit asymmetry, excess of kurtosis and multi-modality, which leads to the rejection of the normality hypothesis. These features are present on financial data when observed with a monthly frequency and become clearer when data are collected with higher frequencies (i.e. weekly or daily). In order to account for these features it is necessary to introduce different assumptions on asset return distribution.

The first aim of this thesis is to remove the Gaussian distribution and the linearity assumptions in some well know financial models. We consider asset portfolio models with shortfall constrains, non-Gaussian asset return distributions and dynamics. The inference process on these models becomes difficult and classical Maximum Likelihood (ML) approach exhibits some limitations.

The second aim is to provide inference tools for nonlinear and non-Gaussian

financial models. We choose to apply Bayesian approach to statistical inference due to its strong probabilistic basis and to the great flexibility of the approach. Moreover the Bayesian framework is strictly related to decision theory and this feature gives a strong economic interpretability to results obtained through Bayesian inference.

The first element of flexibility of the Bayesian framework, with respect to the ML approach, regards the way to put constrains on the parameters of an econometric model. In fact a prior distribution is a versatile instrument to reach this goal. Furthermore in the ML approach, the presence of multiple local maximum makes difficult the choice of the estimator and implies that the standard asymptotic theory for ML estimation and the test theory don't apply. The Bayesian approach avoids these problem as parameters are random variables, with prior and posterior distributions defined on the parameter space. Posterior distribution contains all the information on the parameters. and point estimates are obtained by averaging over the parameter space, weighting by the posterior distribution of the parameters or by the simulated posterior distribution.

Moreover variable dimension problems can be easily treated in the Bayesian approach and this allows to make inference also on the dimension of the model using simulation techniques such as Reversible Jump Markov Chain Monte Carlo or Birth and Death Markov Chain Monte Carlo.

Recently the flexibility of the Bayesian modelling has been rediscovered in making inference on dynamic models, indeed in this context, the ML approach has several limitations. First inference on unobservable variables can be carried out only conditionally on the parameters of the model, thus neglecting the variability of the parameters' estimators. Secondly the state space representation of complex and parameterized models is not always available. The Bayesian approach, which allows to joint estimate hidden states and unknown parameters and provides a general state space representation also for nonlinear and non-Gaussian models, overcome these difficulties. Furthermore, due to this general way of representing stochastic dynamic models, traditional Markov Chain Monte Carlo simulation techniques and recently developed sequential simulation methods, like particle filters, can be easily applied.

Within the Bayesian framework, analytical solutions of the estimation problem do not often exist and numerical solution is needed. Thus Monte Carlo simulation

methods, such as Markov Chain Monte Carlo (MCMC), become a natural way for making inference since they can be included in the inference process in immediate way. The use of simulations introduces a computational complexity instance. However in the last decade, the increase in the computational speed and power of computers gives a strong impulse to the Bayesian framework.
An important aim of this thesis is to show how traditional and recently developed Monte Carlo simulation techniques result particularly useful in the analysis of non-linear and non-Gaussian financial models.

In particular we start analysing (Chapter 4) a portfolio optimisation problem, with shortfall constrains. We remove the traditional hypothesis of Gaussian distributed asset returns introducing heavy tail distributions and suggest to use Monte Carlo simulation method to solve the resulting optimisation problem. Furthermore we study (Chapter 5) the use of mixtures of non-Gaussian distribution in making empirical evidence on financial data. We suggest the use of the mixtures of α-stable distributions and provide a Bayesian model to make inference on the mixture. We show how the use of the Markov Chain Monte Carlo method becomes fundamental to solve the estimation problem. Finally we treat simulation base inference methods on discrete time dynamic models for financial time series. We provide (Chapter 2) an updated review of the Bayesian simulation based inference for dynamic econometric models. Then we focus (Chapter 6) on stochastic volatility models with Markovian jumps and with a non-Gaussian innovation process. Inference on this kind of models is particularly difficult due to the stochastic structure of the model that exhibit two levels of latent processes. We study how sequential Monte Carlo methods apply to this generalised stochastic volatility model in order to make inference on the hidden log-volatility and Markov jump processes and on the parameters of the model.

In the following the give a detailed description of the thesis contents. This thesis can be divided in two parts. The first one corresponds to chapters one, two and three. It reviews simulation methods and serves as introduction to the research work developed in the second part, which corresponds to chapters four, five and six.

The first chapter introduces to Monte Carlo methods and motivates the use of simulation for the numerical solution of integration problems, which arise when treating financial models; for example, when making inference in a Bayesian

framework both on static and dynamic probability models or solving portfolio optimisation problems. We review, also through some examples, the simulation techniques, which will be used in the subsequent chapters. In particular we introduce importance sampling, accept/reject and finally Markov chain Monte Carlo methods.

The second chapter deals with the time evolution of the financial variables, which represents an important issue in financial econometrics. Stochastic dynamic models allow to describe more accurately many features of the financial variables. An accurate financial modelling is relevant not only in forecasting and asset pricing, but also in a risk evaluation and management perspective.

Many time-varying phenomena can be represented through dynamic models, but often there is a trade-off between the modelling accuracy and the complexity of the resulting model. Moreover the degree of complexity is increased by the use of latent factors. In time series analysis, latent factors are often introduced to model the heterogeneous time evolution of the observed process. The presence of unobserved components makes the maximum likelihood inference more difficult to apply. A Bayesian approach is thus preferable since it allows to treat general state space models and makes easier the simulation based approach to parameters estimation and latent factors filtering.

The main aim of the chapter two is to define a Bayesian dynamic model and the related inference problems of parameters estimation and hidden states filtering. Furthermore the chapter analyses the simulation based inference and show how traditional simulation methods, like the single-move Gibbs sampler and the multi-move Gibbs sampler apply to a dynamic model. Finally the chapter introduces recently developed sequential Monte Carlo simulation methods, focusing on Particle Filters. Throughout the chapter, many examples illustrate how Bayesian simulation based inference apply to basic Stochastic Volatility and Business Cycle models.

In Chapter three we make some empirical analysis on financial data in order to evidence some basic features of financial time series. This chapter shows how skewness, excess of kurtosis, multi-modality and volatility clustering characterize the evolution of asset prices of both the stock and bond markets and also of the liquidity market. This chapter motivates the use of non-linear and non-Gaussian models, which will be developed in the subsequent chapters.

Chapter four deals with portfolio models. In asset allocation problems financial managers generally assume normally distributed returns even if extreme realizations usually have an higher frequency than in the Gaussian case. This specification error conducts to an effective risk of the portfolio which is higher than the theoretical one. We propose stochastic optimisation based on a Monte Carlo simulation approach, to solve an asset allocation problem, with shortfall constraint and to evaluate the exact portfolio risk-level, in presence of misspecified tails behaviour. We assume that returns are generated by a multivariate Student-t distribution, while in reality returns come from a multivariate distribution where each marginal is a Student-t but with different degrees of freedom. Stochastic optimisation allows us to value the effective risk for managers. In the specific case analysed, we also found that a multivariate density resulting from different Student-t marginal distributions produces a shortfall probability and a shortfall return level that can be well approximated by assuming a multivariate Student-t with adequately degrees of freedom. The proposed stochastic optimisation approach could be an important instrument for investors who need a qualitative assessment of the reliability and sensitivity of their investment strategies when their models are potentially misspecified.

In Chapter five we propose a flexible parametric model for asset returns and suggest inference tools for the parameter estimation. In financial modelling the Gaussian distribution results unsatisfactory and reveals difficulties in fitting data with skewness, heavy tails and multimodality, thus α-stable distributions have been considered. The use of these distributions allows for modelling skewness and heavy tails but gives rise to inferential problems related to the parameter estimation. We generalise the stable distribution framework by introducing a model that accounts also for multimodality. In particular we introduce a stable mixture model and a suitable reparameterisation of the mixture, which allow us to make inference on the mixture parameters. We develop a full Bayesian approach and MCMC simulation techniques for the estimation of the posterior distribution.

Chapter six deals with dynamic modelling of financial data. First we briefly review some stochastic volatility (SV) models proposed in the literature. These models allow to account for time-varying volatility and for clustering in volatility. Secondly we focus on SV models with jumps and propose a Markov Switching

Stochastic Volatility (MSSV) model, with an heavy tail observable process. Due to the economic interpretation of the hidden volatility regimes, these models have many financial applications like asset allocation, option pricing and risk management. The Markov switching process is able to capture clustering effects and jumps in volatility. Heavy tail innovations account for extreme variations in the observed process. Accurate modelling of the tails is important when estimating quantiles is the major interest like in risk management applications.
The second important aim of the chapter is to deal with inference on heavy tails MSSV models. We follow a Bayesian approach to state filtering and parameter estimation, focusing on recently developed simulation based filtering techniques, called Particle Filter (PF). Simulation based filters are sequential Monte Carlo algorithms, which are useful when assuming non-linear and non-Gaussian latent variable models and when processing data sequentially. We study how auxiliary particle filters behave for heavy tail latent variables model. Furthermore we analyse how PF apply to heavy tails MSSV model in order to update parameter estimates and state filtering as new observations become available.

This thesis has been developed during a one year working experience at GRETA Ass., University of Venice and a two year visiting period at CEREMADE, University Paris IX (Dauphine). The following research works come from this thesis

- Casarin R. and Gobbo M., (2002), "Metodi Monte Carlo per la valutazione di opzioni finanziarie", in *Proceedings of the Summer School in Quantitative Finance*, Auronzo di Cadore 29-31 May 2002, Departments of Applied Mathematics, University Ca' Foscari of Venice.

- Billio M. Casarin R. and Sartore D., (2003), "Bayesian inference in dynamic models with latent factors", forthcoming, *Monography of Official Statistics*, edited by Eurostat. Presented at the 4th Colloquium on Modern Tools for Business Cycle Analysis in Luxembourg on 20-22 October 2003.

- Billio M. and Casarin R., (2003), "Extreme Returns in a Shortfall Risk Framework", in *Atti della giornata di studio Metodi Numerici per la Finanza*, 30 May 2003, Applied Mathematics Department, University "Ca' Foscari", Venice.

- Billio M., Casarin R. and Toniolo G., (2002), "Extreme Returns in a Shortfall Risk Framework", *Working Paper*, GRETA n. 0204 and in *Proceedings of 8th International Conference Forecasting Financial Markets Meeting*, London 2002.

- Casarin, R., (2004), "Bayesian Inference for Mixture of Stable Distributions", forthcoming, *Working Paper* CEREMADE. Presented at the Young Statistician Meeting, Cambridge 14-15 April 2003.

- Casarin, R., (2003), "Bayesian Inference for Mixture of Stable Distributions", in *Atti del Convegno Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione* , 4-6 Semptember 2003, Statistics Department, University " Ca' Foscari", Venice.

- Casarin, R., (2004),"Bayesian Inference for Markov Switching Stochastic Volatility Models", forthcoming, *Working Paper* CEREMADE. Presented at the 4th International Workshop on Objective Bayesian Methodology, CNRS, Aussois, 15-20 June 2003. It received the *Springer's Award* as best poster session.

Chapter 1

Monte Carlo Methods

1.1 Integration Problems

In financial modelling and when making inference on financial data many integration problems arise and the analytical solution is often not available. Moreover it is often necessary to integrate functions over a high dimensional space. In the next section we will show how these kind of problems arise, for example, in a Bayesian inference context. In the following we will adopt a notation similar to the one in Robert and Casella [8].

Assume we are interested in the solution of the following integration problem

$$\mathbb{E}_p\left(g(X)\right) = \int \dots \int g(\mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{1.1}$$

where $p(\mathbf{x})$ is the probability distribution function (p.d.f.) of the random variable X and $g(\mathbf{x})$ is an arbitrary function, whose expectation is finite. If this integral does not admit an analytical solution numerical methods are needed.

The most straightforward approach consists in discretising the domain of the function $g(\mathbf{x})p(\mathbf{x})$, with a finite set of points. Then the multiple integral can be approximated through a multiple summation. Consider a two dimensional integration problem, $\mathbf{x} = (x_1, x_2)$, the numerical approximation of the integral is

$$\mathbb{E}_p\left(g(\mathbf{X})\right) \approx \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} g(\mathbf{x}_{i,j})p(\mathbf{x}_{i,j})\Delta^2 \tag{1.2}$$

where $\mathbf{x}_{i,j} = (\underline{x} + i\Delta, \underline{x} + j\Delta)'$. We assume $\underline{x}$ and $\overline{x}$ are the minimum value and the

maximum value respectively of both x_1 and x_2. Thus the length of the interval is $\Delta = \overline{x} - \underline{x}$.

This basic numerical integration method becomes inefficient for high dimension integration problems and although many modifications of the method have been proposed, they exhibit the same inefficiency problem. Thus an alternative method is needed. In particular Monte Carlo simulation methods allow to perform efficient numerical integration also in high dimension problems.

1.1.1 The Bayesian Paradigm

In the following we briefly review the integration problems which arise when making inference. Due to its recent diffusion in the financial econometric literature, we focus on the Bayesian approach to statistical inference, which represents an alternative to the frequentist approach. Bayesian inference has been recently rediscovered in finance because the approach is strongly based on probability theory and also on decision theory, and this allows to give to inference results also an economic interpretation. Furthermore Bayesian approach allows a greater flexibility in stochastic modelling and in using simulation methods to solve complex problems.

The Bayes' theorem is the probabilistic result which plays a central role in Bayesian inference.

Theorem 1.1.1. *(Bayes' theorem)*
Given the probability space Ω, if $A, B \in \Omega$ are two events such that $P(B) \neq 0$ then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1.3}$$

Let $\mathcal{X}$ be the *observation space* and Θ the *parameter space*. We assume that the random variable $x \in \mathcal{X}$ follows a distribution in the parametric distribution family $p(x|\theta)$, with $\theta \in \Theta$. A Bayesian statistical model is defined by considering the unknown parameters θ as a random variable, with a *prior distribution* $\pi(\theta)$, defined on the parameter space.

By applying the Bayes' theorem for continuous random variables, we obtain the

posterior distribution of the parameters

$$\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int\limits_{\mathcal{X}} p(x|\theta)\pi(\theta)d\theta} = \frac{p(x|\theta)\pi(\theta)}{p(x)} \tag{1.4}$$

where $p(x|\theta)\pi(\theta)$ is the *joint distribution* and $p(x)$ is the *marginal distribution* of x. The Bayesian theorem can be interpreted as a learning process, because it allows to include in the posterior distribution of θ, the information on θ contained in the observation x. One of the advantage in using the Bayesian approach for inference is that the whole posterior distribution of the parameter is available and not only a point estimate of the parameter, as in classical inference.

In order to complete the description of the Bayesian inference approach we use a Bayesian Decision Theory framework. The solution to a statistical inference problem is thus a *decision* $d \in \mathcal{D}$, defined on the *decision space*, $\mathcal{D}$. When making inference on θ, the decision space is just the parameter space: $\mathcal{D} = \Theta$. Moreover we introduce a decision criteria, called *loss function*, $L(\theta, d)$, which is an application from $\Theta \times \mathcal{D}$ in $[0, +\infty)$.

The best decision is called Bayes estimator and is obtained by minimizing the expected value, taken with respect to $p(x)$ and $\pi(\theta|x)$, of the loss function.

Definition 1.1.1. *(Bayes estimator)*
Given a prior distribution π and a loss function L, the Bayes estimator, δ^π of θ is

$$\delta^\pi(x) = \arg\inf_{\delta \in \mathcal{D}} \int_{\mathcal{X}} \int_{\Theta} L(\theta, \delta(x)) p(x|\theta)\pi(x)d\theta dx \tag{1.5}$$

A very commonly used loss function is the squared error loss: $L(\theta, d) = (\theta - d)^2$. As stated in the following theorem, if the loss is a quadratic function then the Bayes estimator is the posterior mean (see Robert [7]).

Theorem 1.1.2. *Given a prior distribution π and a quadratic loss function, the Bayes estimator is the mean of the posterior distribution*

$$\delta^\pi(x) = \mathbb{E}_{\pi(\theta|x)}(\theta) = \frac{\int_\Theta \theta p(x|\theta)\pi(\theta)d\theta}{\int_\Theta p(x|\theta)\pi(\theta)d\theta} \tag{1.6}$$

We conclude that the posterior distribution gives all the information for the parameter inference. However, it is often not easy to find an analytical solution of

the integration problem given in equation (1.4), thus a numerical approximation is needed. In particular Monte Carlo simulation methods (see section 1.2) are widely used integration techniques, which reveal extremely efficient in high dimension problems. In the literature on Bayesian inference, Markov Chain Monte Carlo methods (see section 1.3) play a central role, because they represent flexible simulation techniques that apply also to complex posterior distribution, requiring to know the posterior of interest only up to a proportionality constant.

1.2 Monte Carlo Methods

Monte Carlo methods allow to generate random values from a desired distribution. Through sequences of random numbers it is possible to approximate and to solve numerically many problems involving stochastic quantities, which have not an analytical solution. The main idea is to replace stochastic quantities in the problem by simulated random numbers. The method has been introduced in physics by Ulam and Von Neumann [11] and then becomes popular in other fields like finance (Boyle [1]), statistics (Geman and Geman [3]).

In order to solve the integration problem given in Equation (1.1) through the Monte Carlo simulation method, we generate a sequence of N random values $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ from the distribution $p(\mathbf{x})$ and approximate the integral by the following Monte Carlo estimator

$$\mathbb{E}_p\left(g(\mathbf{x})\right) \approx \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x}_i) \tag{1.7}$$

which is defined for any function $g(\mathbf{x})$ and for every distribution $p(\mathbf{x})$. By the laws of large numbers this estimator converges to $\mathbb{E}_p\left(g(\mathbf{x})\right)$ with $N \to \infty$. The variance of the Monte Carlo estimator can be reduced through variance reduction methods (also called acceleration methods), like antithetic variables, stratified sampling, control variables methods (see Casella and Robert [8] for statistical applications and Dupire B. [2] for a review on financial applications).

One of the main issue in stochastic simulation is the generation of random values from a given distribution.

Uniform random number generation represents the first step in Monte Carlo

simulation, the other random variables simulation methods being based on the generation of uniforms. The key idea in generating random numbers or pseudo-random numbers is to produce, through an algorithm, a sequence of numbers which exhibit all the features of a random phenomena. These features can be tested through some tests, like goodness of fit tests, autocorrelation, runs test, etc. (see Ripley [6] for further details).

A widely used algorithm is the *linear congruential method*. The sequence, $\{X_n\}_n$, of pseudo-random number is generated as follows

$$X_{n+1} = (X_n a + b) \, mod \, M \tag{1.8}$$

where a, b and M are parameters of the algorithm which determines the number of iterations before repeated values are generated (periodicity). If $b = 0$, then the random number generator is called *multiplicative congruential method*. For a right setting of this parameter see Press et al. [5]. From the sequence of random numbers $\{X_n\}_n$, it is possible to generate numbers, $\{U_n\}_n$, uniformly distributed in the interval $[0, 1)$, by means of the following transformation

$$U_n = \frac{X_n}{M} \tag{1.9}$$

Once uniform pseudo-random number have been generated it is possible to simulate random values from any desired distribution by means of other algorithms.

In the one dimensional case the straightforward way to generate a random values from $p(x)$ is to apply the *inverse cumulative distribution function (c.d.f.) method*. If the c.d.f. $F(x)$ of $p(x)$ is known and can be inverted, then a uniform pseudo-random value $u \sim U_{[0,1]}$ can be used to generate a random value x from $p(x)$. By applying the *generalized inverse c.d.f.* to u

$$F^{-}(x) = \inf \{x | F(x) \geq u\} \tag{1.10}$$

it follows that $x \sim p(x)$ (see Ripley [6]). In the following example we show how to generate random values from a Weibull distribution through the inverse cumulative density function method.

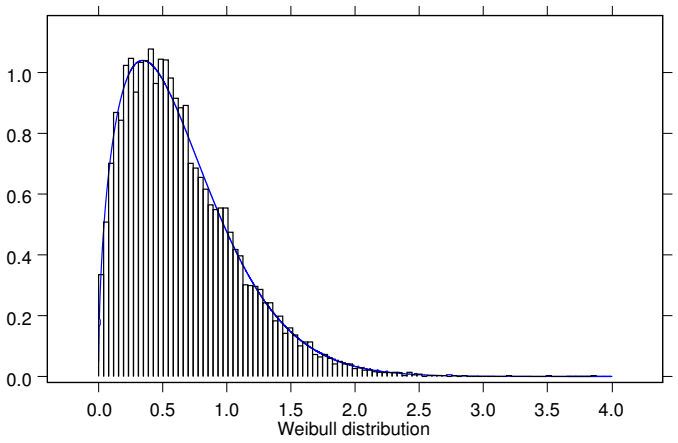

Figure 1.1: The histogram represents 10,000 drawn from a Weibull distribution, $\mathcal{W}e(1.5, 0.5)$. The line indicates the true p.d.f.

Example 1.2.1 - *Weibull distribution*
Consider the Weibull distribution $\mathcal{W}e(\alpha, \lambda)$

$$p(x) = \alpha x^{\alpha-1} e^{\lambda - e^{\lambda} x^{\alpha}} \mathbb{I}_{[0,\infty)}(x) \tag{1.11}$$

then the c.d.f. and its inverse are

$$F(x) = 1 - e^{-e^{\lambda} x^{\alpha}} \quad \Longleftrightarrow \quad F^{-}(u) = \left(-\frac{ln(1-u)}{e^{\lambda}} \right)^{\frac{1}{\alpha}} \tag{1.12}$$

respectively. In order to simulate $\{x_1, \ldots, x_N\}$, with $N = 10.000$, from $\mathcal{W}e(1.5, .0.5)$, we generate uniform pseudo-random numbers $\{u_1, \ldots, u_N\}$ through the *mixed linear congruential* algorithm (with *seed* $= 9$) and then apply the inverse c.d.f. to these uniform numbers. Figure 1.1 shows the true density and the histogram of the simulated random values.

□

If the random variable x does not admit the inverse c.d.f. then it is possible to apply other methods like Accept/Reject and Importance Sampling, which allow to generate values from the distribution of interest by simulating from an instrumental distribution.

1.2.1 Importance Sampling

The key idea used in importance sampling is to generate random values through an instrumental distribution $q(x)$, called *importance distribution*, and to correct the resulting simulated function by means of the ratio between the true density and the instrumental density

$$\mathbb{E}_p\left(g(X)\right) = \int g(x)p(x)dx = \int g(x)\frac{p(x)}{q(x)}q(x)dx = \tag{1.13}$$
$$= \mathbb{E}_q\left(g(X)w(X)\right) \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i)w(x_i)$$

where

$$w(x_i) = \frac{p(x_i)}{q(x_i)}, \qquad i = 1, \ldots, N \tag{1.14}$$

are called *importance weights*. The resulting Monte Carlo estimator is unbiased and converges to $\mathbb{E}_p\left(g(X)\right)$ as $N \to \infty$, whatever the choice of the instrumental distribution q and as long as $supp(q) \supset supp(p)$. Note however that a good choice of the distribution q may reduce the variance of the estimator, which is

$$\mathbb{E}_q(g^2(X)\frac{p^2(X)}{q^2(X)}) = \mathbb{E}_p(g^2(X)\frac{p(X)}{q(X)}) \tag{1.15}$$

Therefore instrumental distributions with tails lighter than those of p are not appropriate for importance sampling. Moreover if the ratio p/q is unbounded then importance weights vary widely giving too much importance to a few values x_i. If the ratio p/q is bounded, then the instrumental distribution can be used also for an Accept/Reject algorithm (see example 1.2.2).

In order to improve the efficiency of the Monte Carlo estimator an alternative is to use the following importance sampling estimator

$$\mathbb{E}_p\left(g(X)\right) \approx \frac{\frac{1}{N}\sum_{i=1}^{N} g(x_i)w(x_i)}{\frac{1}{N}\sum_{i=1}^{N} w(x_i)} \tag{1.16}$$

where $w(x)$ are the importance weights. Note however that the estimator is no more unbiased because the quantity at the denominator is a random variable (see Casella and Robert [8]).

1.2.2 Accept-Reject

This method is similar to the importance sampling and allows to simulate from the *target distribution*, $p(x)$, by simulating from an *instrumental distribution* $f(x)$. The method is particularly useful because the knowledge of the target distribution, up to a multiplicative constant, is required

$$p(x) \leq Mf(x) \tag{1.17}$$

moreover the ratio between the instrumental and the target distributions must be bounded. The algorithm consists in generating $X \sim f(X)$ and $U \sim \mathcal{U}_{[0,1]}$ and to accept the proposed value X if $U \leq p(X)/Mg(X)$. If reject a new pair (X, U) must be simulated. All the accepted values follow the target distribution $p(x)$ (see Casella and Robert [8]). In the literature many extensions of the basic Accept/Reject algorithm have been proposed. For example, the *Envelope Accept/Reject* method (also called *Extended Rejection Method*, see for example Press et al. [5] and Rubinstein [9]), is particularly useful when evaluating the target distribution requires high computing time. By means of a second auxiliary distribution the evaluation of the target distribution is avoided. Another extension is the *Adaptive Rejection Sampling*, which works for log-concave densities and represents a universal sampling method (see Casella and Robert [8]).

We show, through the example (1.2.2), how importance sampling and Accept-Reject algorithms apply to the simulation of a Student-t distribution.

Example 1.2.2 - *A comparison between Monte Carlo estimators*

The Student-t is one of the most used distribution in non-Gaussian financial models, due to the flexibility in modelling the tail behaviour. In this dissertation Student-t has been used in a portfolio model with shortfall probability constraints (see Chapter 4) and in a stochastic volatility model with jumps (see Chapter 6). A general way to simulate the distribution can be found in Ripley [6], see also Chapter 4. In this example we show and compare alternative methods to simulate from the Student-t distribution. Consider a Student-t, $\mathcal{T}(\nu, \mu, \sigma)$, with $\mu = 0$, $\sigma = 1$ and degrees of

freedom, ν. The p.d.f is

$$p(x) = \frac{\Gamma(\frac{\nu+1}{2})/\Gamma(\frac{\nu}{2})}{(\nu\pi)^{1/2}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{1.18}$$

and the instrumental function (Cauchy distribution, $\mathcal{C}(0,1)$)

$$q(x) = \frac{1}{\pi}\left(1 + x^2\right)^{-1} \tag{1.19}$$

Note that the ratio $\phi(x) = p(x)/q(x)$ is bounded

$$\begin{aligned} \phi(x) &= \frac{p(x)}{q(x)} = \frac{\Gamma(\frac{\nu+1}{2})/\Gamma(\frac{\nu}{2})}{(\nu\pi)^{1/2}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \pi(1+x^2) = \\ &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\pi}\nu^{\frac{\nu}{2}} \frac{1+x^2}{(\nu+x^2)^{\frac{\nu+1}{2}}} \leq \sup_{x\in\mathbb{R}} \phi(x) = \\ &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{2\sqrt{\pi}\nu^{\frac{\nu}{2}}}{(\nu+1)^{\frac{\nu+1}{2}}} = M(\nu) < +\infty \end{aligned} \tag{1.20}$$

For $\nu = 3$ and $\nu = 7$ the upper bound $M(\nu)$ is equal to $3\sqrt{3}/4$ and $7^4/(2^7 5\sqrt{7})$ respectively.

We conclude that importance weights and the variance of the associated Monte Carlo estimator are finite. Moreover under this condition Accept-Reject algorithm also applies. In particular, for the following functions

$$h_1(x) = x, \quad h_2(x) = |x|^{0.1}, \quad h_3 = \mathbb{I}_{x>1.96} \tag{1.21}$$

we compare the *Accept/Reject Estimator*, δ^{AR}, with the *Importance Sampling Estimator*, δ^{IS}, and with the *Rescaled Importance Sampling Estimator*, δ^{ISW}, which are defined as follow

$$\delta^{AR} = \frac{1}{n}\sum_{i=1}^{n} h\left(Z_i\right) \tag{1.22}$$

$$\delta^{IS} = \frac{1}{n+t-1}\sum_{i=1}^{n+t-1} \frac{p\left(Z_i\right)}{q\left(Z_i\right)} h\left(Z_i\right) \tag{1.23}$$

$$\delta^{ISW} = \frac{\sum_{i=1}^{n+t-1} \frac{p(Z_i)}{q(Z_i)} h\left(Z_i\right)}{\sum_{i=1}^{n+t-1} p\left(Z_i\right)/q\left(Z_i\right)} \tag{1.24}$$

Table 1.1: A comparison between Importance Sampling and Accept-Reject estimators, on random values from Student-t with 3 and 7 degrees of freedom.

ν		$\mathbb{E}(h_1)$	$\mathbb{E}(h_2)$	$\mathbb{E}(h_3)$	$\mathbb{V}(h_1)$	$\mathbb{V}(h_2)$	$\mathbb{V}(h_3)$
3	δ^{AR}	0.015346	0.961646	0.073609	2.647580	0.012189	0.068199
	δ^{IS}	0.005401	0.967645	0.070808	1.723947	0.130797	0.047196
	δ^{ISW}	0.005363	0.960764	0.070304	$1.7 \cdot 10^{-8}$	$1.3 \cdot 10^{-9}$	$0.5 \cdot 10^{-9}$
7	δ^{AR}	0.016674	0.950258	0.048152	1.412509	0.010560	0.045840
	δ^{IS}	0.002970	0.957760	0.044096	1.273078	0.217395	0.027264
	δ^{ISW}	0.002946	0.949761	0.043727	$1.3 \cdot 10^{-8}$	$2.1 \cdot 10^{-9}$	$0.3 \cdot 10^{-9}$

In order to simulate from the instrumental distribution $q(x)$ we use the inverse c.d.f. of the Cauchy distribution, $x = F^{-}(u) = \tan(\pi(u - 0.5))$, with $u \sim U_{[0,1]}$. The uniform pseudo-random numbers come from a *mixed linear congruential* algorithm. Table 1.1 contains the results with $n + t = 10,000$ simulated values for the Importance Sampling estimators and with $n = 10,000 - t$ simulated values for the Accept/Reject estimator.
We conclude that, in this example, rescaled estimators perform better, in terms of efficiency, than simple importance sampling and Accept/Reject estimators.

□

To conclude this section, note that if the distribution of interest $p(x)$ is too complex, then it can be difficult to find an easily-sampled distribution which is sufficiently similar to $p(x)$. In that case both importance sampling and Accept/Reject methods are difficult to apply and can conduce to Monte Carlo estimators with a high variance. Thus alternative methods, like *Markov Chain Monte Carlo (MCMC)*, are needed.

1.3 Markov Chain Monte Carlo Methods

Markov Chain Monte Carlo (MCMC) methods represent valid alternative sampling techniques that allow to avoid the complexity problem previously discussed, by drawing non-i.i.d. samples from the desired distribution, $p(x)$. These samples are produced using a discrete stochastic process, called *Markov Chain*. Moreover, in this dissertation, Markov chains are used also in a financial modelling perspective, for

capturing jumps in stochastic volatility process (see Chapter 6). Due to the central role of this kind of process, in the following we briefly review some basic properties and results on Markov chains.

1.3.1 Markov Chains

MCMC algorithms are based on the specification of the transition kernel of a Markov Chain. Thus the properties of this kind of stochastic processes are necessary in order to understand MCMC simulation algorithms and in order to study their convergence.

Consider a sequence of $\mathcal{X}$-valued random samples $\{X_t\}_{t=0}^{\infty}$ on the state space $\mathcal{X}$. The elements of $\mathcal{X}$ represent the possible states of a dynamic system, with X_t representing the random state at time t. The time evolution of the system is described trough a *transition kernel*.

Definition 1.3.1. *(Transition kernel, see Casella and Robert [8])*
A transition kernel is a function K defined on $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ such that:

1. $\forall\, x \in \mathcal{X}$, $K(x, \cdot)$ *is a probability measure;*

2. $\forall\, A \in \mathcal{B}(\mathcal{X})$, $K(\cdot, A)$ *is measurable.*

The transition kernel $K(x, x')$ for a Markov chain, defined on the continuous state space $\mathcal{X}$, denotes the conditional probability of transition: $\mathbb{P}(X \in A|x) = \int_A K(x, x')dx'$, which is the probability of the next state being in A given that the current state is x. In the discrete case, the transition kernel is a transition matrix with elements: $P_{x,x'} = \mathbb{P}(X_{t+1} = x'|X_t = x)$, for $x, x' \in \mathcal{X}$.

At time $t = 0$ the state X_0 has the *initial distribution* $\pi(X_0)$. Through the kernel and the initial distribution it is possible to define a stochastic process (Markov chain), which obeys the first-order Markov property in time. That is, given the current state of the process, the distribution of the next state is independent of the past values of the process.

Definition 1.3.2. *(Markov chain)*
Given a transition kernel $K(\cdot, \cdot)$, a sequence of random variables $\{X_t\}_{t=0}^{\infty}$ is a Markov chain if the conditional distribution of X_{t+1} given the past values of chain is the same

as the distribution of X_{t+1} given x_t, that is:

$$\mathbb{P}(X_{t+1} \in A | x_0, x_1, \ldots, x_t) = \mathbb{P}(X_{t+1} \in A | x_t) = \int_A K(x_t, dx) \tag{1.25}$$

for all $x_t \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$.

The transition kernel needs to satisfy some conditions in order to ensure that the associated Markov chain has a desired *invariant distribution*, which is unique and reachable from any starting point of the process. The stability properties of a Markov chain are thus strictly related to the definition of invariant measure.

Definition 1.3.3. *(Invariant distribution)*
We say that a Markov chain has σ-finite invariant measure (or stationary measure) π for the transition kernel K, if

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\pi(dx), \quad \forall B \in \mathcal{B}(\mathcal{X}) \tag{1.26}$$

Once a Markov chain has reached the stationary distribution, all subsequent samples will also be from that distribution. Note however that uniqueness of the invariant distribution is not guaranteed and the Markov chain is not guaranteed to be able to reach it from every starting point.

The property of *irreducibility* plays a crucial role in the stability of a Markov chain and in the convergence of the Markov chain algorithms.

Definition 1.3.4. *(Irreducibility)*
A Markov chain is irreducible with respect a measure ϕ, if $\forall A \in \mathcal{B}(\mathcal{X})$ with $\phi(A) > 0$, there exists n such that $K^n(x, A) > 0, \forall x \in \mathcal{X}$.

In previous definition we set $K^1(x, A) = K(x, A)$, while $K^n(x, A)$ denotes the n-times composition of the transition kernel, *i.e.* $K^n(x, A) = \int_{\mathcal{X}} K^{n-1}(x', A)K(x, dx')$. In other words, a Markov chain is irreducible, if it is possible to eventually get to any other state, from any state to which ϕ assigns positive probability measure. Moreover if $n = 1$ the chain is defined *strongly ϕ-irreducible*.

Define the number of passages of X_t in $A \in \mathcal{B}(\mathcal{X})$ as follow

$$\eta_A = \sum_{t=1}^{\infty} \mathbb{I}_A(X_t) \tag{1.27}$$

then good stability of the Markov chain can be guaranteed by requiring the chain to revisit any subset A of the state space infinitely often, at least from almost all starting points. This property can be formalized through the definition of *recurrence*.

Definition 1.3.5. *(Recurrence)*
An Markov chain $\{X_t\}$ is recurrent if there exists a measure φ such that $\{X_t\}$ is irreducible and $\forall A \in \mathcal{B}(\mathcal{X})$ such that $\varphi(A) > 0$, $\mathbb{E}_x\{\eta_A\} = \infty$ for every $x \in A$.

An irreducible and recurrent chain is *positive recurrent* if it has an invariant probability measure. Otherwise it is *null recurrent*. Note that the recurrence property requires an infinite average number of visits to every small set of $\mathcal{X}$. It is possible to ensure that each state is revisited infinitely often for every path of the Markov chain by the definition of *Harris recurrence*.

Definition 1.3.6. *A set A is Harris recurrent if $P_x(\eta_A = \infty) = 1$ for all $x \in A$. The chain $\{X_t\}$ is Harris recurrent if there exists a measure ψ such that $\{X_t\}$ is ψ-irreducible and for every set A with $\psi(A) > 0$, where A is Harris recurrent.*

Before introducing some important convergence results on Markov chains we give the following definition.

Definition 1.3.7. *(Aperiodicity)*
An irreducible Markov chain is aperiodic if $\forall x \in \mathcal{X}$, g.c.d.$\{t > 0, K^n(x, x) > 0\} = 1$.

The first result is the uniqueness of the invariant distribution and the ergodicity of the chain, that is, the independence on the starting point of the chain.

Theorem 1.3.1. *(Ergodic chain)*
If the Markov chain, $\{X_t\}$ is aperiodic an positive recurrent, then its invariant distribution $\pi(\cdot)$ is the unique probability distribution satisfying $\pi(x') = \int_{\mathcal{X}} \pi(x) K^n(x', x) dx$, $\forall x' \in \mathcal{X}$ and $\forall n \geq 0$ and the chain is said to be ergodic.

For other definitions of ergodicity, like uniform ergodicity and geometric ergodicity, see Casella and Robert [8].
The following convergence result is particularly important when approximating expectation by means of a MCMC simulation algorithm. We denote with $S_n(h) = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$, a sum of transforms of random variables. The following result holds.

Theorem 1.3.2. *(Law of Large Numbers for Markov chains)*
If the chain $\{X_t\}$ has σ-finite invariant measure π, the following statements are equivalent

1. *If $f, g \in L^1(\pi)$ with $\int g(x)d\pi(x) \neq 0$ then*

$$\lim_{n\longrightarrow\infty} \frac{S_n(f)}{S_n(g)} = \frac{\int f(x)d\pi(x)}{\int g(x)d\pi(x)} \tag{1.28}$$

2. *The Markov chain $\{X_t\}$ is Harris recurrent.*

Observe that the previous result does not require the chain to be positive recurrent but only to be positive and this represents a strong results, because the convergence is assured even if the invariant measure π is not finite.

Most Markov chains used in practice satisfy a more restrictive condition, that is the chain has to be *time reversible*.

Theorem 1.3.3. *(Time reversible Markov chains)*
A Markov chain is time reversible with respect to distribution π if and only if $\forall B, C \subset \mathcal{X}$ the detailed balance condition

$$\int_B \int_C \pi(x)K(x', x)dx'dx = \int_C \int_B \pi(x')K(x, x')dxdx' \tag{1.29}$$

is satisfied. Equivalently, the detailed balance condition holds if and only if

$$\pi(dx)K(x, dx') = \pi(dx')K(x', dx), \quad \forall x, x' \in \mathcal{X} \tag{1.30}$$

Further details on Markov chains can be found for example in Meyn and Tweedie [4], other theoretical results on convergence can be found in Tierney [10], while Casella and Robert [8] also provides some techniques for monitoring convergence.

1.3.2 Metropolis Hastings

Metropolis-Hastings (M.-H.) algorithm allows to sample from the *target distribution* $p(x)$, known up to a normalising constant, using an ergodic Markov chain $\{X_t\}$

with stationary distribution $p(x)$. Through a *proposal distribution* $q(x|y)$ the M.-H. algorithm defines a Markov chain with the following transition kernel

$$K(x,y) = \rho(x,y)q(y|x) + (1 - r(x))\delta_x(y) \tag{1.31}$$

where $r = \int \rho(x,y)q(dy|x)$ and $\delta_x(y)$ is the Dirac measure. The steps of the M.-H. algorithm are

Algorithm 1.3.1. *(Metropolis-Hastings)*
Given $x^{(t)}$ *such that* $p(x^{(t)}) > 0$

1. *Draw a value* y_t *from* $Y_t \sim q(y|x^{(t)})$

2. *Take*

$$X^{(t+1)} = \begin{cases} y_t & \textit{with probability} \quad \rho(x^{(t)}, y_t) \\ x^{(t)} & \textit{with probability} \quad 1 - \rho(x^{(t)}, y_t) \end{cases} \tag{1.32}$$

   where

$$\rho(x,y) = \min\left\{ \frac{p(y)}{p(x)} \frac{q(x|y)}{q(y|x)}, 1 \right\} \tag{1.33}$$

3. *Jump to step 1 incrementing* t *of* 1.

It can be shown that the M.-H. chain has stationary or invariant distribution p. In order to assure the chain to be sampling values from p it is necessary to impose some minimal regularity conditions on both the target density, p and the proposal distribution q. In what follows, it will be always supposed that the support of p is connected and that

$$supp\ p(x) \subseteq \bigcup_{x \in\ supp\ p(x)} supp\ q(\cdot|x) \tag{1.34}$$

If the support is not connected, there would be some subsets of the support that could never be visited. Recalling the definition 1.3.3 of invariant probability measure for a Markov chain and theorem (1.3.3) on the time reversibility, it is sufficient to check the M.-H. chain satisfies the detailed balance condition. For the first term of the transition kernel, the detailed balance condition holds, because

$$\begin{aligned}
&\rho(x,y)q(y|x)p(x) = \\
&= \min\left\{\frac{p(y)}{p(x)}\frac{q(x|y)}{q(y|x)}, 1\right\} q(y|x)p(x) = \\
&= \min\left\{p(y)q(x|y), p(x)q(y|x), 1\right\} = \\
&= \min\left\{1, \frac{p(x)}{p(y)}\frac{q(y|x)}{q(x|y)}\right\} q(x|y)p(y) = \\
&= \rho(y,x)q(x|y)p(y)
\end{aligned} \tag{1.35}$$

while the second term of the transition kernel satisfies to

$$(1-\rho(x))\delta_x(y)p(x) = (1-\rho(y))\delta_y(x)p(y) \tag{1.36}$$

since due to the multiplication with the Dirac measure, both sides of the equation are not zero only when $x = y$.
We conclude the section with a convergence result for the Metropolis-Hastings algorithm.

Theorem 1.3.4. *Given a M.-H. Markov chain, $\{X_t\}$, if $q(y|x) > 0$ for every $(x,y) \in \mathcal{X} \times \mathcal{X}$ and*

1. *if $h \in L^1(p)$, then*

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} h(X^{(t)}) = \int h(x)p(x)dx \tag{1.37}$$

   almost everywhere with respect to p;

2. *if the M.-H. satisfies to $P\left(p(X^{(t)})q(Y^{(t)}|X^{(t)}) \leq p(Y^{(t)})q(X^{(t)}|Y^{(t)})\right) < 1$, which means that the proposal can not be the transition kernel of a reversible chain with invariant distribution p, then*

$$\lim_{n\to\infty} \sup_{B} \left| \int K^n(x,B)\mu(dx) - p(B) \right| = 0 \tag{1.38}$$

   for every initial distribution μ.

For a proof of the theorem and for further details on Markov Chain in connection with Monte Carlo method see Tierney [10] and Robert and Casella [8].

The choice of the proposal distribution plays a crucial role in the implementation of the M.-H. algorithm. If the proposal gives low probability to subset of the support of p where p put high probability, then the convergence of the M.-H. to p is very slow. There are many possible choices for the proposal distribution. In this dissertation (see Chapter 5) we use all the following kind of proposals for the implementation of a Metropolis-Hastings step in a Gibbs sampler algorithm for mixture of stable distribution.

When $q(y|X^{(t)}) = q(y)$ the M.-H. algorithm is called *independent Metropolis-Hastings algorithm* and is similar to the Accept/Reject method. However although the proposals are independent, the resulting simulated values are not independent because the probability of acceptance of $Y^{(t)}$ depends on the previous value of the M.-H. chain, $x^{(t)}$.

When the proposal is $q(y|X^{(t)}) = g(y - X^{(t)})$, then the M.-H. algorithm is called *random walk Metropolis-Hastings* because the resulting M.-H. chain is a random walk on $\mathcal{X}$. In fact the algorithm has the equivalent representation

$$Y^{(t)} = X^{(t)} + \varepsilon_t \tag{1.39}$$

where ε_t follows a distribution g independent of $x^{(t)}$.
In random walk M.-H. algorithms it is possible to control the variance of the error term and by choosing the variance of the distribution g it is also possible to obtain the desired level of acceptance rate.

1.3.3 Gibbs Sampling

The Gibbs sampler has been introduced in image processing by Geman and Geman [3] and it is a method of construction of a Markov Chain $\{\mathbf{X}^{(t)}\}_{t=0}^{\infty}$ with multivariate stationary distribution $p(\mathbf{X})$, where $\mathbf{X} \in \mathcal{X}$ is a random vector. In Bayesian inference, this simulation method is particularly useful when the posterior density is defined on a high dimension space. If the random vector $\mathbf{X}$ can be written as

$\mathbf{X} = (X_1, \ldots, X_r)$ and if we can simulate from the full conditional densities:

$$(X_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_r) \sim p_i(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_r) \tag{1.40}$$

then the Gibbs sampling applies the associated transition kernel, from $\mathbf{X}^{(t)}$ to $\mathbf{X}^{(t+1)}$, is described by the following algorithm

Algorithm 1.3.2. *(Gibbs Sampler)*
Given the state $\mathbf{X}^{(t)} = \mathbf{x}^{(t)}$ *at time* t*, generate the state* $\mathbf{X}^{(t+1)}$ *as follows:*

1. $X_1^{(t+1)} \sim p(x_1|x_2^{(t)}, \ldots, x_r^{(t)})$
2. $X_2^{(t+1)} \sim p(x_2|x_1^{(t+1)}, x_3^{(t)}, \ldots, x_r^{(t)})$
3. ...
4. $X_i^{(t+1)} \sim p(x_i|x_1^{(t+1)}, x_2^{(t)}, \ldots, x_{i-1}^{(t)}, x_{i+1}^{(t)}, \ldots, x_r^{(t)})$
5. ...
6. $X_r^{(t+1)} \sim p(x_r|x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_{r-1}^{(t+1)})$

Under some regularity conditions the Markov chain produced by the algorithm converges to the desired stationary distribution (see Robert and Casella [8]).

Each full conditional distribution of the Gibbs sampler can be simulated with a standard method like transformation method, accept/reject or importance sampling. If the full conditional is too much complex and standard methods do not apply, then a Metropolis-Hastings step can be used. The resulting algorithm is called *Hybrid Sampler* (see Robert and Casella [8] for further details on this algorithm).

Bibliography

[1] Boyle P. (1977), Options: a Monte Carlo Approach, *Journal of Financial Economics*, 4, pp.323-338.

[2] Dupire B. (1998), *Monte Carlo*, Methodologies and Applications for Pricing and Risk Management, Risk Books Ed.

[3] Geman S., Geman D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 6, pp. 721-741.

[4] Meyn S.P., Tweedie R.L. (1993), *Markov Chains and Stochastic Stability*, Springer-Verlag, New York.

[5] Press W., Teukolsky S., Vetterling T. and Flannery B.P. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge.

[6] Ripley B.D. (1987), *Stochastic Simulation*, J. Wiley, New York.

[7] Robert C.P. (2001), *The Bayesian Choice, 2nd ed.* Springer Verlag, New York.

[8] Robert C.P. and Casella G. (1999), *Monte Carlo Statistical Methods*,Springer Verlag, New York.

[9] Rubinstein R. Y. (1981), *Simulation and the Monte Carlo Method*, John Wiley & Sons.

[10] Tierney L. (1994), Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, 22, pp. 1701-1786.

[11] Ulam S.M. and Von Neumann J. (1947), On combination of stochastic and deterministic processes (abstract), *Bull. Amer. Math. Soc.*, 53, pp.1120.

Chapter 2

Monte Carlo Methods for Dynamic Models

2.1 Introduction

The analysis of dynamic phenomena, which evolve over time is a common problem to many fields like engineering, physics, biology, statistics, economics and finance. A time varying system can be represented through a *dynamic model*, which is constituted by an observable component and an unobservable internal state. The hidden state vector represents the desired information that we want to extrapolate from the observations.

Several kinds of dynamic models have been proposed in the literature for time series analysis and many approaches have been used for the estimation of these models. The seminal work of Kalman [23] and Kalman and Bucy [24] introduces filtering techniques (Kalman-Bucy filter) for continuous valued, linear and Gaussian dynamic systems. Another relevant work on dynamic model analysis is due to Maybeck [35], [36], [37]. He motivates the use of stochastic dynamic systems in engineering and examines filtering, smoothing and estimation problems for continuous state space models, in both a continuous and a discrete time framework.

[1]Part of this work is in Billio M. Casarin R. and Sartore D., (2003), "Bayesian inference in dynamic models with latent factors", forthcoming, *Monography of Official Statistics*, edited by Eurostat. Presented at the 4th Colloquium on Modern Tools for Business Cycle Analysis in Luxembourg on 20-22 October 2003.

Moreover Harvey [21] extensively studies state space representation of dynamic models for time series analysis and treats the use of Kalman filter for states and parameters estimation, in continuous state space setting. Hamilton [20] analyzes several kinds of time series models and in particular introduces a filter (Hamilton-Kitagawa filter) for discrete time and discrete valued dynamic system. This filter can be used for dynamic models with a finite number of state values.

Bauwens, Lubrano and Richard [2] compare maximum likelihood inference with Bayesian inference on static and dynamic econometric models. Harrison and West [22] treat the problem of the dynamic model estimation in a Bayesian perspective. They give standard filtering and smoothing equations for Gaussian linear models and investigate the estimation problem for conditionally Gaussian linear models and for general nonlinear and non-Gaussian models. They review some Markov Chain Monte Carlo simulation techniques for filtering and smoothing the state vector and for estimating parameters. Moreover, also the problem of processing data sequentially has been examined through the use of the adaptive importance sampling algorithm. Kim and Nelson [26] analyze Monte Carlo simulation methods for non-linear discrete valued model (switching regimes models). Recently, Durbin and Koopman [15] propose an updated review on Markov Chain Monte Carlo methods for estimation of general dynamic models, with both a Bayesian and a maximum likelihood approach.

Sequential simulation methods for filtering and smoothing in general dynamic models have been recently developed to overcome some problems of the traditional MCMC methods. As pointed out by Liu and Chen [31], Gibbs sampler is less attractive when we consider on-line data processing. Furthermore Gibbs sampler may be inefficient when simulated states are very sticky and the sampler has difficulties to move in the state space. In these situations, the use of sequential Monte Carlo techniques and in particular of particle filter algorithms may result more efficient. Doucet, Freitas and Gordon [12] provide the state of the art on sequential Monte Carlo methods. They discuss both applications and theoretical convergence results for these algorithms, with special attention to particle filters.

In the economic and financial literature (for example on business cycle analysis and on asset pricing), dynamic models are used to capture some well known features of the economic time series: comovement, heavy tails and asymmetry. Comovement

of economic variables can be modelled by means of dynamic factor models. Heavy tails and asymmetry denotes an heterogeneous dynamic of the economic variable. If the behavior of the economic or financial time series depends on the phase of the economic cycle or of the financial market, then asymmetry arises. Moreover if the frequency of large deviations from the mean of the economic or financial variable is high, then heavy tails appear. In order to model heavy tails, stochastic volatility models and non-Gaussian innovations can be used. We refer to Chapter 6 for a review on stochastic volatility models and briefly review in the following the literature on economic cycle models. In order to capture asymmetry Goldfeld and Quandt [17] introduced Markov Switching (MS) models for serially uncorrelated data, while Hamilton [19] applies MS to serially correlated time series. In their models parameter are allowed to depend on the hidden state of the economic cycle. This state may assume only two values, which are interpreted as: positive growth trend and negative growth trend.

A different way to model asymmetry in time series can be found in Tong [50] and Potter [41]. They introduce threshold autoregressive models (TAR). In this class of model, the phase of the economic cycle is determined by means of a threshold on the level of the observable variable. Parameters depend on the phase of the cycle.

All above cited approaches and in particular the original work due to Hamilton [19], have been successively extended in many directions.

Kim [25] applies Markov Switching to dynamic linear model in a Bayesian approach. Kim and Nelson [26] analyze general Markov Switching dynamic models and provide Bayesian inference tools together with MCMC simulation techniques.

In his switching model Hamilton [19] assumes that the growth rate of real output depend by an unobserved Markov switching variable. This variable can assume only states accordingly to the two phases of the business cycle: positive trend growth and negative trend growth. This hypothesis seems to be too restrictive when looking at data. In particular transitory and permanent components characterize recession phases. Thus Kim and Murray [27] and Kim and Piger [28] divide economic cycle in three phases: recession, high-growth and normal-growth.

Another kind of extension to the basic model of Hamilton [19] concerns the duration of the phases of the business cycle. Sichel [46], Durland and McCurdy [14], Watson [51] and Diebold and Rudebusch [11] assume that the transition probability

of the Markov switching problem depend on the duration of the current phase of the cycle.

Finally, multivariate extensions to the Hamilton [19] univariate MS model have been suggested by Diebold and Rudebusch [11] and Krolzig [30].

The chapter is structured as follows. Section 2.2 introduces the general representation of a dynamic model in a Bayesian framework and deals with stochastic volatility models and with conditionally normal linear models, which do not admit analytical filtering and smoothing densities. Section 2.3 reviews simulation based methods for inference. In particular the section reviews MCMC methods, presents an adaptive importance sampling algorithm and discusses particle filter algorithms. Through some examples, applications of the particle filter to stochastic volatility and business cycle models are provided. Section 2.4 concludes.

2.2 Bayesian Dynamic Models

In the following we give a quite general formulation of a probabilistic *dynamic model* and we obtain some fundamental relations for Bayesian inference on it. This definition of dynamic model would be general enough to include time series models analyzed in Kalman [23], Hamilton [20], Harrison and West [22] and in Doucet, Freitas and Gordon [12]. Throughout this chapter, we use a notation similar to that one commonly used in particle filter literature (see Doucet, Freitas and Gordon [12]).

We denote by $\{\mathbf{x}_t;\, t \in \mathbb{N}\}$, $\mathbf{x}_t \in \mathcal{X}$, the hidden state vectors of the system, by $\{\mathbf{y}_t;\, t \in \mathbb{N}_0\}$, $\mathbf{y}_t \in \mathcal{Y}$, the observable variables and by $\theta \in \Theta$ the parameter vector of the model. We assume that state space, observation space and parameter space respectively are $\mathcal{X} \subset \mathbb{R}^{n_x}$, $\mathcal{Y} \subset \mathbb{R}^{n_y}$ and $\Theta \subset \mathbb{R}^{n_\theta}$. n_x, n_y and n_θ represent the dimensions of the state vector, of the observable variable and of the parameter vector respectively.

The main advantage in using the general Bayesian state space representation of a dynamic model, is that it accounts also for nonlinear and non-Gaussian models. The Bayesian state space representation is given by an *initial distribution* $p(\mathbf{x}_0|\theta)$, a *measurement density* $p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)$ and a *transition density* $p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}, \theta)$. The dynamic model is

$$\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta) \tag{2.1}$$

$$p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}, \theta) \tag{2.2}$$

$$x_0 \sim p(\mathbf{x}_0|\theta)\,, \quad \text{with } t = 1, \ldots, T \tag{2.3}$$

where $p(\mathbf{x}_0|\theta)$ can be interpreted as the prior distribution on the initial state of the system.
By $\mathbf{x}_{0:t} \stackrel{\Delta}{=} (\mathbf{x}_0, \ldots, \mathbf{x}_t)$ and by $\mathbf{y}_{1:t} \stackrel{\Delta}{=} (\mathbf{y}_1, \ldots, \mathbf{y}_t)$ we denote respectively the collection of state vectors and of observable vectors, up to time t. We denote by $\mathbf{x}_{-t} \stackrel{\Delta}{=} (\mathbf{x}_0, \ldots, \mathbf{x}_{t-1}, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_T)$ the collection of all the state vectors without the t-th element. The same notation is used also for the observable variable and for the parameter vectors.

If the transition density depends on the past, only through the last value of the hidden state vector, the dynamic model is defined first-order *Markovian*. In this case the system becomes

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta) \tag{2.4}$$

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta) \tag{2.5}$$

$$x_0 \sim p(\mathbf{x}_0|\theta)\,, \quad \text{with } t = 1, \ldots, T. \tag{2.6}$$

Assuming that the first-order Markov property holds is not restrictive because a Markov model of order p can always be rewritten as a first-order Markovian model.

The stochastic volatility model given in the example 2.2.1, is a dynamic model, which are now widely used in finance. The chapter 6 of this dissertation also deals with stochastic volatility models, focusing on simulation based inference tools for this kind of models.

Example 2.2.1 - *(Stochastic Volatility Models)*
Two of the main features of the financial time series are time varying volatility and clustering phenomena in volatility. Thus stochastic volatility models have been introduced, in order to account for these features. An example of stochastic log-volatility model is

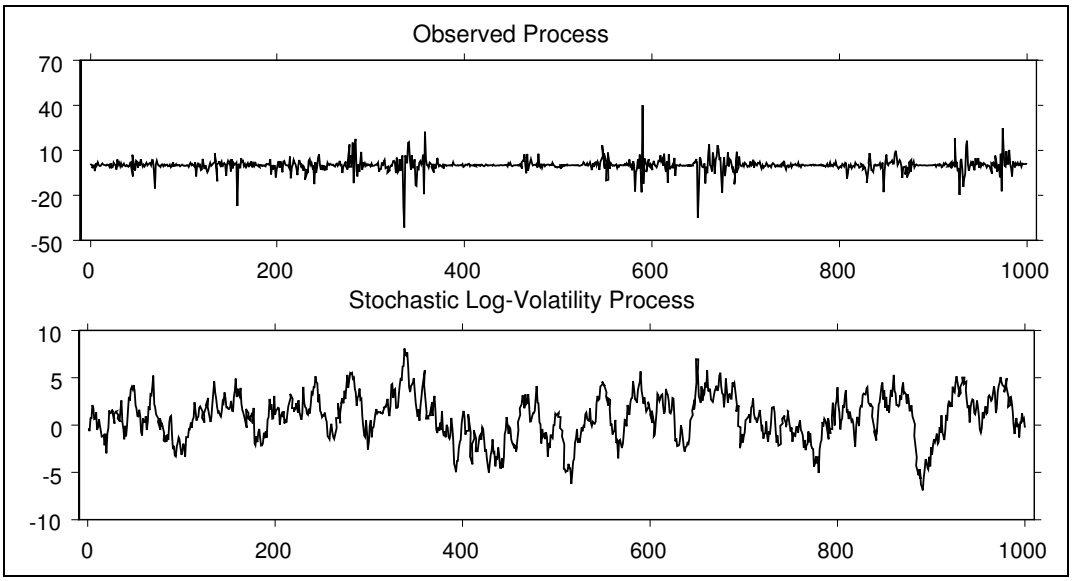

Figure 2.1: Simulated paths of 1,000 observations from the observable process y_t, with time varying volatility and of the stochastic volatility process h_t. We simulate the model given in Example 2.2.1, with $\phi = 0.9$, $\nu = 0.1$ and $\sigma^2 = 1$.

$$y_t \sim \mathcal{N}\left(0, e^{h_t}\right) \tag{2.7}$$

$$h_t \sim \mathcal{N}\left(\nu + \phi h_{t-1}, \sigma^2\right) \tag{2.8}$$

$$h_0 \sim \mathcal{N}\left(0, \sigma^2\right), \quad \text{with } t = 1, \ldots, T \tag{2.9}$$

where y_t is the observable process, with time varying volatility and h_t represents the stochastic log-volatility process. In Fig. 2.1 we exhibit a simulated path of the observable process y_t and of the stochastic volatility process h_t.

□

In the following we discuss the three main issues which arise when making inference on a dynamic model, i.e.: filtering, predicting and smoothing. We present general solutions to these problems, but note that, without further assumptions on the densities, which characterize the dynamic model, these solutions are not yet analytical.

2.2.1 State Estimation

We treat here the problem of estimation of the hidden state vector when parameters are known. We are interested in estimating the density $p(\mathbf{x}_t|\mathbf{y}_{1:s}, \theta)$. If $t = s$ the density of interest is called *filtering density*, if $t < s$ it is called *smoothing density*, finally if $t > s$ it is called *prediction density*.

For the dynamic model given in equations (2.4), (2.5) and (2.6) we assume that at time t the density $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \theta)$ is known. Observe that if $t = 1$ the density $p(\mathbf{x}_0|\mathbf{y}_0, \theta) = p(\mathbf{x}_0|\theta)$ is the initial distribution of the dynamic model.

Applying the Chapman-Kolmogorov transition density, we obtain the one step ahead *prediction density*

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \theta)d\mathbf{x}_{t-1} \tag{2.10}$$

As the new observation $\mathbf{y}_t$ becomes available, it is possible, using the Bayes theorem, to update the prediction density and to filter the current state of the system. The *filtering density* is

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) = \frac{p(\mathbf{y}_t, \mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta)} = \frac{p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta)} \tag{2.11}$$

where $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)$ is the prediction density determined at the previous step and the density at the denominator is the marginal of the current state and observable variable joint density

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta) = \int_{\mathcal{X}} p(\mathbf{y}_t, \mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)d\mathbf{x}_t = \int_{\mathcal{X}} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)d\mathbf{x}_t. \tag{2.12}$$

At each date t, it is possible to determine the K-steps-ahead prediction density, conditional on the available information $\mathbf{y}_{1:t}$. Given the dynamic model described by equations (2.4), (2.5) and (2.6), the *K-steps-ahead prediction density* of the state vector $\mathbf{x}_t$ can be evaluated iteratively. The first step is

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{1:t}, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)d\mathbf{x}_t \tag{2.13}$$

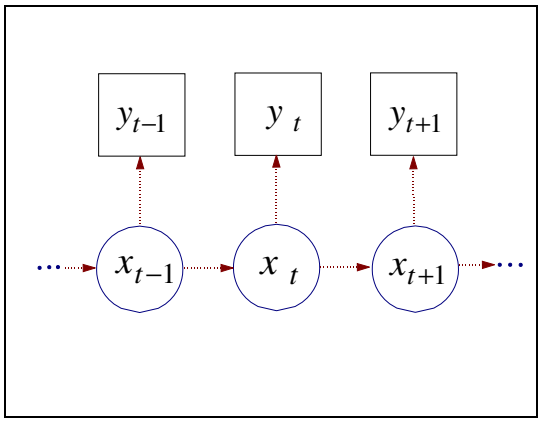

Figure 2.2: The causality structure of a Markov dynamic model with hidden states is represented by means of a *Directed Acyclic Diagram* (DAG). A box around the variable indicates the variable is known, while a circle indicates a hidden variable.

and the k-th step, with $k = 1, \ldots, K$, is

$$p(\mathbf{x}_{t+k}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+k}|\mathbf{x}_{t+k-1}, \mathbf{y}_{1:t}, \theta) p(\mathbf{x}_{t+k-1}|\mathbf{y}_{1:t}, \theta) d\mathbf{x}_{t+k-1} \quad (2.14)$$

where

$$p(\mathbf{x}_{t+k}|\mathbf{x}_{t+k-1}, \mathbf{y}_{1:t}, \theta) = \int_{\mathcal{Y}^{k-1}} p(\mathbf{x}_{t+k}|\mathbf{x}_{t+k-1}, \mathbf{y}_{1:t+k-1}, \theta) p(d\mathbf{y}_{t+1:t+k-1}|\mathbf{y}_{1:t}, \theta) \quad (2.15)$$

where $\mathcal{Y}^k = \otimes_{i=1}^{k} \mathcal{Y}_i$ is the k-times Cartesian product of the state space. Similarly, the *K-steps-ahead prediction density* of the observable variable $\mathbf{y}_{t+K}$ conditional on the information available at time t is determined as follow

$$p(\mathbf{y}_{t+K}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{Y}} p(\mathbf{y}_{t+K}|\mathbf{x}_{t+K}, \mathbf{y}_{1:t+K-1}, \theta) p(d\mathbf{y}_{t+1:t+K-1}|\mathbf{y}_{1:t}, \theta) p(d\mathbf{x}_{t+K}|\mathbf{y}_{1:t}, \theta) \quad (2.16)$$

Due to the high number of integrals that must be solved, previous densities may be difficult to evaluate with general dynamics. From a numerical point of view simulation methods, like MCMC algorithms or Particle Filters allow us to overcome these difficulties; while from an analytical point of view to obtain simpler relations we need to introduce some simplifying hypothesis on the dynamics of the model. For example if we assume that the evolution of the dynamic model does not depend

on the past values of the observable variable $\mathbf{y}_{1:t}$, then equations (2.3), (2.5) and (2.6) become

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t, \theta) \tag{2.17}$$

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \theta) \tag{2.18}$$

$$x_0 \sim p(\mathbf{x}_0|\theta)\,, \quad \text{with } t = 1, \ldots, T. \tag{2.19}$$

The causality structure of this model has been represented through the Directed Acyclic Graph (DAG) exhibited in Fig. 2.2. Under previous assumptions the filtering and prediction densities simplify as follows

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \theta) p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}, \theta) d\mathbf{x}_{t-1} \tag{2.20}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) = \frac{p(\mathbf{y}_t|\mathbf{x}_t, \theta) p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta)}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1}, \theta)} \tag{2.21}$$

$$p(\mathbf{x}_{t+K}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+K}|\mathbf{x}_{t+K-1}, \theta) p(\mathbf{x}_{t+K-1}|\mathbf{y}_{1:t}, \theta) d\mathbf{x}_{t+K-1} \tag{2.22}$$

$$p(\mathbf{y}_{t+K}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{y}_{t+K}|\mathbf{x}_{t+K}, \theta) p(\mathbf{x}_{t+K}|\mathbf{y}_{1:t}, \theta) d\mathbf{x}_{t+K}. \tag{2.23}$$

We conclude this section with two important recursive relations. Both these relations can be proved starting from the definition of joint smoothing density and assuming that the Markov property holds.

The first relation is the sequential filtering equation

$$p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T}, \theta) = p(\mathbf{x}_{0:T-1}|\mathbf{y}_{1:T-1}, \theta) \frac{p(\mathbf{y}_T|\mathbf{x}_T, \theta) p(\mathbf{x}_T|\mathbf{x}_{T-1}, \theta)}{p(\mathbf{y}_T|\mathbf{y}_{1:T-1}, \theta)}. \tag{2.24}$$

which is particularly useful when processing data sequentially and it is fundamental in implementing Particle Filter algorithms. A proof of this relation is given in Appendix A.

The second recursive relation provides factorization of the *smoothing density* of the state vectors $\mathbf{x}_{0:T}$, given the information, $\mathbf{y}_{1:T}$, available at time T

$$p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T}, \theta) = p(\mathbf{x}_T|\mathbf{y}_{1:T}, \theta) \prod_{t=0}^{T-1} p(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{y}_{1:t}, \theta). \tag{2.25}$$

See the proof in Appendix A. The proof differs from that one given in Carter and Köhn [6] because we consider the general model in equations (2.4)-(2.6), with transition and measurement densities depending on past values of $\mathbf{y}_t$. Note that the density $p(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{y}_{1:t}, \theta)$, which appears in the joint smoothing density, can be represented through the filtering and the prediction densities

$$p(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{y}_{1:t}, \theta) = \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{1:t}, \theta)p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)}{p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta)}. \tag{2.26}$$

This factorization of the smoothing density is also relevant when inference is carried out through simulation methods. See for example the multi-move Gibbs sampler of Carter and Köhn [6] and the particle filter algorithms.

Only in some well known cases, these filtering densities admit an analytical form. For the normal linear dynamic model, filtering and smoothing density are given by the Kalman filter. See Harrison and West [22] for a Bayesian analysis of the Kalman filter. See Harvey [21] for a frequentist approach to the Kalman filter.

In the next section and in Chapter 6, we introduce some well known classes of dynamic models, which does not admit a tractable analytical representation of the filtering, prediction and smoothing densities. The next section deals with *conditional normal linear models*, which are widely used in the economic and financial analysis (see Kim and Nelson [26]). In Chapter 6 we analyse inference on Markov switching stochastic volatility models.

2.2.2 Conditionally Gaussian Linear Models

A lot of models used in economic and financial literature belong to the class of the conditionally normal dynamic linear models. These are defined as follows

$$\begin{aligned} \mathbf{y}_t &= F(s_t)\mathbf{x}_t + V(s_t)\epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, I) \\ \mathbf{x}_{t+1} &= G(s_t)\mathbf{x}_t + W(s_t)\eta_t \quad \eta_t \sim \mathcal{N}(0, I) \end{aligned} \tag{2.27}$$

where ϵ_t is independent of η_t and where s_t is a sequence of random variables. Looking at the Bayesian literature, Harrison and West [22] call this model *multi-process* model. In the classification proposed by these authors, if $s_t = s_{t-1} = s$, $\forall t$ the

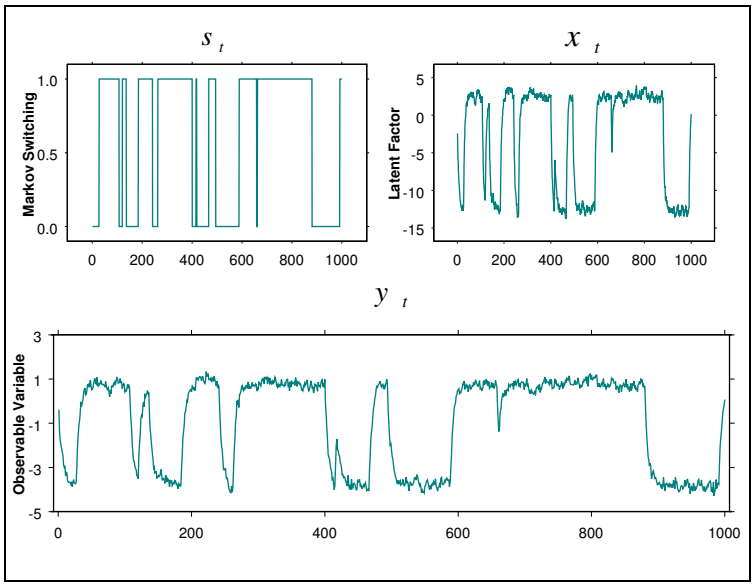

Figure 2.3: Simulation from the Markov switching stochastic trend model given in Example 2.2.2. We set parameters to be $\alpha = 0.3, \sigma_\varepsilon = 0.1, \rho = 0.8, \mu_0 = -2.5, \mu_1 = 0.5$, $\sigma_\eta = 0.1, p_{11} = 0.97, p_{22} = 0.99$.

model is a multi-process of the first kind, while if s_t is a stochastic process, the model is a multi-process of second kind.

Note that if s_t is a discrete time and finite state Markov chain with known transition probabilities, the model is called *jump Markov linear system* or *Markov switching linear model* with parameters evolving over time. In the following we consider an example of Markov switching normal linear model.

Example 2.2.2 - *(Stochastic Latent Factor Model with Markov Switching)*
Many business cycle models can be represented as a Markov switching linear model. Let y_t be the observable variable and x_t the latent factor, which has to be extracted. The switching model is

$$y_t = \alpha x_t + \sigma_\epsilon \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0,1) \tag{2.28}$$

$$x_{t+1} = \mu(s_{t+1}) + \rho\, x_t + \sigma_\eta \eta_{t+1} \qquad \eta_{t+1} \sim \mathcal{N}(0,1) \tag{2.29}$$

$$s_t \sim Markov(\mathbb{P}), \quad \text{with} \qquad s_t \in \{0,1\} \tag{2.30}$$

where $\mu(s_t) = \mu + \nu s_t$ and $\mathbb{P}$ is the transition matrix.

This kind of model can be found in Kim and Nelson [26]. The absence of analytical filtering densities makes Bayesian simulation based inference a possible solution to the filtering problem.

We use parameters, estimated in Kim and Nelson [26], to simulate the MS model. Fig. 2.3 exhibits simulation paths of $1,000$ observations of the Markov switching process, the latent factor and the observable variable, respectively.

□

2.3 Simulation Based Filtering

The main aim of this section is to review both some traditional and recently developed inference methods for nonlinear and non-Gaussian models. We focus on the Bayesian approach and on simulation based methods. First Markov Chain Monte Carlo methods are reviewed with particular attention to the single-move Gibbs sampler due to Carlin, Polson and Stoffer [5] and to the multi-move Gibbs sampler due to Carter and Köhn [6] and Frühwirth-Schnatter [16]. Moreover some basic sequential Monte Carlo simulation methods are introduced. We examine the adaptive importance sampling algorithm due to West [52], [53], the sequential importance sampling algorithm and more advanced sequential Monte Carlo algorithms called Particle Filters (see Doucet, Freitas and Gordon [12]).

Finally, note that another important issue in making inference on dynamic models is the estimation of the parameter vector. In a Bayesian MCMC based approach parameters are estimated together with the hidden states of the model, by simulating from the posterior distribution of the model. Also in a sequential importance sampling approach and following the engineering literature, a common way to solve the parameter estimation problem is to treat parameters θ as hidden state of the system (see Berzuini et al. [3]). The model is restated assuming time dependent parameter vectors θ_t, and imposing a constraint on the evolution: $\theta_t = \theta_{t-1}$. The constraint can be expressed also in terms of transition probability as follows

$$\theta_t \sim \delta_{\theta_{t-1}}(\theta_t), \qquad t = 0, \ldots, T \tag{2.31}$$

where $\delta_{x^*}(x)$ denotes the Dirac delta function. The Bayesian model is then completed by assuming a prior distribution $p(\theta_0)$ on the parameter vector. We refer to Chapter 6 for a wider treatment of the parameter estimation problem in a sequential Monte Carlo approach.

2.3.1 The Gibbs Sampler

In previous sections we examine some estimation algorithms for filtering, predicting and smoothing the state vector of a quite general probabilistic dynamic model. In order to examine Gibbs sampling methods, we consider the following dynamic model

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta) \tag{2.32}$$

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta) \tag{2.33}$$

$$x_0 \sim p(\mathbf{x}_0|\theta) \tag{2.34}$$

$$\theta \sim p(\theta)\,, \quad \text{with } t = 1, \ldots, T. \tag{2.35}$$

The estimation problem is solved in a Bayesian perspective by evaluating the mean of the joint posterior density of the state and parameter vectors $p(\mathbf{x}_{0:T}, \theta|\mathbf{y}_{1:T})$. Tanner and Wong [48] motivates this solution by the *data augmentation principle*, which consists in considering the hidden state vectors as nuisance parameters.

If an analytical evaluation of the posterior mean is not possible, then simulation methods and in particular Monte Carlo Markov Chain apply. The most simple solution is to implement a *single-move Gibbs sampler* (see Carlin, Polson and Stoffer [5] and Harrison and West [22]). This method generates the states one at a time using the Markov property of the dynamic model to condition on the neighboring states. However the first order Markov dependence between adjacent states induces a high correlation between outputs of the Gibbs sampler and causes a slow convergence of the algorithm. To solve this problem Carter and Köhn [6] and Frühwirth-Schnatter [16] simultaneously proposes *multi-move Gibbs sampler*. The main idea of this method is to generate simultaneously all the state vectors using analytical filtering and smoothing relations.

Their approach is less general than that of Carlin, Polson and Stoffer [5], but for linear dynamic models with Gaussian mixture innovations in the observation equation, their approach is more efficient.

In particular the multi-move Gibbs sampler has a faster convergence to the posterior distribution and the posterior moment estimates have smaller variance. These results are supported theoretically by Liu, Wong and Kong [33], [34] and Müller [38], who shows that generating variables simultaneously produces faster convergence than generating them one at a time.

The idea of grouping parameters (or hidden states) when simulating is now commonly in Bayesian inference on stochastic models, with latent factors. For example, multi-move MCMC algorithms have been used for stochastic volatility models by Kim, Shephard and Chib [29] and extended to generalized stochastic volatility models by Chib, Nardari and Shephard [8]. Shephard [44] and Shephard and Pitt [45] discussed multi-move MCMC algorithms to non-Gaussian time series models . Finally, an alternative way of simulating from the smoothing density of the state vectors is discussed in De Jong and Shephard [10].

In the following we briefly present how to implement the single-move Gibbs sampler for parameters and states estimation. On the time interval $\{1, \ldots, T\}$, the conditional posterior distributions of the parameter vector and of the state vectors are

$$p(\theta|\mathbf{x}_{0:T}, \mathbf{y}_{1:T}) \propto p(\theta)p(\mathbf{x}_0|\theta) \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta) \quad (2.36)$$

$$p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T}, \theta) \propto p(\mathbf{x}_0|\theta) \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta). \quad (2.37)$$

A basic Gibbs sampler is obtained by simulating sequentially from the parameter vector posterior (*parameter simulation step*) in equation (2.36) and from the state vectors posterior (*data augmentation step*) in equation (2.37) conditionally on the parameter vector simulated at the previous step. When conditional distributions cannot be directly simulated, the corresponding steps in the Gibbs algorithm can be replaced by Metropolis-Hastings steps. The resulting algorithms are called hybrid sampling algorithms and they are validated in Tierney [49].

A generic Gibbs sampler can be used for simulating the posterior of the parameter vector conditional on the simulated state vectors.

The single-move Gibbs sampler for the state vectors, conditional on the simulated parameter vector, is then obtained by drawing each state vector conditionally on the

other simulated state vectors.

Algorithm 2.3.1. - ***Gibbs sampler for the parameter vector***
Given simulated vectors $\theta^{(i)}$ *and* $\mathbf{x}_{0:T}^{(i)}$*, generate the parameter vector*

1. $\theta_1^{(i+1)} \sim p(\theta_1|\theta_2^{(i)}, \ldots, \theta_{n_\theta}^{(i)}, \mathbf{x}_{0:T}^{(i)}, \mathbf{y}_{1:T})$
2. $\theta_2^{(i+1)} \sim p(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_{n_\theta}^{(i)}, \mathbf{x}_{0:T}^{(i)}, \mathbf{y}_{1:T})$
3. ...
4. $\theta_k^{(i+1)} \sim p(\theta_k|\theta_1^{(i+1)}, \ldots, \theta_{k-1}^{(i+1)}, \theta_{k+1}^{(i)}, \ldots, \theta_{n_\theta}^{(i)}, \mathbf{x}_{0:t}^{(i)}, \mathbf{y}_{1:T})$
5. ...
6. $\theta_{n_\theta}^{(i+1)} \sim p(\theta_{n_\theta}|\theta_1^{(i+1)}, \ldots, \theta_{n_\theta-1}^{(i+1)}, \mathbf{x}_{0:T}^{(i)}, \mathbf{y}_{1:T})$

Algorithm 2.3.2. - ***Single-Move Gibbs Sampler -***
(i) Simulate $\theta^{(i)}$ *through Algorithm 2.3.1;*
(ii) Given $\theta^{(i)}$ *and* $\mathbf{x}_{0:T}^{(i)}$*, simulate state vectors*

7. $\mathbf{x}_0^{(i+1)} \sim p(\mathbf{x}_0|\mathbf{x}_{2:T}^{(i)}, \mathbf{y}_{1:T}, \theta^{(i+1)})$
8. $\mathbf{x}_1^{(i+1)} \sim p(\mathbf{x}_1|\mathbf{x}_0^{(i+1)}, \mathbf{x}_{2:T}^{(i)}, \mathbf{y}_{1:T}, \theta^{(i+1)})$
9. ...
10. $\mathbf{x}_t^{(i+1)} \sim p(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(i+1)}, \mathbf{x}_{t+1:T}^{(i)}, \mathbf{y}_{1:T}, \theta^{(i+1)})$
11. ...
12. $\mathbf{x}_T^{(i+1)} \sim p(\mathbf{x}_T|\mathbf{x}_{0:T-1}^{(i+1)}, \mathbf{y}_{1:T}, \theta^{(i+1)})$

Single-move algorithm can be implemented for general dynamic models. Moreover, note that the dynamic model given in equations (2.32)-(2.35) satisfies

to the Markov property. In this case the full posterior density of the state vector, given in the single-move Gibbs sampler (see the Algorithm 2.3.2), is

$$p(\mathbf{x}_t|\mathbf{x}_{-t}, \mathbf{y}_{1:T}, \theta) \propto p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}, \theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{1:t}, \theta) \quad (2.38)$$

and the implementation of the algorithm becomes easier. For the general model described in Equations (2.32)-(2.35), with measurement and transition densities depending on past values of the observable variable $\mathbf{y}_{1:t-1}$, the proof of Equation (2.38) is given in Appendix B.

Example 2.3.1 - *(Stochastic Volatility Model, Example 2.2.1 continued)*
We illustrate, through this example, how simulation based Bayesian inference applies to dynamic models. In particular, we develop a single-move Gibbs sampler for the stochastic volatility model given in Example 2.2.1. Note that the stochastic volatility model can be rewritten as follows

$$y_t = e^{h_t/2}\varepsilon_t\,, \qquad \varepsilon_t \sim \mathcal{N}(0,1) \quad (2.39)$$

$$h_t = \nu + \phi h_{t-1} + \sigma\eta_t\,, \quad \eta_t \sim \mathcal{N}(0,1) \quad (2.40)$$

$$\text{with} \quad t = 1, \ldots, T. \quad (2.41)$$

The completed likelihood function of this model is

$$\begin{aligned} &L(\phi, \nu, \sigma^2|y_{1:T}, h_{0:T}) = \\ &= \prod_{t=1}^{T} \frac{1}{(2\pi)^{1/2}e^{h_t/2}} \exp\left\{-\frac{y_t^2 e^{-h_t}}{2}\right\} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(h_t - \nu - \phi h_{t-1})^2}{2\sigma^2}\right\} = \\ &= \frac{1}{(2\pi)^T(\sigma^2)^{\frac{T}{2}}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t^2 e^{-h_t} + h_t) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(h_t - \nu - \phi h_{t-1})^2\right\} \end{aligned} \quad (2.42)$$

In order to conclude the description of the Bayesian dynamic model we take the following prior distributions on the parameters ϕ, ν, σ^2 and h_0

$$\phi \sim \mathcal{N}(a, b^2)\mathbb{I}_{(-1,+1)}(\phi) \tag{2.43}$$

$$\nu \sim \mathcal{N}(c, d^2) \tag{2.44}$$

$$\sigma^2 \sim \mathcal{IG}(\alpha, \beta) \tag{2.45}$$

$$h_0 \sim \mathcal{N}(0, \sigma^2) \tag{2.46}$$

where $\mathcal{IG}(\cdot, \cdot)$ indicates the *Inverse Gamma* distribution. Note that for the parameter ϕ we use a truncated Normal distribution, but a Uniform distribution $\mathcal{U}_{[-1,1]}$ could alternatively be used. In the first part of the single-move Gibbs sampler, the parameters are simulated from the posterior distributions, through the following three steps

(i) Simulate $\phi \sim \pi(\phi|\sigma^2, \nu, y_{1:T}, h_{0:T})$, where

$$\begin{aligned}
&\pi(\phi|\sigma^2, \nu, y_{1:T}, h_{0:T}) \propto \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(h_t - \nu - \phi h_{t-1})^2\right\}\frac{1}{(2\pi)^{1/2}b}\exp\left\{-\frac{(\phi-a)^2}{2b^2}\right\}\mathbb{I}_{(-1,+1)}(\phi) \propto \\
&\propto \exp\left\{-\frac{1}{2}\left[\phi^2\left(\frac{1}{b^2} + \frac{\sum_{t=1}^{T} h_{t-1}^2}{\sigma^2}\right) - 2\phi\left(\frac{a}{b^2} + \frac{\sum_{t=1}^{T} h_{t-1}(h_t - \nu)}{\sigma^2}\right)\right]\right\}\mathbb{I}_{(-1,+1)} \\
&\propto \mathcal{N}(\tilde{a}, \tilde{b}^2)\mathbb{I}_{(-1,+1)}(\phi)
\end{aligned} \tag{2.47}$$

with

$$\tilde{a} = \tilde{b}^2\left(\frac{a}{b^2} + \frac{1}{\sigma^2}\sum_{t=1}^{T} h_{t-1}(h_t - \nu)\right), \quad \tilde{b}^2 = \left(\frac{1}{b^2} + \frac{1}{\sigma^2}\sum_{t=1}^{T} h_{t-1}^2\right)^{-1}$$

(ii) Simulate $\nu \sim \pi(\nu|\phi, \sigma^2, y_{1:T}, h_{0:T})$, where

$$\begin{aligned}
&\pi(\nu|\phi, \sigma^2, y_{1:T}, h_{0:T}) \propto \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(h_t - \nu - \phi h_{t-1})^2\right\}\exp\left\{-\frac{(\nu-c)^2}{2d^2}\right\}\frac{1}{(2\pi)^{1/2}d} \propto \\
&\propto \exp\left\{-\frac{1}{2}\left[\nu^2\left(\frac{1}{d^2} + \frac{T}{\sigma^2}\right) - 2\nu\left(\frac{c}{d^2} + \frac{\sum_{t=1}^{T}(h_t - \phi h_{t-1})}{\sigma^2}\right)\right]\right\} \propto \\
&\propto \mathcal{N}(\tilde{c}, \tilde{d}^2)
\end{aligned} \tag{2.48}$$

with

$$\tilde{c} = \tilde{d}^2 \left(\frac{c}{d^2} + \frac{1}{\sigma^2} \sum_{t=1}^{T} (h_t - \phi h_{t-1}) \right), \quad \tilde{d}^2 = \left(\frac{1}{d^2} + \frac{T}{\sigma^2} \right)^{-1}$$

(iii) Simulate $\sigma^2 \sim \pi(\sigma^2 | \phi, \nu, y_{1:T}, h_{0:T})$, where

$$\begin{aligned} &\pi(\sigma^2 | \phi, \nu, y_{1:T}, h_{0:T}) \propto \\ &\propto \left(\frac{1}{\sigma^2} \right)^{\frac{(T+1)}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{T} (h_t - \nu - \phi h_{t-1})^2 - \frac{h_0^2}{2\sigma^2} \right\} \frac{\beta^\alpha e^{-\beta/\sigma^2}}{\Gamma(\alpha)(\sigma^2)^{\alpha+1}} \mathbb{I}_{[0,\infty)}(\sigma^2) \propto \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (2\beta + \sum_{t=1}^{T} (h_t - \nu - \phi h_{t-1})^2 + h_0^2) \right\} \left(\frac{1}{\sigma^2} \right)^{\alpha + \frac{T+1}{2} + 1} \mathbb{I}_{[0,\infty)}(\sigma^2) \propto \\ &\propto \mathcal{IG}(\tilde{\alpha}, \tilde{\beta}) \end{aligned} \tag{2.49}$$

with

$$\tilde{\alpha} = \alpha + \frac{T+1}{2}, \qquad \tilde{\beta} = \beta + \frac{1}{2} \left[\sum_{t=1}^{T} (h_t - \nu - \phi h_{t-1})^2 + h_0^2 \right]$$

In the second part of the single-move Gibbs sampler, the hidden variables, $\{h_t\}_{t=1}^T$, are sampled once at time, from the following posterior distribution

$$\begin{aligned} \pi(h_t | h_{-t}, y_{1:T}, \nu, \phi, \sigma^2) \propto &\exp \left\{ -\frac{(y_t^2 e^{-h_t} + h_t)}{2} \right\} \\ &\exp \left\{ -\frac{(h_t - \nu - \phi h_{t-1})^2}{2\sigma^2} - \frac{(h_{t+1} - \nu - \phi h_t)^2}{2\sigma^2} \right\} \end{aligned} \tag{2.50}$$

for $t = 1, \ldots, T$. Due to the presence of the quantity e^{-h_t} in the exponential any standard simulation method easily applies here. Moreover the posterior distribution is known up to a normalising constant, thus the M.-H. algorithm can be used. The proposal distribution can be obtained by replacing the function $\left(y_t^2 e^{-h_t} + h_t\right)$ in equation (2.50), with $(\log y_t^2 + 1) + (h_t - \log y_t^2)^2/2$, which is its the second order Taylor expansion, around $\log(y_t^2)$. We simulate h_t from the following proposal density

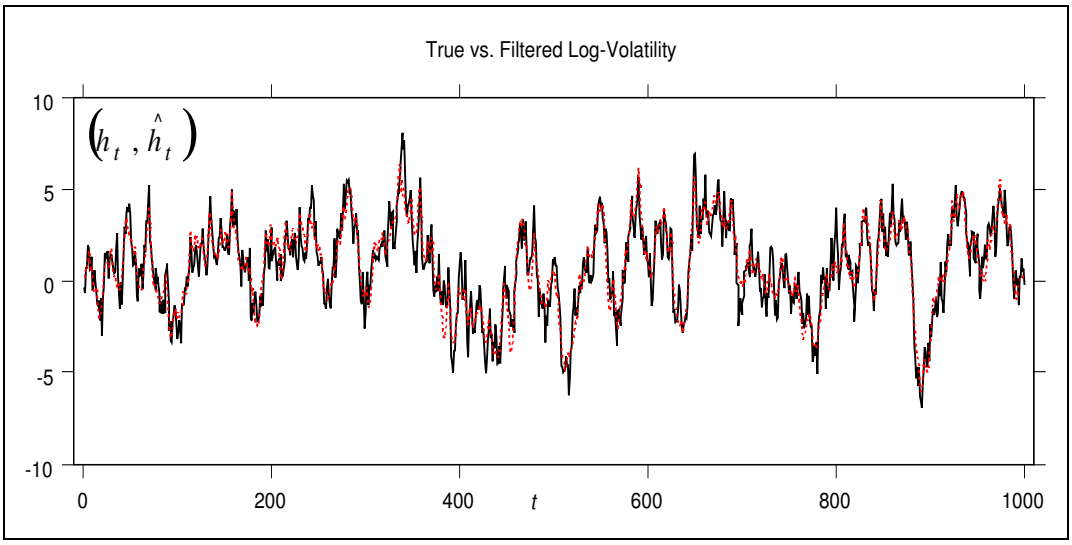

Figure 2.4: Graph of the true (solid line) and filtered (dotted line) densities. The filtered density has been obtained by averaging 50,000 values, simulated through the single-move Gibbs sampler, given in example 2.3.1.

$$
\begin{aligned}
& q(h_t|h_{-t}, y_{1:T}, \nu, \phi, \sigma^2) \propto \\
& \propto \exp\left\{-\frac{(h_t - \log y_t^2)^2}{4} - \frac{(h_t - \nu - \phi h_{t-1})^2}{2\sigma^2} - \frac{(h_{t+1} - \nu - \phi h_t)^2}{2\sigma^2}\right\} \propto \\
& \propto \exp\left\{-\frac{h_t^2}{2}\left(\frac{1+\phi^2}{\sigma^2} + \frac{1}{2}\right) + h_t\left(\frac{\nu(1-\phi) + \phi(h_{t-1} + h_{t+1})}{\sigma^2} + \frac{\log y_t^2}{2}\right)\right\} \propto \\
& \propto \mathcal{N}(\mu, \tilde{\sigma}^2)
\end{aligned}
\tag{2.51}
$$

with

$$
\mu = \left((\nu(1-\phi) + \phi(h_{t-1} + h_{t+1}))\,\sigma^{-2} + \log y_t^2/2\right)\tilde{\sigma}^2\,, \quad \tilde{\sigma}^2 = \left(\frac{1+\phi^2}{\sigma^2} + \frac{1}{2}\right)^{-1}
$$

We apply the single-move Gibbs sampler to the synthetic dataset exhibited in Fig. 2.1. We use a burn-in sample of 10,000 observations to reduce the dependence on the initial conditions of the sampler and take the remaining 40,000 values to calculate the parameter estimates, which are represented in Table 2.1.

Figure 2.4 shows the true stochastic log-volatility versus the filtered volatility. Figures 2.5 and 2.6 show the raw output of the single-move Gibbs sampler and the ergodic means respectively. Figures 2.7 and 2.8 show the autocorrelation of the simulated values and the histograms of the posterior distributions, respectively.

Table 2.1: Bayesian estimates of the parameters ϕ, ν and σ^2. We use a burn-in sample of 10,000 observations and calculate the parameter estimates by averaging 40,000 raw values simulated from the posterior distributions, by means of the Gibbs sampler.

Parameters	True	Estimates	St. Dev.	2.5% percentile	97.5% percentile
ϕ	0.9	0.9096	0.0146	0.8813	0.9370
ν	0.1	0.0754	0.0350	0.0048	0.1382
σ^2	1	1.0075	0.0313	0.9487	1.0712

Simulations have been carried out on a PC with Intel 2400 MHz processor, using routines implemented in Gauss 3.2.8.

□

Although the simplification due to the first order Markov property of the dynamic model makes the single-move Gibbs sampler easier to implement, some problems arise. In particular, the Markovian dependence between neighboring states generates correlation between outputs of the Gibbs sampler and origins slower convergence to the posterior distribution (see Carter and Köhn [6]). A consequence is that if an adaptive importance sampling is carried out by running parallel single-move Gibbs samplers, the number of replications before convergence of the parameter estimates is high.

A general method to solve this autocorrelation problem in the output of the Gibbs sampler is to group parameters (or states) and to simulate them simultaneously. This idea has been independently applied by Carter and Köhn [6] and by Frühwirth-Schnatter [16] to dynamic models and the resulting algorithm is the multi-move Gibbs sampler. Furthermore Frühwirth-Schnatter [16] shows how the use of the multi-move Gibbs sampler improves the convergence rate of an adaptive importance sampling algorithm and makes a comparison with a set of parallel single-move Gibbs samplers. The implementation of the multi-move Gibbs sampler depends on the availability of the analytical form of filtering and smoothing densities.

We give here a general representation of the algorithm, but its implementation is strictly related to the specific analytical representation of the dynamic model. Given the simulated parameter vector obtained through the Gibbs sampler in the Algorithm 2.3.1, the multi-move Gibbs sampler is in Algorithm 2.3.3.

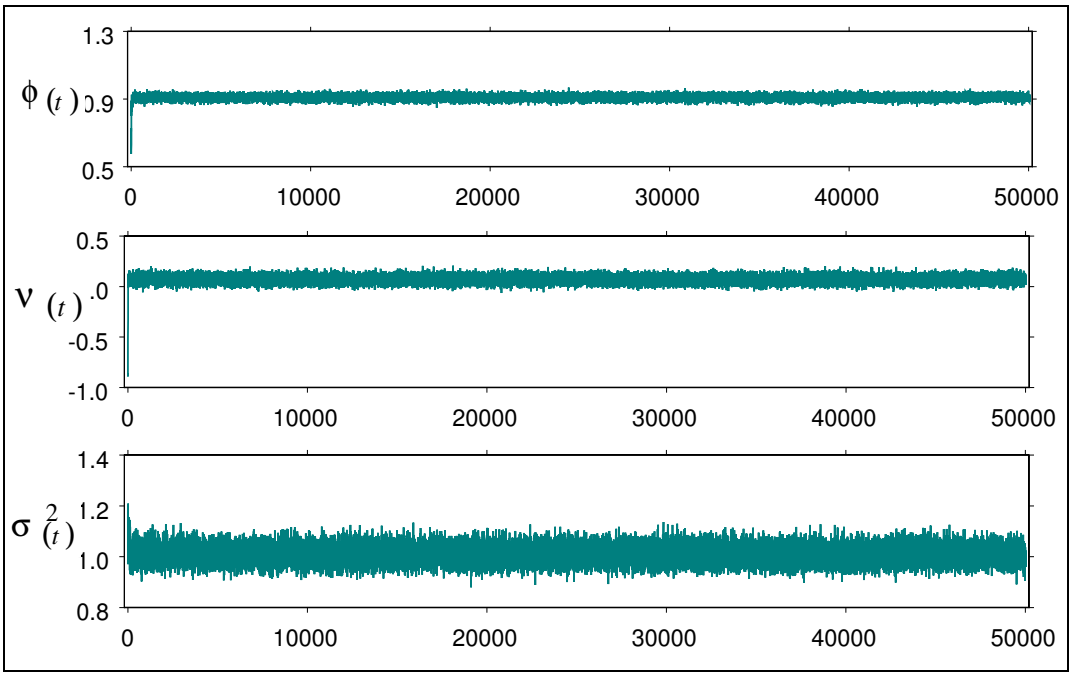

Figure 2.5: Sample of 50,000 parameter values simulated from the posterior distributions of the stochastic volatility model, through the single-move Gibbs sampler.

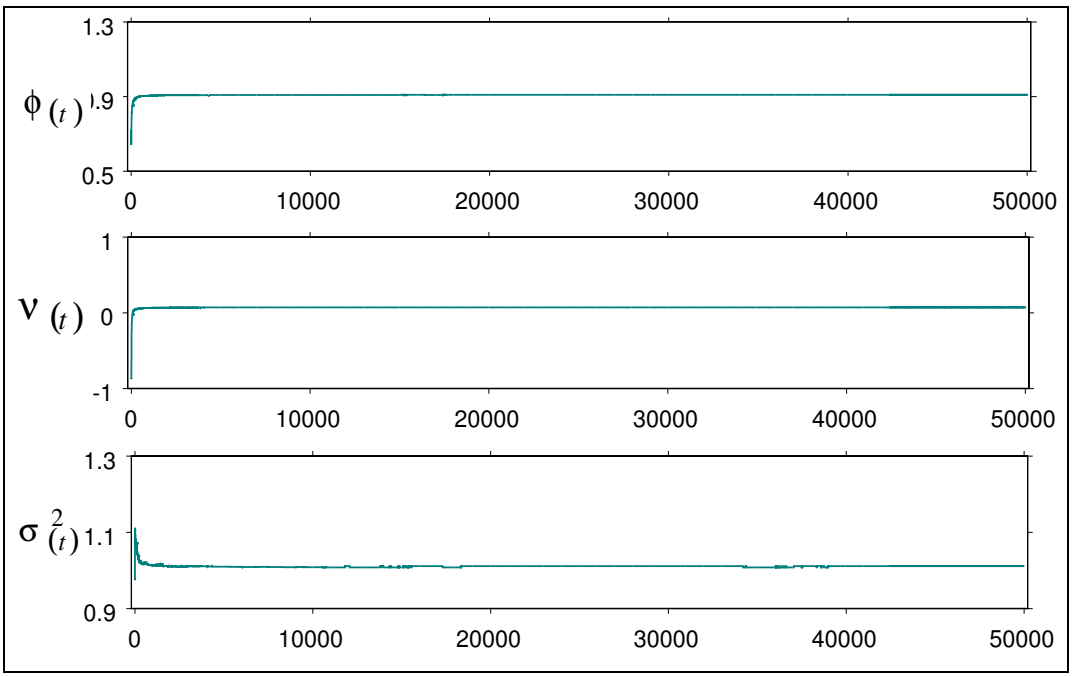

Figure 2.6: Ergodic averages of the parameter values simulated from the posterior distributions of the stochastic volatility model, through the single-move Gibbs sampler.

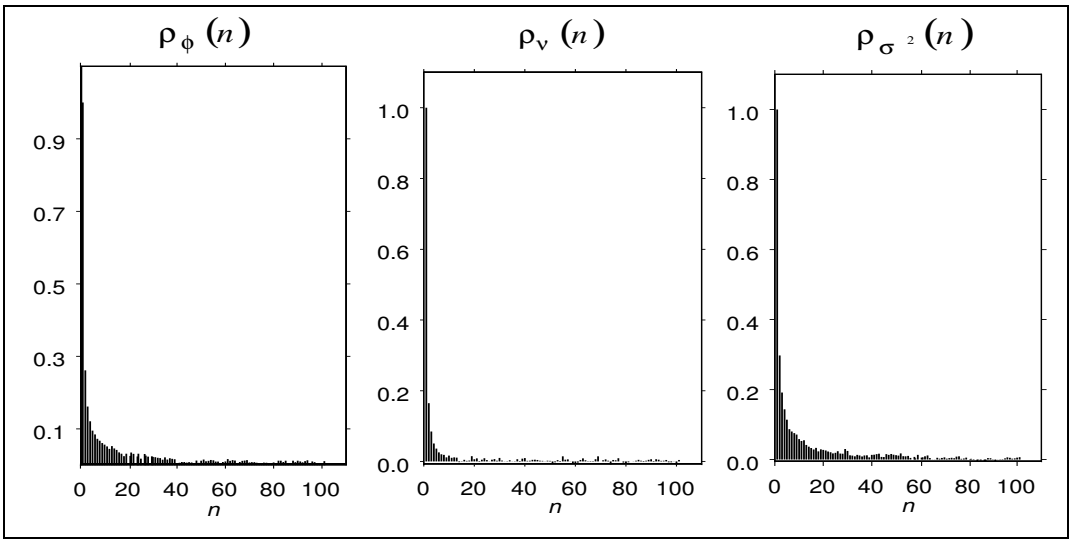

Figure 2.7: The graph exhibits the autocorrelation functions, ρ_ϕ, ρ_ν and ρ_{σ^2}, with lags $n = 1, \ldots, 100$, calculated on the 50,000 parameter values simulated from the posterior distribution.

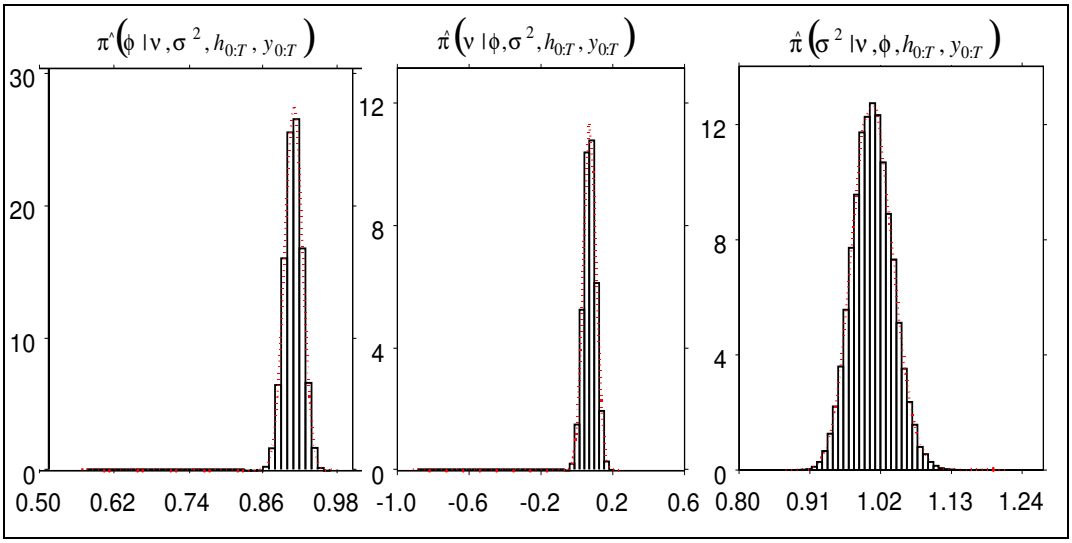

Figure 2.8: The figure show the histogram of 50,000 parameter values simulated from the posterior distributions and the Gaussian kernel density reconstruction.

The algorithm has been derived trough the recursive smoothing relation given in equation (2.25). Moreover, at each simulation step the posterior density is obtained by means of estimated prediction and filtering densities. By applying the fundamental relation given in equation (2.26) we obtain

$$p(\mathbf{x}_t|\mathbf{x}_{t+1}^{(i+1)}, \mathbf{y}_{1:t}, \theta^{(i+1)}) = \frac{p(\mathbf{x}_{t+1}^{(i+1)}|\mathbf{x}_t, \theta^{(i+1)})\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta^{(i+1)})}{\hat{p}(\mathbf{x}_{t+1}^{(i+1)}|\mathbf{y}_{1:t}, \theta^{(i+1)})} \tag{2.52}$$

We stress once more that the multi-move Gibbs sampler does not easily apply to nonlinear and non-Gaussian models. Thus in a MCMC approach, the single-move Gibbs sampler remains the only numerical solution to the estimation problem. Sequential sampling algorithm represents an alternative to MCMC methods.

Algorithm 2.3.3. - *Multi-Move Gibbs Sampler* -
(i) Simulate $\theta^{(i)}$ *through Algorithm 2.3.1;*
(ii) Given $\theta^{(i)}$ *and* $\mathbf{x}_{0:T}^{(i)}$*, run analytical filtering relations in order to estimate prediction and filtering densities for each* $t = 0, \ldots, T$

7. $\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \theta^{(i+1)})$

8. $\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta^{(i+1)})$

(iii) Simulate state vectors by means of the recursive factorization of the smoothing density

9. $\mathbf{x}_T^{(i+1)} \sim p(\mathbf{x}_T|\mathbf{y}_{1:T}, \theta^{(i+1)})$

10. $\mathbf{x}_{T-1}^{(i+1)} \sim p(\mathbf{x}_{T-1}|\mathbf{x}_T^{(i+1)}, \mathbf{y}_{1:T-1}, \theta^{(i+1)})$

11. ...

12. $\mathbf{x}_t^{(i+1)} \sim p(\mathbf{x}_t|\mathbf{x}_{t+1}^{(i+1)}, \mathbf{y}_{1:t}, \theta^{(i+1)})$

13. ...

14. $\mathbf{x}_1^{(i+1)} \sim p(\mathbf{x}_1|\mathbf{x}_2^{(i+1)}, \mathbf{y}_1, \theta^{(i+1)})$

Sequential Monte Carlo algorithms allow us to make inference on general dynamic models. One of the early sequential methods proposed in the literature is Adaptive Importance Sampling, which will be discussed in the next section.

2.3.2 Adaptive Importance Sampling

The adaptive sequential importance sampling scheme is a sequential stochastic simulation method which adapts progressively to the posterior distribution also by means of the information contained in the samples, which are simulated at the previous steps. The adaptation mechanism is based on the discrete posterior approximation and on the kernel density reconstruction of the prior and posterior densities. West [52] proposed this technique in order to estimate parameters of static models. West [53] and West and Harrison [22] successively extended the method in order to estimate parameters and states of dynamic models.

The first key idea is to use importance sampling (see Robert and Casella [43]) in order to obtain a weighted random grid of evaluation points in the state space. Let $\{\mathbf{x}_t^i, w_t^i\}_{t=1}^{n_t}$ be a sample drawn from the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)$ through an importance density g_t. The prediction density, given in equation (2.20), can be approximated as follow

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) \approx \sum_{i=1}^{n_t} w_t^i p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) \tag{2.53}$$

The second key idea, implemented in the adaptive importance sampling algorithm of West [53], is to propagate points of the stochastic grid by means of the transition density and to build a smoothed approximation of the prior density. This approximation is obtained through a kernel density estimation. West [53] suggested to use Gaussian or Student-t kernels due to their flexibility in approximating other densities. For example, the Gaussian kernel reconstruction is

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) \approx \sum_{i=1}^{n_t} w_t^i N(\mathbf{x}_{t+1}|m_t a + \mathbf{x}_t^i(1-a), h^2 V_t) \tag{2.54}$$

The final step of the algorithm consists in updating the prior density and in producing a random sample, $\{\mathbf{x}_{t+1}^i, w_{t+1}^i\}_{i=1}^{n_{t+1}}$, from the resulting posterior density.

The sample is obtained by using the kernel density estimate as importance density. Adaptive importance sampling is represented through algorithm 2.3.4.

Algorithm 2.3.4. - ***Adaptive Sequential Importance Sampling*** -
Given a weighted random sample $\{\mathbf{x}_t^i, w_t^i\}_{t=1}^{n_t}$, *for* $i = 1, \ldots, n_t$

1. *Simulate* $\tilde{\mathbf{x}}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta)$

2. *Calculate* $m_t = \sum_{i=1}^{n_t} w_t^i \tilde{\mathbf{x}}_{t+1}^i$, $V_t = \sum_{i=1}^{n_t} w_t^i (\tilde{\mathbf{x}}_{t+1}^i - m_t)'(\tilde{\mathbf{x}}_{t+1}^i - m_t)$

3. *Generate from the Gaussian kernel*

$$\mathbf{x}_{t+1}^i \sim \sum_{i=1}^{n_t} w_t^i N(\mathbf{x}_{t+1}|(m_t a + \mathbf{x}_t^i(1-a)), h^2 V_t)$$

4. *Update the weights* $w_{t+1}^i \propto w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i)p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i)}{N(\mathbf{x}_{t+1}^i|(m_t a+(1-a)\mathbf{x}_t^i), h^2 V_t)}$

The main advantage of this algorithms relies in the smoothed reconstruction of the prior density. This kernel density estimate of the prior allows to obtain adapting importance densities and to avoid the information loss, which comes from cumulating numerical approximation over time. Furthermore this technique can be easily extended to account for a sequential estimation of the parameter (see the recent work due to Liu and West [32]).

However adaptive importance sampling requires the calibration of parameters a and h, which determines the behavior of the kernel density estimate. The choice of that shrinkage parameters influences the convergence of the algorithm and heavily depends on the complexity of the model studied.

Finally, adaptive importance sampling belongs to a more general class of sequential simulation algorithms, which are particulary efficient for on-line data processing and which have some common problems like sensitivity to outliers and degeneracy. In next paragraph we review some general particle filters.

2.3.3 Particle Filters

In the following we focus on *Particle filters*, also referred in the literature as *Bootstrap filters*, *Interacting particle filters*, *Condensation algorithms*, *Monte Carlo filters* and on the estimation of the states of the dynamic model. See also Doucet, Freitas and Gordon [12] for an updated review on the particle filter techniques, on their applications and on the main convergence results for this kind of algorithms.

The main advantages of particle filters are that they can deal with nonlinear and non-Gaussian noise and can be easily implemented, also in a parallel mode. Moreover in contrast to Hidden Markov Model filters, which work on a state space discretised to a fixed grid, particle filters focuse sequentially on the higher probable regions of the state space. This is feature is common to Adaptive Importance Sampling algorithm exhibited in the previous section.

Assume that the parameter vector θ of the dynamic model given in equations (2.17), (2.18) and (2.19) is known. Different versions of the particle filter exist in the literature and different simulation approaches like rejection sampling, MCMC and importance sampling, can be used for the construction of a particle filter. We introduce particle filters applying the importance sampling reasoning.

At each time step $t+1$, as a new observation $\mathbf{y}_{t+1}$ arrives, we are interested in predicting and filtering the hidden variables and the parameters of a general dynamic model. In particular we search how to approximate prediction an filtering densities given in Equations (2.20) and (2.21) by means of sequential Monte Carlo methods.

Assume that the weighted sample $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$ has been drawn from the filtering density at time t

$$\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) = \sum_{i=1}^{N} w_t^i \delta_{\{\mathbf{x}_t^i\}}(d\mathbf{x}_t) \tag{2.55}$$

Each simulated value $\mathbf{x}_t^i$ is called *particle* and the particles set, $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, can be viewed as a random discretization of the state space $\mathcal{X}$, with associated probabilities weights w_t^i. It is possible to approximate, by means of this particle set, the prediction density given in Eq. (2.21) as follows

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta) p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) d\mathbf{x}_t \simeq \sum_{i=1}^{N} w_t^i p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) \tag{2.56}$$

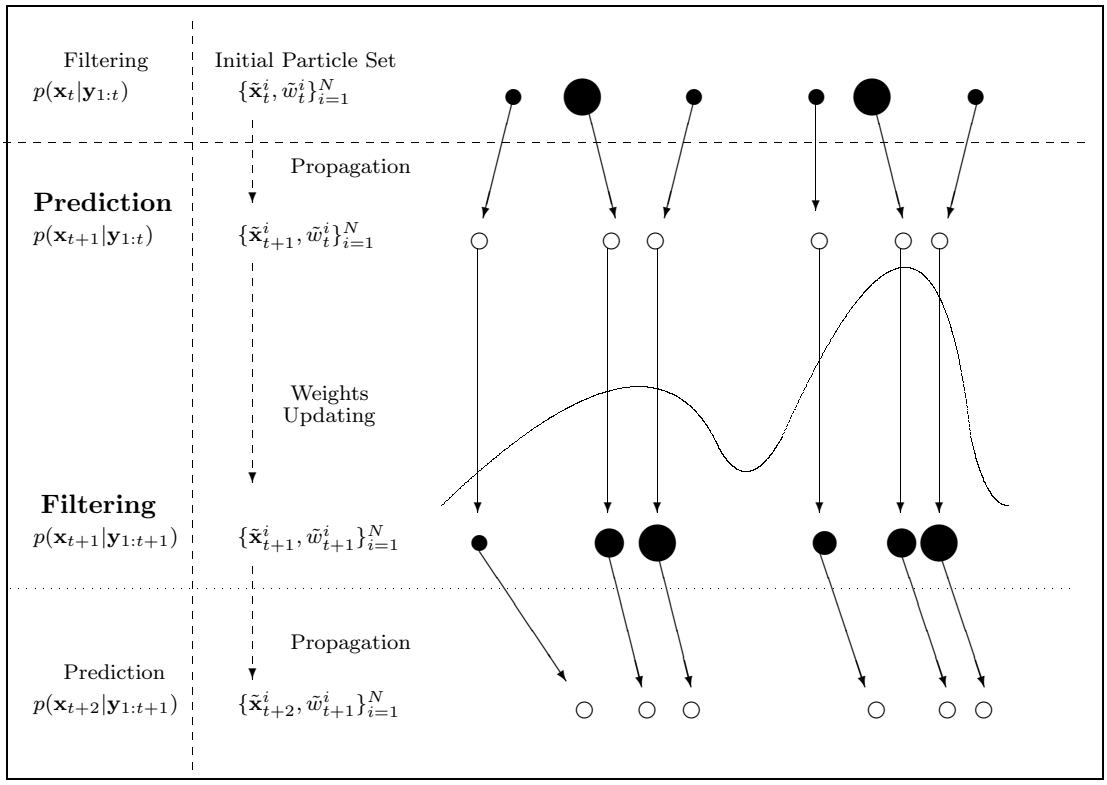

Figure 2.9: Particles evolution in the SIS particle filter.

which is called *empirical prediction density* and is denoted by $\hat{p}(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta)$. By applying the Chapman-Kolmogorov equation it is also possible to obtain an approximation of the filtering density given in Eq. (2.21)

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1}, \theta) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) \simeq \sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) w_t^i \tag{2.57}$$

which is called *empirical filtering density* and is denoted by $\hat{p}(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1}, \theta)$.

Assume now that the quantity $\mathbb{E}(f(\mathbf{x}_{t+1})|\mathbf{y}_{1:t+1})$ is of interest. It can be evaluated numerically by a Monte Carlo sample $\{\mathbf{x}_{t+1}^i, w_{t+1}^i\}_{i=1}^N$, simulated from the filtering distribution

$$\mathbb{E}(f(\mathbf{x}_{t+1})|\mathbf{y}_{1:t+1}) \simeq \frac{\frac{1}{N}\sum_{i=1}^{N} f(x_{t+1}^i) w_{t+1}^i}{\frac{1}{N}\sum_{i=1}^{N} w_{t+1}^i}. \tag{2.58}$$

A simple way to obtain a weighted sample from the filtering density at time $t+1$ is to apply importance sampling to the empirical filtering density given in equation (2.57). This step corresponds to propagate the initial particle set (see Fig. 2.9) through the importance density $q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)$. Moreover if we propagate each particle of the set through the transition density $p(\mathbf{x}_t|\mathbf{x}_{t-1}^i, \theta)$ of the dynamic model, then particle weights update as follows

$$w_{t+1}^i \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta)w_t^i}{q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)} = w_t^i\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta) \tag{2.59}$$

This is the natural choice for the importance density, because the transition density represents a sort of prior at time t for the state x_{t+1}. However, as underlined in Pitt and Shephard [40] this strategy is sensitive to outliers. See also Crisan and Doucet [9], for a discussion on the choice of the importance densities. They focused on the properties of the importance density, which are necessary for the a.s. convergence of the sequential Monte Carlo algorithm.

The basic particle filter developed through previous equations is called *Sequential Importance Sampling* (SIS). In Algorithm 2.3.5, we give a pseudo-code representation of this method.

The Sequential importance sampling permits to obtain recursive updating of the particles weights and is based on the sequential decomposition of the joint filtering density and on a particular choice of the importance density. To evidence these aspects we consider the following smoothing problem.

We want to approximate the smoothing density $p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}, \theta)$, of the state vectors as follows

$$p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}, \theta) \simeq \sum_{i=1}^{N} \tilde{w}_{t+1}^i \delta_{\{\mathbf{x}_{0:t+1}^i\}}(d\mathbf{x}_{0:t+1}) \tag{2.60}$$

by simulating $\{\mathbf{x}_{0:t+1}^i\}_{i=1}^N$ from a proposal distribution $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}, \theta)$ and by correcting the weights of the resulting empirical density. The correction step comes from an importance sampling argument, thus the *unnormalized particles weights*[1] are defined

[1]Note that importance sampling requires to know the importance and the target distributions up to a proportionality constant, thus the unnormalized weights may not sum to one. However *normalized importance sampling weights* can be easily obtained as follows

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_{j=1}^N w_t^j} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T.$$

as follows

$$w_{t+1}^i \triangleq \frac{p(\mathbf{x}_{0:t+1}^i|\mathbf{y}_{1:t+1},\theta)}{q(\mathbf{x}_{0:t+1}^i|\mathbf{y}_{1:t+1},\theta)}. \tag{2.61}$$

The key idea used in SIS consists in obtaining a recursive relation for the weights updating. This property makes them particulary appealing for on-line applications.

Algorithm 2.3.5. - ***SIS Particle Filter*** -
Given the initial set of particles $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

1. *Simulate* $\mathbf{x}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)$
2. *Update the weights:* $w_{t+1}^i \propto w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i,\theta)\, p(\mathbf{x}_{t+1}|\mathbf{x}_t^i;\theta)}{q(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\mathbf{y}_{t+1},\theta)}$

Assume that the dynamic model of interest is the one described by equations (2.17), (2.18) and (2.19) and choose the importance density to factorize as follows: $q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1},\theta) = q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t},\theta)q(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t+1},\theta)$, then the weights can be rewritten in a recursive form

$$w_{t+1}^i = w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i,\theta)p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i,\theta)}{q(\mathbf{x}_{t+1}^i|\mathbf{x}_{t+1}^i,\mathbf{y}_{t+1},\theta)} \tag{2.62}$$

This relation can be proved by using the Bayes rule and the Markov property of the system (see the proof in Appendix B).

Example 2.3.2 - *(SV Model and SIS algorithm, Example 2.2.1 continued)*
The first aim of this example is to show how sequential importance sampling applies to a specific Bayesian dynamic model. We consider the SV model introduced through the example 2.2.1. In the following, we assume that the parameters are known, because we want to study how SIS algorithms work just for filtering the log-volatility. The second aim of this example, is to evidence the degeneracy problem, which arises in using SIS algorithms. Given the initial weighted particle set $\{h_t^i, w_t^i\}$, the SIS filter performs the following steps

The normalization procedure causes the loss of the unbiasness property because the quantity at the denominator is a random variable.

(i) Simulate $h_{t+1}^i \sim \mathcal{N}\left(h_{t+1}|\nu + \phi h_t^i, \sigma^2\right)$

(ii) Update the weights as follow

$$\begin{aligned} \tilde{w}_{t+1}^i &\propto w_t^i p(y_{t+1}|h_{t+1}^i, \theta) \propto \\ &\propto w_t^i \exp\left\{-\frac{1}{2}\left[y_{t+1}^2 e^{-h_{t+1}^i} + h_{t+1}^i\right]\right\} \end{aligned}$$

(iii) Normalize the weights

$$w_{t+1}^i = \frac{\tilde{w}_{t+1}^i}{\sum_{j=1}^N \tilde{w}_{t+1}^j}$$

By applying the SIS algorithm to the synthetic data, simulated in Example 2.2.1, we obtain the filtered log-volatility represented in Fig. 2.10. Note that after some iterations the filtered log-volatility does not fit well to the true log-volatility. The *Root Mean Square Error* (RMSE) defined as

$$RMSE_t = \left\{\frac{\sum_{u=1}^t (\tilde{h}_u - h_u)^2}{t}\right\}^{\frac{1}{2}}, \tag{2.63}$$

measures the distance between the true and the filtered series. In the Fig. 2.11 the RMSE cumulates rapidly over time. Moreover, the same figure exhibits the estimated variance of the particle weights. This indicator shows how the SIS algorithm degenerates after 430 iterations. The discrete probability mass concentrates on one particle, the others particle having null probability.

□

As evidenced in Example 2.3.2 and as it is well known in the literature (see for example Arulampalam, Maskell, Gordon and Clapp [1]), that basic SIS algorithms have a degeneracy problem. After some iterations the empirical distribution degenerates into a single particle, because the variance of the importance weights is non-decreasing over time (see Doucet et al. [13]). In order to solve the degeneracy problem, the *Sampling Importance Resampling* (SIR) algorithm has been introduced by Gordon, Salmond and Smith [18].

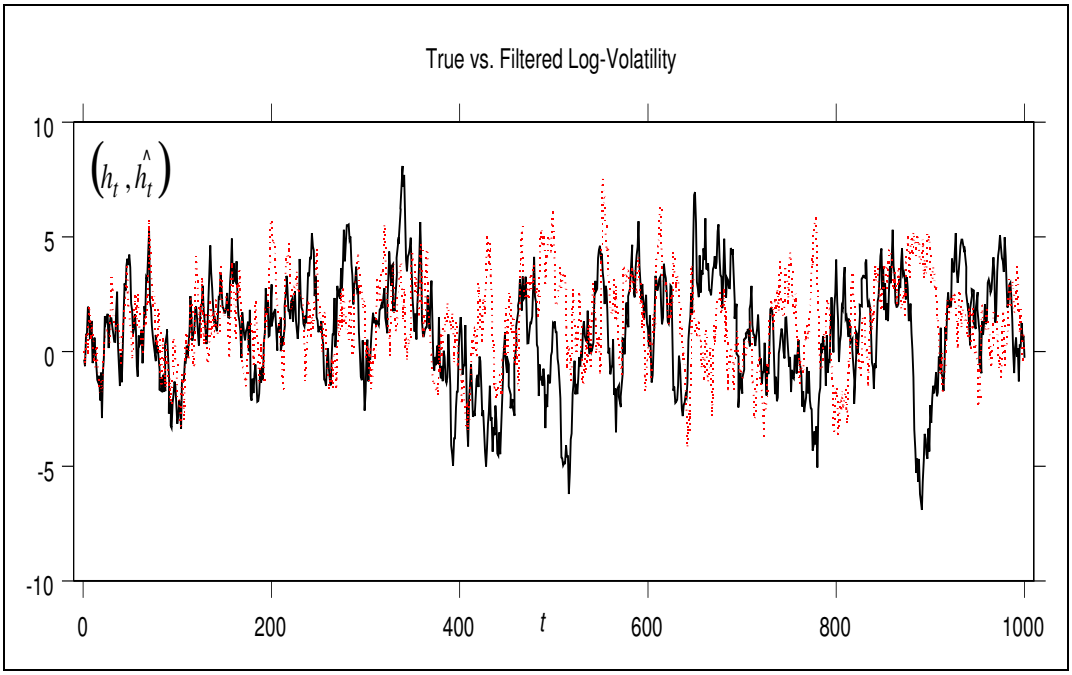

Figure 2.10: True (solid line) versus Filtered (dotted line) log-volatility obtained by applying a SIS algorithm, with $N = 3,000$ particles at each time step.

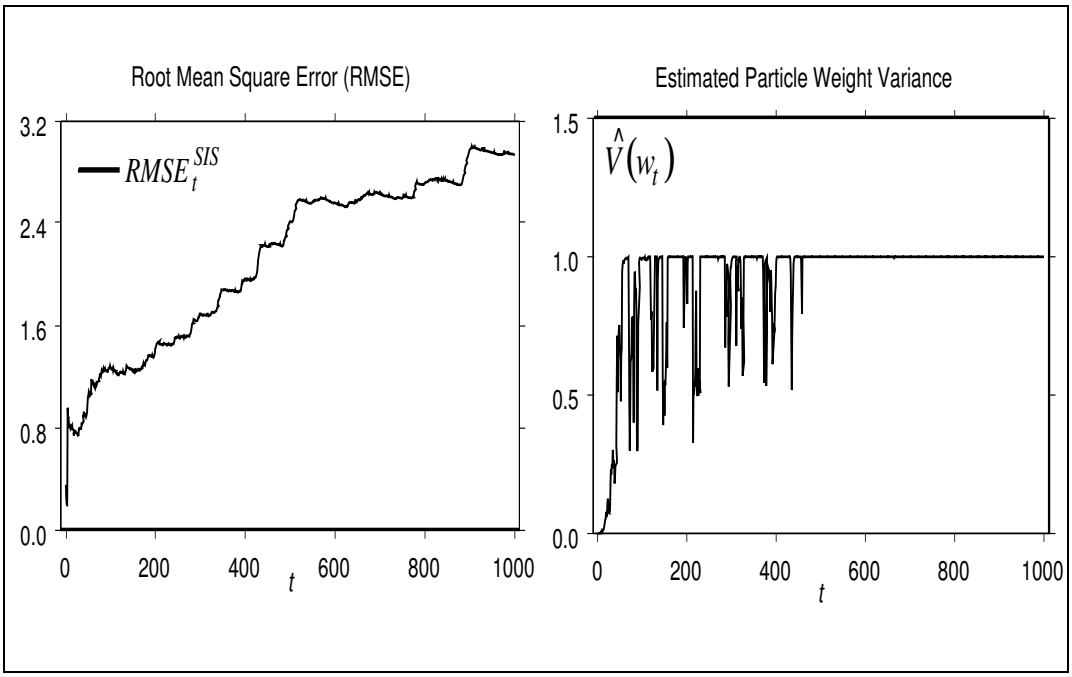

Figure 2.11: The Root Means Square Error between true and filtered log-volatility, and the estimated variance of the particle weights, which allows us to detect degeneracy of the particle weights.

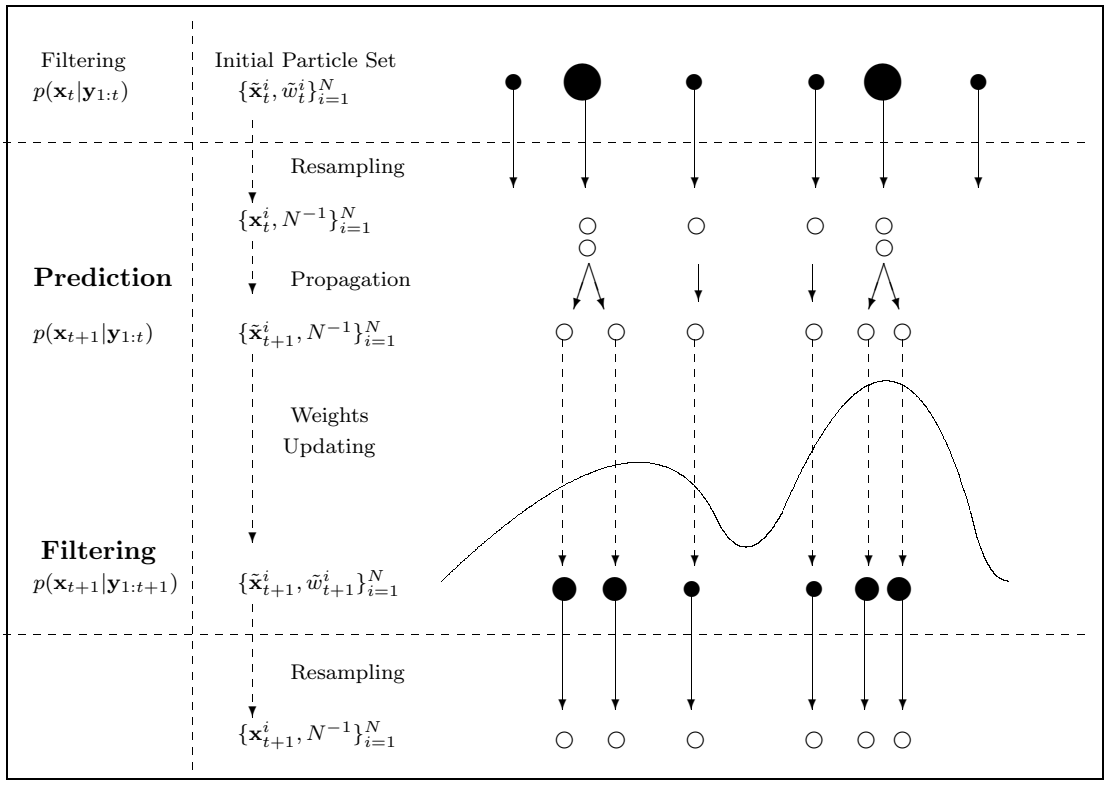

Figure 2.12: Particles evolution in the SIR particle filter.

This algorithm belongs to a wider class of bootstrap filters, which use a resampling step to generate a new set of particles with uniform weights. This step introduces diversity in particle set, avoiding degeneracy. In Algorithm 2.3.6, we give a pseudo-code representation of this method.

Note that in the SIR particle filter, we assumed $q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta) = p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta)$. Moreover, due to the resampling step, the weights are uniformly distributed over the particle set: $w_t^i = 1/N$, thus the weights updating relation becomes: $\tilde{w}_{t+1}^i \propto w_t^i\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta)$.

However, the basic SIR algorithm produces a progressive impoverishment (loss of diversity) of the information contained in the particle set, because of the resampling step and of the fact that particles do not change over filter iterations.

Many solutions have been proposed in literature. We recall the *Regularised Particle Filter* proposed by Musso, Oudjane and LeGland [39], which is based on a discretisation of the continuous state space.

Algorithm 2.3.6. - ***SIR Particle Filter*** -
Given the initial set of particles $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$:

1. *Simulate* $\tilde{\mathbf{x}}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}, \theta)$

2. *Update the weights:* $\bar{w}_{t+1}^i \propto p(\mathbf{y}_{t+1}|\tilde{\mathbf{x}}_{t+1}^i, \theta)$

3. *Normalize the weights:* $\tilde{w}_{t+1}^i = \bar{w}_{t+1}^i \left(\sum_{j=1}^N \bar{w}_{t+1}^j\right)^{-1}$, *for* $i = 1, \ldots, N$.

4. *Simulate* $\{\mathbf{x}_{t+1}^i\}_{i=1}^N$ *from the empirical density* $\{\tilde{\mathbf{x}}_t^i, \tilde{w}_t^i\}_{i=1}^N$

5. *Assign* $w_{t+1}^i = 1/N$, *for* $i = 1, \ldots, N$.

Gilks and Berzuini [4] propose the SIR-Move algorithm, which moves particles after the re-sampling step. Thus, particle value changes and the impoverishment is partially avoided. Finally, Pitt and Shephard [40] introduce the *Auxiliary Particle Filter* (APF) and applied it to a Gaussian ARCH-type stochastic volatility model. They find that the auxiliary particle filter works well and that the sensibility to outliers is lower than in the basic filters. In the following we focus on the APF algorithm.

In order to avoid re-sampling, the APF algorithm uses an auxiliary variable to select most representative particles and to mutate them through a simulation step. Then weights of the regenerated particles are updated through an importance sampling argument. In this way particles with low probability do not survive to the selection and the information contained in the particle set is not wasted. In particular the auxiliary variable μ_t^i contains and resumes the information on the previous particle set and it is used in the selection step to sample the random particle index. Note that the empirical filtering density given in Eq. (2.57) is a mixture of distributions, which can be reparameterised by introducing an auxiliary variable $i \in \{1, \ldots, N\}$, which indicates the component of the mixture. The joint distribution of the hidden state and of the index i is

$$
\begin{aligned}
p(\mathbf{x}_{t+1}, i|\mathbf{y}_{1:t+1}, \theta) &= \frac{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \mathbf{x}_{t+1}, i)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \theta)} p(\mathbf{x}_{t+1}, i|\mathbf{y}_{1:t}, \theta) = \\
&= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \theta)} p(\mathbf{x}_{t+1}|i, \mathbf{y}_{1:t}, \theta) p(i|\mathbf{y}_{1:t}, \theta) = \\
&= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) w_t^i.
\end{aligned} \tag{2.64}
$$

The basic idea of the APF is to refresh the particle set while reducing the loss of information due to this operation. Thus, the algorithm generates a new set of particles by jointly simulating the particle index i (*selection step*) and the selected particle value $\mathbf{x}_{t+1}$ (*mutation step*) from the reparameterised empirical filtering density, according to the following importance density

$$
\begin{aligned}
q(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta) &= q(\mathbf{x}_{t+1}^j|\mathbf{y}_{1:t+1}, \theta) q(i^j|\mathbf{y}_{1:t+1}, \theta) \\
&= p(\mathbf{x}_{t+1}^j|\mathbf{x}^{i^j}, \theta)(p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta) w_t^{i^j})
\end{aligned} \tag{2.65}
$$

for $j = 1, \ldots, N$. Note that the index is sampled using weights which are proportional to the observation density conditional on a summary statistics on the initial particle set. In this way, less informative particles are discarded. The information contained in each particle is evaluated with respect to both the observable variable and the initial particle set. By following the usual importance sampling argument, the updating relation for the particle weights is

$$
\begin{aligned}
w_{t+1}^j &\triangleq \frac{p(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta)}{q(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta)} \\
&= \frac{p(\mathbf{x}_{t+1}^j|\mathbf{x}^{i^j}, \theta) p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta) w_t^{i^j}}{p(\mathbf{x}_{t+1}^j|\mathbf{x}^{i^j}, \theta) p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta) w_t^{i^j}} \\
&= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta)}
\end{aligned} \tag{2.66}
$$

In Algorithm 2.3.7 we give a pseudo-code representation of the Auxiliary Particle Filter.

Algorithm 2.3.7. - ***Auxiliary Particle Filter*** -
Given the initial set of particles $\{\mathbf{x}_t^j, w_t^j\}_{j=1}^N$, *for* $j = 1, \ldots, N$:

1. *Calculate* $\mu_{t+1}^j = \mathbb{E}(\mathbf{x}_{t+1}|\mathbf{x}_t^j, \theta)$
2. *Simulate* $i^j \sim q(i|\mathbf{y}_{1:t+1}, \theta) \propto w_t^i \, p(\mathbf{y}_{t+1}|\mu_{t+1}^i, \theta)$ *with* $i \in \{1, \ldots, N\}$
3. *Simulate* $\mathbf{x}_{t+1}^j \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{i^j}, \theta)$
4. *Update particles weights:* $\tilde{w}_{t+1}^j \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta)}$.
5. *Normalize the weights:* $w_{t+1}^i = \tilde{w}_{t+1}^i \, (\sum_{j=1}^N \tilde{w}_{t+1}^j)^{-1}$, *for* $i = 1, \ldots, N$.

In Examples 2.3.3 and 2.3.4 we show how resampling can improve the performance of the basic SIS algorithm. In particular, we apply SIR and APF algorithms to the stochastic volatility model, illustrated in Example 2.2.1 and make a comparison between them, in terms of Root Means Square Error.

Example 2.3.3 - *(SV Model and SIR algorithm, Example 2.3.2 continued)*
In this example we show how the selection (or resampling) step can improve the performance of the basic Sequential Importance Sampling algorithm when applied to the Stochastic Volatility model given in Example 2.2.1.

In order to implement SIR algorithm we introduce a resampling step after the propagation of the initial set of particle $\{h_t^i, w_t^i = 1/N\}_{i=1}^N$. The steps of the SIR are

(i) Simulate $\tilde{h}_{t+1}^i \sim \mathcal{N}\left(h_{t+1}|\nu + \phi h_t^i, \sigma^2\right)$

(ii) Update the weights

$$\begin{aligned} \bar{w}_{t+1}^i &\propto w_t^i p(y_{t+1}|\tilde{h}_{t+1}^i, \theta) \propto \\ &\propto w_t^i \exp\left\{-\frac{1}{2}\left[y_{t+1}^2 e^{-\tilde{h}_{t+1}^i} + \tilde{h}_{t+1}^i\right]\right\} \end{aligned}$$

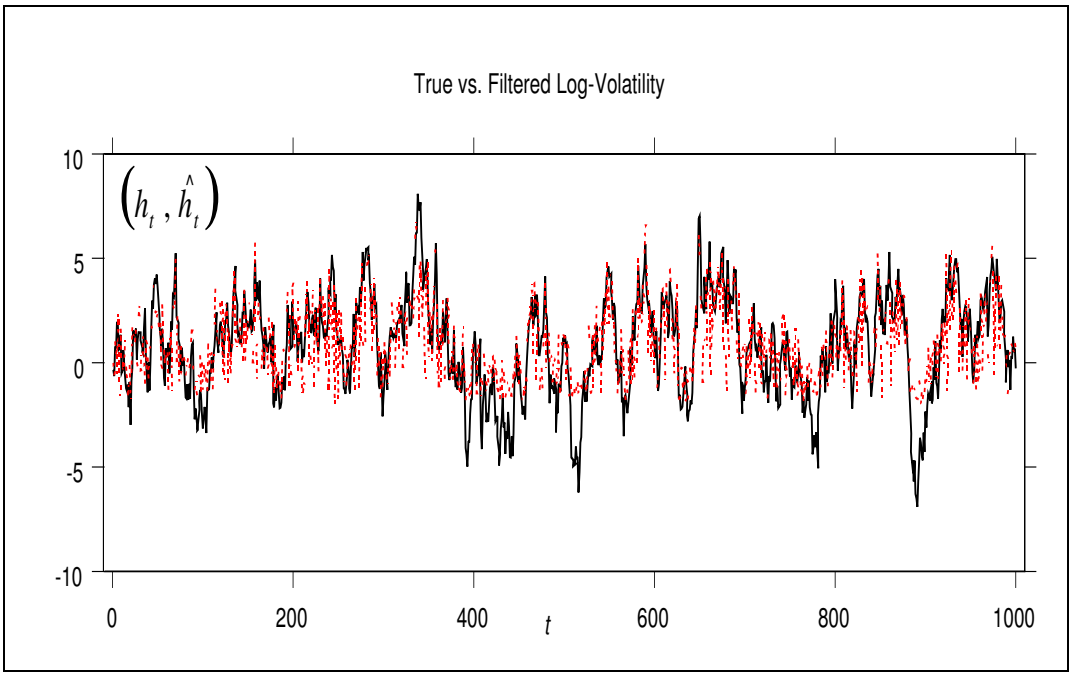

Figure 2.13: True (solid line) versus Filtered (dotted line) log-volatility obtained by applying a SIR algorithm, with $N = 3,000$ particles at each time step.

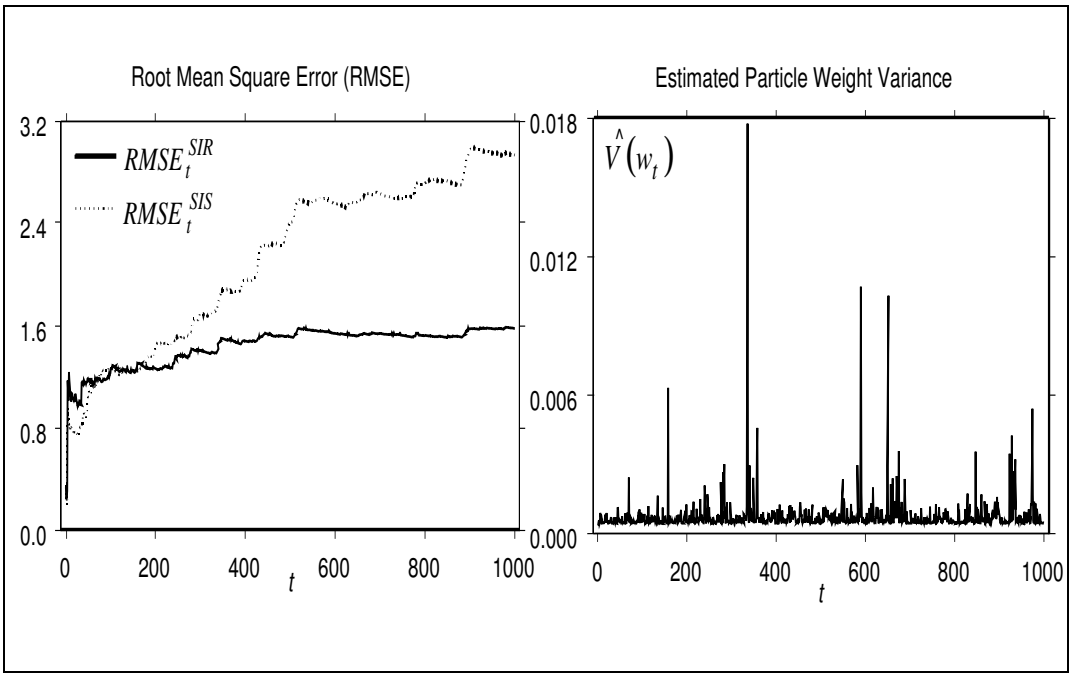

Figure 2.14: The first graph compares SIS and SIR algorithms in terms of Root Means Square Error between true and filtered log-volatility. The second graph shows the estimated variance of the particle weights, which allows us to detect degeneracy of the particle weights.

(iii) Normalize the weights

$$\tilde{w}_{t+1}^{i} = \frac{\bar{w}_{t+1}^{i}}{\sum_{j=1}^{N} \bar{w}_{t+1}^{j}}$$

(iv) Simulate $h_{t+1}^{i} \sim \sum_{i=1}^{N} \tilde{w}_{t+1}^{i} \delta_{\tilde{h}_{t+1}^{i}}$

(v) Set $w_{t+1}^{i} = 1/N\,, \quad \forall i = 1, \ldots, N$

By applying the SIR algorithm to the synthetic data, simulated in Example 2.2.1, we obtain the filtered log-volatility represented in Fig. 2.13. In Fig. 2.14, the first graph on the left compares SIS and SIR algorithm performances evaluated in terms of RMSE. Note how resampling, (selection step), effectively improves the performance of the basic SIS filter, avoiding the degeneracy of the particle weights and reducing the RMSE.

□

Example 2.3.4 - *(SV Model and APF algorithm, Example 2.3.2 continued)* In this example we apply an Auxiliary Particle Filter Algorithm to the Stochastic Volatility model given in Example 2.2.1.

In order to implement the algorithm we introduce a selection step before the propagation of the set of particles $\left\{h_t^j, w_t^j = 1/N\right\}_{j=1}^{N}$. The steps of the APF are

(i) Calculate $\mu_{t+1}^{j} = \phi h_t^j + \nu$ for $j =, \ldots, N$

(ii) Simulate $i^j \sim \sum_{j=1}^{N} w_t^j \mathcal{N}(y_{t+1}|\mu_{t+1}^{j})$ for $j =, \ldots, N$

(iii) Simulate $\tilde{h}_{t+1}^{j} \sim \mathcal{N}\left(h_{t+1}|\nu + \phi h_t^{i^j}, \sigma^2\right)$ for $j =, \ldots, N$

(iv) Update the weights

$$\bar{w}_{t+1}^{j} \quad \propto \quad \frac{\mathcal{N}(y_{t+1}|0, e^{\tilde{h}_{t+1}^{j}})}{\mathcal{N}(y_{t+1}|0, e^{\mu_{t+1}^{i^j}})}$$

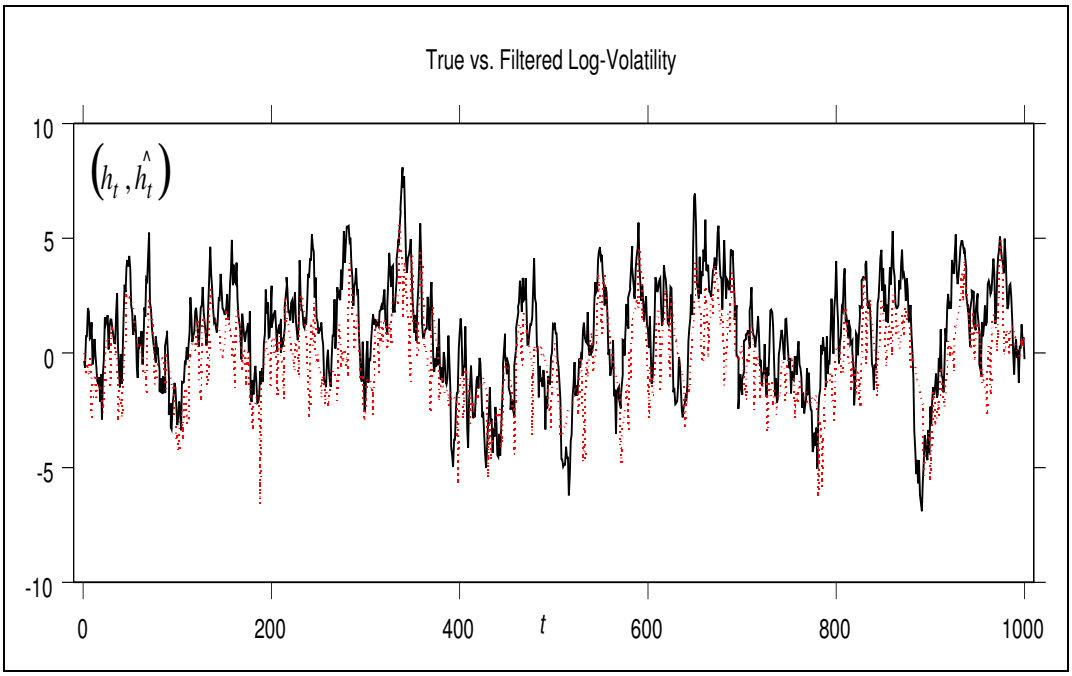

Figure 2.15: True (solid line) versus Filtered (dotted line) log-volatility obtained by applying an APF algorithm, with $N = 3,000$ particles at each time step.

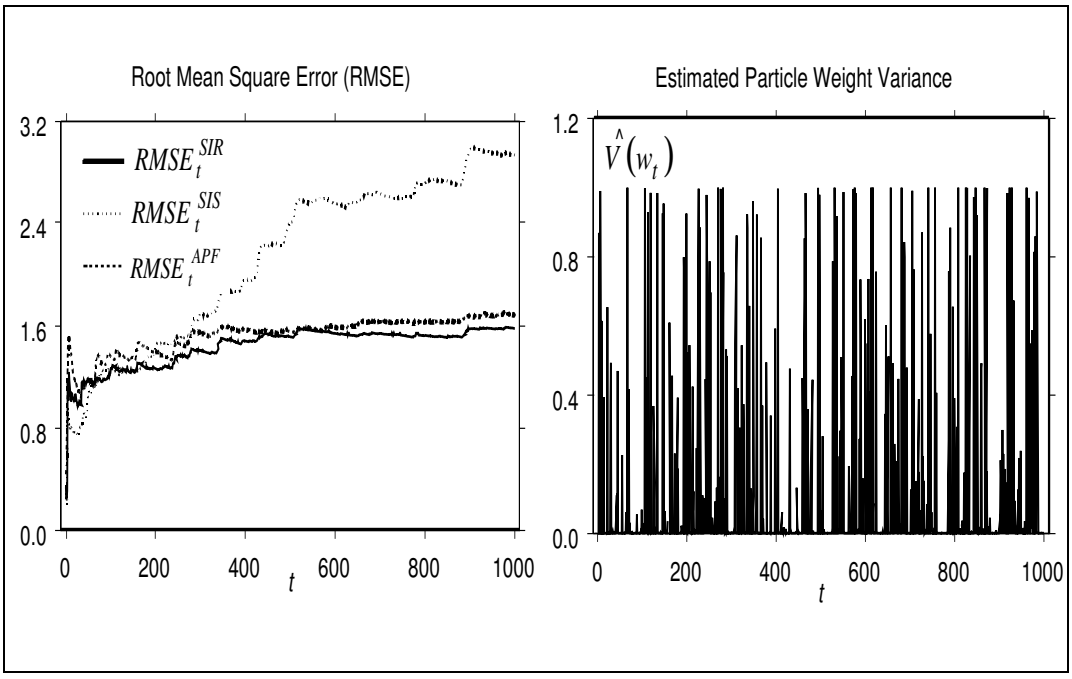

Figure 2.16: The first graph compares SIS, SIR and APF algorithms in terms of Root Means Square Error between true and filtered log-volatility. The second graph shows the estimated variance of the particle weights, which allows us to detect degeneracy of the particle weights.

(iii) Normalize the weights

$$\tilde{w}_{t+1}^{j} = \frac{\bar{w}_{t+1}^{i}}{\sum_{j=1}^{N} \bar{w}_{t+1}^{j}}$$

Note that, for simulating auxiliary variables i^j, Pitt and Shephard [40] suggest to use another proposal distributions, based on the Taylor expansion of the measurement distribution.
By applying the APF algorithm to the synthetic data, simulated in Example 2.2.1, we obtain the filtered log-volatility represented in Fig. 2.15. In Fig. 2.16, the first graph on the left compares SIS, SIR and APF algorithms in terms of RMSE. Although the variability in the weights of the particle set and the RMSE are greater than in the SIR algorithm, there is not degeneracy and the performance of APF algorithm is superior than that one of the basic SIS algorithm. The poor performance of the APF with respect to the SIR algorithm evidence a well known problem of this kind of algorithm. When the transition density exhibits a high noise variance the use of APF does not improve filtering results.

□

We conclude this section with a brief discussion of the problem of parameter estimation (see Chapter 6 for a more detailed analysis), for dynamic models with hidden variables, in a sequential data-processing approach. In principle parameter estimate and state filtering can be treated separately (see Storvik [47]). In many applications of particle filter techniques, parameters are treated as known and MCMC parameter estimates are used instead of the true parameter values. But in this way parameter estimate are not continuously updated as the hidden states. MCMC is typically a off-line approach, it does not allow the sequential updating of parameter estimates, as new observations arrive. Moreover, when applied sequentially, MCMC estimation method is more time consuming than particle filter algorithms.

One of the main issue in researching on particle filter is the inclusion of the parameter estimation procedure in the state filtering algorithm. Some studies have already extended sequential Monte Carlo techniques in order to jointly estimate state vectors and parameter. See Berzuini et al. [3] and Storvik [47] for a general

discussion of the problem and Liu and West [32] for the joint application of the adaptive importance sampling and the auxiliary particle filter. In the following we briefly show how joint parameter estimation and states filtering apply to a Bayesian dynamic model. In particular we apply the algorithm of Liu and West to the stochastic latent factor model given in Example 2.2.2. We refer to Chapter 6 for an updated review of the on-line parameter estimation problem, with applications to Markov switching stochastic volatility models.

Example 2.3.5 - *(APF and Latent Factor Models, Example 2.2.2 continued)*
The aim of this example is to show how particle filter algorithms apply to a widely used class of economic dynamic models: *Markov switching stochastic latent factor models*. In these models latent factor represents the trend of the market, while the switching states are the phases (growth and recession) of the market or of the economy.
We apply APF algorithm to synthetic data in order to show the efficiency of the algorithm and to detect possible degeneracy of the weights.

We refer to the Markov switching model given in Example 2.2.2 and apply the algorithm due to Liu and West [32]. This algorithm combines adaptive importance sampling for sequentially estimating the parameter vector with the auxiliary particle filter for filtering and predicting the hidden state. Observe that the latent structure of the MS model in the example exhibits two levels. The first one is given by the stochastic latent factor x_t and the second one is given by the regime switching process s_t. This stochastic structure makes the inference more difficult than in the simpler Hamilton's MS models.

The adaptation of the algorithm of Liu and West [32] to our MS model give us the following Particle Filter algorithm.

Algorithm 2.3.8. - ***APF for the Business Cycle Model*** -
Given an initial set of particles $\{\mathbf{x}_t^i, s_t^i, \theta_t^i, w_t^i\}_{i=1}^N$*:*

1. *Compute* $V_t = \sum_{i=1}^N (\theta_t^i - \bar{\theta}_t)(\theta_t^i - \bar{\theta}_t)' w_t^i$ *and* $\bar{\theta}_t = \sum_{i=1}^N \theta_t^i w_t^i$

2. *For $i = 1, \ldots, N$ and with a and b well chosen tuning parameters, calculate the following summarizing constant:*

   (a) $\tilde{S}^i_{t+1} = \underset{l \in 1,2}{arg\,max}\, \mathbb{P}(s_{t+1} = l | s_t = s^i_t)$

   (b) $\tilde{X}^i_{t+1} = \mu^i_t + \nu^i_t \tilde{S}^i_{t+1} + \rho^i_t x^i_t$

   (c) $\tilde{\theta}^i_t = a\theta^i_t + (1-a)\bar{\theta}_t$, *where* $\tilde{\theta} = (\tilde{\alpha}, \tilde{\sigma}_\varepsilon, \tilde{\rho}, \tilde{\mu}, \tilde{\nu}, \tilde{\sigma}_\eta, \tilde{p}_{11}, \tilde{p}_{22})$

3. *For $i = 1, \ldots, N$:*

   (a) *Simulate* $k^i \propto q(k|\mathbf{y}_{1:t+1}, \theta) = N(\mathbf{y}_{t+1}|\tilde{\alpha}^k_t \tilde{X}^k_{t+1}, \tilde{\sigma}^k_{\varepsilon\, t}) w^k_t$, *with* $k \in \{1, \ldots, N\}$

   (b) *Simulate* θ^i_{t+1} *from* $N(\tilde{\theta}^{k^i}_t, b^2 V_t)$

   (c) *Simulate* $s^i_{t+1} \in \{1, 2\}$ *from* $\mathbb{P}(s^i_{t+1} = i | s^{k^i}_t)$

   (d) *Simulate* x^i_{t+1} *from* $N(\mu^i_{t+1} + \nu^i_{t+1} s^i_{t+1} + \rho^i_{t+1} x^{k^i}_t, \sigma^i_{\eta\, t+1})$

4. *Update weights* $\tilde{w}^i_{t+1} \propto N(\mathbf{y}_{t+1}|\alpha^i_{t+1} x^i_{t+1}, \sigma^i_{\varepsilon\, t+1}) / N(\mathbf{y}_{t+1}|\tilde{\alpha}^{k^i}_t \tilde{X}^{k^i}_{t+1}, \tilde{\sigma}^{k^i}_{\varepsilon\, t})$

5. *Normalize weights* $w^i_{t+1} = \tilde{w}^i_{t+1} (\sum_{i=1}^N \tilde{w}^i_{t+1})^{-1}$, *for* $i = 1, \ldots, N$.

The tuning parameters a and b are equal to $\frac{3\delta - 1}{2\delta}$ and $\sqrt{1 - a^2}$ respectively, where we chose $\delta = 0.99$ as suggested in West [53].

Figure 2.17 shows on-line estimation of parameters α, σ_ε, ρ, μ_0, μ_1, σ_η, p_{11}, p_{22} obtained by running APF algorithm on the synthetic dataset exhibited in Fig. 2.3. We use a set of $N = 1000$ particles to obtain empirical filtering and prediction densities. All computation have been carried out on a Pentium IV 2.4 Ghz, and the APF algorithm has been implemented in GAUSS 4.0. Figure 2.18 shows on-line estimation of the latent factor x_t and of the switching process s_t.

In order to detect the absence of degeneracy in the output of the APF algorithm we evaluate at each time step the *Survival Rate*. It is defined as the number of particles survived to the selection step over the total number of particles. Particles set degenerates when persistently exhibiting a high number of dead particles from a generation to the subsequent one. Survival rate is calculated as follow

$$SR_t = \{N - \sum_{i=1}^{N} \mathbb{I}_{\{0\}}(Card(I_{i,t}))\} N^{-1} \tag{2.67}$$

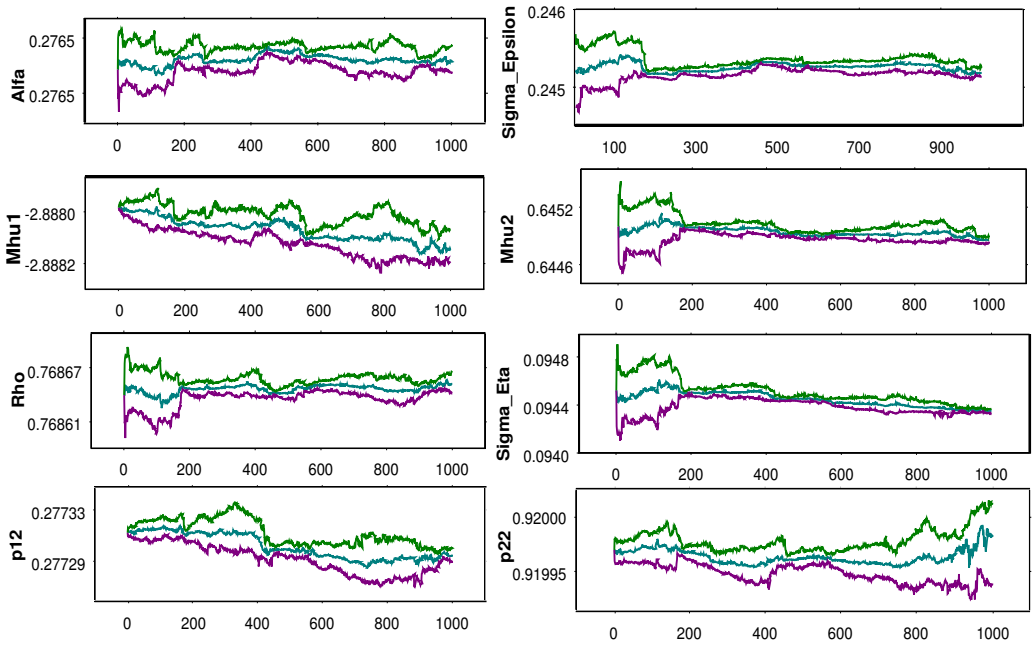

Figure 2.17: On-line parameter estimates. Graphs exhibit at each date the empirical mean and the quantiles at 0.275 and 0.925 for each parameter.

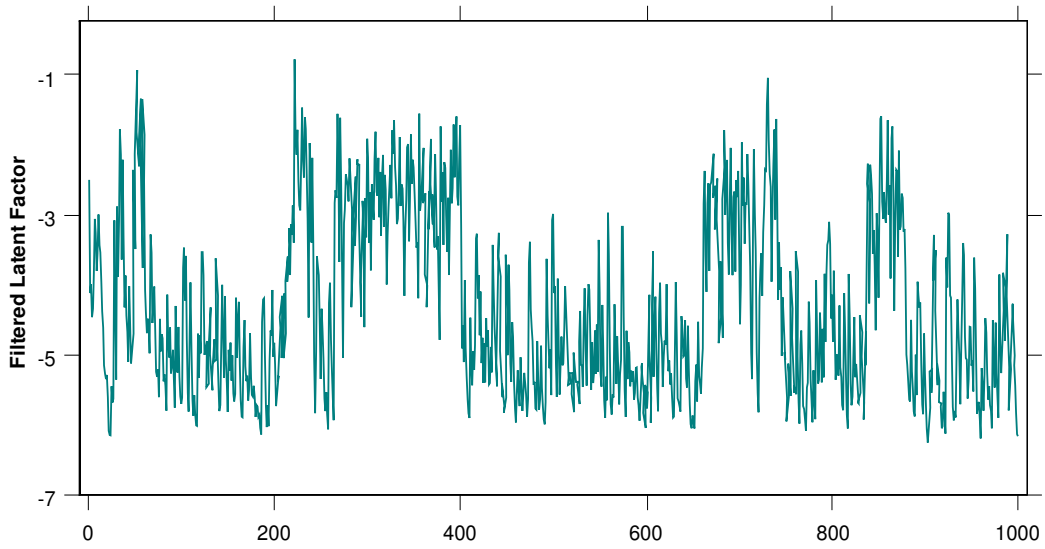

Figure 2.18: Sequentially filtered latent factor, x_t, over $T = 1,000$ observations.

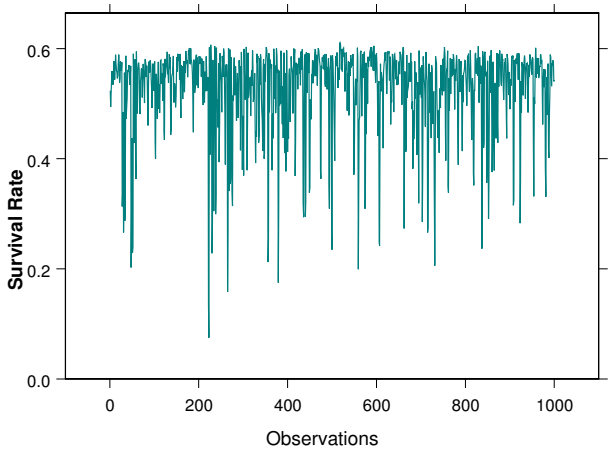

Figure 2.19: Survival Rate of the particle set at each time step.

where $I_{i,t} = \{j \in \{1, \ldots, N\} | i_t^j = i\}$ is the set of all random index values, which are selecting, at time t, the i-*th* particle. If at time t the particle k does not survive to the selection step then the set $I_{k,t}$ becomes empty. Graph 2.19 shows the survival rate at each time step. The rate does not decrease thus we conclude that the APF algorithm does not degeneracy in our study.

□

2.4 Conclusion

In this work we describe the Bayesian approach to general dynamic models analysis. We briefly review the literature on the latent factor dynamic models, focusing on the Bayesian approach and recognizing the importance of simulation based methods in Bayesian inference.

To deeply review simulation based methods in latent factor dynamic models we analyze the problems of state filtering and parameter estimation for quite general dynamic models used in time series analysis, such as business cycle and stochastic volatility models. We discuss general filtering, predicting and smoothing relations and give a proof of these relations.

In the second part of the work, we provide a review of the Bayesian simulation based inference and develop through some examples the application of the simulation based inference to switching latent factor models and to stochastic volatility models. Traditional MCMC methods, like single-move and multi-move Gibbs sampler have been discussed. Moreover, sequential Monte Carlo methods have been introduced by means of the adaptive importance sampling algorithm. Finally we analyze basic sequential techniques and recently developed sequential techniques like particle filter algorithms.

Appendix A - General Filtering

Proof. **- Recursive filtering relation given in Equation (2.24) -**
Consider the joint posterior density of the state vectors, conditional on the information available at time T

$$\begin{aligned} p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T},\theta) &= \frac{p(\mathbf{x}_{0:T},\mathbf{y}_T|\mathbf{y}_{1:T-1},\theta)}{p(\mathbf{y}_T|\mathbf{y}_{1:T-1},\theta)} = \\ &= p(\mathbf{x}_{0:T-1}|\mathbf{y}_{1:T-1},\theta)\frac{p(\mathbf{x}_T,\mathbf{y}_T|\mathbf{x}_{0:T-1},\mathbf{y}_{1:T-1},\theta)}{p(\mathbf{y}_T|\mathbf{y}_{1:T-1},\theta)} = \\ &= p(\mathbf{x}_{0:T-1}|\mathbf{y}_{1:T-1},\theta)\frac{p(\mathbf{y}_T|\mathbf{x}_{0:T},\mathbf{y}_{1:T-1},\theta)p(\mathbf{x}_T|\mathbf{x}_{0:T-1},\mathbf{y}_{1:T-1},\theta)}{p(\mathbf{y}_T|\mathbf{y}_{1:T-1},\theta)} \quad (2.68) \\ &= p(\mathbf{x}_{0:T-1}|\mathbf{y}_{1:T-1},\theta)\frac{p(\mathbf{y}_T|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)p(\mathbf{x}_T|\mathbf{x}_{T-1},\mathbf{y}_{1:T-1},\theta)}{p(\mathbf{y}_T|\mathbf{y}_{1:T-1},\theta)} = \\ &= p(\mathbf{x}_{0:T-1}|\mathbf{y}_{1:T-1},\theta)\frac{p(\mathbf{y}_T|\mathbf{x}_T,\theta)p(\mathbf{x}_T|\mathbf{x}_{T-1},\theta)}{p(\mathbf{y}_T|\mathbf{y}_{1:T-1},\theta)}. \end{aligned}$$

where the last line is due to the Markov property of the measurement and transition densities.

□

Proof. **- Recursive smoothing density given in Equation (2.25) -**
Consider the joint posterior density of the state vectors, conditional on the available information $\mathbf{y}_{1:T}$

$$\begin{aligned} & p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T},\theta) = \\ = \; & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{0:T-1}|\mathbf{x}_T,\mathbf{y}_{1:T},\theta) = \\ = \; & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{0:T-2}|\mathbf{x}_{T-1:T},\mathbf{y}_{1:T},\theta) = \\ \overset{Bayes}{=} \; & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)\frac{p(\mathbf{y}_T|\mathbf{x}_{T-1:T},\mathbf{y}_{1:T-1},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)}{p(\mathbf{y}_T|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)}p(\mathbf{x}_{0:T-2}|\mathbf{x}_{T-1:T},\mathbf{y}_{1:T},\theta) = \\ \overset{Markov}{=} \; & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)\frac{p(\mathbf{y}_T|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)}{p(\mathbf{y}_T|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)}p(\mathbf{x}_{0:T-2}|\mathbf{x}_{T-1:T},\mathbf{y}_{1:T},\theta) \quad (2.69) \\ = \; & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)p(\mathbf{x}_{0:T-2}|\mathbf{x}_{T-1},\mathbf{y}_{1:T},\theta). \end{aligned}$$

By applying iteratively Bayes theorem and Markov property of the dynamic model we obtain the recursive smoothing relation

$$
\begin{aligned}
& p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T},\theta) = \\
= \quad & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)\prod_{t=0}^{T-2} p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:T},\theta) \\
= \quad & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)\prod_{t=0}^{T-2} p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\mathbf{y}_{t+1:T},\theta) \qquad (2.70) \\
\overset{Bayes}{=} \quad & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)\prod_{t=0}^{T-2} \frac{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{t:t+1},\mathbf{y}_{1:t},\theta)p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta)}{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta)} \\
\overset{Markov}{=} \quad & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)p(\mathbf{x}_{T-1}|\mathbf{x}_T,\mathbf{y}_{1:T-1},\theta)\prod_{t=0}^{T-2} \frac{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta)p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta)}{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta)} \\
= \quad & p(\mathbf{x}_T|\mathbf{y}_{1:T},\theta)\prod_{t=0}^{T-1} p(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{y}_{1:t},\theta).
\end{aligned}
$$

□

Appendix B - Simulation Based Filtering

Proof. **- Single-Move Gibbs Sampler, Equation (2.38) -**
Consider the full posterior density of the t-th state vector and apply the independence assumption between $\mathbf{y}_{t+1:T}$ and $\mathbf{x}_t$

$$\begin{aligned}
p(\mathbf{x}_t|\mathbf{x}_{-t},\mathbf{y}_{1:T},\theta) &= p(\mathbf{x}_t|\mathbf{x}_{-t},\mathbf{y}_{1:t},\mathbf{y}_{t+1:T},\theta) \\
&= \frac{p(\mathbf{x}_t,\mathbf{y}_{t+1:T}|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta)}{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta)} = \\
&= \frac{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{0:T},\mathbf{y}_{1:t},\theta)p(\mathbf{x}_t|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta)}{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta)} = \\
&= \frac{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta)p(\mathbf{x}_t|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta)}{p(\mathbf{y}_{t+1:T}|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta)} = \\
&= p(\mathbf{x}_t|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta).
\end{aligned}$$

We can simplify the last density as follow

$$\begin{aligned}
p(\mathbf{x}_t|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta) &= p(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{x}_{t+1:T},\mathbf{y}_{1:t},\theta) = \\
&= \frac{p(\mathbf{x}_{t:T},\mathbf{y}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1},\theta)}{p(\mathbf{x}_{t+1:T},\mathbf{y}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1},\theta)} = \\
&= \frac{p(\mathbf{x}_{t+1:T},\mathbf{y}_t|\mathbf{x}_{0:t},\mathbf{y}_{1:t-1},\theta)p(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1},\theta)}{p(\mathbf{x}_{t+1:T},\mathbf{y}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1},\theta)} = \\
&= \frac{p(\mathbf{x}_{t+1:T}|\mathbf{x}_{0:t},\mathbf{y}_{1:t},\theta)p(\mathbf{y}_t|\mathbf{x}_{0:t},\mathbf{y}_{1:t-1},\theta)p(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1},\theta)}{p(\mathbf{x}_{t+1:T},\mathbf{y}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1},\theta)} = \\
&\overset{Markov}{=} \frac{p(\mathbf{x}_{t+1:T}|\mathbf{x}_{0:t},\mathbf{y}_{1:t},\theta)p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{y}_{1:t-1},\theta)p(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{y}_{1:t-1},\theta)}{p(\mathbf{x}_{t+1:T},\mathbf{y}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1},\theta)}.
\end{aligned}$$

The full posterior density of the t-th state vector is thus proportional to

$$\begin{aligned}
& p(\mathbf{x}_t|\mathbf{x}_{-t},\mathbf{y}_{1:t},\theta) \propto \\
& p(\mathbf{x}_{t+1:T}|\mathbf{x}_{0:t},\mathbf{y}_{1:t},\theta)p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{y}_{1:t-1},\theta)p(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{y}_{1:t-1},\theta) = \\
= \quad & p(\mathbf{x}_{t+2:T}|\mathbf{x}_{0:t+1},\mathbf{y}_{1:t},\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t},\theta)p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{y}_{1:t-1},\theta)p(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{y}_{1:t-1},\theta) = \\
\overset{Markov}{=} \quad & p(\mathbf{x}_{t+2:T}|\mathbf{x}_{0:t+1},\mathbf{y}_{1:t},\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{1:t},\theta)p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{y}_{1:t-1},\theta)p(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{y}_{1:t-1},\theta) \propto \\
\propto \quad & p(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{1:t},\theta)p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{y}_{1:t-1},\theta)p(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{y}_{1:t-1},\theta).
\end{aligned}$$

□

Proof. **- Recursive weights updating relation given in Equation (2.62) -** Starting from the definition of importance weights

$$
\begin{aligned}
w_{t+1} &\stackrel{\Delta}{=} \frac{p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1},\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1},\theta)} = \\
&\stackrel{Bayes}{=} \frac{p(\mathbf{x}_{0:t+1},\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1},\theta)p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\theta)} = \\
&= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t},\theta)p(\mathbf{x}_{t+1},\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\mathbf{x}_{0:t},\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1},\theta)p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\theta)} = \\
&= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t},\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1},\theta)} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{0:t+1},\mathbf{y}_{1:t},\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t},\theta) = \quad (2.71) \\
&\stackrel{Markov}{=} \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t},\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1},\theta)} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta) = \\
&= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t},\theta)}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t},\theta)} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\theta)q(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t+1},\theta)} = \\
&= w_t \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\theta)q(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t+1},\theta)}.
\end{aligned}
$$

Thus particle weights updating recursive relation is

$$w_{t+1} \propto w_t \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta)}{q(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{t+1},\theta)}. \quad (2.72)$$

Moreover, if we assume that the importance density for the state $\mathbf{x}_{t+1}$ is the transition density: $q(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{t+1},\theta) = p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta)$, then equation (2.72) simplifies to

$$w_{t+1} \propto w_t\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta). \quad (2.73)$$

□

Bibliography

[1] Arulampalam S., Maskell S., Gordon N. and Clapp T. (2001), A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking, *Technical Report*, QinetiQ Ltd., DSTO, Cambridge.

[2] Bauwens L., Lubrano M. and Richard J.F., (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, New York.

[3] Berzuini C., Best N.G, Gilks W.R. and Larizza C., (1997), Dynamic conditional independence models and Markov chain Monte Carlo Methods, *Journal of the American Statistical Association*, Vol. 92, pp. 1403-1441.

[4] Berzuini C., Gilks W.R. (2001), Following a moving average-Monte Carlo inference for dynamic Bayesian models, *J.R. Statist. Soc. B*, vol. 63, pp.127-146.

[5] Carlin B.P., Polson N.G. and Stoffer D.S., (1992), A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modelling, *Journal of American Statistical Association*, Vol. 87, n.418, pp.493-500.

[6] Carter C.K. and Köhn R., (1994), On Gibbs Sampling for State Space Models, *Biometrika*, Vol. 81, n.3, 541-553.

[7] Casarin R., (2003), Bayesian Inference for Markov Switching Stochastic Volatility Models, *working paper*, CEREMADE, forthcoming.

[8] Chib S., Nardari F. and Shephard N. (2002), Markov chains Monte Carlo methods for stochastic volatility models, *Journal of Econometrics*, 108(2002), pp. 281-316.

[9] Crisan D. and Doucet A. (2000), Convergence of sequential Monte Carlo methods, *Technical Report 381*, CUED-F-INFENG.

[10] De Jong P. and Shephard N. (1995), The Simulation Smoother for Time Series Models, *Biometrika*, Vol. 82, Issue 2, pp. 339-350.

[11] Diebold F.X. and Rudebusch G.D., (1996), Measuring Business Cycles: A Modern Perspective, *The Review of Economics and Statistics*, 78, 67-77.

[12] Doucet A., Freitas J.G. and Gordon J., *Sequential Monte Carlo Methods in Practice*, Springer Verlag, New York.

[13] Doucet A., Godsill S. and Andrieu C. (2000), On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*,Vol. 10, pp. 197-208.

[14] Durland J. and McCurdy T., (1994), Duration-dependent transitions in a markov model of u.s. gnp growth, *Journal of Business and Economic Statistics*, 12, 279-288.

[15] Durbin J. and Koopman S.J., (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

[16] Frühwirth-Schnatter S., (1994), Data augmentation and dynamic linear models, *Journal of Time Series Analysis*, Vol. 15, n.2, 183-202.

[17] Goldfeld S.M. and Quandt R.E., (1973), A Markov Model for Switching Regression, *Journal of Econometrics*, 1, 3-16.

[18] Gordon N., Salmond D. and Smith A.F.M., (1993), Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation, *IEE Proceedings-F*, 1993, Vol. 140, pp. 107-113.

[19] Hamilton J.D., (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357-384.

[20] Hamilton J.D., (1994), *Time Series Analysis*, Princeton University Press.

[21] Harvey A.C., (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.

[22] Harrison J. and West M., (1997), *Bayesian Forecasting and Dynamic Models* 2nd Ed., Springer Verlag, New York.

[23] Kalman R.E., (1960), A new approach to linear filtering and prediction problems, *Journal of Basic Engineering, Transaction ASME, Series D*, 82, 35-45.

[24] Kalman R.E. and Bucy R.S., (1960), New results in linear filtering and prediction problems, *Transaction of the ASME-Journal of Basic Engineering, Series D*, 83, 95-108.

[25] Kim C.J., (1994), Dynamic linear models with Markov switching, *Journal of Econometrics*, 60, 1-22.

[26] Kim C.J. and Nelson C.R., (1999), *State-Space Models with Regime Switching*, Cambridge, MIT press.

[27] Kim C.J. and Murray C.J., (2001), Permanent and Transitory Components of Recessions, forthcoming, *Empirical Economics*.

[28] Kim C.J. and Piger J., (2000), Common stochastic trends, common cycles, and asymmetry in economic fluctuations, Working paper, n. 681, International Finance Division, Federal Reserve Board,Semptember 2000.

[29] Kim S., Shephard N. and Chib S. (1998), Stochastic volatility: likelihood inference and comparison with arch models, *Review of Economic Studies*, 65, pp. 361-393.

[30] Krolzig H.M., (1997), *Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*, Springer-Verlag.

[31] Liu J.S. and Chen R., (1998), Sequential Monte Carlo Methods for Dynamical System., *Journal of the American Statistical Association*, 93, pp. 1032-1044.

[32] Liu J. and West M., (2001), Combined Parameter and State Estimation in Simulation Based Filtering, in *Sequential Monte Carlo Methods in Practice* eds. Doucet A., Freitas J.G. and Gordon J., (2001), Springer-Verlag, New York.

[33] Liu J.S., Wong W.H. and Kong A., (1994), Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes, *Biometrika*, 81, 27-40.

[34] Liu J.S., Wong W.H. and Kong A., (1995), Correlation structure and convergence rate of the Gibbs sampler with various scans, *Journal of the Royal Statistical Society B*, 57, 157-169.

[35] Maybeck P.S., (1982), *Stochastic Models, Estimation and Control*, vol. 1, Academic Press.

[36] Maybeck P.S., (1982), *Stochastic Models, Estimation and Control*, vol. 2, Academic Press.

[37] Maybeck P.S., (1982), *Stochastic Models, Estimation and Control*, vol. 3, Academic Press.

[38] Müller P., (1992), Alternatives to the Gibbs sampling scheme, *Tech. report*, Institute of Statistics and Decision Sciences, Duke University.

[39] Musso C, Oudjane N. and LeGland F., (2001), Improving Regularised Particle Filters, in *Sequential Monte Carlo in Practice, eds Doucet A., Freitas J.G. and Gordon J.*, (2001), Springer Verlag, New York.

[40] Pitt M. and Shephard N., (1999), Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, Vol. 94(446), pp. 590-599.

[41] Potter S.M., (1995), A Nonlinear Approach to U.S. GNP, *Journal of Applied Econometrics*, 10, 109-125.

[42] Robert C.P., (2001), *The Bayesian Choice*, 2nd ed. Springer Verlag, New York.

[43] Robert C.P. and Casella G. (1999), *Monte Carlo Statistical Methods*, Springer Verlag, New York.

[44] Shephard N., (1994), Partial non-Gaussian state space, *Biometrika*, 81, 115-131.

[45] Shephard N. and Pitt M. K. (1997), Likelihood Analysis of Non-Gaussian Measurement Time Series, *Biometrika*, Vol. 84, Issue 3, pp. 653-667.

[46] Sichel D.E., (1991), Business cycle duration dependence: A parametric approach, *Review of Economics and Statistics*, 73, 254-256.

[47] Storvik G. (2002), Particle filters for state space models with the presence of unknown static parameters, *IEEE Trans. on Signal Processing*, 50, pp. 281-289.

[48] Tanner M. and Wong W., (1987), The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-550.

[49] Tierney L., (1994), Markov chains for exploring posterior distributions, Ann. of Statist., 22, 1701-1786.

[50] Tong H., (1983), *Threshold Models in Non-Linear Time-Series Models*, New York, Springer-Verlag, 1983.

[51] Watson J., (1994), Business cycle durations and postwar stabilization of the u.s. economy, *American Economic Review*, 84, 24-46.

[52] West M., (1992), Mixture models, Monte Carlo, Bayesian updating and dynamic models, *Computer Science and Statistics*, 24, pp. 325-333.

[53] West M., (1993), Approximating posterior distribution by mixtures, *Journal of Royal Statistical Society*, B, 55, pp. 409-442.

Chapter 3

Financial Modelling

3.1 Financial Data

3.1.1 Returns

Before starting a brief empirical analysis of some well known financial time series, we give some basic definitions and notation. Financial analysis concentrates on returns rather than raw prices because a return is a complete scale free measure of an investment opportunity and they display better statistical properties.

Let P_t indicate the raw price of an asset at time t. The *simple net return* R_t when holding the asset between $t-1$ and t is

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

and the *simple gross return* is $1 + R_t$. The gross return over the holding period $t-k, \ldots, t$ is obtained compounding the single period simple gross returns

$$1 + R_t(k) = (1 + R_t)(1 + R_{t-1}) \cdot \ldots \cdot (1 + R_{t-k+1})$$

Note that each simple net return must be associated to an interval of time. Multiperiod returns may be annualised to make investments with different horizon comparable

$$AR_t(k) = \left(\prod_{j=0}^{k-1} (1 + R_{t-j}) \right)^{\frac{1}{k}} - 1$$

Another way to compound return is based on the natural logarithm transformation of gross return. The *continuously compounded return* or *log-return* is

$$r_t = \log(1 + R_t) = \log\left(\frac{P_t}{P_{t-1}}\right) = p_t - p_{t-1}$$

where lower case indicates log has been taken. The multiperiod returns are defined as follow

$$r_t(k) = \log(1 + R_t(k)) = r_t + \ldots + r_{t-k+1}$$

The main disadvantage in using continuously compounded return is that the simple return of a portfolio, R_{pt}, is a weighted average of simple returns of the assets, R_{it},

$$R_{pt} = \sum_{i=1}^{N} w_{ip} R_{it}$$

while continuously compounded returns have not this property since the log of the sum is different from the sum of the log

$$r_{pt} \neq \sum_{i=1}^{N} w_{ip} r_{it}$$

Finally we consider the case when the asset pays dividend, D_t, just before the price at time t, P_t. The simple and the continuously compounded *dividend adjusted returns* are

$$\begin{aligned} R_t &= \frac{P_t + D_t}{P_{t-1}} - 1 \\ r_t &= \log(P_t + D_t) - \log(P_{t-1}) \end{aligned}$$

3.1.2 Financial Indexes

In the construction of stock (or bond) price indices such as the S&P500, S&P100, FTSE100, CAC40, NIKKEI, MSCI Indices, JP Morgan indices, etc. the effect of capitalisation changes on return measurement is taken into account through

renormalisation factors. Resulting returns are thus adjusted for dividend and capitalisation changes due to spin off and splits. Simple returns are

$$R = \frac{P_t f_t + D_t}{P_{t-1}} - 1 \tag{3.1}$$

where f_t is an renormalisation factor for changes in capitalisation. The weight of each stock in the index involves its market capitalisation and is obtained as follow

$$w_{i,t} = \frac{P_{i,t}Q_{i,t}}{\sum_{i=1}^{N} P_{i,t}Q_{i,t}} \tag{3.2}$$

where N is the number of assets in the index, $P_{i,t}$ and $Q_{i,t}$ are the price and number of shares outstanding for the i-th asset at time t.
The resulting index is called *value weighted* because variations in the index value are determined overall by the stocks with a high number of outstanding share, which are stocks with a high market value.
In order to assure that a value weighted index remains comparable over time, a renormalisation factor must be used. When events, such as stock splits occur, the price of the stock and the market value do not change, thus no adjustment to the index is needed. When events, such as spin-offs, share repurchase and issuances, special cash dividends, company changes occur, then market value and stock index value change. This variation is not due to market movements and thus the stock index must be adjusted through a renormalisation factor.

In this chapter we will use some indices which have already been adjusted for dividends and capitalisation changes.

3.2 Graphical and Statistical Analysis

In this section we briefly review some useful tools for time series analysis. In particular we focus on some tools which allows to detect skewness, excess of kurtosis and more generally the absence of normality in the observed variables.

3.2.1 Q-Q Plot

A very useful graphical tool for comparing two distribution is the Quantile-Quantile Plot (Q-Q Plot). This graphical techniques is based on the definition of quantiles.

Define the p-th quantile as the value x that satisfies to

$$Q(p) = \inf \{x : F(x) \geq p\} \tag{3.3}$$

or in other words the smallest value of x such that

$$\mathbb{P}r \{X < Q(p)\} \leq p \tag{3.4}$$

Through the Q-Q plot the empirical quantiles from data can be compared to the theoretical quantiles from a given distribution, say the normal. If the underlying data comes from the assumed theoretical distribution then the two set of quantiles should be the same and the plot of the empirical quantiles against those obtained from the theoretical distribution must lie on the 45 degree line. If the Q-Q plot does not lie on a straight line, the two distributions differ along some dimension.
The interpretation of the Q-Q plot can reveal many interesting features of the empirical distribution. If there are a few points above or below the straight line then these may be outliers from the base distribution. An S shaped curve means that the empirical distribution is leptokurtic and has thicker tails then the theoretical distribution. An inverted S indicates that the empirical distribution is platykurtic and exhibits thinner tails than the theoretical one.

3.2.2 Jarque-Bera Normality Test

The mean, μ, and the variance, σ^2, capture the first two moments, μ_1 and μ_2, of a distribution, in other words the location and the dispersion of the random variable, X, around the central value. Higher order moments capture other features of the shape of the density. In particular *Skewness* (left and right) is important when a distribution is asymmetric and has the left (right) tail longer than the other. One measure of the degree of skewness is the index of asymmetry

$$\alpha_3 = \frac{\mathbb{E}(X - \mu)^3}{\sigma^3} = \frac{\mu^3}{\sigma^3}. \tag{3.5}$$

Kurtosis measures the concentration of the values around the mean or the peakedness of the distribution. If the values are highly concentrated around the

mean the distribution is called platykurtic. If there is a high concentration on the tails the distribution is leptokurtic. A measure of kurtosis is

$$\alpha_4 = \frac{\mathbb{E}(X-\mu)^4}{\sigma^4} = \frac{\mu^4}{\sigma^4}. \tag{3.6}$$

For a normal distribution, the asymmetry index is zero and the kurtosis is 3. The Jarque-Bera test allows to verify if a random variable follow a normal distribution. The Jarque-Bera statistic involves the empirical estimation of asymmetry index and kurtosis and compares it with those that we would expect from a normal distribution

$$JB = \frac{T-k}{6}\left(S^2 + \frac{1}{4}(K-3)\right) \tag{3.7}$$

where S and K are the sample estimates of α_3 and α_4, T represents the number of observations and k the number of parameters that have been estimated. Under the null hypothesis of a normal distribution, the JB statistic asymptotically follows a χ^2 distribution with 2 degrees of freedom. Null hypothesis is rejected at 95% if the JB statistic is greater than 5.99 and at 99% if the JB statistic is greater than 9.21.

3.2.3 Goodness of Fit Test

Goodness of Fit (GoF) tests are used to verify if a sample of data comes from a given distribution. In the literature many GoF tests have been proposed. In the following the briefly review some well know tests, which will be used in this chapter.

The Chi-Square test groups data in k classes and compares the observed frequencies with that one obtained with a theoretical distribution. The test statistic proposed by K. Pearson is

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - n\,p_i^0)^2}{n\,p_i^0} \tag{3.8}$$

where N_i is the observed frequency for the i-th class and $n\,p_i^0$ the frequency we would expect for the theoretical distribution. The test is sensitive to the choice of the number of classes. Under the null hypothesis, that data follow the theoretical distribution the χ^2 statistic asymptotically follows a χ^2_{k-1}. If q parameters of the theoretical distribution have been estimated then the statistic follows a χ^2_{k-q-1}

Another important test is the Kolmogorov-Smirnov (KS) test. Given a sample $X_1, \ldots, X_n$, the KS statistic, D_n, requires the estimation of the empirical cumulative density function (EDF)

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(X_i)$$

and evaluates the maximum distance between the EDF and the theoretical cumulative density function $F(x)$

$$D_n = \sup_x |F_n(x) - F(x)|$$

If the random variables $X_1, \ldots, X_n$ are i.i.d then the limiting distribution of $K_n = \sqrt{n} D_n$ is known and can be used for a GoF test. If the statistic is greater than the critical value the null hypothesis will be rejected.
If the theoretical distribution belongs to a parametric family $F(x, \theta),\ \theta \in \Theta$ and the parameters θ is unknown, then it can be estimated and the resulting distribution used in the KS test. The limiting distribution under the null hypothesis has been tabulated in D'Agostino and Stephens [1].

In this chapter we use two extensions of the Kolmogorov-Smirnov statistics: the Cramér von Mises and the Anderson Darling. These tests belong to the general class of quadratic statistics

$$Q = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \varphi(x) dF(x) \tag{3.9}$$

where $\varphi(x)$ is a weighting function. If $\varphi(x) = 1$, we obtain the Cramér von Mises statistic, denoted by W^2. If $\varphi(x) = F(x)\,(1 - F(x))^{-1}$ we obtain the Anderson Darling statistic, denoted with A^2. The statistics W^2 and A^2 can be approximated as follow

$$\begin{aligned} W^2 &= \sum_{i=1}^{n} \left\{ F[x_i - \frac{(2i-1)}{2n}] \right\}^2 \\ A^2 &= -n - \left\{ \sum_{i=1}^{n} \frac{(2i-1)}{n} (\ln F(x_i) + \ln(1 - F(x_{n+1-i}))) \right\} \end{aligned}$$

In both cases if the test statistics result greater than a given critical value, the null hypothesis (the sample come from the theoretical distribution) will be rejected.

3.3 Some Empirical Results

The main aim of this section is to verify the absence of normality on financial data. The analysis concerns the international market, since we want to show that the high frequency of extreme returns is a phenomenon widely diffused and independent of a specific geographical area. Moreover we will study financial time series from various markets, i.e. stock, bonds and liquidity markets. We will use both statistical and graphical analysis to find evidence also of skewness, kurtosis, multimodality and clustering volatility.

3.3.1 Stock Indexes

First we analyse the S&P 500 index in order to evidence, through graphical and statistical analysis, the existence of some well known stylized facts for financial time series, i.e. excess of kurtosis, skewness, multimodality and volatility clustering. We consider annual returns on the S&P500 index, including dividend, observed with a monthly frequency on the time interval 01/1970-04/2001 and with a daily frequency on the time interval 02/01/1990-24/01/2003.

For the monthly frequency, the Jarque Bera test given in Tab. 3.1 allows to accept the null hypothesis of normally distributed returns. Also Goodness of Fit tests (i.e. χ^2, Kolmogorov-Smirnov, Cramer von Mises and Anderson Darling) given in Tab. 3.2 confirm normality. Nevertheless S&P500 index returns exhibit positive excess of kurtosis, 3.14, and negative skewness, -0.207. These elements suggest that data may also come from a distribution different from a Gaussian one. One of the most used alternative to the normal distribution is the Student-t. Thus we test the hypothesis Student-t distributed returns for various degrees of freedom. The Anderson Darling statistic captures the ability of a theoretical distribution to fit the tails of the empirical distribution. Thus we use this statistic to choose the best Student-t. Tab. 3.2 shows that the Student-t which exhibits the highest Anderson Darling statistic's p-value, has 7 degrees of freedom.

Table 3.1: Jarque Bera test on S&P 500 annual returns. Monthly observations over the period 01/1970-04/2001. The null hypothesis has been accepted with a p-Value of 0.247. E-Views software has been used for the test.

J.-B. Test	S&P 500 Index
Observations	364
Mean	0.147
Standard Deviation	0.158
Skewness	-0.207
Excess of Kurtosis	0.14
Min Value	-0.389
Max Value	0.611
Test Statistic	2.794
Critical Value 1%	9.21
Critical Value 5%	5.99
p-Value	0.247
Normality Hypoth.	Accepted

Table 3.2: Goodness of Fit tests on S&P 500 annual returns. Monthly observations over the period 01/1970-04/2001. For χ^2 test we use S-Plus, the other tests have been implemented in GAUSS. We use critical values of the modifies test statistics (in D'Agostino and Stephens [1]) for testing normality and Student-t (7 d.f.) hypothesis when parameters are estimated.

	χ^2	Kolm. Smir.	Cr.v. Mises	A.Darling
Test Statistic	16.769	0.675	0.08	0.495
Critical Value 1%	38.932	1.035	0.179	1.035
Critical Value 5%	32.671	0.895	0.126	0.752
Critical Value 10%	-	0.819	0.104	0.631
p-Value	0.725	-	-	0.211
Normality Hypoth.	Accepted	Accepted	Accepted	Accepted
Test Statistic	27.044	0.046	0.142	0.974
Critical Value 1%	38.932	0.085	0.743	3.857
Critical Value 5%	32.671	0.071	0.461	2.492
Critical Value 10%	-	0.064	0.347	1.933
p-Value	0.169	0.438	-	0.371
Student-t (7 d.f.) Hypoth.	Accepted	Accepted	Accepted	Accepted

Table 3.3: Jarque Bera test on MSCI stock index annual returns. Monthly observations on the period 01/1986-12/1999. E-Views software has been used for the test.

J.-B. Test	MSCI Europe	MSCI North America
Mean	0.140	0.156
Standard Deviation	0.175	0.133
Skewness	-0.050	-0.338
Excess of Kurtosis	-0.617	-0.186
Min Value	-0.197	-0.216
Max Value	0.532	0.496
Test Statistic	2.635	4.375
Critical Value 1%	9.210	9.210
Critical Value 5%	5.991	5.991
p-Value	0.2679	0.1122
Normality Hypoth.	Accepted	Accepted

Table 3.4: Goodness of Fit tests on MSCI stock index annual returns. Monthly observations on the period 01/1986-12/1999. For χ^2 test we use S-Plus, the other tests have been implemented in GAUSS. We use critical values of the modifies test statistics (in D'Agostino and Stephens [1]) for testing normality hypothesis when parameters are estimated.

	χ^2	Kolm. Smir.	Cr.v. Mises	A.Darling
MSCI Europe				
Test Statistic	14.095	0.709	0.051	0.461
Critical Value 1%	30.578	1.0350	0.179	1.035
Critical Value 5%	24.996	0.8950	0.126	0.752
Critical Value 10%	-	0.8190	0.104	0.631
p-Value	0.518	-	-	0.256
Normality Hypoth.	Accepted	Accepted	Accepted	Accepted
MSCI North America				
Test Statistic	16.952	0.92649	0.14139	0.8937
Critical Value 1%	30.578	1.0350	0.179	1.035
Critical Value 5%	24.996	0.8950	0.126	0.752
Critical Value 10%	-	0.8190	0.104	0.631
p-Value	0.322	-	-	0.0211
Normality Hypoth.	Accepted	Rejected 5%	Rejected 5%	Rejected 5%

As exhibited in Tab. 3.3, elements of skewness and kurtosis are present also in MSCI stock index returns when observed with a monthly frequency, although the Jarque Bera statistics and the Goodness of Fit tests (see Tab. 3.4) suggest to accept the null hypothesis of Gaussian distributed returns.

We can conclude that returns of the analysed stock indexes (i.e. S&P500, MSCI North America and MSCI Europe) exhibit kurtosis and skewness which cannot be modelled through the Gaussian distribution. Moreover an accurate modelling of the tails becomes important in risk measurement and management, thus heavy tail distributions should be used instead of the Gaussian one.

In the following we show how skewness and excess of kurtosis become markedly evident, on stock market indexes, when these variables are observed at higher frequency (i.e. daily frequency). Moreover we find evidence of time varying volatility and volatility clustering (see Fig. 3.1).

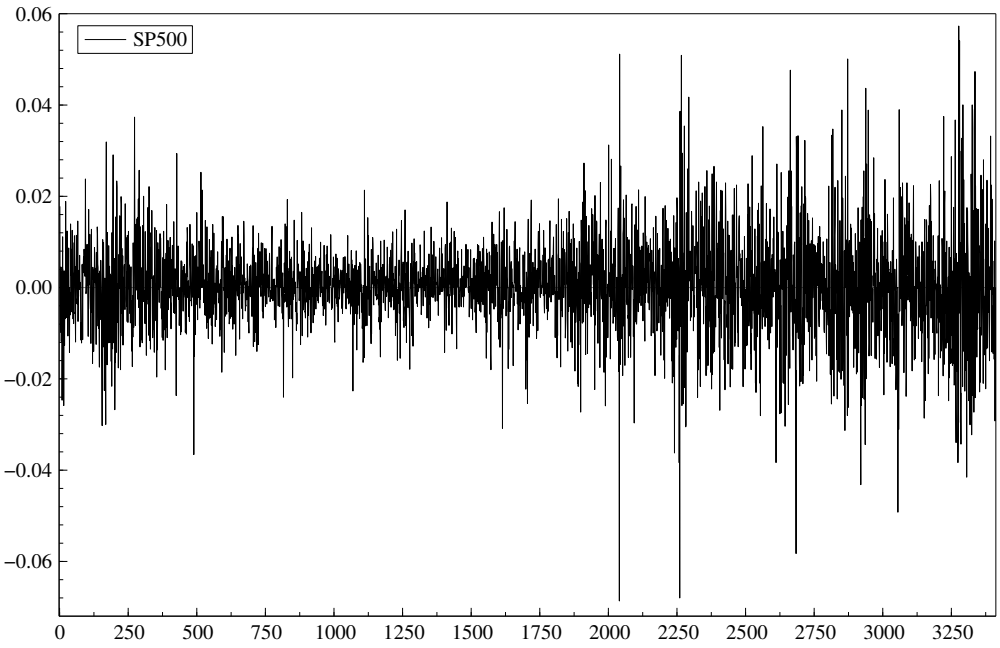

Figure 3.1: Daily returns on S&P500 stock index.

The first graph in Fig. 3.2 shows the histogram of the S&P500 daily returns, the Gaussian kernel density estimation and the best normal distribution. The second graph evidences that the left tail of best normal distribution do not fit well the tail of the empirical density. Daily returns exhibit a positive excess of kurtosis and this

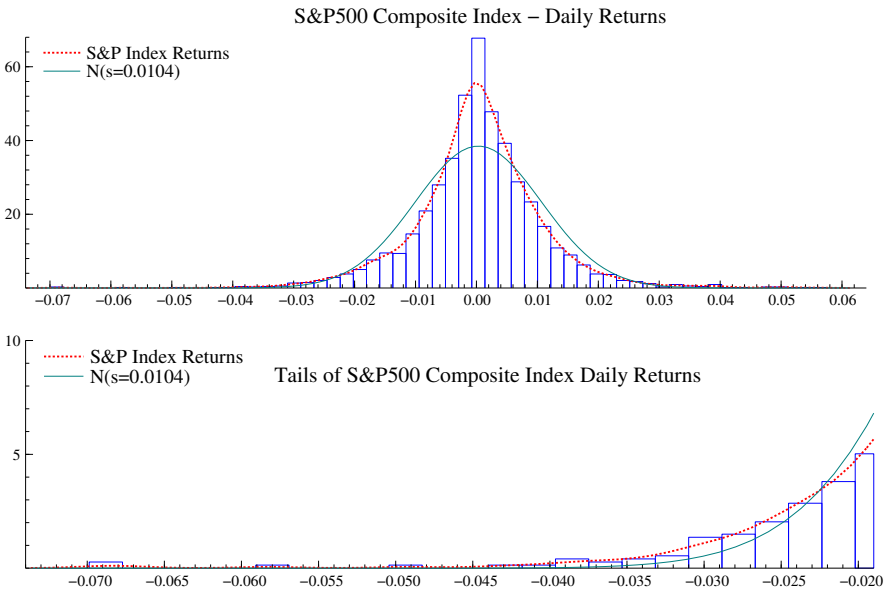

Figure 3.2: Histogram, Gaussian kernel density estimation and the best normal for daily returns on S&P500 stock index.

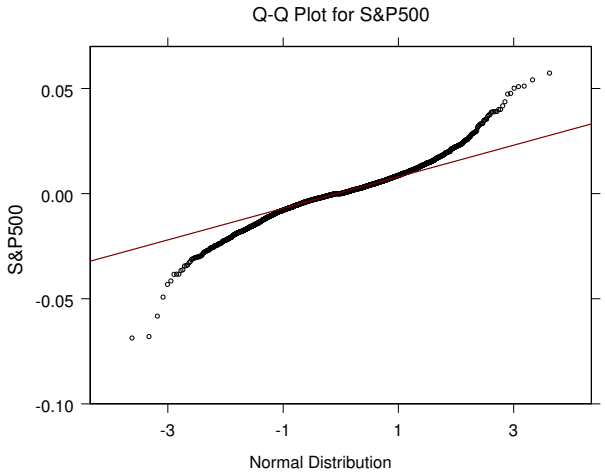

Figure 3.3: Q-Q Plots for daily returns on S&P500 stock indexes.

Table 3.5: Jarque Bera test on S&P 500 annual returns observed with daily frequency over the period 02/01/1990-24/01/2003. The null hypothesis has been rejected with a p-Value of 0. E-Views software has been used for the test.

J.-B. Test	S&P 500 Index
Observations	3410
Mean	0.000315
Standard Deviation	0.010368
Skewness	-0.016920
Excess of Kurtosis	0.896477
Min Value	-0.068657
Max Value	9.957315
Test Statistic	2157.347
Critical Value 1%	9.210
Critical Value 5%	5.991
p-Value	0.0000
Normality Hypothesis	Rejected

can be detected also in the Q-Q plot (Fig. 3.3). Tab. 3.5 evidences the presence of positive excess of kurtosis and negative asymmetry. The Jarque Bera statistic allows to reject the Gaussian distribution hypothesis.

We conclude that on stock market heavy tails and asymmetry are two marked features of the asset returns when data are observed both with a monthly and a daily or higher frequency. Therefore models alternative to the Gaussian distribution are needed. For this reason in this thesis we analyse non-Gaussian and nonlinear models. In particular in Chapter 4 we use Student-t distributions in a portfolio model, in Chapter 5 we propose mixtures of α-stable distributions for asset returns and finally in Chapter 6 we analyse a stochastic volatility model with non-Gaussian innovations and a Markovian jump components.

3.3.2 Bond Indexes

In this paragraph we analyse the features of time series on the bond market. In particular we consider the returns on the JP Morgan bond indexes.
In Tab. 3.6 we analyse the normality of the returns in the following countries: Great Britain, Japan, France and Belgium, over the period 01/1986-01/2000. For these markets the Jarque Bera test allows to accept the Gaussian distribution hypothesis

for the returns series of all the bond indexes. Nevertheless all time series exhibit negative skewness and excess of kurtosis.

Table 3.6: Jarque Bera test on JP Morgan bond indexes annual return. Monthly observation over the period 01/1986-01/2000. E-Views software has been used for the test.

J.-B. Test	Great Britain	Japan	France	Belgium
Observations	157	157	157	157
Mean	0.114405	0.064719	0.094355	0.092792
Standard Deviation	0.067214	0.055158	0.060213	0.047616
Skewness	-0.352192	-0.37995	-0.22498	-0.3093
Excess of Kurtosis	-0.094495	-0.18288	-0.0127	-0.3114
Min Value	-0.07032	-0.0895	-0.05496	-0.02507
Max Value	0.27148	0.185556	0.256071	0.189196
Test Statistic	3.94	5.3117	1.4112	4.1472
Critical Value 1%	9.21	9.21	9.21	9.21
Critical Value 5%	5.991	5.991	5.991	5.991
p-Value	0.1395	0.0702	0.4938	0.1257
Normality Hypothesis	Accepted	Accepted	Accepted	Accepted

We extend the analysis to the daily frequency data and consider JP Morgan index returns in the following countries: Great Britain, Japan, France, Germany, Italy and Belgium, on the period (04/11/1988-13/01/2003).

The time series are represented in Fig. 3.4. All indexes exhibit time varying volatility and clustering in volatility, which contribute to rise the kurtosis of the empirical distribution. Fig. 3.5 shows the histograms, the Gaussian kernel estimation of the empirical density function and the best normal approximation. For each histogram a detailed representation of the tails behavior is given. Note that in all the analysed series the estimated Gaussian density is not able to fit the tails of the estimated empirical density which always exhibits fatter tails. This feature is evident also in the Q-Q plots given in Fig. 3.6. In order to statistically detect heavy tails we conduct the Jarque Bera test. From Tab. 3.8, it results that returns exhibit negative asymmetry and positive excess of kurtosis.

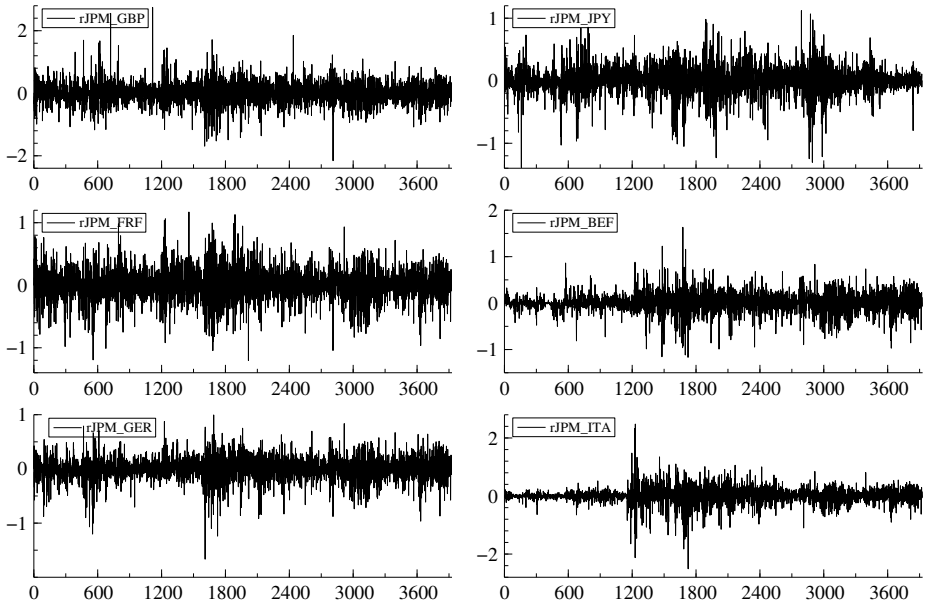

Figure 3.4: Time series of daily returns on JPM Bond indexes.

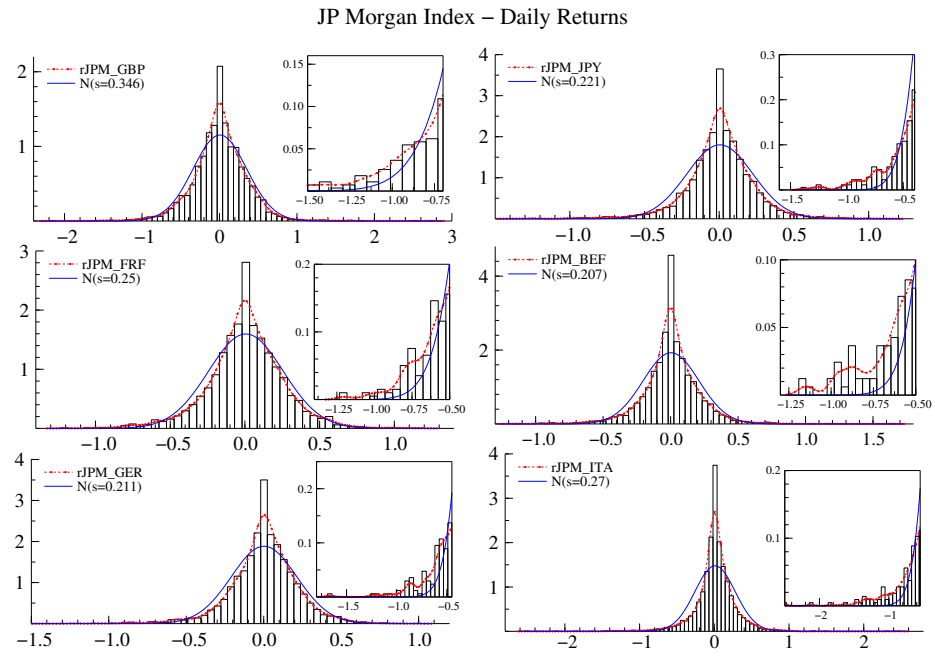

Figure 3.5: Histogram, Gaussian kernel density estimation and the best normal for daily returns on JPM Bond indexes.

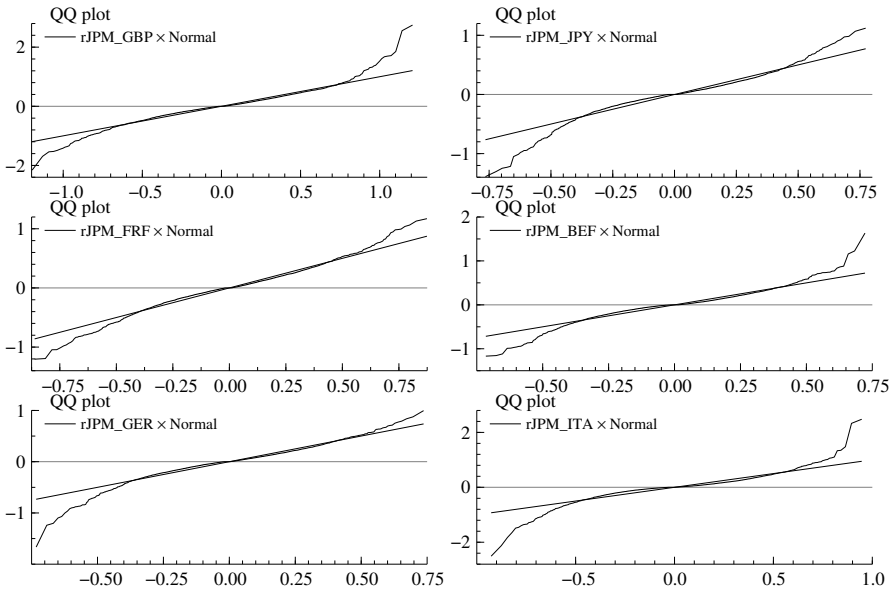

Figure 3.6: Q-Q Plots for daily frequency returns on JPM Bond indexes.

Table 3.7: Jarque Bera Test on JPM bond indexes annual return. Daily frequency observations over the period 04/11/1988-13/01/2003. E-Views software has been used for the test.

J.-B. Test	Belgium	France	G. Britain	Germany	Italy	Japan
Observations	3922	3922	3922	3922	3922	3922
Mean	0.00323	0.0062	0.0055	0.0020	0.0095	0.0040
Std. Dev.	0.2068	0.2497	0.3458	0.2113	0.26955	0.22125
Skewness	-0.2017	-0.1561	0.1322	-0.5948	-0.5100	-0.4643
Exc. of Kurt.	4.6296	1.8485	4.1773	3.6658	10.4277	4.0600
Min Value	-1.1681	-1.2044	-2.1581	-1.6660	-2.5066	-1.3889
Max Value	1.6323	1.1707	2.7469	0.9962	2.4778	1.1211
Test Statistic	3529.28	574.37	2863.10	2427.36	17939.53	2834.71
Crit. Val. 1%	9.21	9.21	9.21	9.21	9.21	9.21
Crit. Val. 5%	5.991	5.991	5.991	5.991	5.991	5.991
p-Value	0	0	0	0	0	0
Norm. Hyp.	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected

3.3.3 Liquidity Market

In this section we show how the Gaussian distribution hypothesis can be sometimes rejected also due to multimodality of the empirical distribution. In particular we consider 3-months interest rate on Eurodeposits observed with daily frequency over the period 01/11/1986-13/01/2003, for the following countries: USA, UK, Japan, France, Italy and Germany. Fig. 3.7 shows the interest rate time series, which have been analysed.

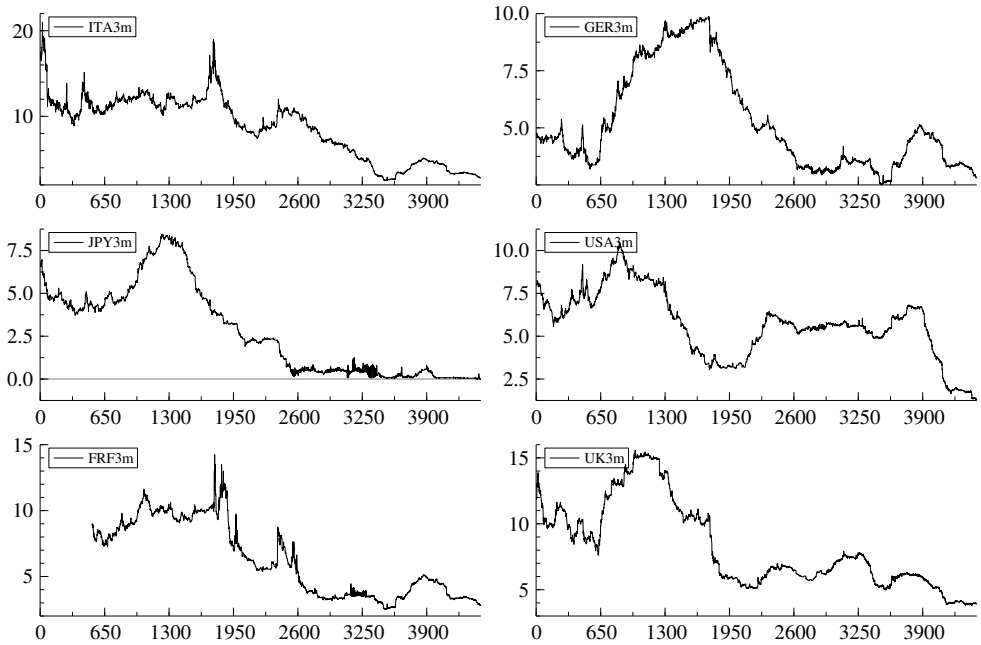

Figure 3.7: Daily Eurodeposits interest rates. Daily observations over the period 01/11/1986-13/01/2003.

Also in this case, graphical analysis gives us quite clear results. The histogram of these series are represented in Fig. 3.8 and the existence of various modes (multimodality) in all the interest rate distributions is clear. Empirical distributions show multimodality and are moreover platykurtic, as result from the QQ-Plots given in Fig. 3.9.

Fixed income investment, i.e. Eurodeposits, are less sensitive to shock in financial markets. Thus negative events are less frequent and the decay rate of the tails is sub-exponential. The joint effect of these features on the Jarque Bera normality test are given in Tab. 3.8. For all the series there are negative skewness, negative

Table 3.8: Jarque Bera test on 3-month Eurodeposits interest rates (daily frequency). E-Views software has been used for the test.

J.-B. Test	Italy	Germany	Japan	USA	France	UK
Observations	4444	4444	4444	4444	3922	4444
Mean	8.6910	5.2204	2.7654	5.6791	6.2529	8.1976
Std. Dev.	3.6405	2.1643	2.5873	1.9396	2.7820	3.2053
Skewness	-0.0818	0.8487	0.5606	-0.1242	0.3690	0.7314
Ex. of Kurt.	-0.7807	-0.6889	-0.9508	-0.264	-1.3272	-0.6236
Min Value	2.5	2.5	-0.0312	1.296	2.5	3.765
Max Value	21	9.875	8.4375	10.5	14.25	15.5625
Test Statistic	117.82	621.46	400.22	24.40	376.92	468.33
Crit. Val. 1%	9.21	9.21	9.21	9.21	9.21	9.21
Crit. Val. 5%	5.991	5.991	5.991	5.991	5.991	5.991
p-Value	0	0	0	0	0	0
Norm. Hypoth.	Rejected	Rejected	Rejected	Rejected	Rejected	Rejected

excess of kurtosis and the hypothesis of normal distributed interest rates has been rejected.

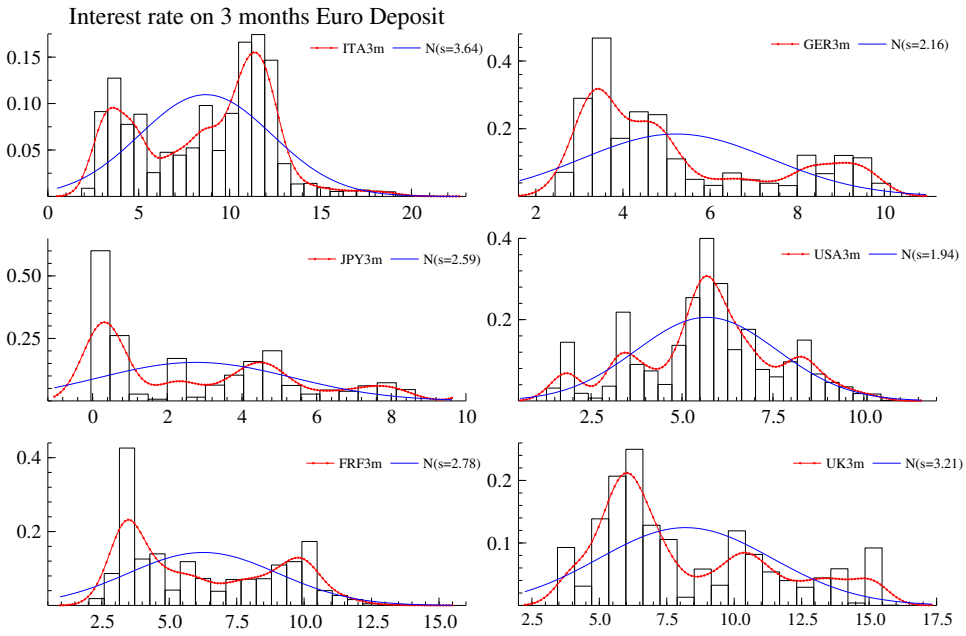

Figure 3.8: Histogram, Gaussian kernel density estimation and the best normal for daily 3-month Eurodeposits interest rates.

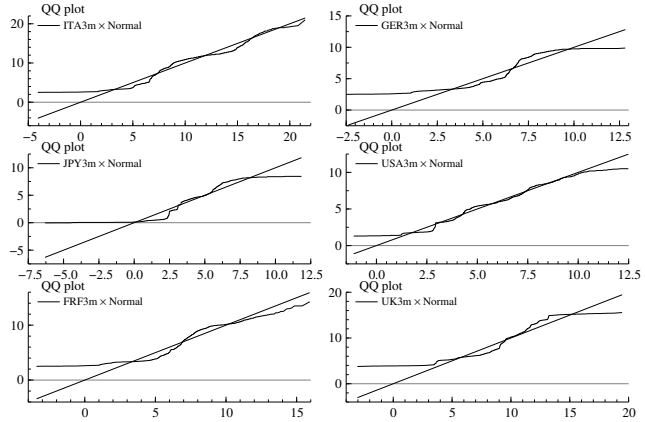

Figure 3.9: Q-Q Plots for daily 3-month Eurodeposits interest rates.

Bibliography

[1] D'Agostino, R. D. and Stephens, M. A. (1986), *Goodness-of-fit tecniques*, Marcel Dekker, Inc.

Chapter 4

Extreme Returns in a Shortfall Risk Framework

4.1 Introduction

In asset allocation problems, the assumption on the probability distribution of future returns is an important aspect. In many theoretical and empirical works, a normal or log-normal distribution is usually assumed. It is well known that the normal distribution has several attractive properties: it is easy to use and produces tractable results in many analytical exercises; all moments of positive order exist, and it is completely characterized by its first two moments, thus establishing the link with the mean-variance optimization theory. Normal distribution arises as the limiting distribution of a whole class of statistical testing and estimation procedures, and therefore plays a central role in empirical modelling exercises. One of the main characteristics of the normal distribution is that its tails decay exponentially toward zero; thus extreme realizations are very unlikely. However, this seems to contradict empirical findings on asset returns, which evidence that

[1]Part of this work is in:

Billio M. and Casarin R., (2003), "Extreme Returns in a Shortfall Risk Framework", in *Atti della giornata di studio Metodi Numerici per la Finanza*, 30 May 2003, Applied Mathematics Department, University "Ca' Foscari", Venice.

Billio M., Casarin R. and Toniolo G., (2002), "Extreme Returns in a Shortfall Risk Framework", *Working Paper*, GRETA n. 0204 and in *Proceedings of 8th International Conference Forecasting Financial Markets Meeting*, London 2002.

returns' distribution generally exhibits leptokurtic behaviour, i.e. has fatter tails than the normal distribution. This means that extreme returns of either sign occur far more often in practice than predicted by the normal model. For financial managers, who are interested in risk management, these are crucial aspects. It is typically suggested that the use of leptokurtic instead of normal distributions in asset allocation problems leads to more prudent asset portfolios. In other words, it is commonly believed that optimal asset allocations under the assumption of normally distributed returns have a higher Value at Risk (VaR) than the model suggests if returns in reality follow a leptokurtic distribution.

By means of Monte Carlo simulations we study an asset allocation problem under a given shortfall constraint, and we show that this idea is not generally valid. The use of simulation methods allows testing the effects of assuming different kinds of returns distribution in modelling asset class returns. To illustrate this point, we consider a simple one-period asset allocation problem with one *shortfall constraint*[1] (see [25], Telser [28] and Kataoka [14]).

As is known from the *safety first principle*, the shortfall constraint reflects the investors typical desire to limit downside risk by putting a (probabilistic) upper bound on the maximum loss. In other words, the investor wants to determine an optimal asset allocation for a given Value at Risk. Results obtained reveal that the degree of shortfall probability plays a crucial role in determining the effects of the choice between a fat tailed and a normal distribution. These effects concern the composition of optimal asset allocations as well as consequences of misspecification of the degree of fat-tailedness for the downside risk measure[2].

If the shortfall is moderately large, say 5%, then the assumption of fat tails

[1] An extensive literature exists since the fifties, known as *Downside Risk Approach*. It tries to explain the risk associated to an investment, exclusively evaluating downside oscillations of returns. The *Downside Risk* is an alternative to the more common *standard deviation* concept, generally used by financial managers in asset allocation problems. Although these approaches were already developed at the beginning of the fifties, they were followed in the mid seventies with the introduction of the *lower partial moment framework* (see Harlow [7]). A special case of lower partial moments is the *safety first principle* (see Roy [25]). This principle allows investors to identify portfolios revealing a minimum probability to fall under a specified return level.

[2] A first analysis of these important aspects was provided by Lucas and Klassen [18] which studied an analogous problem in an analytical way. The main aim of this work is to examine closely these aspects valuing the effects on optimal financial portfolio when each asset studied follows a different marginal probability distribution. To make this possible the Monte Carlo method seems to be the most suitable instrument.

results in more aggressive asset allocations. As a consequence, if reality is fat-tailed, optimal asset allocations that are based on the normal distribution may be far too prudent for a given 95% confidence level VaR.

If the shortfall is small, say, 1%, then the use of leptokurtic distribution leads to more prudent asset allocations. Consequently, an optimal asset mix that is based on a normality assumption, will violate a 99% confidence level VaR if reality is leptokurtic. We show that the true VaR may in that case exceed the VaR obtained by simulation by over 30%. After these first results, we analyse the effects, on the portfolio management, when the distribution of each asset returns considered, shows a different behaviour. In fact, it is quite unrealistic to suppose that the probability distribution of each financial class shows the same degrees of leptokurtosis. Through the Monte Carlo method it is possible to study the effects on the true risk when the data come from a mixed distribution while the manager uses a multivariate distribution with identically distributed marginal distributions.

The main result is that a correct estimate of the degrees of freedom for each of them is a necessary condition in order to have no excessive loss of information, an adequate formulation of the optimal strategy and, consequently, a correct perception of the true risk. We have also noted that for the particular combination we used, it is possible to find a multivariate distribution with identically distributed marginal distributions able to approximate the empirical one, with a loss of information that could be minimal. This means that we can simplify the problem with no excessive loss of generalities, which may be very useful when we solve complex mathematical models. The analysis we carry out may be very important to understand which approach a manager could follow to identify the most suitable distribution of probability to use. We concentrate on the interaction between different distribution assumptions made by the manager on the one hand and, on the other, the resulting optimal financial management decisions and downside risk measures. In particular, we pay close attention to the effect of financial policies and VaR if the probability of extreme returns is underestimated.

The structure of the work is as follows. In §4.2 we present the asset allocation problem and the class of probability distributions considered. In Section 4.3 we examine how the problem can be solved through a Monte Carlo approach, and in particular we concentrate on the stochastic optimization aspects. After this

analysis, we give some general characterizations of the theoretical effect of fat tails on the problem at hand (see Section 4.4) and a numerical illustration of the model, initially using the returns of three financial assets class (cash, stocks and bonds) identically distributed (see Section 4.5) and then differently distributed (see Section 4.7). Parameters of the probability distribution (mean, variance, correlation matrix), are estimated on three U.S. asset categories. Following the empirical example, in Sections 4.6 and 4.7, we study the effect of misspecification of the return distribution on downside risk.

4.2 The Portfolio Model

We consider a one-period model with n asset categories. At the beginning of the period, the manager can invest the money available in any of the n asset categories and short positions are not allowed.

The objective of the investment manager is to maximize the expected return on the portfolio, subject to a shortfall constraint. This shortfall constraint states that with a sufficiently high probability $1-\alpha$ (with α being a small number), the return on the portfolio will not fall below the *threshold return* r^{low}.

Formally, the asset allocation problem can be written as follows:

$$\underset{x\in\mathbb{R}^n}{Max} \quad \mathbb{E}\left(\sum_{i=1}^{n} x_i r_i\right) \tag{4.1}$$

s.t.

$$\mathbb{P}\left(\sum_{i=1}^{n} x_i r_i < r^{low}\right) \leq \alpha \tag{4.2}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{4.3}$$

$$x_i \geq 0 \tag{4.4}$$

where x_i and r_i , (with $i = 1, 2, \ldots, n$), denote the fraction of capital invested in the asset category i, and the (stochastic) return on asset category i, respectively. The operator $\mathbb{E}(\cdot)$ is the expectations operator with respect to the probability distribution $\mathbb{P}$ of the asset returns. The probabilistic constraint in (4.2) fixes the permitted VaR for feasible asset allocation strategies. We know that Value at Risk is the maximum

amount that can be lost with a certain confidence level in a given period. In the setting of (4.2) with $r^{low} < 0$, the VaR per Euro invested is $-r_{low}$ with a confidence level of $1 - \alpha$.

Our aim is to study the effect of extreme returns on the solution of the asset allocation problem in Equations (4.1) to (4.4). We need therefore to introduce a stochastic optimisation technique by simulation and then a class of probability distribution that allows for fat tails. The class of Student-t distributions meets these requirements.

The probability density function of n-dimensional multivariate Student-t distribution is given by:

$$\mathcal{T}_n(r; \mu, \Omega, \nu) = \frac{\Gamma((\nu + n)/2)|\Omega|^{-1/2}}{\Gamma(\nu/2)(\pi\nu)^{n/2}} \left(1 + \frac{(r - \mu)'\Omega^{-1}(r - \mu)}{\nu}\right)^{-(\nu+n)/2} \tag{4.5}$$

where $\Gamma(\cdot)$ denotes the gamma function, $r = (r_1, r_2, \ldots, r_n)'$ denotes the vector of stochastic asset returns, and μ, Ω^{-1} and ν denote the mean, the precision matrix and the degrees of freedom parameter of the Student-t distribution, respectively. For sake of simplicity, in the following we will denote the Student-t distribution with the alternative notation: $t(\nu)$, omitting the specification of scale and location parameters.

It's important to note that Ω satisfies the following relation:

$$\Omega = \left(1 - \frac{2}{\nu}\right) V \tag{4.6}$$

where V denotes the variance-covariance matrix.
The degrees of freedom parameter ν determines the degrees of leptokurtosis. ν has to be strictly positive. The smaller ν, the fatter the tails of the Student-t distribution. The Student-t distribution has the normal distribution as special case: (4.5) reduces to the normal density with mean μ and covariance matrix Ω if $\nu \to \infty$.

The first two moments of the Student-t distribution play an important role in the subsequent analysis. These moments are given by $\mathbb{E}(r) = \mu$ and $\mathbb{E}((r-\mu)(r-\mu)') = \nu\Omega/(\nu - 2)$, and they require $\nu > 1$ and $\nu > 2$, respectively.

Figure (4.1) displays several univariate Student-t distributions. The distributions are scaled in such a way that they all have zero mean and unit variance. It is clearly

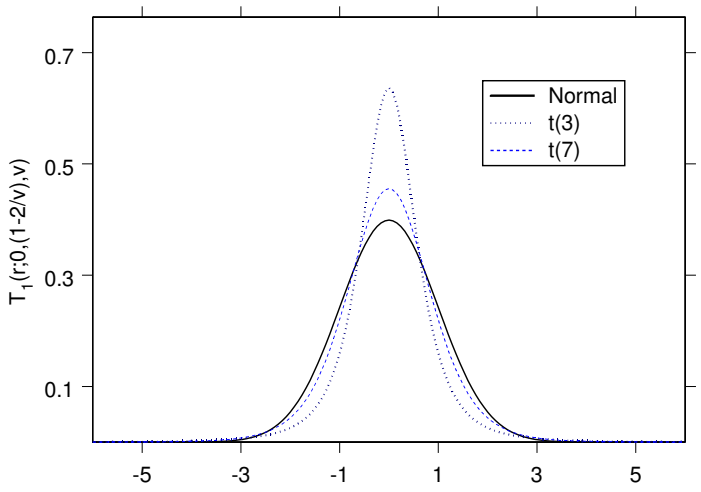

Figure 4.1: Student-t distribution for various values of degrees of freedom parameter.

seen that for lower values of ν, the tails of the distribution become fatter and the distribution becomes more peaked near the centre $\mu = 0$.

We have to note that μ and V are usually unknown and they are therefore estimated (m, V) from historical time series. So we have:

$$\mu \cong m \tag{4.7}$$

$$\Omega \cong \left(1 - \frac{2}{\nu}\right)\tilde{V} \tag{4.8}$$

In our analysis m and $\tilde{V}$ are considered fixed. In the following sections we use the Monte Carlo approach to analyse an asset allocation problem with shortfall constraint, from the manager's point of view.

Our aim is to consider the effects on the portfolio, when the probability distribution of asset returns shows a leptokurtic behaviour, and the concrete risk borne by the financial manager.

4.3 Monte Carlo Simulation Approach to Stochastic Optimisation

We use a Monte Carlo simulation approach in order to solve the optimization problem, and to compute numerically nontrivial integrals. There are many features that distinguish this method from most of the others generally used. First it can handle problems of far greater complexity and size than most other methods. The robustness and simplicity of the Monte Carlo approach are its strengths. Second, the Monte Carlo method is intuitively based on the law of large numbers and central limit theorem. The probabilistic nature of the Monte Carlo method has important implications. The result of any Monte Carlo procedure is a random variable. Any numerical method has errors, but the probabilistic nature of the Monte Carlo errors puts structure on the errors that we can exploit. In particular, the accuracy of the Monte Carlo method can be controlled by adjusting the sample size. The Monte Carlo method uses pseudo random numbers to solve a given problem; that is, deterministic sequences generated using pseudo random generators, such as *linear congruential generators* (Ripley [22]), that seem to be random [3]. We also know that these generators give an identical sequence of pseudorandom numbers if the same *seed* is set. The random numbers generated are good ones if they are uniformly distributed, statistically independent, and reproducible (Rubinstein [26]). In order to solve our financial problem, we need to simulate assets returns from different probability distributions. Initially we generate random numbers from a normal distribution with mean μ and variance-covariance matrix V and then from a multivariate Student-t distribution with ν degrees of freedom [4].

Let Z have a standard multivariate normal distribution, let Y have a multivariate chi-square distributions with ν degrees of freedom, and let Z and Y be independent,

[3] In this work we have used the *Mixed Congruential Generator* in the following form: $X_{i+1} = a\,X_i + c\ (mod\, m)$ for $i = 1, 2, \ldots, N$ and with $c = 0$, $m = 2^{31} - 1$ and $a = 397204094$

[4] There exist many algorithms that allow the transformation of random numbers extracted from a uniform distribution into normally distributed random numbers. Many of these techniques are very well explained in Rubinstein [26] and Ripley [22] (see, for example, *Box Müller algorithm*, *Monro algorithm*, etc.).

then:

$$X = \frac{Z}{\sqrt{Y/\nu}} \tag{4.9}$$

has a multivariate Student-t distribution with ν degrees of freedom.

It is known that the Monte Carlo method allows numerical solutions of complex mathematical problems to be obtained, where mathematical procedures seem inadequate. Stochastic optimization problems are an example. We consider a *stochastic optimisation problem* in the following form:

$$x^* = \underset{x \in U}{arg\,max}\, \mathbb{E}\{g(x, Z)\} \tag{4.10}$$

where Z is a random variable with p.d.f. $f_Z(z)$ and U is the set of admissible solutions. The solution of the problem (4.10) needs the computation of $\mathbb{E}(\cdot)$. A numerical solution can be performed by simulation of the objective function and then by applying standard optimisation techniques. The most simple idea (see Judd [12] for some examples of application in economics, Robert [23], Robert and Casella [24] for an introduction to other stochastic optimisation methods) is to take a sample of size D of the random variable Z, and to approximate $\mathbb{E}\{g(x, Z)\}$ by its sample mean:

$$\frac{1}{D} \sum_{i=1}^{D} g(x, Z_i) \tag{4.11}$$

where $Z_i \sim f_Z(z)$. Then all standard numerical optimisation techniques can be applied.

We use this approach to solve our portfolio problem. In fact the use of Monte Carlo integration is quite natural for such problems since we are essentially simulating the problem. The solution is denoted by $\tilde{x}^*$ and approximates the true solution x^*, it gives us the fraction of capital has to be invested in each asset class to maximise portfolio's expected return.

The quality of this procedure depends on the size D and how well the integral is approximated by the random sample mean. We are therefore interested in knowing the sample size D and the number of samples N of D draws, necessary to obtain a

"good" estimate. To do that, we control the error from the analytical solution [5] and the standard deviation of each estimate. In our optimisation problem we have found that for $D = 10,000$ and $N = 100$, we can approximate the underlying distribution adequately and obtain an accurate estimate with a very small standard error.
Although there exist analytical techniques to choose the optimal number of simulation N (see Rubinstein [26]), graphical techniques are often preferred (see Robert [23], Robert and Casella [24]). They permit the choice of the adequate simulations number, necessary to obtain the stabilisation of the solution. The higher N, the better the solution approximation, because the variance of sample mean reduces to zero. In Appendix A, we can find the convergence in the estimated solution for portfolio fractions invested in cash, bonds and stocks, when data come from the normal distribution. For cash and bonds, the stabilisation is quite evident. For bonds, the interval in which the fraction moves is very small, therefore the volatility is imperceptible.

4.4 Theoretical Effects

We have seen that the parameter ν plays a prominent role in our asset allocation problem, through its presence in the shortfall constraint. Decreasing ν has two effects. First of all, the tails of the distribution become fatter, resulting in a higher probability of extreme events for fixed precision matrix Ω^{-1}. As can be seen in (4.6), the precision matrix is not independent of ν if the variance of the returns is held fixed. As ν decreases, the eigenvalues of the precision matrix increase. As a result, the distribution becomes more concentrated around the mean μ. The composite effect on the shortfall constraint of altering ν depends critically on the shortfall probability α.

It can be shown that for a sufficiently small value of α, the shortfall constraint becomes less binding if the distribution used tends to normal. The reverse holds if we consider sufficiently large values of α.

It is interesting to present the break-even shortfall probability for the normal distribution obtained by simulation, i.e., the value of α, as a function of ν such that

[5] The *analytical solution* is obtained by calculating integrals "analytically" instead of "numerically". For multivariate distributions (such as normal or Student-t), it can be obtained by using a common calculator.

the shortfall constraint for that value of ν is as binding as the shortfall constraint for the corresponding normal distribution (results are similar to those obtained by Lucas and Klaassen [18]). More precisely, given the random variables $u \sim \mathcal{N}(\mu, \sigma)$ and $w \sim \mathcal{T}(\mu, \Omega, \nu)$ and a shortfall probability α, then the quantile associated to α, for the two distributions is

$$\Phi_{\mathcal{N}}(r_{\mathcal{N}}^{low}) = \int_{-\infty}^{r_{\mathcal{N}}^{low}} u\mathcal{N}(u; \mu, \sigma)du = \alpha \quad \Leftrightarrow \quad r_{\mathcal{N}}^{low} = \Phi_{\mathcal{N}}^{-1}(\alpha) \tag{4.12}$$

$$\Phi_{\mathcal{T}}(r_{\mathcal{T}}^{low}) = \int_{-\infty}^{r_{\mathcal{T}}^{low}} w\mathcal{T}(w; \mu, \Omega, \nu)dw = \alpha \quad \Leftrightarrow \quad r_{\mathcal{T}}^{low} = \Phi_{\mathcal{T}}^{-1}(\alpha) \tag{4.13}$$

In order to obtain the critical shortfall probability $\alpha(\nu)$, for a fixed value of ν we take $r_{\mathcal{N}}^{low} = r_{\mathcal{T}}^{low}$ and solve w.r.t. α the resulting non linear equation

$$\Phi_{\mathcal{T}}^{-1}(\alpha) = \Phi_{\mathcal{N}}^{-1}(\alpha) \tag{4.14}$$

Such a value of α, produces the same solution under either normality assumption or the assumption of a Student-t distribution with ν degrees of freedom for the asset returns. The values of $\alpha(\nu)$, are given in Figure 4.2. This graph indicates the critical shortfall probability ranges from $\alpha = 1.8\%$ for $\nu = 3$ to $\alpha = 3.6\%$ for $\nu = 10$. For values of α below these critical levels, the effect of fat tails on the shortfall constraint dominates the effect caused by increased precision. In these cases, the probability restriction in (4.2) for the Student-t distribution is more binding for a given asset allocation than in the case of normally distributed asset returns.

Again, the reverse holds for values of α above the critical level. In our empirical study, we use $\alpha = 0.5\%$, $\alpha = 1\%$, $\alpha = 5\%$ and $\alpha = 10\%$ in order to illustrate both settings.

4.5 Results

To illustrate our results, we present and solve through a Monte Carlo approach an asset allocation problem (similar to which studied by Lucas and Klaassen [18]), involving three U.S. asset categories: cash, stocks, and bonds. For cash, we use the return on one month Eurodollar deposits. Stock returns are based on the S&P 500 and include dividends. Bond returns are computed using holding period returns on

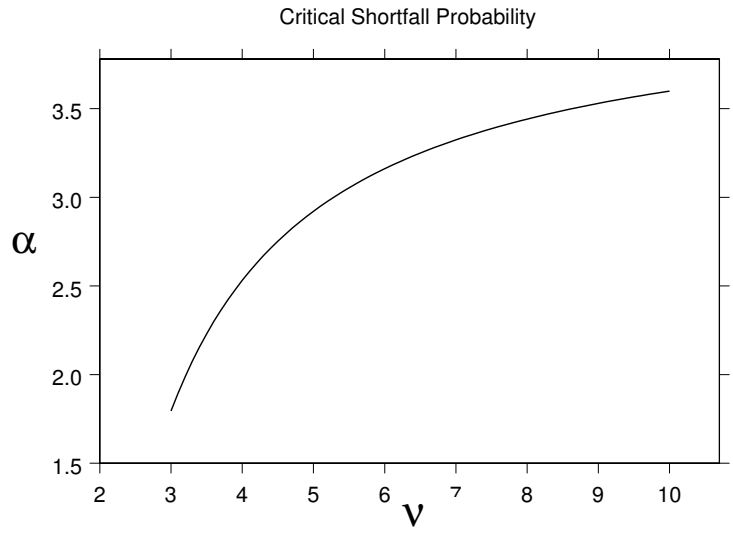

Figure 4.2: Critical Shortfall Probability $\alpha(\nu)$ for Student-t distribution with ν degrees of freedom (benchmark is the normal distribution).

10-year Treasury bonds. We consider annual returns over the period 1983-1994. All data are obtained from Datastream. Initially, we need to compute the mean and variance of the returns series. Let x_1, x_2 and x_3 denote the amount invested in cash, stocks, and bonds, respectively, and let the corresponding returns be denoted by r_1, r_2 and r_3. Then $r = (r_1, r_2, r_3)'$ has mean and standard deviation:

$$\begin{array}{rcccc} & & Cash & Stocks & Bond \\ \mu' & = & (6.8\% & 17\% & 12.3\%) \\ & & Cash & Stocks & Bond \\ \sigma' & = & (2.3\% & 14.7\% & 10.5\%) \end{array}$$

and the correlation matrix:

$$\rho = \begin{pmatrix} 1 & 0.01 & 0.18 \\ 0.01 & 1 & 0.73 \\ 0.18 & 0.73 & 1 \end{pmatrix} \tag{4.15}$$

In Table 4.1 we consider results obtained for two values of shortfall probability

($\alpha = 0\%,\ \alpha = 5\%$) and for several values of the shortfall return r^{low} at the levels of 0%, −5%, −10% [6]

Thus, for example, the combination $(\alpha, r^{low}) = (1\%, 0\%)$, means that the manager requires an asset mix that results in no loss with a 99% probability. Similarly, the combination $(\alpha, r^{low}) = (5\%, -5\%)$, means that the manager is satisfied with a 5% *Value at Risk* per Euro invested with a confidence level of 95% probability.

Table 4.1: Optimal asset allocation obtained through Monte Carlo simulation (E* indicates the expected portfolio returns).

ν	$r^{low} = 0\%$							
	Cash	Stock	Bond	E^*	Cash	Stock	Bond	E^*
	Shortfall probability 5%				Shortfall probability 1%			
3	28.10%	66.70%	5.10%	13.90%	80.80%	19.20%	0.00%	8.70%
5	46.30%	50.10%	3.50%	12.10%	80.00%	20.00%	0.00%	8.80%
7	49.00%	48.50%	2.50%	11.90%	78.50%	21.50%	0.00%	9.00%
10	50.30%	47.20%	2.50%	11.70%	77.70%	22.30%	0.00%	9.10%
∞	51.60%	45.90%	2.50%	11.60%	75.60%	24.40%	0.00%	9.30%
	$r^{low} = -5\%$							
	Cash	Stock	Bond	E^*	Cash	Stock	Bond	E^*
	Shortfall probability 5%				Shortfall probability 1%			
3	0.00%	100.00%	0.00%	17.00%	58.50%	40.00%	1.50%	11.00%
5	3.80%	88.30%	7.90%	16.30%	58.20%	40.20%	1.60%	11.00%
7	8.00%	84.50%	7.60%	15.90%	56.30%	42.00%	1.80%	11.10%
10	10.20%	82.80%	7.00%	15.60%	54.60%	43.50%	1.90%	11.30%
∞	12.30%	80.60%	7.00%	15.40%	50.70%	46.70%	2.60%	11.70%
	$r^{low} = -10\%$							
	Cash	Stock	Bond	E^*	Cash	Stock	Bond	E^*
	Shortfall probability 5%				Shortfall probability 1%			
3	0.00%	100.00%	0.00%	17.00%	38.80%	57.20%	4.00%	12.90%
5	0.00%	100.00%	0.00%	17.00%	38.70%	57.30%	4.00%	12.90%
7	0.00%	100.00%	0.00%	17.00%	36.20%	59.80%	4.00%	13.10%
10	0.00%	100.00%	0.00%	17.00%	33.40%	62.40%	4.30%	13.40%
∞	0.00%	100.00%	0.00%	17.00%	27.90%	67.00%	5.10%	13.90%

Using the optimisation library available in *GAUSS 3.1.4*, we compute in simulation the optimal values of x_i satisfying the shortfall constraint in (4.2) for several values of ν. The main results are presented in Table 4.1.

Some obvious effects in Table 4.1 are that the optimal asset mixes become more aggressive if the shortfall constraint is loosened. This can be done by increasing the allowed shortfall probability α or by lowering the required shortfall return r^{low} , i.e. increasing the *Value at Risk* per Euro invested. If we focus on the effects of ν, we note the difference between the $\alpha = 5\%$ and the $\alpha = 1\%$ case.

[6] The studies we made use in particular four values of shortfall probability (0.5%, 1%, 5%, 10%), and five values of shortfall return (0%, −3%, −5%, −7%, −10%). Table 4.1 only indicates the main results obtained in our work.

In the 5% case, increasing the fatness of the tails of the asset returns' distribution $\mathbb{P}$ then leads to more aggressive asset allocation. The optimal asset mixes involve less of the relatively safe cash and more of the risky assets, stocks and bonds. The effect is more pronounced if the required shortfall return r^{low} is lower.

Although the results that fat tails lead to more aggressive asset mixes may seem counterintuitive at first sight, it is easily understood, given the results we have seen analysing Figure 4.2. In fact, decreasing ν while keeping the variance fixed has two opposite effects. First, the probability of extreme (negative) returns increases, leading to more prudent asset allocation strategies. Second, the precision of the distribution increases, leading to more certainty about the spread of the outcome and, thus, to a more aggressive strategy.

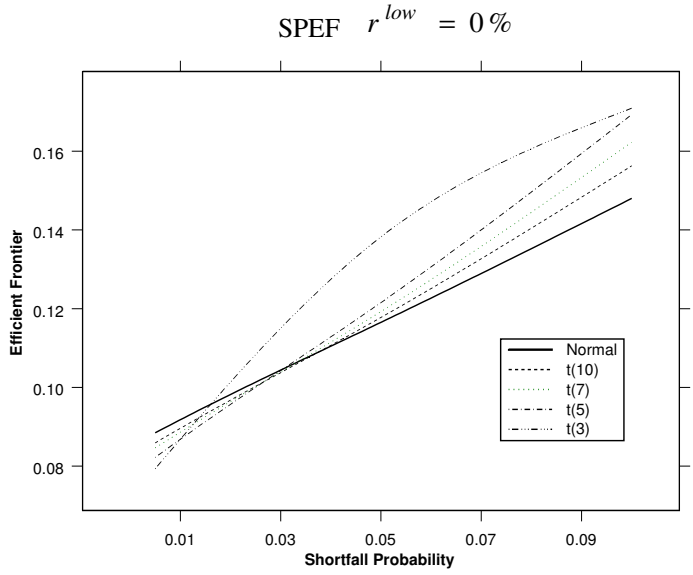

Figure 4.3: Shortfall Probability Efficient Frontiers obtained in simulation.

For a shortfall probability of 5%, the latter of these two effects dominates. In the case of 1% shortfall probability, the opposite occurs. Decreasing ν now leads to more prudent asset mixes. Again the effect is more pronounced if the required shortfall return r^{low} is lower.

Figure 4.3 shows the effects on expected portfolio returns, for different distributions, when the shortfall probability vary[7]. Curves obtained by simulations,

[7] r^{low} has been set equal to 0%.

are labelled *Shortfall Probability Efficient Frontiers (SPEF)*[8] (see Rudolf [27]), and represent all the efficient portfolios, given a certain level of risk (expressed by the shortfall probability). On the left side of the graph, we obtain lower expected returns decreasing the degrees of freedom. The reverse holds on the right side. In particular, the SPEF obtained from the normal distribution intercepts all the others in different points for a level of α that is the same decrypted in Figure 4.2.

Through the Monte Carlo study, it is also possible to determine the effects of a variation of the shortfall probability on the fractions invested in each asset. If we observe Figure 4.9 (Appendix B), we can note that the higher the level of α, the lower the fraction invested in cash, since to obtain more aggressive portfolios the manager directs capitals in riskier assets. However, while for a sufficiently small value of α (i.e. 0.5%, 1%), a normal distribution shows lower value invested in cash than the other distributions, for higher levels of shortfall probability (i.e. 5%, 10%), the lowest percentages in liquid assets are obtained augmenting the degrees of leptokurtosis. The reverse holds for stocks. In fact for this asset category, the behaviour is exactly reverse to which showed by cash, since there exists a trade-off between liquid and risky asset in order to obtain efficient portfolios. Fractions invested in bonds shows instead a more complex behaviour. In general, we can say that initially the percentage is increasing for smaller value of shortfall probability, and it is decreasing for higher value of shortfall probability.

The intersection between the curve of the normal distribution and the others in this case too, occurs for the values of showed in Figure 4.2. However, in this particular case each curve intercepts the others in more than one point, so they are more difficult to analyse.

It is also interesting to observe the expected returns behaviour when the shortfall return varies (given the shortfall probability level). We have choose to graph the expected returns behaviour for $\alpha = 1\%$ and for $\alpha = 5\%$. In Appendix C we can note that the higher the shortfall return value, the lower the expected return, but while for $\alpha = 1\%$ the *fat tail* effect dominates, for an $\alpha = 5\%$ the highest values are obtained when asset returns are leptokurtic. For very high losses levels (ex. 10%), it makes no difference if we use the normal distribution or a Student-t with

[8]They are an alternative representations of the Efficient Frontier, known from the Portfolio Theory.

different degrees of freedom, since the portfolio always contains only risky assets (100% stocks).

We can make the same analysis, observing the effects on the fractions invested in each asset, varying r^{low} for different levels of shortfall probability. The results are indicated in Appendix D.

As we can note, the lower the losses, the higher the fraction invested in cash, since the manager is more "*conservative*"; for the same reason, the fraction invested in stocks is decreasing. For bonds, the behaviour is not regular, and the fraction shows it is be decreasing for $\alpha = 1\%$, while it appears initially increasing and then decreasing for $\alpha = 5\%$. As we know, for $\alpha = 1\%$ the fat tail effect dominates and therefore the fraction invested in risky assets is much higher if data come from the normal distribution; the reverse holds for $\alpha = 5\%$.

4.6 Effects of Misspecified Tail Behaviour

The probability distribution $\mathbb{P}$ is taken by the investment manager as a description of the true distribution of the asset returns. We label the distribution used by the manager $\mathbb{P}_m$ and the true distribution $\mathbb{P}_t$. These distributions are characterized by the parameter specifications (μ_m, Ω_m, ν_m) and (μ_t, Ω_t, ν_t) respectively.

Obviously, the manager would do best by matching μ_m, Ω_m and ν_m to μ_t, Ω_t and ν_t, respectively. However, the manager can fail to match all the parameters of the distribution used to solve (4.1) and (4.2) to those of the true distribution $\mathbb{P}_t$.

The effects of misspecification of means μ_m and/or covariance Ω_m has been investigated in the literature (see, e.g. Chopra and Ziemba [1]). We concentrate here on the possible mismatch between the true degree of leptokurtosis and the degree of leptokurtosis used by the investment manager, while assuming that means and covariance of the returns are precisely estimated.

The most obvious example of such a situation is the use of the normal distribution for solving (4.1), while the asset returns are actually fat-tailed. The mismatch between ν_m and ν_t can have important effects for the feasibility and efficiency of the optimal asset mixes. We assume that for a given value of ν_m , the manager chooses μ_m and Ω_m, in such a way that the mean and variance of $\mathbb{P}_m$ match the corresponding moments of the true distribution $\mathbb{P}_t$. This amounts to setting

$$\mu_m = \mu_t \tag{4.16}$$

$$\Omega_m = (1 - 2\nu_m^{-1})\frac{\nu_t \Omega_t}{\nu_t - 2} \tag{4.17}$$

$$\tag{4.18}$$

We assume, that the true mean μ_t and variance $\nu_t\Omega_t/(\nu_t - 2)$ are observed without error. Of course (4.16) and (4.17) are only estimates of the underlying true parameters. We abstract from the associated estimation error for exposition purposes and in order to fully concentrate on the effects of fat tails (the results would be very similar if slightly different values for the means and variances were used, thus allowing for estimation error). Let x_m denote the optimal strategy of the investment manager using the distribution $\mathbb{P}_m$ with ν_m degrees of freedom. The appropriate values of x_m can be found in Table 4.1.

We want to compute the effect of using x_m when the data follow the distribution $\mathbb{P}_t$ instead of $\mathbb{P}_m$. In particular, we are interested in the effect of a discrepancy between ν_m and ν_t on the shortfall constraint.

We can quantify this effect in at least two different ways. First, we can use the strategy x_m while keeping the required shortfall return r^{low} constant and compute the actual shortfall probability α_* under the true probability measure $\mathbb{P}_t$. Alternatively, we can use the strategy x_m while keeping the required shortfall probability α constant and compute the corresponding shortfall return $r^{*,low}$, i.e., the (negative) Value at Risk per Euro invested.

First consider the case of fixed r^{low}. We then compute

$$\alpha^* = \mathbb{P}_t\left(\sum_{i=1}^{3} x_{m,i}(1 + r_i) \leq 1 + r^{low}\right) \tag{4.19}$$

where $x_{m,i}$, is the optimal asset allocation to category i for ν_m (see Table 4.1). So $(1 - \alpha^*)$ is the true confidence level of the investment manager's value at risk, given the asset allocation x_m.

In particular, we generate in simulation the asset returns from a distribution with ν_t degrees of freedom, and evaluate the shortfall constraint under the strategy x_m. This allows us to estimate the true risk for the financial manager.

We have seen that different values for r^{low} produce similar results, so we present only the case with $r^{low} = 0\%$. The results are given in Figure 4.4. The right panel in the figure gives the results if the optimal strategy is computed with $\alpha = 5\%$. The first thing to note is that, as expected, the true shortfall probability α^* is equal to α, if and only if the investment manager uses the correct distribution, i.e., $\nu_m = \nu_t$.

Second, if the investment manager uses a distribution that has thinner tails than those of the true distribution, then the manager is conservative in the sense that the shortfall constraint in (4.2) is not binding.

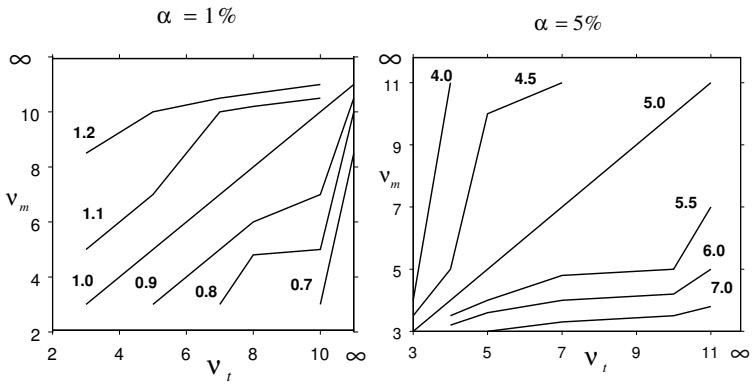

Figure 4.4: True Shortfall Probability for several combinations of ν_m and ν_t, obtained by simulation.

This holds even though the manager may believe the constraint to be binding based on the (misspecified) distribution $\mathbb{P}_m$ of the asset returns. As a result, efficiency could be gained by using the correct degree of leptokurtosis. By contrast, if the manager uses a distribution with a fatter tail than reality, the shortfall constraint is violated.

If we consider the case $\alpha = 1\%$, the results are reversed. If a thin-tailed distribution is assumed for the asset returns, e.g., the one based on normality, and if reality is leptokurtic, then the shortfall constraint is violated. Moreover, if $\nu < \nu_t$, the shortfall probability constraint is not binding. These results are directly relevant for risk management, because one minus the true shortfall probability equals the manager's required confidence level for the value at risk $-r^{low} > 0$. For example, for a 99% confidence level VaR, our results imply that the true confidence level of the manager's computed VaR is smaller than 99% if the manager uses the normal

Table 4.2: Differences $(r^{*,low} - r^{low})$ in basis points.

ν_m	$r^{low} = 0\%$			$r^{low} = -5\%$			$r^{low} = -10\%$		
	ν_t			ν_t			ν_t		
	3	7	∞	3	7	∞	3	7	∞
Shortfall probability 5%									
3	0	-249	-296	202	-153	-218	-	-	-
7	178	0	-33	313	0	-57	-	-	-
∞	201	32	0	352	54	0	-	-	-
Shortfall probability 1%									
3	0	43	88	0	53	164	0	71	243
7	-45	0	51	-56	0	117	-74	0	181
∞	-105	-60	0	-210	-135	0	-301	-205	0

distribution while reality is fat-tailed. Note that, although the absolute difference between α^* and α in Figure 4.4 is smaller for $\alpha = 1\%$ than for $\alpha = 5\%$, the relative differences are approximately equal for different combinations of (ν_m, ν_t).

To illustrate the effect on the shortfall return r^{low}, for fixed α, we compute the required shortfall return $r^{*,low}$ such that:

$$\alpha = \mathbb{P}_t \left(\sum_{i=1}^{3} x_{m,i}(1 + r_i) \leq 1 + r^{*,low} \right) \tag{4.20}$$

The differences $(r^{*,low} - r^{low})$, in basis points, obtained in simulation, are presented in Table 4.2. Remember that r^{low} is the Value at Risk per Euro invested if $r^{low} < 0$. Therefore, $r^{*,low}$ in (4.20), is the managers true VaR if the investment policy x_m based on $\mathbb{P}_m$ is used.

The qualitative results are similar to those in Figure 4.4. For high values of α, using a distribution $\mathbb{P}_m$, which has thin tails compared to reality $\mathbb{P}_t$, produces a conservative strategy. Again, the opposite holds for small values of the shortfall probability α.

The impact of using the normal distribution for $\mathbb{P}_m$ if reality is fat-tailed is quite substantial. Assume the postulated required minimum return r^{low} is -5%, i.e., a Value at Risk of 5 cents per Euro invested. That is, with a maximum probability of α, the manager is willing to take losses exceeding 5% of the invested notional principal.

If $\alpha = 5\%$, the true shortfall return can be as much as 352 basis points above

the postulated level, implying a shortfall return of about -1.45% instead of -5%, with a probability of 95%. Exploiting the fat tail property in this case can lead to more aggressive asset allocations and, therefore, efficiency gains for a given level of shortfall.

Alternatively, consider the case $\alpha = 1\%$. Using a normal scenario generator ($\nu_m = \infty$), for a reality with $\nu_t = 3$ now leads to a violation of the shortfall constraint. While the manager believes the maximum loss with a 99% probability is 5% of the invested notional, the actual loss, given that probability, may be about 210 basis points higher, or 7.12%, an increase of over 40%. In this case a correct assessment of the degree of leptokurtosis will lead to a more correct assessment of risk and to the exclusion of infeasible strategies. All these effects are even more pronounced if the manager is, *a priori*, willing to take higher losses, i.e., if r^{low} is lower. These results have obvious important consequences for Value at Risk analyses. Using a distribution with an incorrect tail behaviour may lead to portfolios with a minimum required return $r^{*,low}$ that can be as much as 301 basis points below the minimum return r^{low} imposed by the model. In value at risk calculations, this implies that the true VaR may deviate by more than 30% from what an incorrectly specified Extreme Returns in a Shortfall Risk Framework model suggests. This illustrates the importance of trying to get the tail behaviour of the distribution used to solve (4.1) and (4.2) right.

4.7 Asset Returns and Tails Behaviour

The analysis carried out, allowed us to value the true risk for the financial manager, when the degrees of freedom of the empirical distribution are not correctly estimated. Nevertheless, the results we have showed up to now are based on the idea that returns of the three asset classes considered are identically distributed. As we know, this hypothesis is quite unrealistic, given the heterogeneity of the financial indexes studied. We therefore decide to develop our analysis, using a different probability distribution for each index, remaining in the Student-t class. The use of stochastic optimisation techniques allow us to extend the study of the asset allocation problems for particular multivariate distribution for which in general it is not possible to obtain the result in a closed form. This approach makes it necessary to determine which

Student-t combination seems able to fit the empirical distribution adequately. To do this, we have analysed, through several tests [9], the behaviour of some distributions of financial indexes [10], for each asset category (cash, stocks, bonds), valuing for each of them the existence of *fat-tails*, and the parametric distribution that allows us to obtain the best fit for the empirical distribution on the tails. The test results show that the normal distribution is often unable to capture the behaviour of the tails, since financial time series are usually leptokurtic, while the Student-t distribution seems more adequate to capture the fat-tails effect.

In particular, tests indicate that for stocks, the Student-t with 7 degrees of freedom seems a good approximation of the empirical distribution in the extreme returns area. Otherwise, bonds seem to prefer Student-t with 8 degrees of freedom. Finally, cash behaviour does not appear unimodal, and for this reason no Student-t appears suitable to fit the empirical distribution adequately. Yet, some statistical tests seem to accept the null hypothesis that some of the cash indexes studied follow a Student-t distribution with 30 degrees of freedom (well approximated by a normal). For this reason we use this marginal distribution to simulate returns even if we know that it is not the most suitable. In fact our objective is to study the effects on the portfolio model when the empirical distribution is fat-tails, and in particular when each asset class presents a different degree of leptokurtosis. To do this, we generate asset returns simulating from the Student-t class, which present leptokurtic but not plurimodal or asymmetric characteristics.

Making use of tests results, we now calculate the optimal investment strategy and expected portfolio returns, generating asset returns from a multivariate distribution [11], where marginal distributions are $t(30)$, $t(7)$ and $t(8)$. We label this new

[9] We have used two categories of tests: graphical tests and statistical tests. The first category includes QQ-plots, and the study of the empirical density function and empirical distribution function with respect to theoretical ones. The second category includes Jaque Bera normality tests, Chi-square tests, Anderson Darling tests, Kolmogorov-Smirnov tests and Cramer von Mises tests (see DAgostino and Stephens [2]).

[10] We consider annual returns. For stocks, indexes analysed are: S&P 500 over the period (1/1/1970- 1/4/2001), MSCI Europe and MSCI North America over the period (1/1/1986-1/12/1999); for bonds: JP Morgan Great Britain, JP Morgan Japan, JP Morgan France, JP Morgan Belgium over the period (1/1/1986-1/1/2000); for cash: Euro-Mark 3 mth, Euro-Franc 3 mth, Euro-Lire 3 mth, Euro-Yen 3 mth, Euro-£ 3 mth, and Euro-$ 3 mth over the period (1/1/1985-1/4/2001). All data are obtained from Datastream.

[11] Note that it is not a multivariate Student-t distribution, because each t has different degrees of freedom. Moreover the covariance between marginal random variables is no more proportional to the covariance matrix used in equation 4.15 to simulate from that multivariate distribution.

distribution as "Mixed" with mean μ and variance-covariance matrix V. To impose the desired correlative structure to the simulated series, we use the *calibration method*.

In particular, given two variables x and w, where $x \sim t(\nu_x)$ and $w \sim t(\nu_w)$, we say that the correlation between x and w is ρ^* if there exists a value (τ^*) for the parameter τ such that $\rho^* = f(\tau^*, \nu_x, \nu_w)$, where f is the correlation between the components of the random vector given in (4.9). It is very difficult to obtain an analytical solution, and therefore we use the Monte Carlo method to calibrate τ step by step, until we obtain the desired value for $\rho(\rho^*)$[12]. In our case, we impose a correlation matrix τ to the multivariate normal in Equation 4.9 and simulate a sample from the multivariate "Mixed" distribution. Then we vary τ until the correlation matrix ρ estimated on simulated data is equal to the desired correlation matrix ρ^*. This calibration method has been recently used also in Palmitesta and Provasi [20] in order to fit the parameters of the Koehler-Symanowski distributions on real data. They minimize the distance between the correlation matrix simulate from the multivariate Koehler-Symanowski distribution and correlation matrix estimated on real data.

We now examine the effects on asset allocation strategies. In particular, we can note that results obtained combining Student-t with different degrees of freedom (see Appendix D) are intermediate between those obtained using a multivariate Student-t with 7 degrees of freedom and with 10 degrees of freedom. This aspect is very important in a risk management framework because the use of different marginal distributions gives more information than previous analysis. This means that financial manager can invest with higher precision in the estimates of optimal strategies and of expected returns. Furthermore, it is important to note that for different combinations (for example $t(15)$, $t(3)$ and $t(9)$), where the leptokurtosis degree is very different, the loss of information could be very high if we choose a distribution with identically distributed marginal distributions.

Following the same techniques used in previous sections, we now concentrate our analysis on the study of effective risk associated with the investment if the financial manager uses a distribution $\mathbb{P}_m \neq \mathbb{P}_t$.

We have seen that we can quantify these effects in at least two different ways:

[12] ρ^* used is indicated in (4.15).

first, we can use the strategy x_m while keeping the required shortfall return r^{low} constant and compute the actual shortfall probability α^* a under the true probability measure $\mathbb{P}_t$. Alternatively, we can use the strategy x_m while keeping the required shortfall probability a constant and compute the corresponding shortfall return $r^{*,low}$.

Figure 4.5 indicates the level of α^* if the manager uses the distribution $t(\nu_m)$ and the data follow the Mixed distribution. This analysis has been made, using $\alpha = 5\%$ and $\alpha = 1\%$ [13]

For $\alpha = 5\%$, if the investment manager uses a distribution that has thinner tails than reality, then the manager is conservative, in the sense that the shortfall constraint in (4.2) is not binding, while if the investment manager uses a distribution that has fatter tails than reality, the shortfall constraint is violated.

As we know, if we consider the case $\alpha = 1\%$, the results are reversed. If a thin-tailed distribution is assumed for the asset returns, e.g., the one based on normality, and if reality is leptokurtic, then the shortfall constraint is violated. Moreover, if $\nu_m < \nu_t$, the shortfall constraint is not binding.

It is interesting to note the relationship that exists between Mixed distribution, $t(7)$ and $t(10)$.

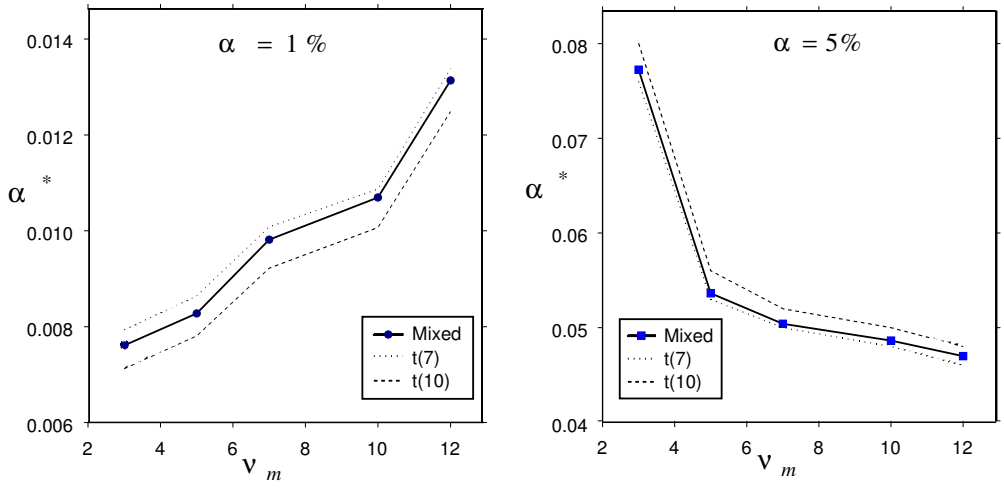

Figure 4.5: True Shortfall Probability for Mixed distribution, $t(7)$ and $t(10)$; $\alpha = 1\%$ and $\alpha = 5\%$.

If we use a shortfall probability $\alpha = 5\%$ in the model, then the values of true

[13] We have set $r^{low} = 0\%$

Table 4.3: Differences $(r^{*,low} - r^{low})$ in basis points. Data generated from Mixed distribution.

Shortfall probability=5%			
ν_m	$r^{low} = 0\%$	$r^{low} = -5\%$	$r^{low} = -10\%$
3	-253	-157	-
5	-37	-59	-
7	-4	-3	-
10	23	21	-
∞	27	45	-
Shortfall probability= 1%			
3	43	47	71
5	33	39	70
7	3	10	13
10	-16	-39	-67
∞	-62	-137	-198

shortfall probability $\alpha^*(\nu_m)$ are not only intermediate to those obtained using the other two distributions, but we can demonstrate that they are very close to what we could have by generating data from a $t(7)$, for every level of ν_m. If we use a shortfall probability $\alpha = 1\%$ in the model, when the data come from "Mixed" distribution, the behaviour of $\alpha^*(\nu_m)$ seems initially intermediate for $\nu_m = 3$, and very close to $t(7)$ curve, for higher degrees of freedom. For example, if the manager uses a distribution $t(3)$, when the data follow a Mixed distribution, we obtain $\alpha^* = 0.75\%$, while if the manager uses a distribution with thinner tails, the true shortfall probability for Mixed distribution tends to the true shortfall probability for $t(7)$ one. It is important to consider that the interval in which the curve oscillates is very small, and for this reason every variation seems imperceptible. The second analysis we can do is to study the differences $r^{low} - r^{*,low}$ in basis points, when data come from Mixed distribution. Results obtained are indicated in Table 4.3.

For $\alpha = 5\%$, if the investment manager uses a distribution with tails heavier than reality, the loss could exceed r^{low}. For $\alpha = 1\%$, the loss could exceed r^{low} only if the investment manager uses a distribution with tails thinner than reality. Furthermore, for $\alpha = 1\%$ and $r^{low} = -5\%$, the impact of using the normal distribution to explain the empirical distribution behaviour if reality is "Mixed", is quite substantial. In fact the manager is willing to take losses exceeding 5% of 137 basis points, i.e. an increase of over 30%. It is interesting to note that for $\nu_m = 7$, the difference

$r^{low} - r^{*,low}$ is very small and for $\alpha = 5\%$ it tends to zero, decreasing the r^{low} level.

For $\alpha = 1\%$, the difference becomes smaller with the reduction of losses. From this study we can conclude that for the case analysed, a multivariate Student-t distribution with 7 degrees of freedom seems a good approximation of the Mixed distribution. For the financial manager it could be simpler to use a multivariate $t(7)$ with identically distributed marginal distributions, without excessive loss of information on the true risk. However, if this result held for the specified combination used $t(30)$, $t(7)$, $t(8)$, in general we can not extend our conclusions for the enormous number of combinations of marginal distributions. This means that before any application, it is wise for the financial manager to analyse whether an approximation of the mixed distribution can cause losses of information and thus undesired effects for risk management.

4.8 Conclusion

In our work, we have investigated the effects of extreme returns on the optimal asset allocation problem with a shortfall constraint, using a Monte Carlo simulation approach. We have seen that financial markets usually show a "non-normal" behaviour, since the tails of returns distributions often appear very heavy. Extreme returns can be modelled by using a statistical distribution with fatter tails than those of a normal distribution. We have used the Student-t distribution to show the salient effects of fat tails in financial decision context. Initially, we have analysed the effects on asset allocation choices when all asset returns are identically distributed. This analysis is not very realistic, but has allowed us to discover the true risk for the financial manager when the behaviour of empirical distribution is incorrectly estimated. We have then tried to simulate data from a multivariate distribution where each marginal distribution has a different degree of leptokurtosis. For each of these two analyses, we may conclude that a correct assessment of the fattailedness of asset returns is important for the determination of optimal asset allocation. If asset allocations are based on the normal distribution, the resulting allocation may be either inefficient or unfeasible if reality is non-normal. Both effects can be quite substantial. Then, it appears that the shortfall probability set by the investment manager plays a crucial role for the nature of the effect

of leptokurtic asset returns. If the shortfall probability is set sufficiently high, using normal scenarios for the leptokurtic asset return leads to overly prudent and therefore inefficient asset allocations. If the shortfall probability is sufficiently small, however, the use of normal scenarios leads to unfeasible strategies if reality is fat-tailed. This second result implies that the Value at Risk of a given portfolio may be underestimated if the tail behaviour of asset returns is not captured adequately. Our results show that the actual VaR can be substantially higher than the model suggests in such cases. Extreme Returns in a Shortfall Risk Framework In studying the "Mixed" distribution we have also shown by simulation that another important aspect in financial analysis is the correct specification of the marginal distribution for each asset analysed. This aspect becomes crucial when there is a shortfall constraint in the asset allocation problem and at the same time the asset class return has a different tails behaviour. For our particular case, we have seen that the resulting optimal allocation obtained with the use of Mixed distribution could be adequately approximated by a multivariate Student-t at certain level of shortfall probability and shortfall return. Furthermore we obtain the following result of interest. When asset classes returns have a "Mixed distribution", the assumption of Student-t distribution with misspecified degrees of freedom produces effective shortfall probability and effective shortfall return which differ from the desired ones. However these errors have known upper and lower bounds. Even if this result can be very useful for the financial manager, since he can enormously simplify the mixed problem, in general, given any "mixed distribution", we do not know what its adequate approximation is and so we can not conclude that an adequate approximation always exists for all combinations of marginal distributions. For this reason, the use of stochastic simulation in the study we have performed in this article, is a very effective instrument to choose the optimal strategy to apply in asset allocation problems, and in particular when we use a shortfall constraint. In a more general sense, our findings imply that a good characterisation of the distribution of asset returns is needed in a financial decision context involving downside risk. Such a characterisation may require not only the specification of usual measures like mean and variance, but also a correct specification of additional features of the distribution of asset returns, such as the tail behaviour or degree of leptokurtosis. Specification and estimation of such additional features can proceed along familiar

lines, for example using parametric or non-parametric methods. Regardless of the method chosen, our results provide insights into the general effects of different type of leptokurtic distributions on optimal asset allocations and associated risk measures. The results are therefore valuable to investors who require a qualitative assessment of the reliability and sensitivity of their adopted investment strategies in case their models are potentially misspecified.

A.Convergence Analysis of the Solution Increasing the Simulation's Number

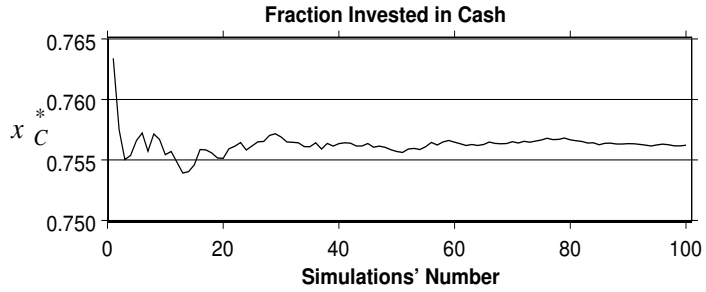

Figure 4.6: Fraction invested in Cash increasing the simulations' number

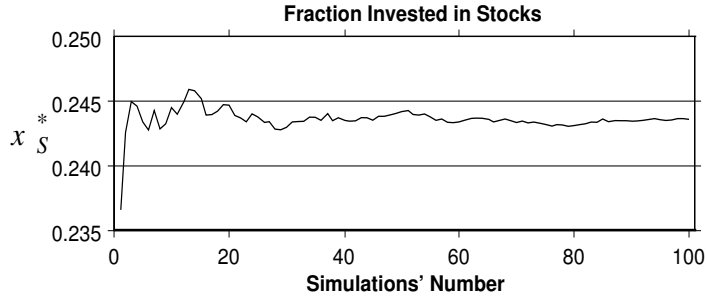

Figure 4.7: Fraction invested in Stocks increasing the simulations' number

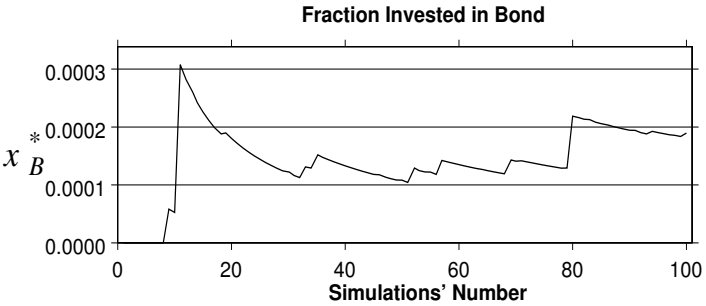

Figure 4.8: Fraction invested in Bonds increasing the simulations' number

B.Fraction Invested, Varying the Shortfall Probability

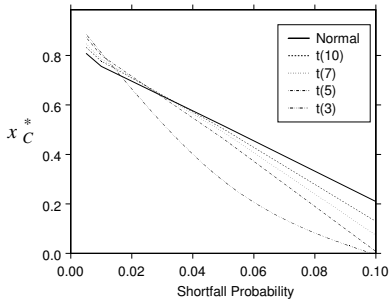

Figure 4.9: Fraction invested in Cash varying the shortfall probability level.

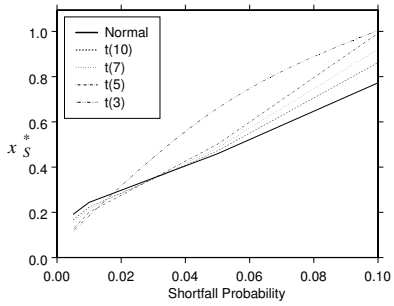

Figure 4.10: Fraction invested in Stocks varying the shortfall probability level.

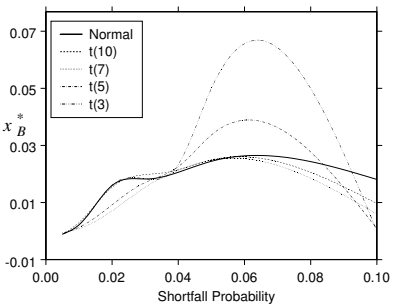

Figure 4.11: Fraction invested in Bonds varying the shortfall probability level.

C.The Effects of the Shortfall Return Level on the Optimal Allocation

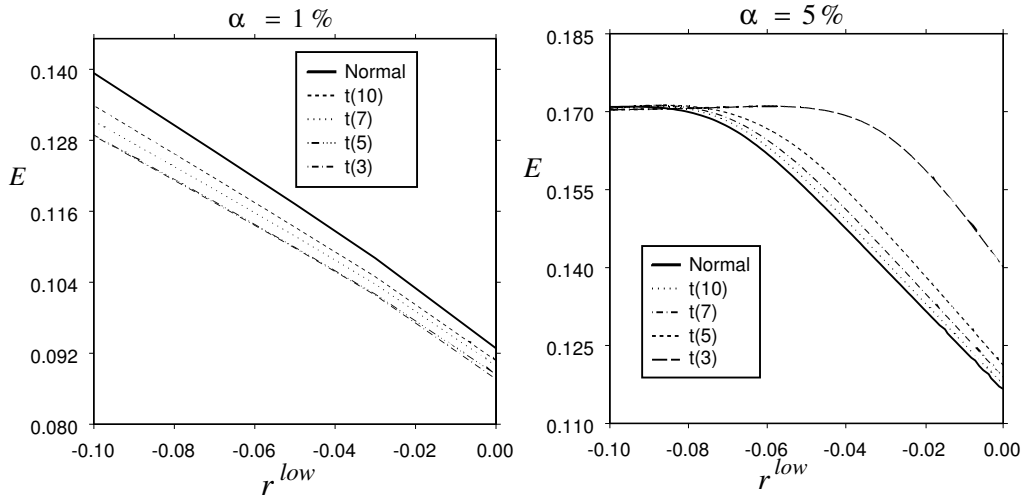

Figure 4.12: Expected return, varying the shortfall return for $\alpha = 5\%$ and $\alpha = 1\%$.

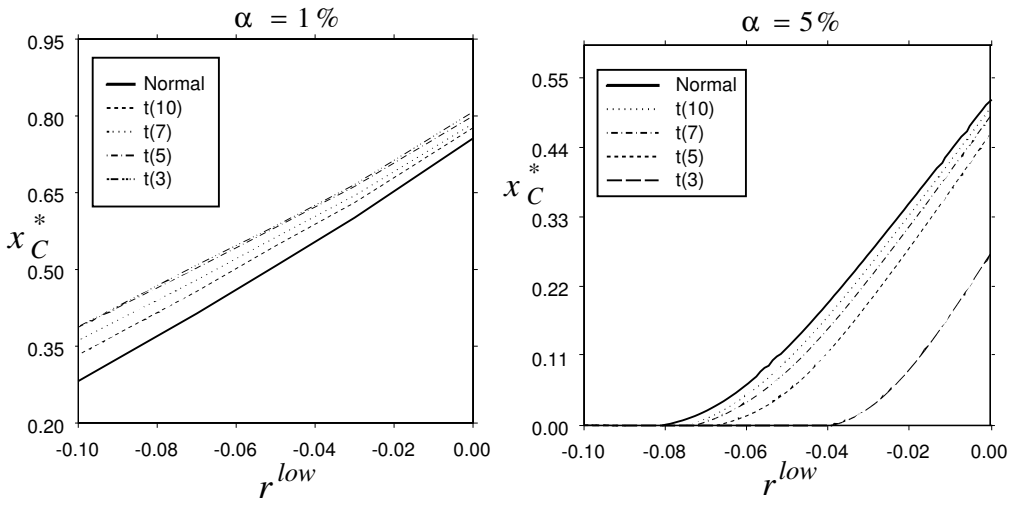

Figure 4.13: Fraction invested in Cash for different distributions ($\alpha = 1\%$, $\alpha = 5\%$).

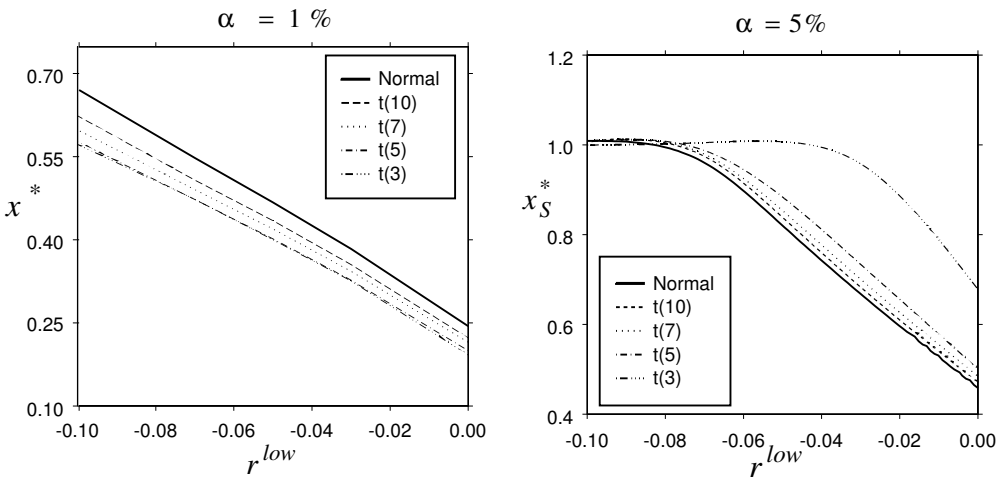

Figure 4.14: Fraction investd in Stocks for different distributions ($\alpha = 1\%$, $\alpha = 5\%$).

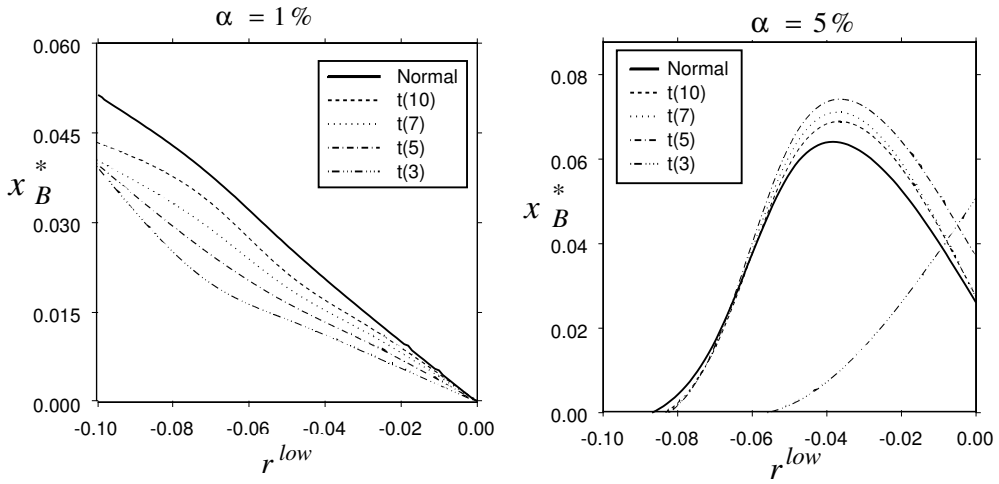

Figure 4.15: Fraction invested in Bonds for different distributions ($\alpha = 5\%$, $\alpha = 1\%$).

D.Rersults Obtained for Mixed Distribution

Table 4.4: Expected portfolio return (Mixed distribution).

	α			
r^{low}	0.5%	1%	5%	10%
0%	8.442	8.980	11.825	16.182
	(0.012)	(0.013)	(0.026)	(0.043)
-3%	9.612	10.313	14.190	17.000
	(0.016)	(0.016)	(0.034)	(0.000)
-5%	10.306	11.127	15.760	17.000
	(0.019)	(0.020)	(0.043)	(0.000)
-7%	10.955	11.971	16.982	17.000
	(0.021)	(0.025)	(0.017)	(0.0000)
-10%	11.938	13.151	17.000	17.000
	(0.026)	(0.024)	(0.000)	(0.000)

Table 4.5: Fraction invested in each asset class (Mixed distribution)

	α				
r^{low}	0.5%	1%	5%	10%	Asset
0%	83.878	78.665	49.254	7.813	CASH
	(0.111)	(0.112)	(0.203)	(0.343)	
	16.092	21.261	47.643	91.695	STOCK
	(0.110)	(0.111)	(0.216)	(0.361)	
	0.0285	0.0731	3.101	0.491	BONDS
	(0.010)	(0.024)	(0.015)	(0.026)	
-3%	72.302	64.982	25.277	0.000	CASH
	(0.139)	(0.139)	(0.232)	(0.000)	
	27.179	33.776	67.382	1.000	STOCK
	(0.136)	(0.139)	(0.232)	(0.000)	
	0.518	1.238	7.340	0.000	BONDS
	(0.021)	(0.006)	(0.094)	(0.000)	
-5%	65.245	56.596	8.516	0.000	CASH
	(0.164)	(0.167)	(0.239)	(0.000)	
	33.485	41.261	83.476	1.000	STOCK
	(0.165)	(0.172)	(0.463)	(0.000)	
	1.269	2.142	8.011	0.000	BONDS
	(0.021)	(0.013)	(0.224)	(0.000)	
-7%	57.928	48.348	0.144	0.000	CASH
	(0.185)	(0.202)	(0.055)	(0.000)	
	39.164	48.593	99.778	1.000	STOCK
	(0.195)	(0.214)	(0.081)	(0.000)	
	2.907	3.058	0.077	0.000	BONDS
	(0.011)	(0.017)	(0.027)	(0.000)	
-10%	47.797	35.883	0.000	0.000	CASH
	(0.216)	(0.200)	(0.000)	(0.000)	
	48.208	60.088	1.000	1.000	STOCK
	(0.237)	(0.219)	(0.000)	(0.000)	
	3.994	4.028	0.000	0.000	BONDS
	(0.030)	(0.023)	(0.000)	(0.000)	

Bibliography

[1] Chopra, V. K., and Ziemba, W. T. (1993), The effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice, *Journal of Portfolio Management*, 19, p.6-11.

[2] D'Agostino, R. D. and Stephens, M. A. (1986), *Goodness-of-fit tecniques*, Marcel Dekker, Inc.

[3] Fang, K. T., Kotz, S. and Wang, NG K. (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

[4] Gill M. and Kellezi E. (2000), Extreme Value Theory For Tail-Related Risk Measures, *Working paper*, University of Geneva, Department of Economics and FAME, Switzerland.

[5] Glasserman P., Heidelberger P., and Shahabuddin P. (2000), Portfolio Value-at-Risk with Heavy-Tailed Risk Factors, *Working Paper*, Columbia University.

[6] Haas, M. and Kondratyev A. (2000), Value-at-Risk and Expected Shortfall with Confidence Bands: An Extreme Value Theory Approach.

[7] Harlow W. V. (1991), Asset Allocation in a Downside Risk Framework, *Financial Analysis Journal*, 28.

[8] Huisman, R., Koedijk K. G. and Pownall, R. A. (1998), VaR-x: Fat Tails in Financial Risk Management, *Journal of Risk*, 1, 47-61.

[9] Huisman R., Koedijk K. G. and Pownall R. A. J. (1999), Asset Allocation in a Value-at-Risk Framework, *Working Paper*, Erasmus University of Rotterdam.

[10] Jaeger S., Rudolf M. and Zimmermann H. (1993),Efficient Shortfall Frontier, *Working paper*, Swiss Institute of Banking and Finance-University of St. Gallen, June.

[11] Johnson, L. N. and Kotz, S. (1972), Distributions in Statistics: Continuous Multivariate Distributions, John Wiley & Sons, Inc., New York, London.

[12] Judd, K. L. (1998), *Numerical Methods in Economics*, Cambridge Massachusetts Institute of Technology

[13] Kalin D. and Zagst R., (1999), Portfolio optimisation: volatility constraints versus shortfall constraints, *Working paper*, University of Ulm, Helmholtzstrabe 18 Germany, Department of Mathematics and Economics.

[14] Kataoka S., (1963), A Stochastic Programming Model, *Econometrica*, 31, 181-196.

[15] Leibowitz M. L. and S. Kogelman (1991), Asset Allocation under Shortfall constraints, *Journal of Portfolio Management* 25, 71-79.

[16] Leibowitz M., Bader N. L. and Kogelman S. (1992), Asset Performance and Surplus Control: A dual Shortfall Approach, *Journal of Portfolio Management*, Winter, pp. 28-37.

[17] Leibowitz M., Bader N. L. and Kogelman S. (1996), *Return Targets and Shortfall Risks*, IRWIN.

[18] Lucas R. and Klaassen P. (1998), Extreme Returns Downside Risk, and Optimal Asset Allocation, *Journal of Portfolio Management*, 25, 71-79.

[19] Niederreiter H. (1992), *Random Number Generation and Quasi Monte Carlo Methods*, Society for industrial and applied mathematics, Philadelphia Pennsylvania.

[20] Palmitesta P. and Provasi C. (2003), Multivariate Input Modelling with the Family of Koehler-Symanowski Distributions, in *Atti del Convegno Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione*, 4-6 Semptember 2003, Statistics Department, University " Ca' Foscari", Venice.

[21] Press W., Teukolsky S., Vetterling T. and Flannery B.P. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge.

[22] Ripley B.D. (1987), *Stochastic Simulation*, J. Wiley, New York.

[23] Robert, C. P. (1996), *Methodes de Monte Carlo par chaines de Markov*, Economica, Paris.

[24] Robert C.P. and Casella G. (1999), *Monte Carlo Statistical Methods*, Springer Verlag, New York.

[25] Roy A. D. (1952), Safety First and the Holding of Assets, *Econometrica*, 20, 431-449.

[26] Rubinstein R. Y. (1981), *Simulation and the Monte Carlo Method*, John Wiley & Sons.

[27] Rudolf M. (1994), Efficient frontier and Shortfall Risk, *Finanzmarkt und Portfolio Management*, 8 Jahrgang No. 1, pp. 88-101.

[28] Telser L. G. (1955), Safety first and Hedging, *Review of Economics and Statistics*, Vol. 23, pp. 1-16.

Chapter 5

Bayesian Inference for Mixtures of Stable Distributions

5.1 Introduction

In many different fields such as hydrology, telecommunications, physics and finance, Gaussian models reveal difficulties in fitting data that exhibits a high degree of heterogeneity; thus *stable distributions* have been introduced as a generalisation of the Gaussian model. Stable distributions allow also for infinite variance, skewness and heavy tails. The tails of a stable distribution decay like a power function, allowing extreme events to have higher probability mass than in Gaussian model. For a summary of the properties of the stable distributions see Zoloratev [42] and Samorodnitsky and Taqqu [36], which provide a good theoretical background on heavy-tailed distributions. The practical use of heavy-tailed distributions in many different fields is well documented in the book of Adler, Feldman and Taqqu [1], which also reviews the estimation techniques.

In finance, the first studies on the hypothesis of stable distributed stock prices

[1]Part of this work is in:

Casarin, R., (2004), "Bayesian Inference for Mixture of Stable Distributions", forthcoming, *Working Paper* CEREMADE. Presented at the Young Statistician Meeting, Cambridge 14-15 April 2003.

Casarin, R., (2003), "Bayesian Inference for Mixture of Stable Distributions", in *Atti del Convegno Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione* , 4-6 Semptember 2003, Statistics Department, University " Ca' Foscari", Venice.

can be attributed to Mandelbrot [22], Fama [13], [14] and Fama and Roll [15], [14]. They propose stable distributions and give some statistical instruments for the inference on the characteristic exponent. The use of stable distributions has been motivated also on the basis of empirical evidence from financial markets. Brenner [5] uses the notion of stationarity of the time series to explain stability of stock prices. An illuminating work on inference for stable distributions is due to Buckle [6], who makes also an empirical analysis on daily stock prices, using a full Bayesian approach to estimate stable distributions parameters and finding significant evidence of the stable distribution hypothesis.

There are many recent works treating the use of stable distributions in finance. For example see Bradley and Taqqu [4] and Mikosch [26] for an introduction to the use of stable distributions in financial risk modelling. The work of Mittnik, Rachev and Paolella [25] and of Rachev and Mittnik [31] provides a quite complete analysis of the theoretical and empirical aspects of the stable distributions in finance.

Other early studies, performing empirical analysis on stocks prices, suggest to use mixtures of distributions in order to modelling the financial markets heterogeneity. Barnes and Downes [2] use the same estimation techniques of Fama and Roll [16] in order to discuss the results of Teichmoeller [39]. They find that for some stock the property of stability does not hold and that the characteristic exponent varies across the stocks. In order to account for this kind of heterogeneity of the stock prices the authors suggest mixture of stable distributions as an alternative hypothesis. Simkowitz and Beedles [3] perform an empirical analysis focusing on the asymmetry of stock returns. They find that the skewness of the stock returns is frequently positive and dependent on the level of the characteristic exponent. They conclude that securities distributions may be better modelled through mixtures of stable distributions. Finally an extensive empirical analysis due to Fieltz and Rozelle [17] shows that mixtures of Gaussian, or non-Gaussian distributions can better describe stock prices. In particular the authors suggest to use non-Gaussian stable mixtures model with changing scale parameter because it directly accounts for skewness. We can conclude that the problem of multimodality and in general of heterogeneity is well documented in the financial literature, also from the earlier studies on the stable distributions. Thus an appropriate modelling is needed.

Observe that, in order to account for heterogeneity and non-linear dependencies

exhibited by the data, stable distributions have been already introduced in different kind of statistical models. For instance in *survival models*, the heterogeneity within survival times of a population are modelled trough common latent factors, which follow stable distributions, see for example Qiou, Ravishanker and Dey [29]. Stable distributions are also used to model heterogeneity over time. For an introduction to *time series models* with stable noises, see Qiou and Ravishanker [30] and Mikosch [26].

Different estimation methods for stable distributions have been proposed in the literature. For a full Bayesian approach see Buckle [6], for a maximum likelihood approach see DuMouchel [11] and for MCEM approach with application to time series with symmetric stable innovations see Godsill [21].

The first aim of our work is to propose a stable distributions mixture model in order to capture the heterogeneity of data. In particular we want to account for multimodality, which is present, for example, in financial data. The second goal of the work is to provide some inferential tools for stable distributions mixtures. As suggested in the literature on Gaussians mixtures (see for example Robert [34]), we propose a particular reparameterisations of the mixture model in order to make more easy the statistical inference on the mixture parameters. Furthermore we use both a full Bayesian approach and MCMC simulation techniques in order to estimate the parameters.

The maximum likelihood approach (see for example McLachlan and Peel [23]) to the mixture model implies numerical difficulties, which rely on the fact that for many parametric density family the likelihood surface has singularities. Furthermore, as pointed out by Stephens [38], the likelihood may have several local maximum and it will be difficult to justify the choice of one of these point estimates. The presence of several local maximum and of singularities implies that the standard asymptotic theory for maximum likelihood estimation and the test theory do not apply in the mixture context. The Bayesian approach avoids these problem as parameters are random variables, with prior and posterior distributions defined on the parameter space. Thus it is no more necessary to choose between several local maximum, because point estimates are obtained by averaging over the parameter space, weighting by the posterior distribution of the parameters or by the simulated posterior distribution.

The structure of the work is as follows. Section 5.2 defines a stable distribution and the method to simulate from a stable. Section 5.3 provides an introduction to some basic Markov Chain Monte Carlo(MCMC) methods for mixtures and exhibits the Bayesian model and the Gibbs sampler for a stable distribution. Section 5.4 describes the Bayesian model for stable mixtures, with particular attention to the missing data structure of the stable mixture model and the Gibbs sampler for stable mixture is developed in the case where the number of components is fixed. Section 5.5 provides some results of the Bayesian stable mixture model on financial dataset. Section 5.6 concludes.

5.2 Simulating from a Stable Distribution

The existence of simulation methods for stable distributions opens the way to Bayesian inference on the parameters of this distribution family. In this section we define a stable random variable and briefly describe the method to simulate from a stable distribution, first proposed by Chamber, Mallows and Stuck [8] and then discussed also in Weron [41]. We use this method in our work, to generate dataset to test the efficiency of the MCMC based Bayesian inference approach. In the following we denote a stable distribution by $S_\alpha(\beta, \delta, \sigma)$. Stable distributions do not generally have an explicit probability density function and are thus conveniently defined through their characteristic function. The most well known parametrisation is defined in Samorodnitsky and Taqqu [36].

Definition 5.2.1. *(Stable distribution)*
A random variable X has stable distribution $S_\alpha(\beta, \delta, \sigma)$ if its parameters are in the following ranges: $\alpha \in (0, 2]$, $\beta \in [-1, +1]$, $\delta \in (-\infty, +\infty)$, $\sigma \in (0, +\infty)$ and if its characteristic function can be written as

$$E[exp(i\,\vartheta\,x)] = \begin{cases} exp(-|\sigma\vartheta|^\alpha)(1 - i\,\beta(sign(\vartheta))tan(\pi\alpha/2) + i\delta\vartheta) & if\ \alpha \neq 1; \\ exp(-|\sigma\vartheta|(1 + 2\,i\,\beta ln|\vartheta|sign(\vartheta)/\pi) + i\delta\vartheta) & if\ \alpha = 1. \end{cases} \tag{5.1}$$

where $\vartheta \in \mathbb{R}$.

The stable distribution is thus completely characterised through the following four parameters: the *characteristic exponent* α, the *skewness parameter* β, the *location*

parameter δ and finally the *scale parameter* σ. An equivalent parametrisation is proposed by Zoloratev [42]. For a review on all the equivalent definitions of stable distribution and on all their properties see Samorodnitsky, Taqqu [36]. The distribution $S_\alpha(\beta, 0, 1)$ is usually called *standard stable* and when $\alpha \in (0, 1)$ it is called *positive stable* because the support of the density is the positive half of the real line. In this case the characteristic function reduces to

$$E\left[\exp(i\theta x)\right] = e^{-|\theta|^\alpha} \tag{5.2}$$

Stable distributions admit explicit representation of the density function only in the following cases: the Gaussian distribution $S_2(0, \sigma, \delta)$, the Cauchy distribution $S_1(0, \sigma, \delta)$ and the Lévy distribution $S_{1/2}(1, \sigma, \delta)$.
The algorithm we used for simulating a standard stable (see Chamber, Mallows and Stuck [8] and Weron [41]) is the following

(i) Simulate

$$V \sim U_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \tag{5.3}$$

$$W \sim \varepsilon xp(1); \tag{5.4}$$

(ii) If $\alpha \neq 1$ then

$$\begin{aligned} Z &= S_{\alpha,\beta} \frac{sin(\alpha(V + B_{\alpha,\beta}))}{cos(V)^{1/\alpha}} \left(\frac{cos(V - \alpha(V + B_{\alpha,\beta}))}{W} \right)^{\frac{1-\alpha}{\alpha}} \\ B_{\alpha,\beta} &= \frac{arctan(\beta tan(\frac{\pi\alpha}{2}))}{\alpha} \\ S_{\alpha,\beta} &= (1 + \beta^2 tan^2(\frac{\pi\alpha}{2}))^{\frac{1}{2\alpha}} \end{aligned} \tag{5.5}$$

If $\alpha = 1$ then

$$Z = \frac{2}{\pi} \left[(\frac{\pi}{2} + \beta V) \tan(V) - \beta \log(\frac{W \cos(V)}{\frac{\pi}{2} + \beta V}) \right] \tag{5.6}$$

Once a value Z from a standard stable $S_\alpha(\beta, 0, 1)$ has been simulated, in order to obtain a value X from a stable distribution with scale parameter σ and location parameter δ, the following transformation is required

$$X = \begin{cases} Z + \delta & if\ \alpha \neq 1 \\ \sigma Z + \frac{2}{\pi}\beta\sigma\, log(\sigma) + \delta & if\ \alpha = 1 \end{cases}$$

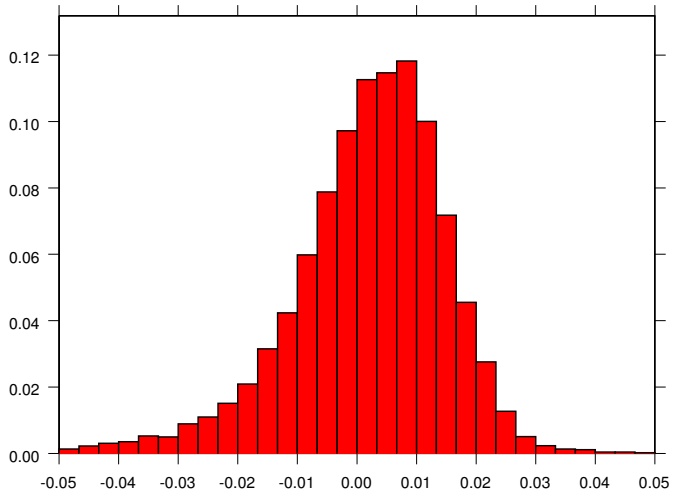

Figure 5.1: Simulation from stable distribution $S_{1.65}(-0.8, 0.00053, 0.0079)$.

Fig. 5.1 exhibits simulated data from a stable distribution. We use the parameters estimated by *Buckle* [6] on financial time series.

5.3 Bayesian Inference for Stable Distributions

In order to make inference on the parameters of a stable distribution in a Bayesian approach it is necessary to specify a *hierarchical model* on the parameters of the distribution. Often, the resulting posterior distribution of the Bayesian model cannot be calculated analytically, thus it is necessary to chose a numerical approximation method. Monte Carlo simulation techniques provide an appealing solution to the problem because, in high dimensional space, they are more efficient than traditional numerical integration methods and furthermore they require the densities involved in the posterior to be known only up to a normalising constant. In the following the basic Markov Chain Monte Carlo (*MCMC*) techniques will be introduced and the Gibbs sampler for a stable distribution will be discussed.

5.3.1 MCMC Methods for Bayesian Models

As evidenced in Chapter 1, in Bayesian inference many quantities of interest can be represented in the integral form

$$I(\theta) = \mathbb{E}_{\pi(\theta|\mathbf{x})}\{f(\theta)\} = \int_{\mathcal{X}} f(\theta)\pi(d\theta|\mathbf{x}) \tag{5.7}$$

where $\pi(\theta|\mathbf{x})$ is the posterior distribution of the parameter $\theta \in \mathcal{X}$ given the observed data $\mathbf{x} = (x_1, \ldots, x_k)$. In many cases to find an analytical solution to the integration problem is difficult and a numerical approximation is needed. A way to approximate the integral is to simulate the posterior distribution and to average the simulated values of $f(\theta)$. In particular the MCMC methods consist in the construction of a Markov chain $\left\{\theta^{(t)}\right\}_{t=1}^{n}$ and in the following approximation of the integral given in Eq. (5.7)

$$I_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} f(\theta^{(t)}) \tag{5.8}$$

which is a consistent estimator of the quantity of interest

$$I_n(\theta) \xrightarrow{a.s.} \mathbb{E}_{\pi(\theta|\mathbf{x})}\{f(\theta)\} \tag{5.9}$$

In some cases, as in mixture models, is not possible to simulate directly from the posterior distribution and a further simulation step (*completion step*) is needed. All MCMC algorithms are based on the construction of a discrete time Markov Chain, through the specification of its transition kernel. Thus the properties of this kind of stochastic process are useful in order to study the convergence of the MCMC simulation algorithms.

We recall that the irreducibility of the chain is a sufficient condition in order to guarantee the convergence of I_n to the quantity of interest given in Eq. (5.7).

Theorem 5.3.1. *(Law of Large Numbers)*
If the Markov chain $\{\Theta\}_{t=0}^{\infty}$ is irreducible and has σ-finite invariant measure π, then $\forall f, g \in L^1(\pi)$, with $\int g(\theta)d\pi(\theta) \neq 0$

$$\lim_{n\to\infty} \frac{I_n(f)}{I_n(g)} = \frac{\int f(\theta)d\pi(\theta)}{\int g(\theta)d\pi(\theta)} \tag{5.10}$$

where $I_n(h) = \frac{1}{n}\sum_{i=1}^{n} h(\Theta_i)$.

For a brief introduction to Markov chains and to Markov Chain Monte Carlo methods we refer to Chapter 1. Further details on Markov chains can be found for example in Meyn and Tweedie [24], other theoretical results on convergence are in Tierney [40], finally Robert and Casella [35] provides some techniques for monitoring convergence.

5.3.2 The Gibbs Sampler

The Gibbs sampler has been introduced in image processing by Geman and Geman [19] (see also Chapter 1 for a general introduction to MCMC methods and to Gibbs sampling) and it is a method of construction of a Markov Chain $\{\Theta^{(t)}\}_{t=0}^{\infty}$ with multivariate stationary distribution $\pi(\theta|(x))$, where $\theta \in \chi$. This simulation method is particularly useful when the posterior density is defined on a high dimension space. If the random vector θ can be written as $\theta = (\theta_1, \ldots, \theta_p)$ and if we can simulate from the full conditional densities

$$(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p) \sim \pi_i(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p, \mathbf{x}) \tag{5.11}$$

then the associated Gibbs sampling algorithm is given by the following transition kernel from $\theta^{(t)}$ to $\theta^{(t+1)}$:

Definition 5.3.1. *(Gibbs Sampler)*
Given the state $\Theta^{(t)} = \theta^{(t)}$ at time t, generate the state $\Theta^{(t+1)}$ as follows

1. $\Theta_1^{(t+1)} \sim \pi(\theta_1|\theta_2^{(t)}, \ldots, \theta_p^{(t)}, \mathbf{x})$
2. $\Theta_2^{(t+1)} \sim \pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_p^{(t)}, \mathbf{x})$
3. ...
4. $\Theta_p^{(t+1)} \sim \pi(\theta_p|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \ldots, \theta_{p-1}^{(t+1)}, \mathbf{x})$

Under some regularity conditions the Markov chain produced by the algorithm converges to the desired stationary distribution (see Robert and Casella [35]).

5.3.3 The Gibbs Sampler for Univariate Stable Distributions

In this paragraph we give a description of the Gibbs sampler proposed by Buckle [6] in order to estimate the characteristic exponent α of a stable distribution. It is known (see Section 5.1) how to simulate values from a stable distribution; furthermore it is possible to represent the stable density in integral form, by introducing an auxiliary variable y, as suggested by Buckle [6]. The stable density is obtained by integrating with respect to y the bivariate density function of the pair (x, y)

$$f(x,y|\alpha,\beta,\sigma,\delta) = \frac{\alpha}{|\alpha-1|}\exp\left\{-|\frac{z}{\tau_{\alpha,\beta}(y)}|^{\alpha/(\alpha-1)}\right\}\left|\frac{z}{\tau_{\alpha,\beta}(y)}\right|^{\alpha/(\alpha-1)}\frac{1}{|z|} \tag{5.12}$$

$$(x,y) \in (-\infty,0)\times(-1/2,l_{\alpha,\beta})\cup(0,\infty)\times(l_{\alpha,\beta},1/2) \tag{5.13}$$

$$\tau_{\alpha,\beta}(y) = \frac{\sin(\pi\alpha y+\eta_{\alpha,\beta})}{\cos(\pi y)}\left[\frac{\cos(\pi y)}{\cos(\pi(\alpha-1)y)+\eta_{\alpha,\beta}}\right]^{(\alpha-1)/\alpha} \tag{5.14}$$

$$\eta_{\alpha,\beta} = \beta\, min(\alpha, 2-\alpha)\pi/2 \tag{5.15}$$

$$l_{\alpha,\beta} = -\eta_{\alpha,\beta}/\pi\alpha \tag{5.16}$$

where $z = \frac{x-\delta}{\sigma}$. Previous elements allow to perform simulation based Bayesian inference on the parameters of the stable distribution. The Bayesian model is described through the *Directed Acyclic Graph* (DAG) in Fig. 5.2. Suppose to observe n realizations $\mathbf{x} = (x_1, \ldots, x_n)$ from a stable distribution $S_\alpha(\beta, \sigma, \delta)$ and simulate a vector of auxiliary variables $\mathbf{y} = (y_1, \ldots, y_n)$, then the *completed likelihood* and the completed posterior distribution are respectively

$$L(\mathbf{x},\mathbf{y}|\theta) = \prod_{i=1}^{n} f(x_i, y_i|\theta) \tag{5.17}$$

$$\pi(\theta|\mathbf{x},\mathbf{y}) = \frac{L(\mathbf{x},\mathbf{y}|\theta)\pi(\theta)}{\int_\Theta L(\mathbf{x},\mathbf{y}|\theta)\pi(\theta)d\theta} \propto \prod_{i=1}^{n} f(x_i, y_i|\theta)\pi(\theta) \tag{5.18}$$

where $\theta = (\alpha, \beta, \delta, \sigma)$ is the stable parameter vector varying in the parameter space Θ.

In the following we suppose to observe n values from a standard stable distribution $S_\alpha(\beta, 1, 0)$ and we assume the other parameters to be known. Parameters α and β are estimated by simulating from the complete posterior distribution and by

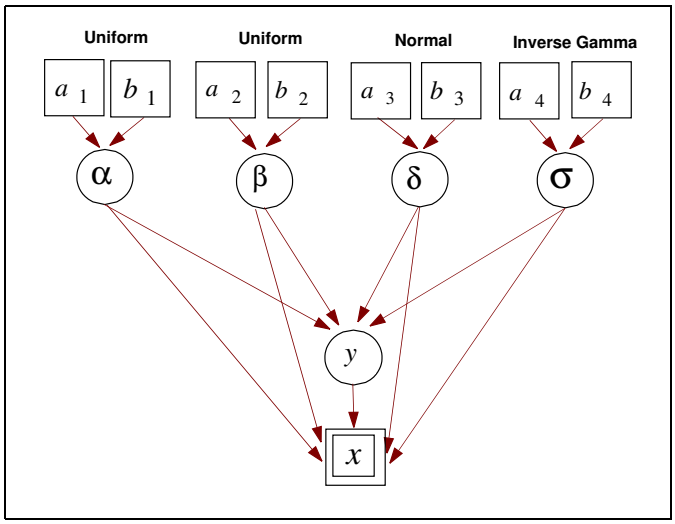

Figure 5.2: DAG of the Bayesian model for inference on stable distributions. It exhibits the hierarchical structure of priors, and hyperparameters. A single box around a quantity indicates that it is a known constant, a double box indicates the variable is observed and a circle indicates the random variable is not observable. The directed arrows show the dependence structure of the model. We use the the prior suggested by Buckle [6].

averaging simulated values. Simulations from the posterior distribution are obtained by iterating the following steps of the Gibbs sampler

(i) *Update the completing variable*

$$\pi(y_i|\alpha,\beta,\delta,\sigma,z_i) \propto \exp\left\{1-\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}} \tag{5.19}$$

with $i = 1,\ldots,n$.

(ii) *Simulate from the complete full conditional distributions*

$$\pi(\alpha|\beta,\delta,\sigma,\mathbf{x},\mathbf{y}) \propto \frac{\alpha^n}{|\alpha-1|^n}\exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\pi(\alpha) \tag{5.20}$$

$$\pi(\beta|\alpha,\delta,\sigma,z_i,y_i) \propto \exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{1}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\pi(\beta) \tag{5.21}$$

$$\pi(\delta|\alpha,\beta,\sigma,z_i,y_i) \propto \exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\frac{1}{|x_i-\delta|}\,\pi(\delta) \quad (5.22)$$

$$\pi(\sigma|\alpha,\beta,\sigma,z_i,y_i) \propto \left(\frac{1}{\sigma^{\alpha/(\alpha-1)}}\right)^{n}\exp\left\{-\frac{1}{\sigma^{\alpha/(\alpha-1)}}\sum_{i=1}^{n}\left|\frac{(x_i-\delta)}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\right\}\pi(\sigma) \quad (5.23)$$

where $\pi(\alpha)$, $\pi(\beta)$, $\pi(\delta)$, $\pi(\sigma)$ are the prior distributions on the parameters, $\mathbf{y}$ is a vector of auxiliary variables $(y_1,\ldots,y_n)$ and $\tau_{\alpha,\beta}$ is a function of y defined in Eq. (5.14).

In order to simulate from the density function given in equation (5.19) we apply the accept reject method (see Devroye [9]), because the density is proportional to a function which has finite support $(-\frac{1}{2},\frac{1}{2})$ and which is bounded with value 1 at the maximum y^*, where y^* is such that $\tau_{\alpha,\beta}(y^*)=x$. To emphasize numerical problems which arise in making inference on stable distributions, we plot in Fig. 5.4 the density function of y for different values of x. Note that for all values of $\alpha\in(0,1)$, high values of x make the density function spiked around the mode. Thus the basic accept method performs quite poorly. A way to improve the simulation method is to build a histogram with the rejected values and to use it as an envelope in the accept reject algorithm.

Due to the way the parameter α enters in the likelihood, the densities given in Equations (5.20), (5.21), (5.22) and (5.23) are undulating and rather concentrated, therefore as suggested by Buckle [6] and Ravishanker and Qiou [30] we introduce the following reparametrisations which give a more manageable form of the conditional posteriors of α, β and δ

$$v_i = \tau_{\alpha,\beta}(y_i) \quad (5.24)$$

$$\phi_i = \frac{\tau_{\alpha,\beta}}{x_i-\delta} \quad (5.25)$$

the resulting posteriors are

$$\pi(\alpha|\beta,\delta,\sigma,\mathbf{x},\mathbf{v}) \propto \frac{\alpha^n}{|\alpha-1|^n}\exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{v_i}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{z_i}{v_i}\right|^{\frac{\alpha}{(\alpha-1)}}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=v_i}\pi(\alpha) \quad (5.26)$$

$$\pi(\beta|\alpha,\delta,\sigma,z_i,v_i) = \prod_{i=1}^{n}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=v_i}\pi(\beta) \quad (5.27)$$

$$\pi(\delta|\alpha,\beta,\sigma,z_i,v_i) \propto \prod_{i=1}^{n}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=\phi_i(x_i-\delta)}\pi(\delta) \quad (5.28)$$

At each step of the reparametrised Gibbs sampler, the Jacobian of the transformation, $\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}$, must be evaluated in $y_i = \tau^{-1}_{\alpha,\beta}(v_i)$. Due to the complexity of the function $\tau_{\alpha,\beta}$, its inverse has not an analytical expression. Therefore, following Buckle [6], the inverse transformation is determined numerically. We use the modified safeguard Newton algorithm proposed in Press et al. [28].
In order to simulate from the posteriors given in Equations (5.26), (5.27) and (5.28) we use Metropolis-Hastings algorithm (see Chapter 1 for an introduction to Monte Carlo Markov Chain methods)

Given $\theta_i^{(k-1)}$

(*i*) *Generate from the proposal distribution* $\theta_i^* \sim q(\theta_i|\theta_i^{(k-1)})$

(*ii*) *Take:*

$$\theta_i^{(k)} = \begin{cases} \theta_i^* & \textit{with probability} \quad \rho(\theta_i^{(k-1)},\theta_i^*) \\ \theta_i^{(k-1)} & \textit{with probability} \quad 1-\rho(\theta_i^{(k-1)},\theta_i^*) \end{cases}$$

where: $\rho(\theta_i^{(k-1)},\theta_i^*) = 1 \wedge \left\{\frac{\pi(\theta_i^*|\theta_{-i})}{\pi(\theta_i^{(k-1)}|\theta_{-i})}\frac{q(\theta_i^{(k-1)})}{q(\theta_i^*)}\right\}$.

In order to simulate the full conditional posteriors given in Eq. (5.26) and (5.27), we use a beta distribution, $\mathcal{B}e(a,b)$ as proposal. The sample generated from the beta distribution is not independent because in order to simulate the k-th value of the M.-H. chain, we pose the mean of the beta distribution to be equal to the $(k-1)$-th value of the chain. In setting a and b, the parameters of the proposal distribution we distinguish the following cases

$$\begin{cases} a = \frac{\alpha_{k-1}^2 (1-\alpha_{k-1}) - v\,\alpha_{k-1}}{v} \\ b = a\frac{1-\alpha_{k-1}}{\alpha_{k-1}} \end{cases} \qquad when \quad \alpha \in [0,1)$$

and

$$\begin{cases} a = \frac{(\alpha_{k-1}-1)^2 (2-\alpha_{k-1}) - v\,(\alpha_{k-1}-1)}{v} \\ b = a\frac{2-\alpha_{k-1}}{(\alpha_{k-1}-1)} \end{cases} \qquad when \quad \alpha \in (1,2]$$

where α_{k-1} is the value generated by the Metropolis-Hastings chain at step $(k-1)$ and v is the variance of the proposal distribution. This parameters choice allows us also to avoid numerical problems related to the evaluation of the Metropolis-Hastings acceptance ratio in the presence of fat tailed and quite spiked likelihood functions.

We use a gaussian random walk proposal to simulate the full conditional posterior of the location parameter (Eq. (5.28)).

In order to complete the description of the hierarchical model and of the associated Gibbs sampler, we consider the following joint prior distribution

$$\mathbb{I}(\alpha)_{(1,2]}\frac{1}{2}\mathbb{I}(\beta)_{[-1,1]}\frac{1}{\sqrt{2\pi b_3}}e^{-\frac{(\delta-a_3)^2}{2b_3}}\mathbb{I}(\delta)_{(-\infty,+\infty)}\frac{b_4^{a_4}}{\Gamma(a_4)}\frac{e^{-\frac{b_4}{\theta}}}{\theta^{a_4+1}}\mathbb{I}(\theta)_{[0,\infty)} \tag{5.29}$$

where $\theta = \sigma^{\frac{\alpha}{\alpha-1}}$. We use informative priors for the location and scale parameters. For δ we assume a normal distribution. Note that the prior distribution of θ is the *inverse gamma* distribution $IG(a_4, b_4)$, which is a conjugate prior of the distribution given in equation (5.23). Simulations of the parameter σ can be obtained from the simulated values of θ by a simple transformation. Finally for parameters α and β we assume non informative priors.

We show the efficiency of the MCMC based Bayesian inference, running the Gibbs sampler on simulated dataset. In the following examples we discuss the numerical results and also some computational remarks related to different values of the characteristic exponent.

Example 5.3.1 - *(Positive α-Stable Distributions)*

Fig. 5.5, in *Appendix A*, exhibits the dataset of 1,000 observations generated from the positive stable distribution $S_{0.5}(0.9, 0, 1)$. Figures 5.6, 5.7 give the results of

$15,000$ iterations of the Gibbs sampler relative to the characteristic exponent α, and the asymmetry parameter β, for an increasing number of simulations. Figures 5.8, 5.9 exhibit the ergodic averages. Note that convergence has been achieved after 6,000 iterations. In the first experiment we use uniform priors for α and β. We set the variance of the proposal distribution (see Eq. 5.46) at $v = 0.0001$ for α, at $v = 0.0009$ for β and the starting values of the Gibbs sampler at $\alpha = 0.2$ and $\beta = 0.5$. Figures 5.10 and 5.11 exhibit the estimated acceptance rate for the M.-H. steps in the Gibbs sampler, for an increasing number of simulations. The histograms of the posteriors are depicted in figures 5.12 and 5.13. Note that through the Gibbs sample it is also possible to obtain the confidence intervals for the estimated parameters and to perform a goodness of fit test.

□

Example 5.3.2 - *(α-Stable Distributions with $\alpha > 1$)*
Note that in the first dataset α is less than 1, therefore the moment of order less than two are infinite. In some applications like in finance, in order to give an interpretation to the result it is preferable to work at least with finite first order moments. Therefore we verify the efficiency of the Gibbs sampler also on a sample generated from a stable distribution with $\alpha \in (1, 2]$. Fig. 5.14 in *Appendix A* exhibits the dataset of 1,000 values simulated from the stable distribution $S_{1.5}(0.9, 0, 1)$. Gibbs sampler realisations and ergodic averages are represented respectively in figures 5.15, 5.16 and in figures 5.17, 5.18. Convergence has been achieved after 6,000 iterations. In the experiment we use uniform priors for α and β and set the variance of the proposal distribution at $v = 0.0005$ for α, at $v = 0.0009$ for β and the starting values of the Gibbs sampler at $\alpha = 1.2$ and $\beta = 0.2$. Figures 5.19, 5.20 exhibit the estimated acceptance rate, for an increasing number of simulations. Finally, posterior distributions are depicted in figures 5.21 and 5.22.

□

For each dataset, Table 5.1 summarizes the estimated parameters, the standard deviations and the estimated acceptance rates of the M.-H. steps of the Gibbs sampler. Results are obtained on a PC with Intel *1063 MHz* processor, using

routines implemented in *C/C++*. We validate the MCMC code by checking that without any data the estimated joint posterior distribution correspond to the joint prior.

Table 5.1: Numerical results - Ergodic Averages over 15,000 Gibbs realisations.

	First Dataset: $S_{0.5}(0.9, 0, 1)$				
Parameter	True Value	Starting Value	Estimate$^{(*)}$	Std.Dev.	Acc. Rate
α	0.5	0.2	0.54	0.02	0.11
β	0.9	0.5	0.95	0.07	0.14
	Second Dataset: $S_{1.5}(0.9, 0, 1)$				
Parameter	True Value	Starting Value	Estimate$^{(**)}$	Std.Dev.	Acc. Rate
α	1.5	1.2	1.46	0.04	0.49
β	0.9	0.2	0.96	0.04	0.21

(*) *Time (sec): 4259*
(**) *Time (sec): 4128*

5.4 Bayesian Inference for Mixtures of Stable Distributions

In this section we extend the Bayesian framework, introduced in the previous section, to the mixtures of stable distributions. In many situations data may exhibit simultaneously: heavy tails, skewness, and multimodality. In time series analysis, the multimodality of the empirical distribution can also find a justification in a heterogeneous time evolution of the observed phenomena. For example, the distribution of financial time series like prices or prices volatility may have many modes because the stochastic process evolves over time following different regimes. Stable distributions allow for skewness and heavy tails, but not for multimodality. Thus a way to model these features of the data, is to introduce stable mixtures. Furthermore the use of stable mixtures is appealing also because they have normal mixtures as special case, which is a widely studied topic (see for example Stephens [38], Richardson and Green [32]). Other relevant works on the Bayesian approach to the mixture models estimation are Diebolt and Robert [10], Escobar and West [12] and Robert [34], [33]. In *Appendix C* some examples of two components stable

mixtures are exhibited. We simulate stable mixtures with different parameters setting, in order to understand the influence of each parameter on the shape of the mixture's distribution.

5.4.1 The Missing Data Model

In the following we define a stable mixture model, while assuming to known the number of mixture components. Under a practical point of view the number of components may be detected by looking at the number of modes in the distribution or by performing a statistical test, see section 5.5. Let L be the finite number of mixture components and $f(x|\alpha_l, \beta_l, \delta_l, \sigma_l)$ the l-th stable distribution in the mixture, then the mixture model $m(x|\theta, p)$ is

$$m(x|\theta, p) = \sum_{l=1}^{L} p_l f(x|\theta_l) \tag{5.30}$$
$$\text{with} \quad \sum_{l=1}^{L} p_l = 1, \quad p_l \geq 0, \quad l = 1, \ldots, L$$

where $\theta_l = (\alpha_l, \beta_l, \delta_l, \sigma_l)$, $l = 1, \ldots, L$ are the parameter vector and $\theta = (\theta_1, \ldots, \theta_L)$. In the following we suppose L to be known. In order to perform Bayesian inference two steps of completion are needed. First, we adopt the same completion technique used for stable distributions. The auxiliary variable, y, is introduced in order to obtain an integral representation of the mixture distribution

$$m(x|\theta, p) = \sum_{l=1}^{L} p_l \int_{-1/2}^{1/2} f(x, y|\theta_l)\, dy. \tag{5.31}$$

The second step of completion is introduced in order to reduce the complexity problem, which arises in simulation based inference for mixtures. The completing variable (or *allocation variable*), $\nu = \{\nu_1, \ldots, \nu_L\}$ is defined as follow

$$\nu_l = \begin{cases} 1 & if\ x \sim f(x, y|\theta_l) \\ 0 & otherwise \end{cases} \qquad \text{with } l = 1, \ldots, L \tag{5.32}$$

and is used to select the mixture component. The allocation variable is not observable and this missing data structure can be estimated by following a simulation

based approach. Simulations from the mixture model can be performed in two step: first, simulating the allocation variable; second, simulating a mixture component conditionally on the allocation variable. The resulting demarginalized mixture model is

$$m(x,\nu|\theta,p)=\prod_{l=1}^{L}\left(p_l\int_{-1/2}^{1/2}f(x,y|\theta_l)\,dy\right)^{\nu_l}, \qquad \sum_{l=1}^{L}\nu_l=1 \tag{5.33}$$

This completion strategy is now quite popular in Bayesian inference for mixtures (see Robert [33], Robert and Casella [35], Escobar and West [12] and Diebolt and Robert [10]). For an introduction to Monte Carlo methods in Bayesian inference from data modeled by mixture of distributions see also Neal [27] and for a discussion of the numerical and identifiability problems in mixtures inference see Richardson and Green [32], Stephens [37] and Celeux, Hurn and Robert [7].

5.4.2 The Bayesian Approach

The Bayesian model for inference on stable mixtures is represented through the DAG in Fig. 5.3. Before specifying the Bayesian model we introduce two distributions that are quite useful in Bayesian inference form mixtures: the *multinomial distribution* and the *Dirichlet distribution*.

Definition 5.4.1. *(Multinomial distribution)*
The random variable $X=(x_1,\ldots,x_L)$ has L dimensional multinomial distribution if its density function is

$$\mathcal{M}(n,p_1,...,p_L)=\begin{pmatrix} n \\ x_1\ldots x_L \end{pmatrix} p_1^{x_1}\cdot\ldots\cdot p_L^{x_L}\mathbb{I}_{\sum x_l=n} \tag{5.34}$$

where $\sum_{l=1}^{L}p_l=1$ and $p_l\geq 0,\ l=1,\ldots,L$.

As suggested in the literature on gaussian mixtures, in the following we assume a multinomial prior distribution for the completing variable ν: $V\sim f_V(\nu)=\mathcal{M}_L(1,p_1,\ldots,p_L)$.

Definition 5.4.2. *(Dirichlet distribution)*

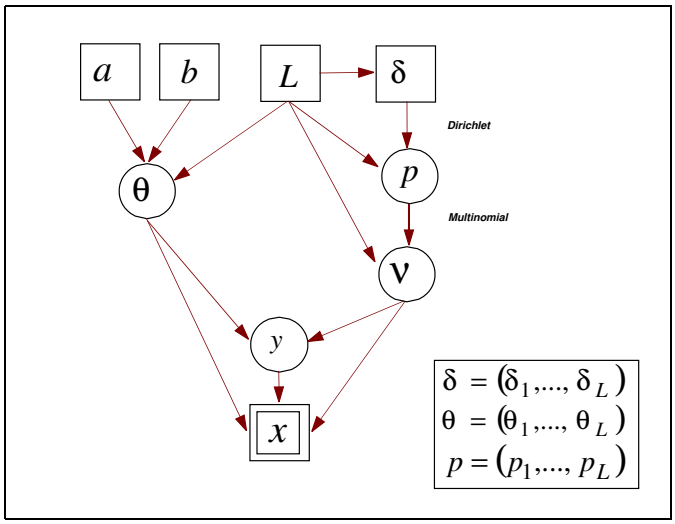

Figure 5.3: DAG of the Bayesian hierarchical model for inference on stable mixtures. Note that the completing variable ν is not observable. Thus, two levels of completion, y and ν, are needed for a stable mixture model.

The random variable $X = (x_1, \ldots, x_L)$ *has* L *dimensional Dirichlet distribution if its density function is*

$$\mathcal{D}(\delta_1, ..., \delta_L) = \frac{\Gamma(\delta_1 + ... + \delta_L)}{\Gamma(\delta_1) \cdot \ldots \cdot \Gamma(\delta_L)} x_1^{\delta_1 - 1} \cdot \ldots \cdot x_L^{\delta_L - 1} \mathbb{I}_S(x) \tag{5.35}$$

where $\delta_l \geq 0, l = 1, \ldots, L$ *and* $S = \{x = (x_1, \ldots, x_L) \in \mathbb{R}^L | \sum x_l = 1, x_l > 0 \; l = 1, \ldots, L\}$ *is the simplex of* $\mathbb{R}^L$.

We assume that the parameters of the discrete part of the mixture distribution has the standard conjugate Dirichlet prior: $(p_1, \ldots, p_L) \sim \mathcal{D}_L(\delta_1, \ldots, \delta_L)$, with hyperparameters $\delta_1 = \ldots = \delta_L = \frac{1}{L}$.
Observing n independent values, $\mathbf{x} = (x_1, \ldots, x_n)$, from a stable mixture, the likelihood and the completed likelihood are respectively

$$L(\mathbf{x}, \mathbf{y}|\theta, p) = \prod_{i=1}^{n} \sum_{l=1}^{L} p_l \int_{-1/2}^{1/2} f(x_i, y_i|\theta_l) dy_i \tag{5.36}$$

$$L(\mathbf{x}, \mathbf{y}, \nu|\theta, p) = \prod_{i=1}^{n} \prod_{l=1}^{L} (p_l f(x_i, y_i|\theta_l))^{\nu_{il}} \tag{5.37}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ and $\nu = (\nu_1, \ldots, \nu_n)$ are respectively the auxiliary variable and the allocation variable vectors and $\theta = (\theta_1, \ldots, \theta_L)$ and $p = (p_1, \ldots, p_L)$ are the mixture's parameters vectors. From the completed likelihood and from the priors it follows that the complete posterior distribution of the Bayesian mixture model is:

$$\pi(\theta, p|\mathbf{x}, \mathbf{y}, \nu) \propto \prod_{i=1}^{n} \left\{ \prod_{l=1}^{L} \left(f(x_i, y_i|\theta_l) \right)^{\nu_{il}} \pi(\nu_i) \right\} \pi(\theta)\pi(p). \tag{5.38}$$

Bayesian inference on the mixture parameters requires the calculation of the expected value from the posterior distribution. A closed form solution of this integration problem does not exist, thus numerical methods are needed. The introduction of auxiliary variables, that are not observable, simplifies inference for mixtures and also suggests the way to approximate numerically the problem. In fact the auxiliary variables can be replaced by simulated values and the simulated completed likelihood can be used for calculating the posterior distributions. Furthermore in order to approximate numerically the posterior means is necessary to perform simulations from the posterior distributions of the parameters and to average the simulated values.

5.4.3 The Gibbs Sampler for Mixtures of Stable Distributions

Gibbs sampling allows us to simulate from the posterior distribution avoiding computational difficulties due to the dimension of the parameter vector. Due to the ergodicity of the Markov chain generated by the Gibbs sampler, the choice of the initial values is arbitrary. In particular we choose to simulate them from the prior. The steps of the Gibbs sampler for a mixture model can be grouped in: simulation of the full conditional distributions and augmentation by the completing variables

(i) *Simulate initial values:* $\nu_i^{(0)}, y_i^{(0)}, \quad i = 1, \ldots, n$ *and* $p^{(0)}$ *respectively from*

$$\nu_i^{(0)} \sim \mathcal{M}_L(1, p_1, \ldots, p_L) \tag{5.39}$$

$$y_i^{(0)} \sim f(y_i|\theta, \nu, x_i) \propto \exp\{1 - \left| \frac{z_i}{\tau_{\alpha,\beta}(y_i)} \right|^{\alpha/(\alpha-1)} \} \left| \frac{z_i}{\tau_{\alpha,\beta}(y_i)} \right|^{\alpha/(\alpha-1)} \tag{5.40}$$

$$p^{(0)} \sim \mathcal{D}_L(\delta, \ldots, \delta). \tag{5.41}$$

(ii) *Simulate from the full conditional posterior distributions*

$$\pi(\theta_l|\theta_{-l}, p, \mathbf{x}, \mathbf{y}, \mathbf{v}) \propto \prod_{i=1}^{n} \{f(x_i, y_i|\theta_l)\, p_l\}^{\nu_{il}}\ \pi(\theta_l) \quad l = 1, \ldots, L \tag{5.42}$$

$$\pi(p_1, \ldots, p_L|\theta, \mathbf{x}, \mathbf{y}, \mathbf{v}) = \mathcal{D}(\delta + n_1(\nu), \ldots, \delta + n_L(\nu)) \tag{5.43}$$

(iii) *Update the completing variables*

$$\pi(y_i|\theta, p, \mathbf{x}, \mathbf{y}_{-i}, \mathbf{v}) \propto \exp\left\{1 - \left|\frac{x_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\right\} \left|\frac{x_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)} \tag{5.44}$$

$$\pi(\nu_i|\theta, p, \mathbf{x}, \mathbf{y}, \mathbf{v}_{-i}) = \mathcal{M}_L(1, p_1^*, \ldots, p_L^*) \tag{5.45}$$

for $i = 1, \ldots, n$, where

$$z = \frac{x - \delta}{\sigma}$$

$$n_l(\nu) = \sum_{i=1}^{n} \nu_{il}\,, \quad l = 1, \ldots, L$$

$$p_l^* = \frac{p_l f(x_i, y_i|\theta_l)}{\sum_{l=1}^{L} f(x_i, y_i|\theta_l) p_l}\,, \quad l = 1, \ldots, L.$$

Steps (5.43) and (5.45) of the Gibbs sampler are proved in *Appendix D*. Observe that simulations from the conditional posterior distribution of Eq.(5.43) can be obtained by running the Gibbs sampler given in equations (5.20)-(5.23), conditionally to the value of the completing variable ν. To simulate from the Dirichlet posterior distribution given in Eq. (5.43), we use the algorithm proposed by Casella and Robert [35], while to draw value from the multinomial posterior distribution of Eq. (5.45), we use the algorithm proposed by Fishman [18].

In Examples 5.4.1 and 5.4.2, we verify the efficiency of the Gibbs sampler on some test samples simulated from stable mixtures. For each mixture's component we assume the joint prior distribution given in equation (5.29). Furthermore, for the shake of simplicity, we consider $L = 2$. Because of the quite irregular form of the density $f(x_i, y_i|\theta_l)$, during the MCMC based estimation, some computational difficulties were encountered in evaluating the probability p_l^* of each mixture's component. Thus we introduce the following useful reparameterisation and approximation

$$q_l^* = \begin{cases} \frac{e^{g_1(x_i,y_i)-g_2(x_i,y_i)}}{e^{g_1(x_i,y_i)-g_2(x_i,y_i)}+1} & \text{if} \quad g_1 > g_2 \quad \text{or} \quad g_1 \leq g_2 \\ 1 & \text{if} \quad g_1 \gg g_2 \\ 0 & \text{otherwise} \end{cases}$$

where $g_l(x_i, y_i) = log(p_l\, f(x_i, y_i|\theta_l))$, with $l = 1, 2$.

Table 5.2: Numerical results - Ergodic Averages over 15,000 Gibbs realisations.

	Dataset: $0.5S_{1.7}(0.3,1,1) + 0.5S_{1.3}(0.5,30,1)$				
Par.	True Value	Starting Value	Estimate$^{(*)}$	Std.Dev.	Acc. Rate
α_1	1.7	1.9	1.66	0.09	0.32
α_2	1.3	1.9	1.36	0.07	0.41
β_1	0.3	0.8	0.28	0.09	0.41
β_2	0.5	0.8	0.37	0.10	0.42
p_1	0.5	0.4	0.52	0.02	-
	Dataset: $0.5S_{1.3}(0.3,1,1) + 0.5S_{1.3}(0.8,30,1)$				
Par.	True Value	Starting Value	Estimate$^{(**)}$	Std.Dev.	Acc. Rate
α	1.3	1.7	1.25	0.08	0.23
β_1	0.3	0.5	0.15	0.03	0.11
β_2	0.8	0.5	0.95	0.05	0.13
p_1	0.5	0.5	0.75	0.09	-
p_2	0.5	0.5	0.25	0.09	-

(*) *Time (sec):9249*
(**) *Time (sec):9525*

Example 5.4.1 - *(α-Stable Mixture with varying β and α)*
In this example, we apply the Gibbs sampler to a synthetic dataset of 1,000 observations generated from the stable mixture: $0.5S_{1.7}(0.3, 1, 1)+0.5S_{1.3}(0.5, 30, 1)$. Fig. 5.29 in *Appendix E* exhibits the dataset. In the M.-H. step of the Gibbs sampler, we set v=0.0001 for β and v=0.005 for α.
Result are briefly represented in Table 5.4.3 and graphically described in Figg. 5.30-5.33. Note that the presence in the mixture model, of distribution with different tails behaviour, produces some problems in the convergence of the ergodic averages, due to the label switching of the observations.

□

Example 5.4.2 - *(α-Stable Mixture with constant α and varying β)*
In this experiment we keep α fixed over the mixture components. First we generate 1,000 observations from the following mixture $0.5S_{1.3}(0.3, 1, 1) + 0.5S_{1.3}(0.8, 30, 1)$. Secondly we apply the Gibbs sampler for stable mixture and obtain the results given in Table 5.4.3. The graphical description of the results is in Figg. 5.35-5.38 of *Appendix E* and exhibits a more appreciable mixing of the chain associated to the Gibbs sampler.

□

To conclude this section, we remark that in developing the Gibbs sampler for α-Stable mixtures and also in previous Monte Carlo experiments the number of components of the mixture is assumed to be known. Thus our research framework can be extended in order to make inference on the number of components. For example, *Reversible Jump MCMC* (RJMCMC) or *Birth and Death MCMC* (BDMCMC) could be applied in this context.

5.5 Application to Financial Data

Introducing two level of auxiliary variables in the stable mixture model allows us to infer all the parameters of mixture from the data. Gaussian distribution is usually assumed in modelling financial time series, but it performs poorly when data are heavy-tailed and skewed. Moreover the assumption of unimodal distribution becomes too restrictive for some financial time series. In this section, we illustrate how stable mixtures may result particularly useful in modelling different kind of financial variables and present estimates obtained with the MCMC based inferential technique proposed in the previous section.

Example 5.5.1 - *(Stock Market)*
In this example we analyse the return rate of the S&P500 composite index from 01/01/1990 to 27/01/2003. The return on the index is defined as: $r_t = (p_t - p_{t-1})/p_{t-1}$. Alternatively, logarithmic returns could be used. The number of observations is 3410. Fig. 5.39 shows the data histogram and the best normal

which is possible to estimate. The QQ-plot in Fig. 5.40 reveals that data are not normally distributed. We apply the Gibbs sampler for α-Stable mixtures to this dataset. The result is in Tab. 5.5. Parameter estimates are ergodic averages over the last 10,000 values of the 15,000 Gibbs sampler realisations. Note that index return distribution has tails heavier than Gaussian, because $\hat{\alpha} = 1.674$.

□

Table 5.3: Parameter Estimates on S&P 500 Index daily returns.

	Starting Value	Estimate	Std.Dev.	Acc. Rate
α	1.8	1.674	0.005	0.2
β	0.2	0.159	0.004	0.1
σ	0.01	0.070	0.002	-
δ	0.0001	0.000091	0.0001	0.1

Example 5.5.2 - *(Bond Market)*
Our second dataset (source: *DataStream*) contains daily price returns on the J.P. Morgan's indexes concerning following countries: France, Germany, Italy, United Kingdom, USA and Japan, between 01/01/1988 and 13/01/2003. Denoting with p_t the price index at time t. The return on the index is defined as: $r_t = (p_t - p_{t-1})/p_{t-1}$. Fig.5.41 in *Appendix E* exhibits jointly the histogram, the best Gaussian approximation and the density line of the returns distribution. All time series exhibit a certain degree of kurtosis and skewness. Estimation result on the J.P. Morgan Great Britain index is in Tab. 5.5.2.

Table 5.4: Parameter Estimates on JP Morgan - Great Britain index daily returns.

	Starting Value	Estimate	Std.Dev.	Acc. Rate
α	1.5	1.95	0.004	0.2
β	0.02	0.013	0.001	0.2
σ	0.01	0.270	0.003	-
δ	0.005	0.0062	0.02	0.1

□

Example 5.5.3 - *(3 Months Euro-Deposits Interest Rate)*
Our third dataset (source: *DataStream*) contains daily 3-Months Interest Rates on Euro-Deposits concerning following countries: France, Germany, Italy, United Kingdom, USA and Japan, between 01/01/1988 and 13/01/2003. Fig.5.43 in *Appendix E* exhibits jointly the histogram, the best Gaussian approximation and the density line of the returns distribution. Quite all time series of this dataset exhibit multimodality.
Estimation result on the 3-month Interest Rate for France is in Tab. 5.5.3

Table 5.5: Parameter Estimates on Interest Rates - France. Two components α-stable mixture.

	Starting Value	Estimate	Std.Dev.	Acc. Rate
α_1, α_2	1.5	1.2	0.003	0.15
β_1	0.01	0.02	0.001	0.1
β_2	0.01	0.04	0.001	0.1
σ_1	1.5	0.307	0.002	-
σ_2	1.5	0.873	0.001	-
δ_1	4	3.012	0.02	0.1
δ_2	10	7.301	0.03	0.1

□

5.6 Conclusion

In this work we propose a α-Stable mixture model. As result in the literature from many empirical evidences, α-Stable distributions are particularly adapted for modelling financial variables. Moreover some financial empirical studies evidence that mixture models are needed in many cases, due to the presence of multi-modality in the asset returns distribution. We chose Bayesian inference due to the flexibility of the approach, that allows to simultaneously estimate all the parameters of the model. Furthermore we introduce a suitable reparameterisation of the α-Stable mixture in order to perform Bayesian inference. The proposed approach to α-Stable mixture models estimation is quite general and worked well in our simulation analysis, but

it needs much more evaluation, with a particular attention to the case of symmetric stable mixtures. Furthermore the Bayesian approach used in this work allows to perform goodness of fit tests and also to use RJMCMC and BDMCMC techniques in order to make inference on the number of components of the mixture.

Appendix A - Bayesian Inference for Stable Distributions

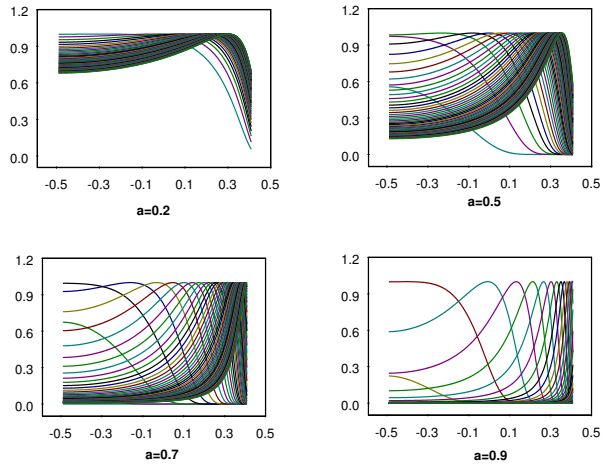

Figure 5.4: The density function of y for different values of x (0.04,...,0.14) and of α (0.2,...,0.9).

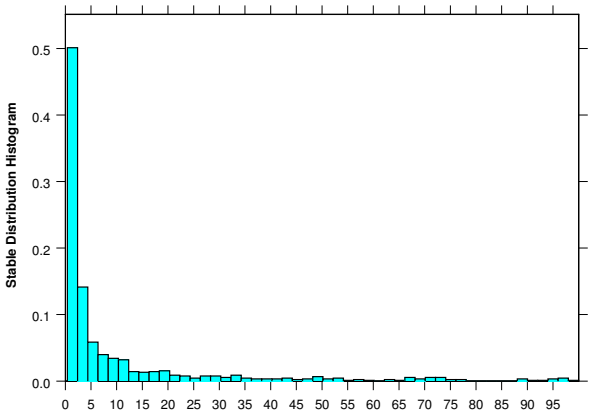

Figure 5.5: Simulated dataset, 1,000 values from $S_{0.5}(0.9, 0, 1)$.

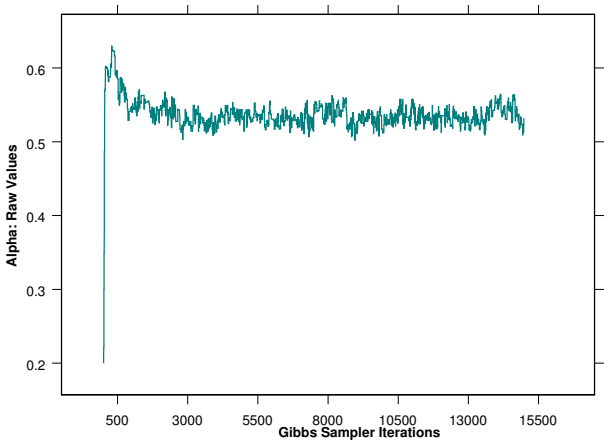

Figure 5.6: Gibbs sampler realisations, characteristic exponent α.

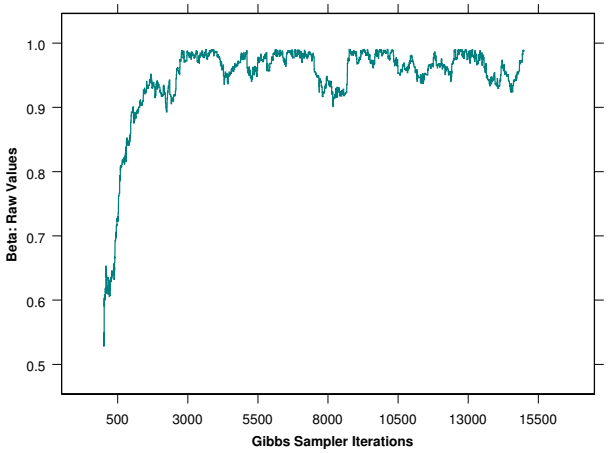

Figure 5.7: Gibbs sampler realisations, asymmetry parameter β.

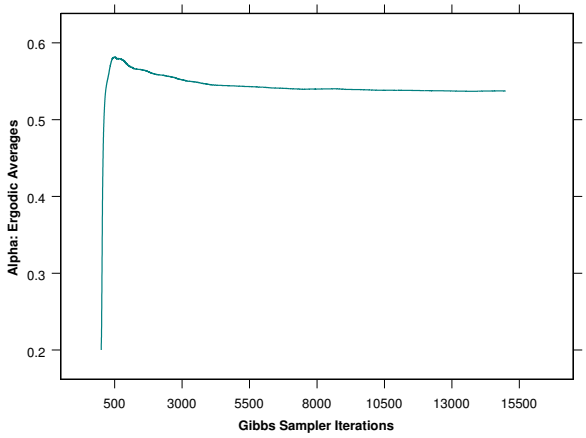

Figure 5.8: Gibbs sampler, ergodic averages for the characteristic exponent α.

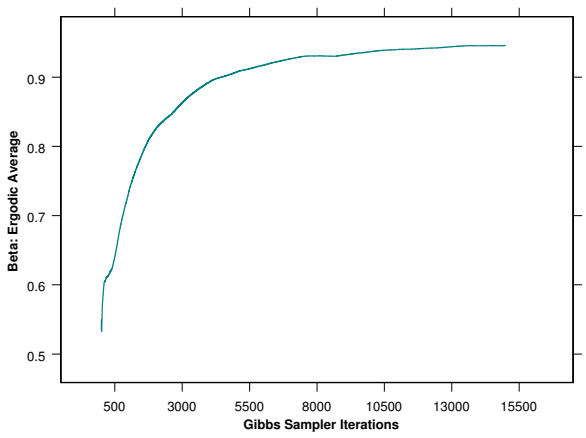

Figure 5.9: Gibbs sampler, ergodic averages for the asymmetry parameter β.

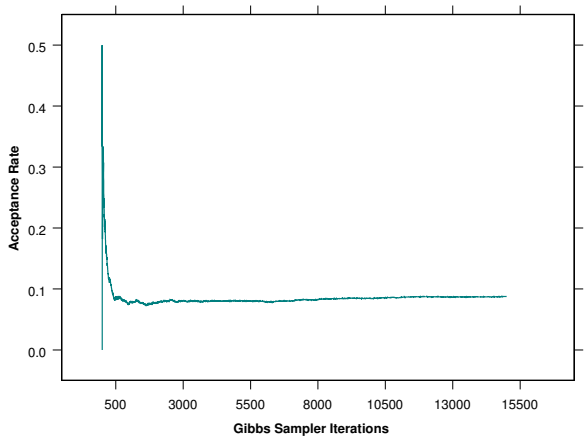

Figure 5.10: Gibbs sampler, acceptance rate of the M.-H. step, parameter α

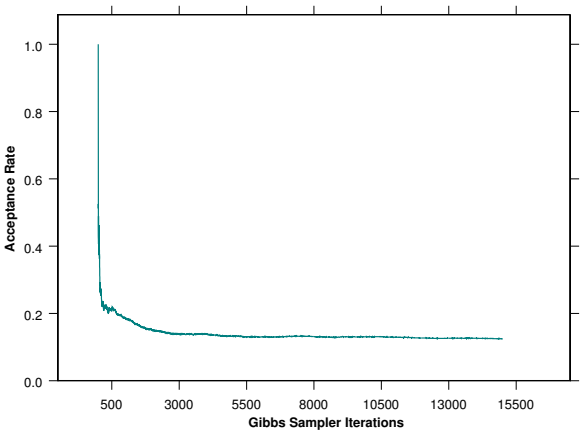

Figure 5.11: Gibbs sampler, acceptance rate of the M.-H. step, parameter β

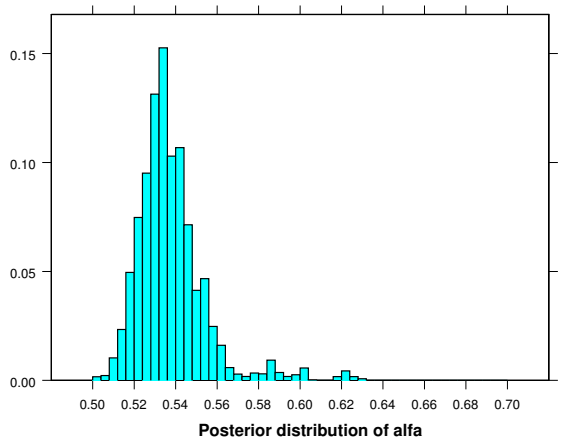

Figure 5.12: Gibbs sampler, histogram of the posterior $\pi(\alpha|\mathbf{x})$

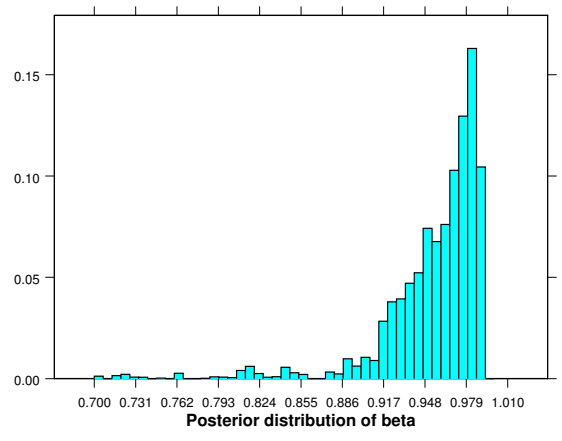

Figure 5.13: Gibbs sampler, histogram of the posterior $\pi(\beta|\mathbf{x})$

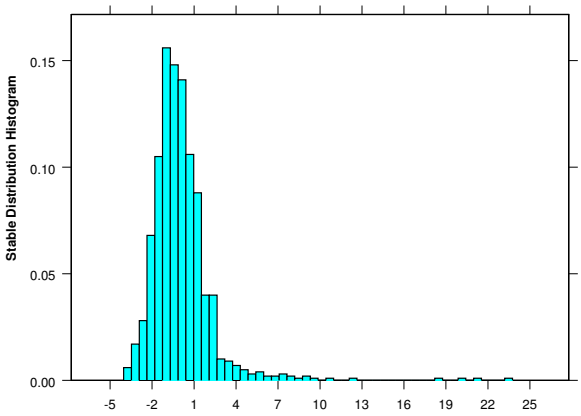

Figure 5.14: Simulated dataset, 1,000 values from $S_{1.5}(0.9, 0, 1)$.

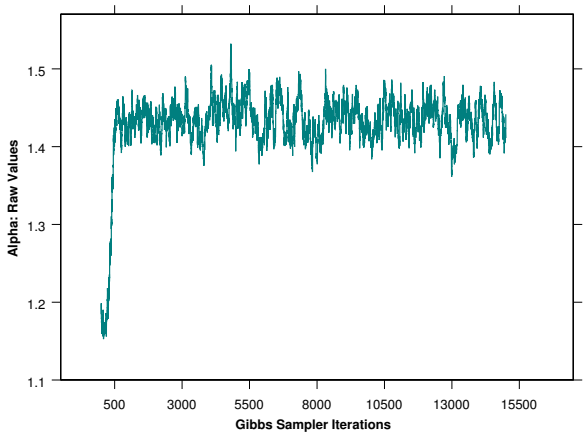

Figure 5.15: Gibbs sampler realisations, characteristic exponent α.

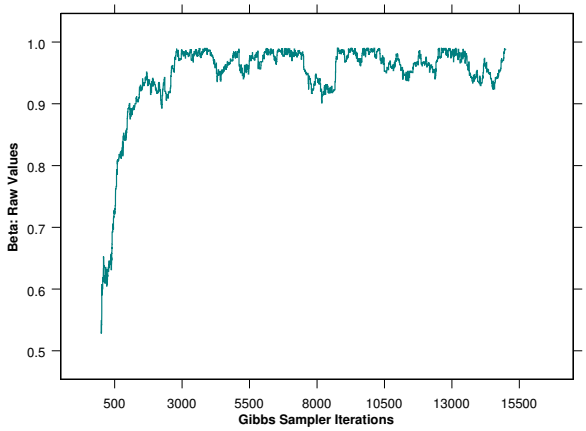

Figure 5.16: Gibbs sampler realisations, asymmetry parameter β.

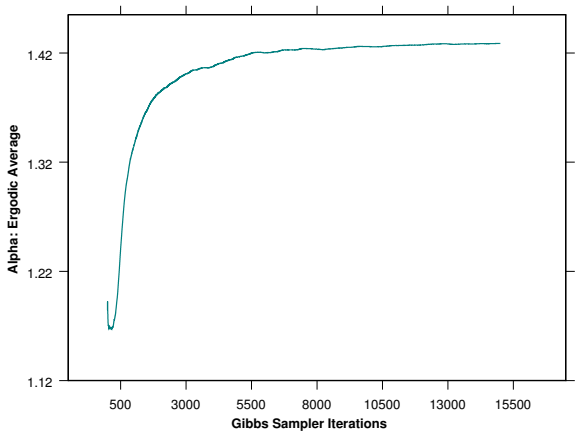

Figure 5.17: Gibbs sampler, ergodic averages for the characteristic exponent α.

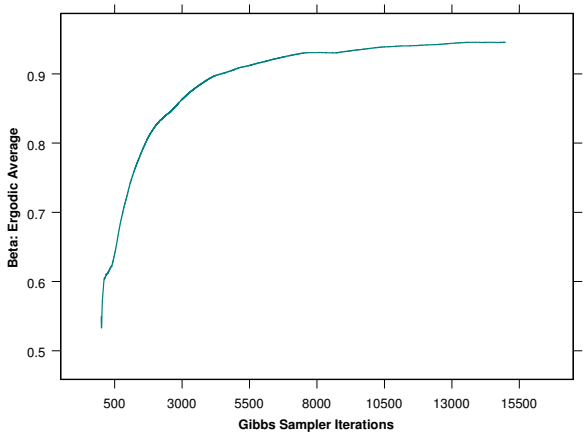

Figure 5.18: Gibbs sampler, ergodic averages for the asymmetry parameter β.

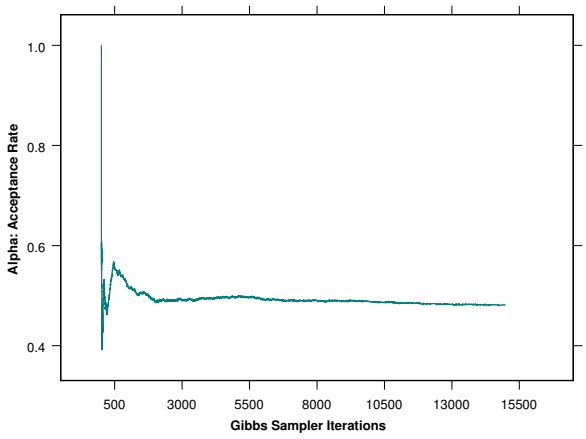

Figure 5.19: Gibbs sampler, acceptance rate of the M.-H. step, parameter α

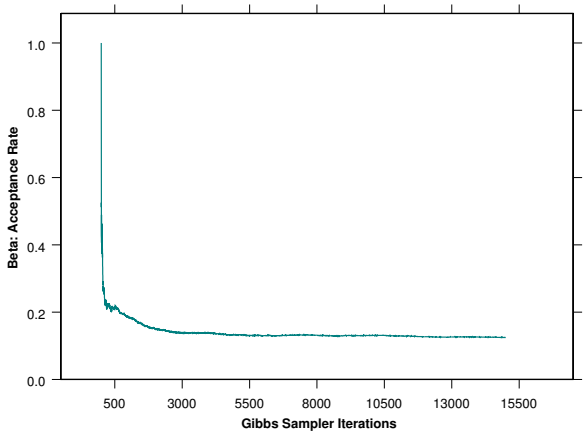

Figure 5.20: Gibbs sampler, acceptance rate of the M.-H. step, parameter β

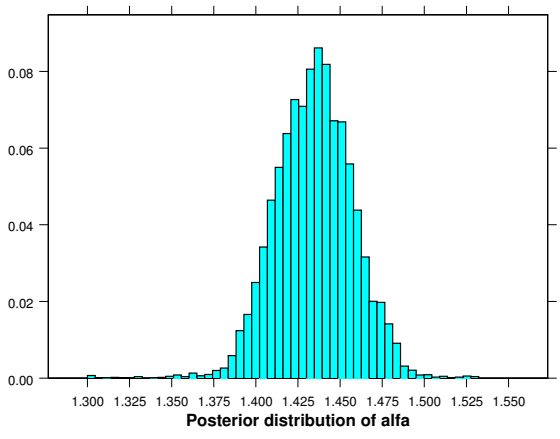

Figure 5.21: Gibbs sampler, histogram of the posterior $\pi(\alpha|\mathbf{x})$

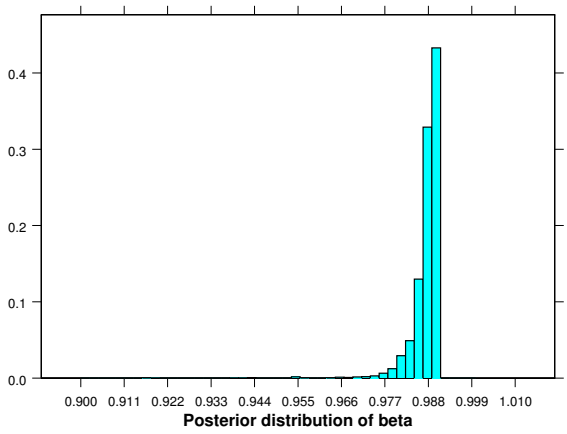

Figure 5.22: Gibbs sampler, histogram of the posterior $\pi(\beta|\mathbf{x})$

Appendix B - Proposal Distributions for the Metropolis-Hastings Algorithm

The shape of the stable distribution and the presence of skewness suggest us to use a Beta distribution $\mathcal{B}e(a,b)$ as proposal for the Metropolis-Hastings algorithm

$$\alpha|\alpha_{k-1} \sim \mathcal{B}e(a,b) = \frac{1}{B(a,b)}\alpha^{a-1}(1-\alpha)^{b-1}\mathbb{I}(\alpha)_{(0,1)}. \tag{5.46}$$

We assume that the mean of the distribution is equal to the $(k-1)$-th value of the M.-H. chain and set exogenously the variance equal to v. Through the parameter v it is thus possible to control the acceptance rate of the Metropolis-Hastings algorithm. When $\alpha \in (0,1)$ the value of the parameters of the proposal is

$$\begin{cases} \frac{a}{a+b} = \alpha_{k-1} \\ \frac{ab}{(a+b)^2(1+a+b)} = v \end{cases} \quad \Leftrightarrow \quad \begin{cases} a = \frac{\alpha_{k-1}^2(1-\alpha_{k-1}) - v\,\alpha_{k-1}}{v} \\ b = a\frac{1-\alpha_{k-1}}{\alpha_{k-1}} \end{cases}$$

where α_{k-1} is the $(k-1)$-th value of the M.-H. chain and v is the variance. In addition to the previous system of equations, also the positivity constraint on the Beta's parameters: $a > 0$ and $b > 0$ must hold. Thus at each iteration of the M.-H. algorithm the following constraint must be satisfied

$$\alpha_{k-1} \in \left(\frac{3-v}{2} - \frac{\sqrt{v^2-8v+1}}{2}, \frac{3-v}{2} + \frac{\sqrt{v^2-8v+1}}{2}\right). \tag{5.47}$$

When $\alpha \in (1,2]$ we use a translated Beta distribution

$$\alpha|\alpha_{k-1} \sim \mathcal{B}e(a,b) = \frac{1}{B(a,b)}(\alpha-1)^{a-1}(2-\alpha)^{b-1}\mathbb{I}(\alpha)_{(1,2)}. \tag{5.48}$$

By imposing the usual constraints on the mean and the variance we obtain the values of the proposal's parameters

$$\begin{cases} \frac{2a+b}{a+b} = \alpha_{k-1} \\ \frac{ab}{(a+b)^2(1+a+b)} = v \end{cases} \quad \Leftrightarrow \quad \begin{cases} a = \frac{(\alpha_{k-1}-1)^2(2-\alpha_{k-1}) - v\,(\alpha_{k-1}-1)}{v} \\ b = a\frac{2-\alpha_{k-1}}{(\alpha_{k-1}-1)} \end{cases}$$

Also in this case the positivity constraints on the Beta's parameters must be considered. We proceed in a similar way for the proposal distribution of the skewness parameter β.

Appendix C - Mixtures of Stable Distributions

C.1 Mixtures with varying α

Observe that in all dataset exhibited in the histograms, N=100,000 values from right skewed ($\beta = 1$) standard stables have been simulated.

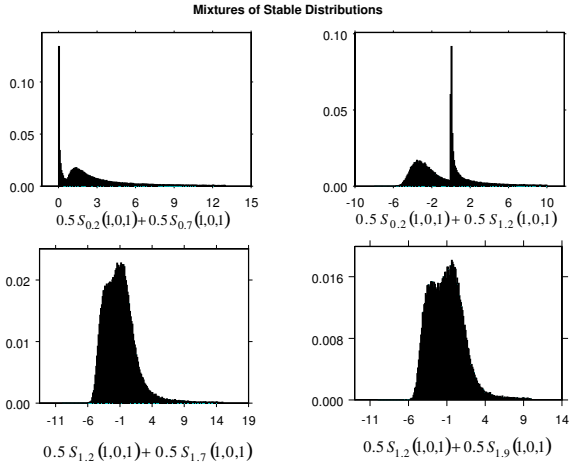

Figure 5.23: Stables mixtures, with equally weighted components.

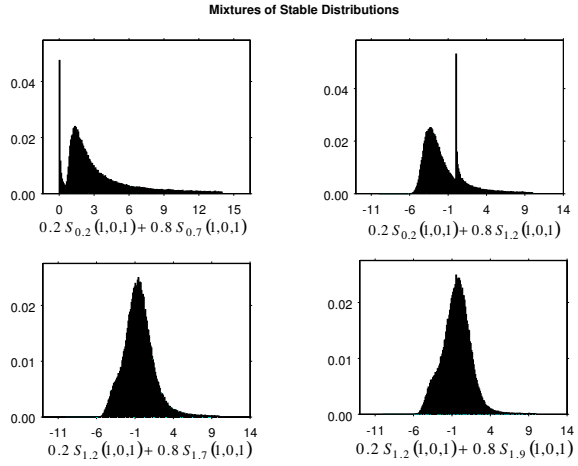

Figure 5.24: Stables mixtures, with unequally weighted components.

C.2 Mixtures with varying β

Simulated samples of $N = 100,000$ stable values are exhibited in the following histograms. In all the samples the location and the scale parameters of the mixture are: $\delta_1 = 1$, $\delta_2 = 40$, $\sigma_1 = \sigma_2 = 7$.

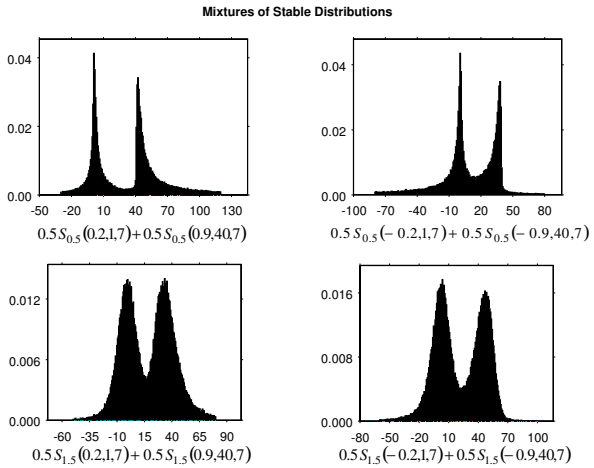

Figure 5.25: Stables mixtures, with equally weighted components.

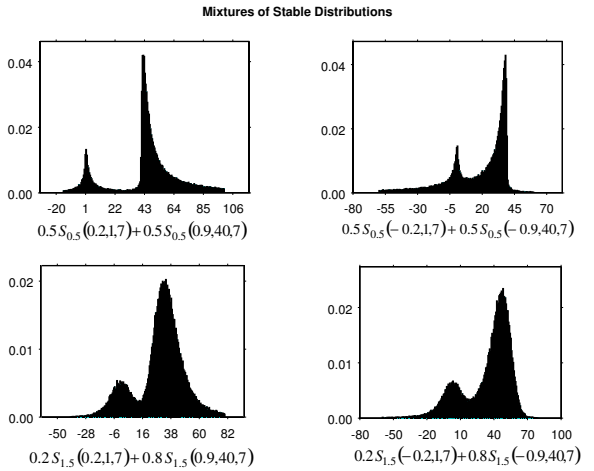

Figure 5.26: Stables mixtures, with unequally weighted components.

C.3 Mixtures with varying σ

Simulated samples of N=100,000 stable values are exhibited in the following histograms. In all the samples the location and the skewness parameters of the mixture are: $\delta_1 = 1$, $\delta_2 = 40$, $\beta_1 = \beta_2 = 1$.

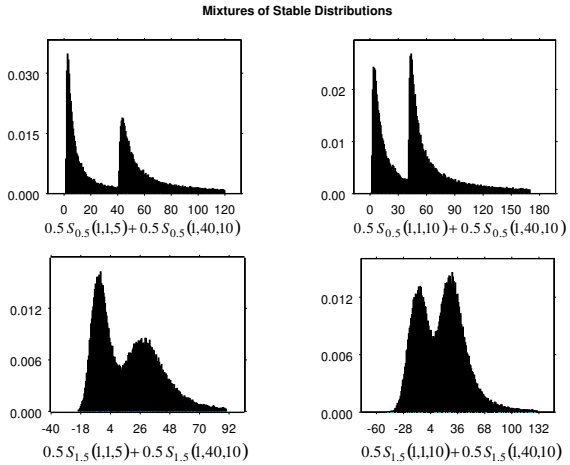

Figure 5.27: Standard stables mixtures, with equally weighted components.

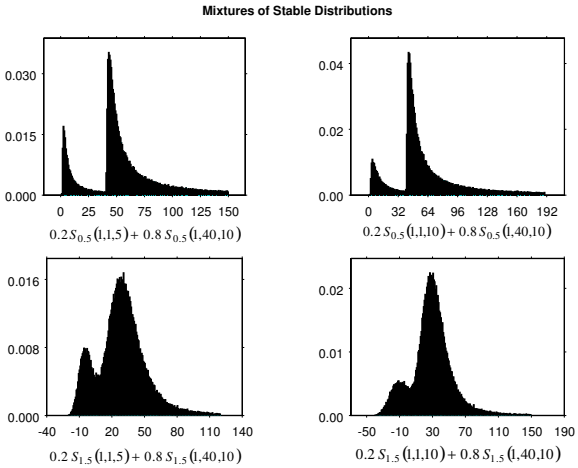

Figure 5.28: Standard stables mixtures, with unequally weighted components.

Appendix D - The Gibbs Sampler for a Stable Distributions Mixture

Proof. (Allocation Probabilities Posterior Distribution)

The posterior distribution of probabilities $(p_1, \ldots, p_L)$, given in Eq. (5.43), is a *Dirichlet* and is derived in the following

$$
\begin{aligned}
&\pi(p_1, \ldots, p_L|\theta, \mathbf{x}, \mathbf{y}, \nu) = \frac{L(\mathbf{x}, \mathbf{y}, \nu|\theta, p)\pi(\theta)\pi(p)}{\int L(\mathbf{x}, \mathbf{y}, \nu|\theta, p)\pi(\theta)\pi(p)dp} = \\
&= \frac{\prod_{i=1}^{n}\prod_{l=1}^{L}\left(p_l f(x_i, y_i|\theta_l)\right)^{\nu_{il}}\pi(\theta)\pi(p)}{\int \prod_{i=1}^{n}\prod_{l=1}^{L}(p_l f(x_i, y_i|\theta_l))^{\nu_{il}}\pi(\theta)\pi(p)dp} = \\
&= \frac{\prod_{i=1}^{n}\prod_{l=1}^{L} p_l^{\nu_{il}}\pi(p)}{\int \prod_{i=1}^{n}\prod_{l=1}^{L} p_l^{\nu_{il}}\frac{\Gamma(\delta)\cdots\Gamma(\delta)}{\Gamma(L\,\delta)}p_1^{\delta-1}, \cdots, p_L^{\delta-1}dp_1, \cdots, dp_L} = \\
&= \frac{\prod_{i=1}^{n}\prod_{l=1}^{L} p_l^{\nu_{il}}\pi(p)}{\int \prod_{i=1}^{n}\frac{\Gamma(\delta)\cdots\Gamma(\delta)}{\Gamma(L\,\delta)}\prod_{l=1}^{L} p_l^{\nu_{il}+\delta-1}dp_1\cdots dp_L} = \\
&= \frac{\prod_{i=1}^{n}\prod_{l=1}^{L} p_l^{\nu_{il}}\pi(p)}{\int \prod_{l=1}^{L} p_l^{\sum_{i=1}^{n}\nu_{il}+\delta-1}\frac{\Gamma(\delta)\cdots\Gamma(\delta)}{\Gamma(L\,\delta)}dp_1, \cdots, dp_L} = \\
&= \frac{\Gamma\left(\delta + \sum_{i=1}^{n}\nu_{i1}\right)\cdots\Gamma\left(\delta + \sum_{i=1}^{n}\nu_{iL}\right)}{\Gamma\left(\delta\, L + \sum_{i=1}^{n}(\nu_{i1} + \ldots + \nu_{iL})\right)} p_1^{\sum_{i=1}^{n}\nu_{i1}+\delta-1}\cdots p_L^{\sum_{i=1}^{n}\nu_{iL}+\delta-1} = \\
&= \mathcal{D}_L(\sum_{i=1}^{n}\nu_{i1} + \delta, \ldots, \sum_{i=1}^{n}\nu_{iL} + \delta) = \\
&= \mathcal{D}_L(n_1(\nu) + \delta, \ldots, n_L(\nu) + \delta)
\end{aligned}
\tag{5.49}
$$

where $n_l(\nu) = \sum_{i=1}^{n}\nu_{il}$, with $l = 1, \ldots, L$.

□

Proof. (Allocation Variables Posterior Distribution)

The posterior distribution of the allocation variables given in Eq.(5.45) is a *Multinomial* and follows from

$$
\begin{aligned}
\pi(\nu_1,\ldots,\nu_n|\theta,p,\mathbf{x},\mathbf{y}) &= \frac{L(\mathbf{x},\mathbf{y},\nu|\theta,p)\pi(\theta)\pi(p)}{\int L(\mathbf{x},\mathbf{y},\nu|\theta,p)\pi(\theta)\pi(p)d\nu} = \\
&= \frac{\prod_{i=1}^{n}\left\{\prod_{l=1}^{L}(f(x_i,y_i|\theta_l))^{\nu_{il}}\prod_{l=1}^{L}p_l^{\nu_{il}}\right\}\pi(\theta)\pi(p)}{\int\prod_{i=1}^{n}\prod_{l=1}^{L}(f(x_i,y_i|\theta_l))^{\nu_{il}}\prod_{l=1}^{L}p_l^{\nu_{il}}\pi(\theta)\pi(p)d\nu} = \\
&= \prod_{i=1}^{n}\frac{\prod_{l=1}^{L}(f(x_i,y_i|\theta_l)p_l)^{\nu_{il}}}{\int\prod_{l=1}^{L}(f(x_i,y_i|\theta_l)p_l)^{\nu_{il}}\,d\nu_i} = \\
&= \prod_{i=1}^{n}\prod_{l=1}^{L}\left(\frac{f(x_i,y_i|\theta_l)p_l}{\sum_{l=1}^{L}f(x_i,y_i|\theta_l)p_l}\right)^{\nu_{il}} = \\
&= \prod_{i=1}^{n}\mathcal{M}_L(1,p_1^*,\ldots,p_L^*)
\end{aligned}
\tag{5.50}
$$

where $p_l^* = \frac{f(x_i,y_i|\theta_l)p_l}{\sum_{l=1}^{L}f(x_i,y_i|\theta_l)p_l}$, for $l = 1,\ldots,L$.

□

Appendix E - Bayesian Inference for Stable Distributions Mixtures

E.1 Mixtures with varying α and β

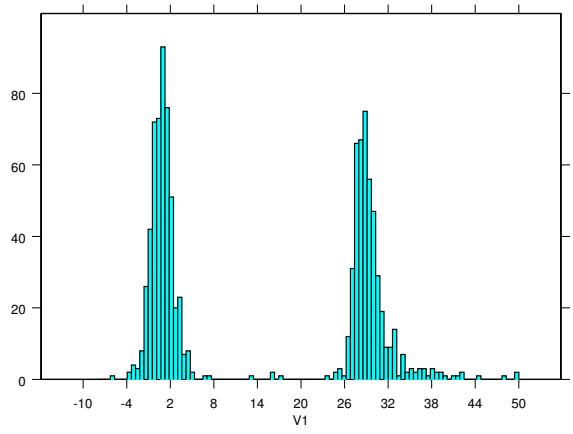

Figure 5.29: Simulated dataset, 1,000 values from $0.5S_{1.7}(0.3, 1, 1)+0.5S_{1.3}(0.5, 30, 1)$

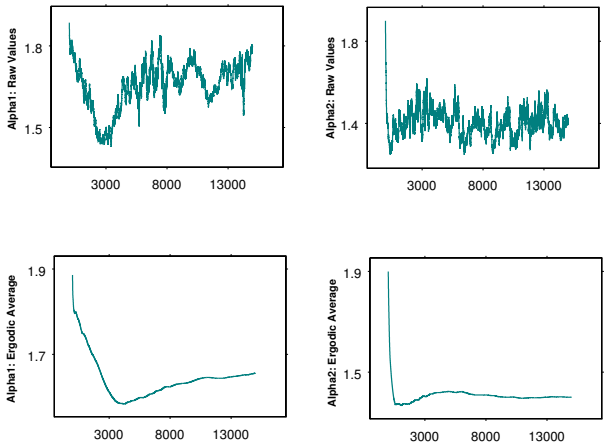

Figure 5.30: Gibbs sampler realisations and ergodic averages for α_1 and α_2.

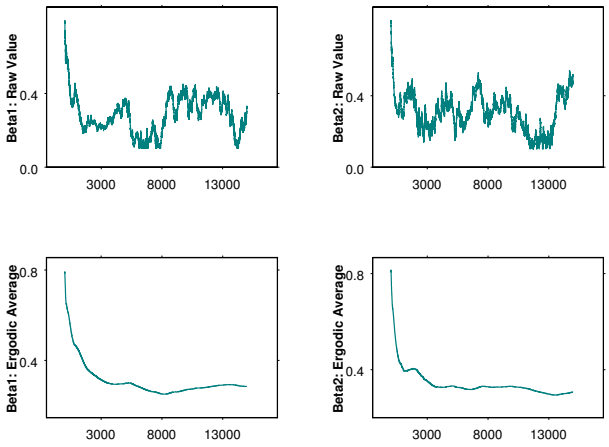

Figure 5.31: Gibbs sampler realisations and ergodic averages for β_1 and β_2.

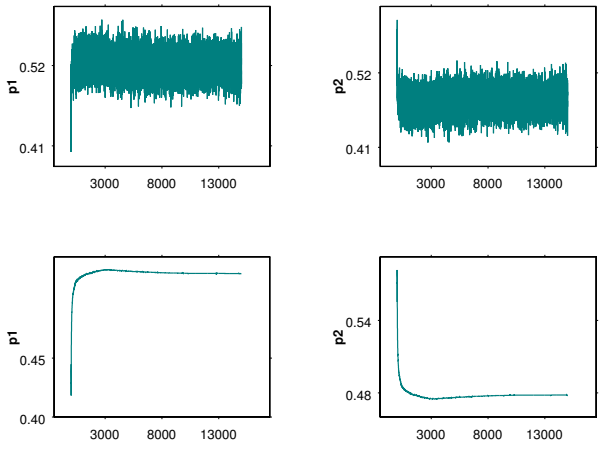

Figure 5.32: Gibbs sampler realisations and ergodic averages for p_1 and p_2.

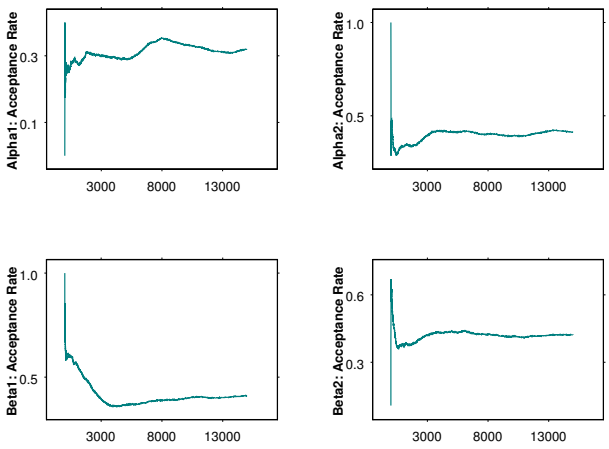

Figure 5.33: Acceptance rates for α_1, α_2, β_1 and β_2

E.2 Mixtures with fixed α and varying β

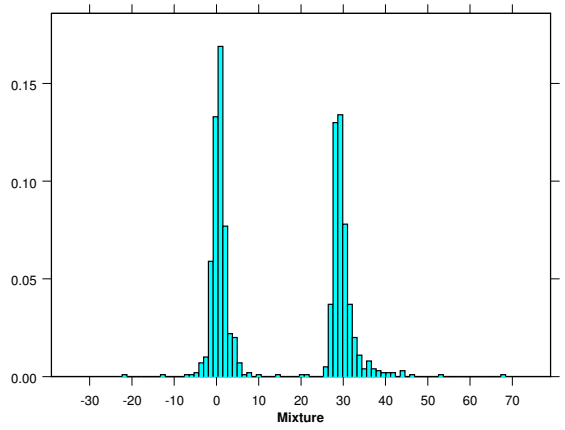

Figure 5.34: Simulated dataset, 1,000 values from $0.7S_{1.3}(0.3, 1, 1)+0.3S_{1.3}(0.8, 30, 1)$

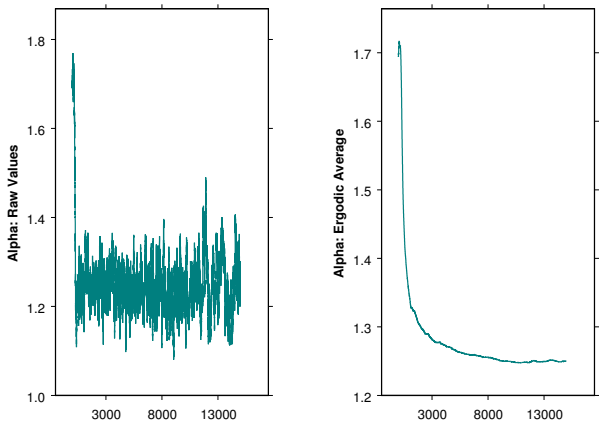

Figure 5.35: Gibbs sampler realisations and ergodic averages for α

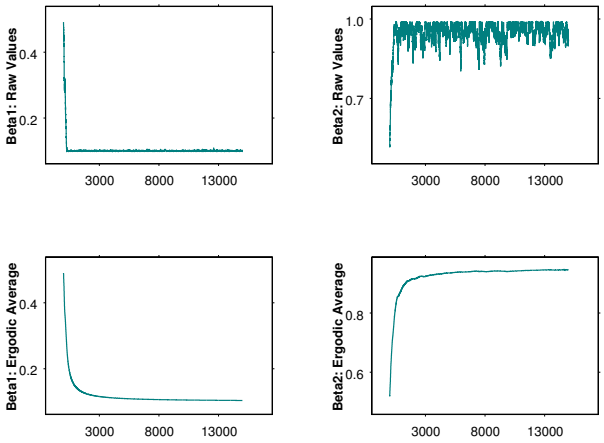

Figure 5.36: Gibbs sampler realisations and ergodic averages for β_1 and β_2.

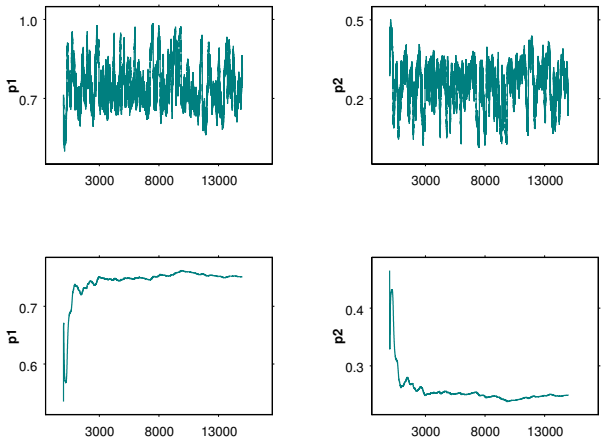

Figure 5.37: Gibbs sampler realisations and ergodic averages for p_1 and p_2.

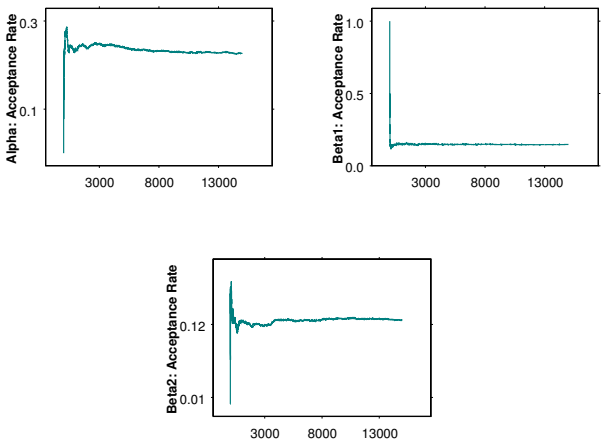

Figure 5.38: Acceptance rates for α, β_1 and β_2

E.3 Stock price indexes

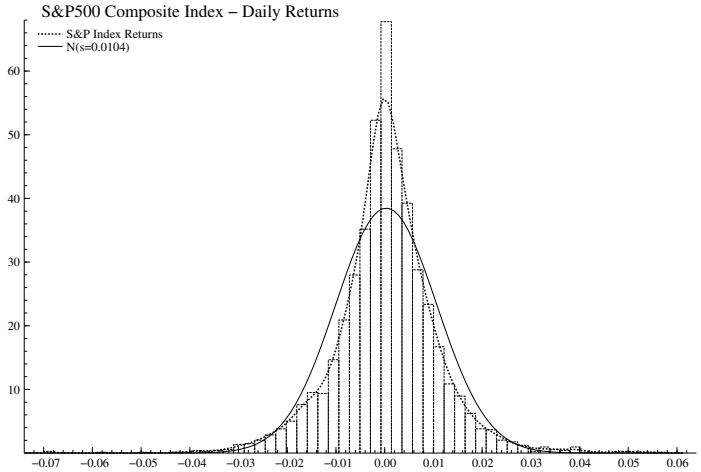

Figure 5.39: Dataset of daily price returns

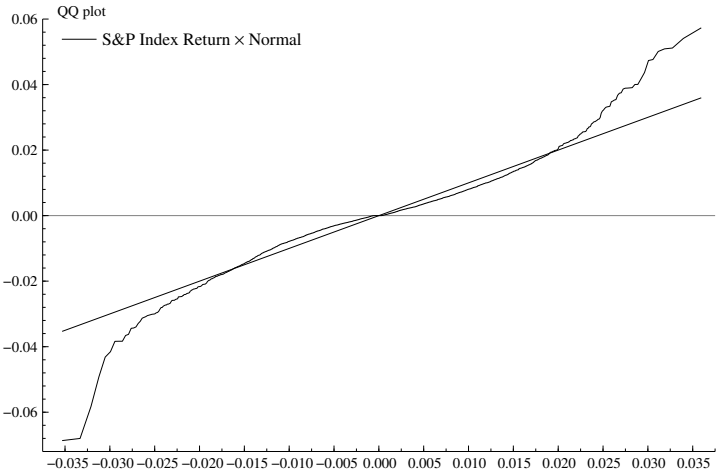

Figure 5.40: QQ-plot of daily price returns distribution against a normal distribution with the same mean and variance

E.4 Bond Indexes

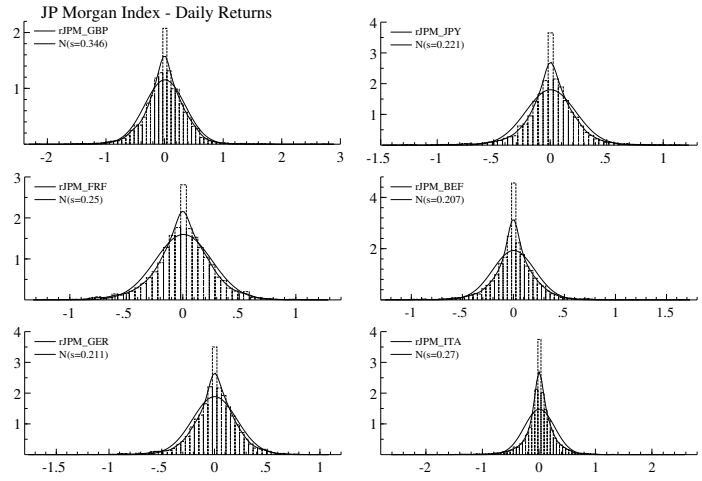

Figure 5.41: Dataset of daily price returns

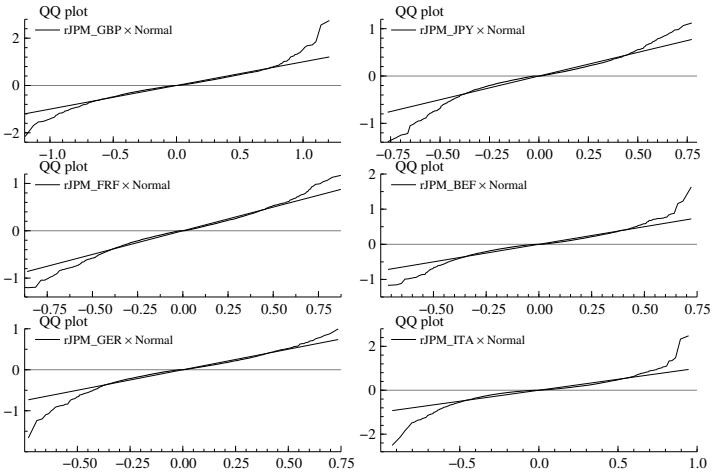

Figure 5.42: QQ-plots of daily price returns distribution against a normal distribution with the same mean and variance

E.5 3-Months Interest Rates

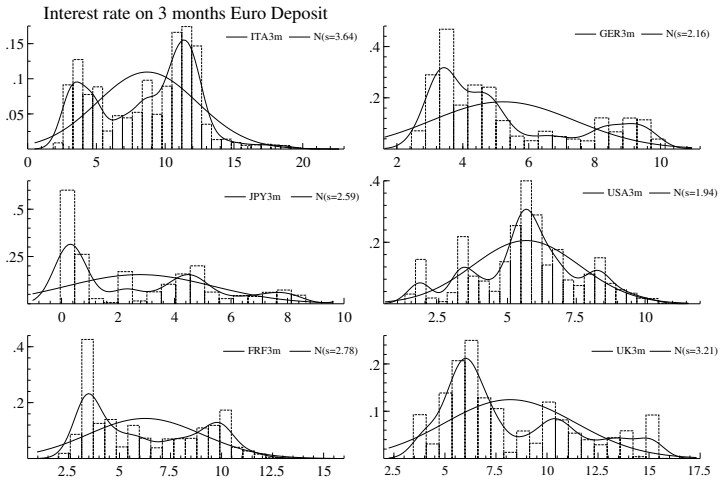

Figure 5.43: Dataset of daily interest rates

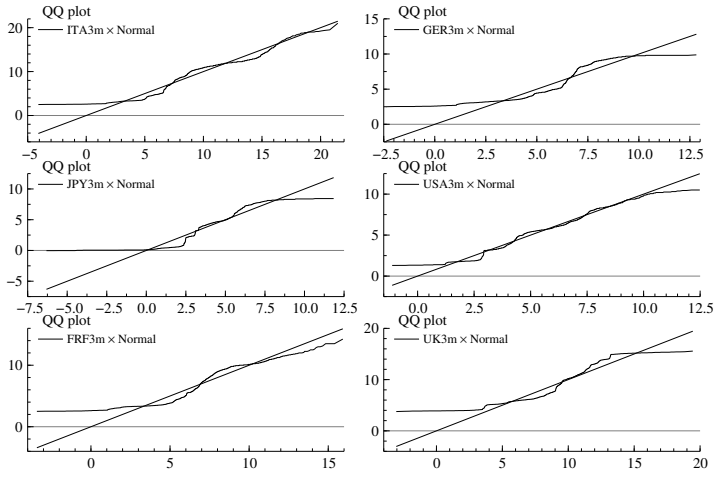

Figure 5.44: QQ-plots of daily interest rates distribution against a normal distribution with the same mean and variance

Bibliography

[1] Adler R.J., Feldman R.E., Taqqu M.S. (1998), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkauser Ed., Boston.

[2] Barnea A., Downes D.H. (1973), A Reexamination of the Empirical Distribution of Stock Price Changes, *Journal of the American Statistical Association*, Vol. 68, Issue 342, pp. 348-350.

[3] Beedles W.L., Simkowitz M.A. (1980), Asymmetric Stable Distributed Security Returns, *Journal of the American Statistical Association*, Vol. 75, Issue 370, pp. 306-312.

[4] Bradley B.O., Taqqu M.S. (2001), Financial Risk and Heavy Tails, Preprint, Boston University.

[5] Brennen M. (1974), On the Stability of the Distribution of the Market Component in Stock Price Changes, *Journal of Financial Quantitative Analysis*, Vol. 9, Issue 6, pp. 945-961.

[6] Buckle D.J. (1995), Bayesian inference for stable distribution, *Journal of American Statistical Association*, Vol. 90, pp. 605-613.

[7] Celeux G., Hurn M., Robert C.P. (2000), Computational and inferential difficulties with mixture posterior distributions, *Journal of American Statistical Association*, 95, pp. 957-979.

[8] Chambers J.M., Mallows C.L., Stuck B.W. (1976), A Method for Simulating Stable Random Variables, *Journal of the American Statistical Association*, Vol. 71, Issue 354, pp. 340-344.

[9] Devroye L. (1986), *Nonuniform Random Variate Generation*, Springer Verlag, New York.

[10] Diebolt J., Robert C.P. (1994), Estimation of Finite Mixture Distributions through Bayesian Sampling, *Journal of the Royal Statistical Society*, Series B (Methodological), Vol. 56, Issue 2, pp. 363-375.

[11] DuMouchel W. (1973), On the Asymptotic Normality of the Maximum-Likelihood Estimate when Sampling from a Stable Distribution, *Annals of Statistics*, 3, pp.948-957.

[12] Escobar M.D., West M. (1995), Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association*, 90, pp. 577-588.

[13] Fama E.F. (1965), The Behavior of Stock Market Prices, *Journal of Business*, 34, pp. 420-429.

[14] Fama E.F. (1965), Portfolio Analysis in a Stable Paretian Market, *Management Science*, B 2, pp. 409-419.

[15] Fama E.F., Roll R. (1968), Some Properties of Symmetric Stable Distributions, *Journal of the American Statistical Association*, 68, pp. 817-836.

[16] Fama E.F., Roll R. (1971), Parameter Estimates for Symmetric Stable Distributions, *Journal of the American Statistical Association*, 66, pp. 331-338.

[17] Fielitz B.D., Rozelle J. P. (1983), Stable Distributions and the Mixtures of Distributions Hypotheses for Common Stock Returns, *Journal of the American Statistical Association*, Vol. 78, Issue 1983, pp. 28-36.

[18] Fishman G.S. (1996), *Monte Carlo: concepts, algorithms, and applications*, Springer Verlag, New York.

[19] Geman S., Geman D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 6, pp. 721-741.

[20] Gilks W.R., Richardson S. and Spiegelhalter D.J. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.

[21] Godsill S. (1999), MCMC and EM-based methods for inference in heavy-tailed processes with α-stable innovations, *Working paper*, Signal Processing Group, Dep. of Engineering, University of Cambridge.

[22] Mandelbrot B., (1963),The Variation of Certain Speculatives Prices, *Journal of Business*,36,394-419.

[23] McLachlan G., Peel D. (2000), *Finite Mixture Models*, John Wiley & Sons.

[24] Meyn S.P., Tweedie R.L. (1993), Markov Chains and Stochastic Stability, Springer-Verlag, New York.

[25] Mittnik S., Rachev T., Paolella M.S. (1998), Stable Paretian modelling in finance: Some empirical and theoretical aspects. In Adler *et. al.*, editor, *A Practical Guide to Heavy Tails*, Birkhauser, Boston, 1998.

[26] Mikosch T. (2000), Modelling Dependence and Tails of Financial Time Series, *Preprint*.

[27] Neal R. M. (1991), Bayesian Mixture Modeling by Monte Carlo Simulation, *Technical Report*, CRG-TR-91-2, Department of Computer Science, University of Toronto.

[28] Press W., Teukolsky S., Vetterling T. and Flannery B.P. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge.

[29] Qiou Z., Ravishanker N., and Dey D.K. (1999), Multivariate Survival Analysis with Positive Stable Frailties, *Biometrics*, vol. 55 pp. 81-88.

[30] Qiou Z., Ravishanker N. (1998), Bayesian inference for time series with infinite variance stable innovations. In Adler R., Feldman R. and Taqqu M. editors, *A Practical Guide to Heavy Tails*, pp. 259-282, Birkhauser, Boston, 1998.

[31] Rachev S., Mittnik S. (2000), *Stable Paretian Models in Finance*, John Wiley & Sons.

[32] Richardson S., Green P.J. (1997), On Bayesian analysis of mixtures with an unknown number of components, *Journal of Royal Statistics Society*, Ser. B, 59, pp. 731-792.

[33] Robert C.P. (2001), *The Bayesian Choice, 2nd ed.* Springer Verlag, New York.

[34] Robert C.P. (1996), Mixtures of distributions: Inference and estimation. In *Markov Chain Monte Carlo in Practice* (Eds W.R.Gilks, S.Richardson and D.J.Spiegelhalter). London: Chapman & Hall.

[35] Robert C.P. and Casella G. (1999), *Monte Carlo Statistical Methods*, Springer Verlag, New York.

[36] Samorodnitsky G. and Taqqu M.S. (1994), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinte Variance*, Chambrdige University Press.

[37] Stephens M. (2000), Bayesian Analysis of Mixture Models with an Unknown Number of Components: an alternative to reverstible jump methods, *Annals of Statistics*, 28, pp. 40-74.

[38] Stephens M. (1997), Bayesian Methods for Mixtures of Normal Distributions, *Ph.D. thesis*, University of Oxford.

[39] Teichmoeller J. (1971), A Note on the Distribution of Stock Price Changes, *Journal of the American Statistical Association*, 66, pp. 282-284.

[40] Tierney L. (1994), Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, 22, pp. 1701-1786.

[41] Weron R. (1996), On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables, *Working Paper*, Technical University of Wroclaw, Poland.

[42] Zoloratev V.M. (1966), On Representation of Stable Laws by Integrals, *Selected Translations in Mathematical Statistics and Probability*, 6, pp. 84-86.

Chapter 6

Bayesian Inference for Markov Switching Stochastic Volatility Models

6.1 Introduction

Stochastic volatility (SV) models find many financial applications, for example option pricing, asset allocation and risk management. The first work on time series with time changing volatility is due to Clark [12]. The most simple continuous SV model has been proposed by Taylor [61], [62], while Hamilton [35] considers a simple discrete SV model. Hull and White [39] introduce continuous time SV models in the modern theory of finance. Other results in continuous time asset pricing under the assumption of time varying stochastic volatility are due to Melino and Turnbull [49] and Wiggins [66]. Barndorff-Nielsen and Shephard [6], [7] develop continuous time stochastic volatility modelling through Lévy and α-stable stochastic processes.

Many extensions to the basic SV models have been proposed in the literature. In particular *Markov Switching Stochastic Volatility* models (*MSSV*), studied in So, Lam and Li [59], are continuous SV models with a jump component in the

[1]Part of this work is in Casarin, R., (2004),"Bayesian Inference for Markov Switching Stochastic Volatility Models", forthcoming, *Working Paper* CEREMADE. Presented at the 4th International Workshop on Objective Bayesian Methodology, CNRS, Aussois, 15-20 June 2003. It received the *Springer's Award* as best poster session.

mean of the volatility process. They result quite appealing because of the financial interpretation of the hidden Markov process, which drives the volatility. Chib, Nardari and Shephard [13] propose a *Generalized Stochastic Volatility* (GSV) models characterized by heavy tail innovations of the observable process. Moreover they study a GSV model with a jump process, which drives the mean of the observed process. Following the suggestion of Chib, Nardari and Shephard [13], we extend their jump GSV model by considering a Markov jump component in the mean of the volatility process. The models proposed in our work represent also an extension to the MSSV model of So, Lam and Li [59], because the observable process is characterized by heavy tail innovations.

Estimation of SV is difficult due to the latent variable structure of the model. In particular MSSV models are more difficult to estimate than simple continuous SV models because there are two hidden levels in the latent structure. In the following we briefly describe the current state of the art of SV estimation techniques. The Method of Moments (MM) has been applied by Taylor [61], [62], [63], by Andersen [1] and Andersen and Sørensen [2]. The MM and the Generalized MM (GMM) avoid the integration problem associated to the evaluation of the likelihood function. But MM reveals to be inefficient when compared with Maximum Likelihood method (ML). In particular in SV models score function cannot be evaluated and the choice of the moments is thus impossible. An alternative approach is the Quasi-Maximum Likelihood method (QML). It is based on the maximization of the approximated likelihood function. Nelson [51], Harvey and Shephard [37], Harvey, Ruiz and Shephard [38] and So, Lam and Li [58] employ a linearized filtering method (Extended Kalman Filter) to obtain QML estimation.

Simulation based methods are more time consuming, but represent a valid alternative to GMM and to QML. In the literature we find the following approaches. The indirect inference method (see Gourieroux, Monfort and Renault [34]) uses an auxiliary model and a calibration procedure to simulate from the correctly specified model. The Efficient Methods of Moments (see Gallant and Tauchen [28] and Gallant, Hsieh and Tauchen [29]) uses the score of the auxiliary model to improve the indirect inference method. Strictly related to the QML approach is the Simulated Maximum Likelihood method (SML). The method approximates through Monte Carlo simulation the likelihood function. Danielson [18], Danielson and Richard [19]

and Durbin and Koopman [23] apply importance sampling to simulate the likelihood function and then maximize the approximated function.

Our work is based on particle filter techniques and belongs to the more general Bayesian framework for time series analysis. For an introduction to estimation methods for dynamic Bayesian models see Harrison and West [36], moreover we refer to Chapter 2 for a brief review on simulation based methods in a Bayesian perspective. Bayesian inference represents an alternative framework to the above cited estimation methods and in the following we discuss the main estimation approaches within this framework.

A first approach is the Monte Carlo Markov Chain-Expectation Maximization method (MCMC-EM). It uses MCMC simulation techniques to evaluate the likelihood function and to calculate the expectation with respect the latent variables. The resulting approximated expectation is then maximized to obtain the ML estimator. Shephard [56], Geyer [30], [31] apply MCMC-EM to stochastic volatility methods. Andrieu and Doucet [3] propose and compare different on-line MCMC-EM algorithms, which allow to process data sequentially. On-line MCMC-EM reveals efficient also for non-linear models if a set of sufficient statistics exists. As example, they evaluate the efficiency of this estimation method also on a basic continuous SV model.

A second approach, in a Bayesian framework, is the Monte Carlo Markov Chain (MCMC) method. It is based on a data completion (or augmentation) principle. It allows to obtain a simulated sample from the posterior distribution of parameters and hidden states, given the available information. In Chapter 2 we show through Example 2.3.1, how data augmentation principle and MCMC, i.e. single-move Gibbs sampler, apply to the basic continuous SV model. Jacquier, Polson and Rossi [40] develop a Bayesian approach to SV model estimation. Their method is based on a hybrid MCMC algorithm and the superiority of the Bayes estimator is exhibited through a comparison with QML and MM estimation methods. De Jong and Shephard [20] apply MCMC approach to SV models and propose a simulation smoother and a multi-move Gibbs sampler to simulate from the disturbances of a time series rather than from the hidden states. The algorithm effectively improves the efficiency of the MCMC method for time series. Shephard and Pitt [57] provide estimation methods for non-Gaussian time series model with application to SV. They

analyse MCMC methods for simulation smoothing and parameters estimation and compare them with maximum likelihood estimation. The likelihood function has been approximated through importance sampling. Kim, Shephard and Chib [44] compare continuous SV models with ARCH models and with GARCH t-Student model. They provide also an analysis of MCMC method for parameters inference and volatility filtering when applied to an approximated likelihood function. In particular they linearized the measurement equation by taking the logarithm of the square and by approximating the resulting innovation distribution with a mixture of distribution. The same approximation technique is used in So, Lam and Li [59]. They generalize the usual continuous SV model by introducing a Markov jump process in the volatility mean. Through this switching process the model accounts for both persistence effects and tilts in volatility. They adopt MCMC approach with a data augmentation principle and take into account the works of Harvey, Ruiz and Shephard [38] and of De Jong and Shephard [20]. Recently, Chib, Nardari and Shephard [13] introduce GSV models, with Student-t innovations and with a jump process in the mean of the measurement equation. They use a MCMC approach for estimating parameters and Particle Filter for approximating the likelihood function in order to perform model diagnostic. Many recent papers focus on the use of MCMC methods in financial models estimation. Johannes and Polson [41] review financial applications of MCMC methods. They discretize the continuous time diffusion process and apply MCMC for parameters estimation and hidden state filtering. Particle filter are then used for model diagnostic. Eraker [24] follows the same framework. See Johannes, Polson and Stroud [42] for a Bayesian approach to state filtering and parameter estimation to jump and diffusion stochastic processes.

In this work, we follow a third Bayesian approach, which has been recently developed and which reveals efficient for general dynamic models. This is sequential simulation based filtering and is called *Particle Filter*. This filtering method is particularly useful in financial applications, when processing data sequentially. As a new observation becomes available, the hidden states and the parameters of the dynamic model can be updated and a new prediction can be performed. Particle filter allows also to perform model diagnostic and parameter inference. For a review of the state of the art see Doucet, Freitas and Gordon [22]. Pitt and Shephard [52] improve standard Sequential Importance Sampling filtering techniques

by introducing the Auxiliary Particle Filter (APF). They apply APF to stochastic volatility models and find that the method performs better than other simulation based techniques and that it is particularly sensitive to outliers. Kim, Shephard and Chib [44] and Chib, Nardari and Shephard [13] apply particle filter for stochastic volatility extraction but not for parameter estimation. Polson, Stroud and Müller [54] apply a practical filter for sequential parameter estimation and state filtering. They show the superiority of their method when compared to the APF with the sequential parameter learning algorithm due to Storvik [60]. Lopes and Marino [48] and Lopes [47] apply APF to a MSSV model for sequential parameter learning and state filtering.

The first aim of our work is to develop the idea of Chib, Nardari and Shephard [13], which propose to extend their jump GSV model by introducing a Markov jump process in the volatility.

The second aim is to develop the joint estimation of the states and the parameters of Markov switching SV model. Recently Storvik [60] analyses this problem and reviews main approaches in the literature. Our work refers to the algorithm of Liu and West [46]. They suggest to combine the APF algorithm with the kernel reconstruction of the parameters posterior distribution. Sequential filtering techniques introduce approximation errors in estimation of the states and parameters. Moreover these errors cumulate over time. Thus, for financial applications of the dynamic Bayesian models and of the particle filtering, it is necessary to take into account and to correct approximation errors.

In the following we apply particle filters techniques to both the Gaussian SV models and the GSV models. The work is structured as follows. In section 6.2 we state the SV models and discuss some useful reparameterisations. Section 6.3 focuses on the particle filter for the joint estimation of states and parameters. Section 6.4 presents some simulation results. Section 6.5 concludes.

6.2 The Markov Switching Stochastic Volatility Models

Financial time series are often characterised by heavy tails, asymmetry and time varying volatility. In particular they may exhibit jumps in volatility, volatility

persistence effects, also called volatility clustering and leverage effects. In this work we focus on the joint modelling of heavy tails of the observable process and on the clustering effects in volatility dynamic.

The hypothesis of Gaussian evolution of the observable process seems to be quite restrictive in many financial applications. This is the reason why some authors proposed generalised stochastic volatility models characterised by Student-t innovations (see Harvey, Ruiz and Shephard [38], Shephard and Pitt [57] and Chib, Nardari and Shephard [13]). In our work we consider MSSV heavy tails processes and make a comparison between models with Student-t or α-stable innovations and the usual Gaussian innovations model.

Another aspect of interest is volatility clustering. It is possible to capture volatility persistence by introducing a jump component in the volatility dynamic. So, Lam and Li [59] extend the simple continuous volatility model of Taylor [62], by adding a Markov jump process to the drift of the stochastic volatility. Following them, in our Markov switching stochastic volatility model ($MSSV$), we assume that the log-volatility h_t is a continuous Markov process, conditionally to a discrete homogeneous Markov process s_t. This process is called *switching process* and determines the regime of volatility. Moreover we assume the switching process varies in a finite and known set of states. See Chopin [14] for an application of particle filters to switching model with a varying number of states. In the following we give some examples of MSSV models under different assumptions on the distribution of the observable process. We will consider both a Gaussian innovations process and heavy tail processes like Student-t and α-stable innovations processes, with unknown degrees of freedom and unknown characteristic exponent respectively.

6.2.1 The Gaussian MSSV Model

The assumption of Gaussian innovations is quite common in practice, thus in this section, we define a basic MSSV model ($\mathcal{M}_1$), which is completely Gaussian

$$(y_t|h_t) \sim \mathcal{N}(0, e^{h_t}) \tag{6.1}$$

$$(h_t|h_{t-1}, s_t) \sim \mathcal{N}(h_t|\alpha_{s_t} + \phi\, h_{t-1}, \sigma^2) \tag{6.2}$$

for $t = 1, \ldots, T$, where s_t is a homogeneous discrete Markov's process, with

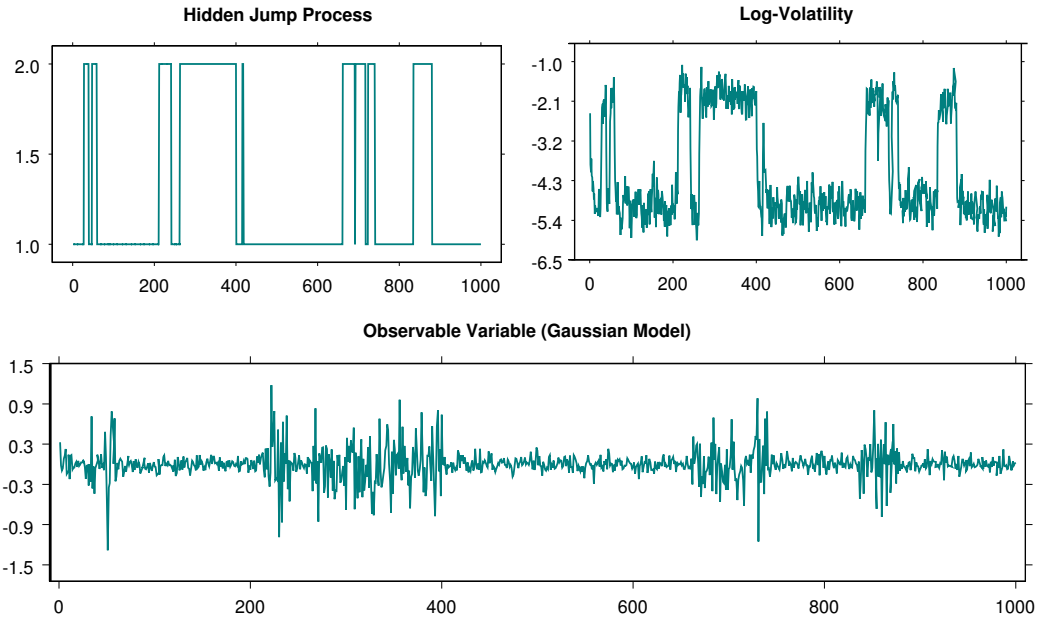

Figure 6.1: Simulation of the Markov switching stochastic volatility model $\mathcal{M}_1$ ($\alpha_1 = -2.5$, $\alpha_2 = -1$, $\phi = 0.5$, $\sigma^2 = 0.1$, $p_{11} = 0.99$, $p_{22} = 0.975$). The left upper graph exhibits the evolution of the hidden jump process, the right upper graph shows the log-volatility of the observable process, which is represented in the third graph.

transition probabilities

$$\mathbb{P}(s_t = j | s_{t-1} = i, s_{t-2} = i_2 ..., s_0 = i_t) = \mathbb{P}(s_t = j | s_{t-1} = i) = p_{ij} \tag{6.3}$$

with $i, j = 1, \ldots, L$, L denoting the number of unobservable states.

For the sake of simplicity we introduce the following notation: $\theta = ((\alpha_1, \ldots, \alpha_L), \phi, \sigma^2, (\mathbf{p}_1, \ldots, \mathbf{p}_L))$, with $\mathbf{p}_i = (p_{i1}, \ldots, p_{iL})$, for the parameters and $s_{0:t} = (s_0, \ldots, s_t)$ and $h_{0:t} = (h_0, \ldots, h_t)$, for the two hidden Markov processes. For estimation purposes, in order to impose the positivity constraint on σ^2 and to constrain p_{ij} to be in $(0, 1)$, we adopt the following reparameterisation: $log(\sigma^2)$ and $log(\frac{p_{ij}}{1-p_{ij}})$.

Fig. 6.1 exhibits a sample of T=1,000 values, simulated from the MSSV Gaussian model $\mathcal{M}_1$, with parameters: $\alpha_1 = -2.5$, $\alpha_2 = -1$, $\phi = 0.5$, $\sigma^2 = 0.1$, $p_{11} = 0.99$, $p_{22} = 0.975$. We use the parameters values estimated by So, Lam and Li [59]. Note that the value of the transition probabilities induces in this simulation example a high degree of persistence in the volatility regimes of the observed process.

6.2.2 Heavy tails MSSV Models

Due to the high degree of heterogeneity of the time series, the assumption of Gaussian observable process seems to be restrictive in many real contexts and for this reason it has been removed by many recent studies. Moreover a common way to model heterogeneous dynamics in time series is to include a stochastic latent structure in the model. For example Chib, Nardari and Shephard [13] propose a Student-t discrete time GSV model and a similar model with a jump component in the mean of the observable process. In a continuous time setting Barndorff-Nielsen and Shephard [6] study heavy tail processes.

Financial time series often exhibit volatility tilts and clustering behaviour. In order to capture these features of the volatility dynamic, we study the following non-Gaussian Markov switching stochastic volatility models. We assume that the observable variable follows a heavy tail process, which will alternatively be a *Student-t* process or a *stable* process. Note that both of them have the Gaussian model as particular case.

The first GSV model ($\mathcal{M}_2$), is

$$(y_t|h_t) \sim \mathcal{T}_\nu(y_t|0, e^{h_t}) \tag{6.4}$$

$$(h_t|h_{t-1}, s_t) \sim \mathcal{N}(h_t|\alpha_{s_t} + \phi\, h_{t-1}, \sigma^2) \tag{6.5}$$

$$(s_t|s_{t-1}) \sim \mathcal{M}_L(s_t|1, p_{s_{t-1}1}, \ldots, p_{s_{t-1}L}) \tag{6.6}$$

for $t = 1, \ldots, T$, where $\mathcal{M}_L$ is the multinomial distribution and $\mathcal{T}_\nu(y|\delta, \sigma)$ represents the density of a *Student-t* distribution

$$\mathcal{T}_\nu(y|\delta, \sigma) = \frac{\Gamma((\nu+1)/2)\Gamma(\nu/2)}{(\nu\pi\sigma^2)^{1/2}} \left(1 + \frac{1}{\nu\sigma^2}(y-\delta)^2\right)^{-(1+\nu)/2}. \tag{6.7}$$

The distribution is characterised by three parameters: ν the *degrees of freedom* parameter, δ the *location parameter* and σ the *scale parameter*. Note that the heaviness of the tails is controlled by the parameter ν and that when $\nu \to \infty$ the distribution converges to a Gaussian distribution.

The second GSV model ($\mathcal{M}_3$) also is characterised by an heavy tail observable process

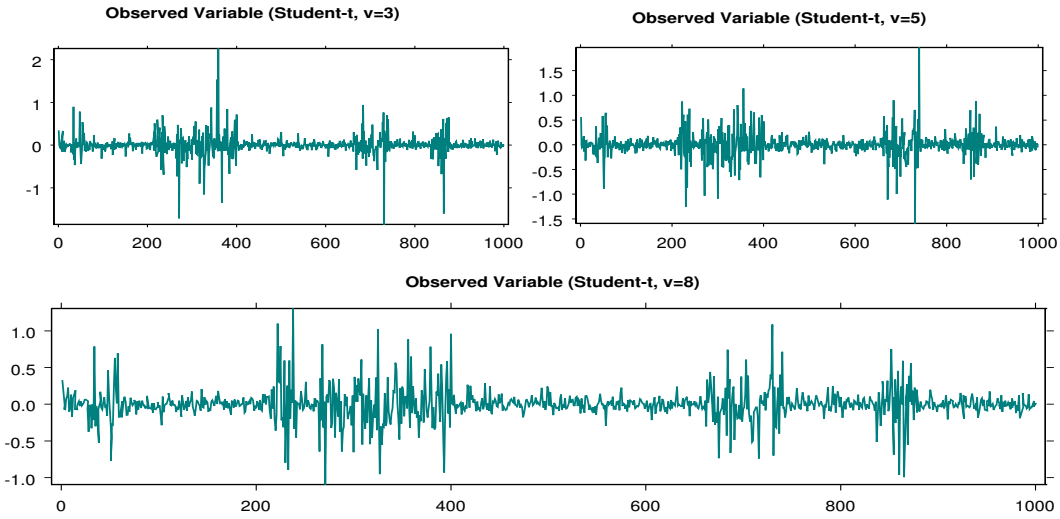

Figure 6.2: Simulation of the Markov switching stochastic volatility model $\mathcal{M}_2$ ($\alpha_1 = -2.5$, $\alpha_2 = -1$, $\phi = 0.5$, $\sigma^2 = 0.1$, $p_{11} = 0.99$, $p_{22} = 0.975$ and $\nu = 3,\ 5,\ 8$). The hidden jump process realisations are the same depicted in Fig. 6.1

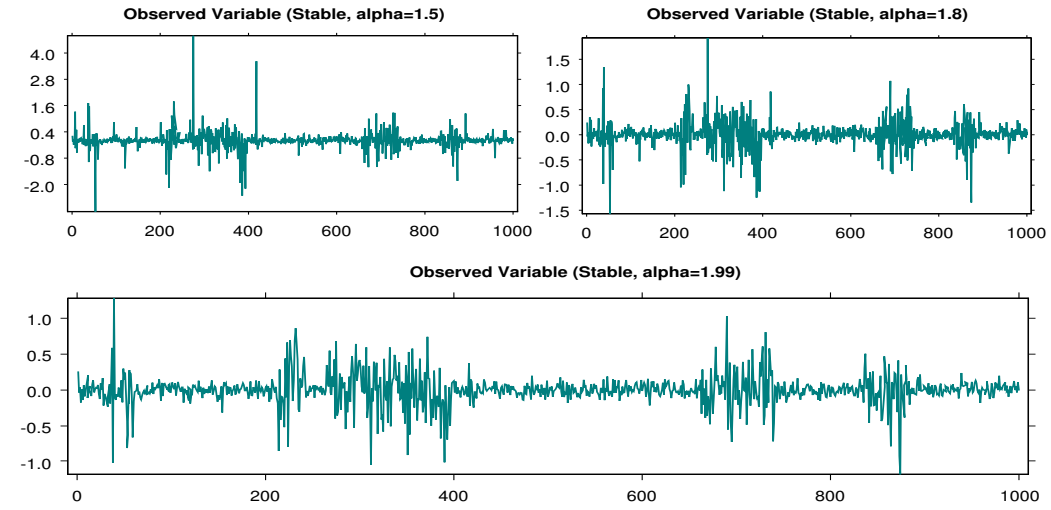

Figure 6.3: Simulation of the Markov switching stochastic volatility model $\mathcal{M}_3$ ($\alpha_1 = -2.5$, $\alpha_2 = -1$, $\phi = 0.5$, $\sigma^2 = 0.1$, $p_{11} = 0.99$, $p_{22} = 0.975$, $\beta = 0$ and $\alpha = 1.5,\ 1.8,\ 1.99$). The realisations of the hidden jump process are the same depicted in Fig. 6.1

$$(y_t|h_t) \sim S_\alpha(y_t|0,0,e^{h_t}) \tag{6.8}$$

$$(h_t|h_{t-1},s_t) \sim \mathcal{N}(h_t|\alpha_{s_t}+\phi\, h_{t-1},\sigma^2) \tag{6.9}$$

$$(s_t|s_{t-1}) \sim \mathcal{M}_L(s_t|1,p_{s_{t-1}1},\ldots,p_{s_{t-1}L}) \tag{6.10}$$

for $t = 1,\ldots,T$, where $\mathcal{M}_L$ is the multinomial distribution and $S_\alpha(y|\beta,\delta,\sigma)$ represents the density of a *stable* distribution, which is completely characterised by the following four parameters: the *characteristic exponent* α, the *skewness parameter* β, the *location parameter* δ and finally the *scale parameter* σ. We assume for simplicity that $\beta = 0$. Moreover we take $\alpha \in (1,2]$ in order to have a finite first order moment. Note that stable distributions have the Gaussian distribution as a particular case, when $\alpha = 2$.

The stable distribution density can not generally be written in an analytic form, thus it is conveniently defined through its characteristic function. The most well known parametrisation is defined in Samorodnitsky and Taqqu [55]

$$\mathbb{E}\left[e^{i\,\vartheta\,y}\right] = \begin{cases} \exp(-|\sigma\vartheta|^\alpha)(1 - i\,\beta(sign(\vartheta))\tan(\pi\alpha/2) + i\delta\vartheta) & if\ \alpha \neq 1; \\ \exp(-|\sigma\vartheta|(1 + 2\,i\,\beta\ln|\vartheta|sign(\vartheta)/\pi) + i\delta\vartheta) & if\ \alpha = 1. \end{cases} \tag{6.11}$$

where $\vartheta \in \mathbb{R}$. In the parameter setting of our model the characteristic function reduces to

$$E[\exp(i\,\vartheta\,y)] = e^{-|\sigma\,\vartheta|^\alpha}. \tag{6.12}$$

In order to obtain an analytic representation of the density of a stable random variable an auxiliary variable has to be introduced. The same strategy is used in Buckle [10] for α-stable distributions and in Godsill [32] for inference on time series with α-stable innovations.

For the sake of simplicity we introduce the following notations. The parameter vector is $\theta = (\nu,(\alpha_1,\ldots,\alpha_L),\phi,\sigma^2,(\mathbf{p}_1,\ldots,\mathbf{p}_L))$ for the model $\mathcal{M}_2$ and $\theta = (\alpha,(\alpha_1,\ldots,\alpha_L),\phi,\sigma^2,(\mathbf{p}_1,\ldots,\mathbf{p}_L))$ for the model $\mathcal{M}_3$. As usual $\mathbf{p}_i = (p_{i1},\ldots,p_{iL})$, $s_{0:t} = (s_0,\ldots,s_t)$ and $h_{0:t} = (h_0,\ldots,h_t)$. For estimation purposes, in order to constrain the parameter α to be into $(1,2]$ we consider the following invertible transformation: $\log((\alpha-1)/(2-\alpha))$. For the Student-$t$ distribution we let ν

vary uniformly in the interval $[2, 100]$, thus we use the transformation: $\log(((\nu - 2)/98)/(1 + ((\nu - 2)/98)))$.

Fig. 6.2 and 6.3 exhibit some samples of T=1,000 values, simulated respectively from the MSSV models $\mathcal{M}_2$ and $\mathcal{M}_3$, with parameters: $\nu = 3,\ 5,\ 8$, for the Student-t model, $\alpha = 1.5,\ 1.8,\ 1.99$, for the stable model and $\alpha_1 = -2.5$, $\alpha_2 = -1$, $\phi = 0.5$, $\sigma^2 = 0.1$, $p_{11} = 0.99$, $p_{22} = 0.975$.

6.2.3 Stationarity Conditions for SV models

The standard continuous SV process is often assumed in order to model the excess of kurtosis in the unconditional distribution of the observable process. Moreover continuous SV is able to capture volatility clustering, but many financial time series exhibit also a multi-modal unconditional distribution. This feature can be explained by a volatility process with a mean changing over time. In MSSV models a hidden jump process (i.e. *Markov Switching process*) is added to the mean parameter of the log-volatility process. A first consequence of including a hidden Markov Switching process in the log-volatility is to increase furthermore the degree of kurtosis of the observable process. Moreover the MSSV model is able to capture both volatility persistence and volatility tilts.

While stationarity of the continuous SV model is well discussed in the literature, some considerations on the stationarity of the MSSV models are needed. Define the following reparameterisation of the MSSV model

$$y_t = e^{h_t/2}\varepsilon_t\,, \quad \varepsilon_t \sim \mathcal{N}(0,1) \tag{6.13}$$

$$h_t = \alpha_{s_t} + \phi h_{t-1} + \sigma_\eta \eta_t\,, \quad \eta_t \sim \mathcal{N}(0,1) \tag{6.14}$$

$$\alpha_{s_t} = \alpha + \beta s_t \tag{6.15}$$

$$\mathbb{P}(s_t = i | s_{t-1} = j) = p_{ij} \tag{6.16}$$

with $p_{ij} \leq 0$, $\forall\, i, j \in E$ and $\sum_{l=1}^{L} p_{il} \leq 1$. Moreover $\{s_t\}_{t\in\mathbb{N}}$ is a Markov jump process, which takes value in the finite countable state space $E = \{0, \ldots, L\}$. In the following we assume that $E = \{0, 1\}$, the initial state s_0 of the process has

probability measure μ_0 and finally that the transition matrix of s_t is

$$\mathbb{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \tag{6.17}$$

Note that through the transition matrix and the initial probability measure, the Markov jump process is well defined (see Theorem 6.5.1 in Appendix 6.5). As stated in Theorem 6.5.2, the second order stationarity of the observable process y_t is guaranteed by the first order stationarity of the process h_t.

In the following we focus on the stationarity conditions for the hidden Markov process $\{h_t, s_t\}_{t\in\mathbb{N}}$. Due to the causality relations between s_t and h_t, it is possible to study first the unconditional stationarity of $\{s_t\}_{t\in\mathbb{N}}$ and secondly the stationarity of $\{h_t\}_{t\in\mathbb{N}}$ conditionally on $\{s_t\}_{t\in\mathbb{N}}$.

Stationarity conditions for $\{s_t\}_{t\in\mathbb{N}}$ follow from the properties of the n-times composition of the transition matrix, which is given in Theorem 6.5.3 (see Appendix 6.5). When $n \to +\infty$, the transition probability $\mathbb{P}^n$ tends to a finite quantity if and only if $1 - p_{10} - p_{01} < 1$.

Observe that the autoregressive structure of the log-volatility process, see the equation (6.14), makes it dependent on the past history of the Markov jump process. This feature becomes evident from the ergodic solution of the system of stochastic difference equation (6.14), (6.15) and (6.16)

$$h_t = \alpha \sum_{i=0}^{+\infty} \phi^i + \beta \sum_{i=0}^{+\infty} \phi^i s_{t-i} + \sigma_\eta \sum_{i=0}^{+\infty} \phi^i \eta_{t-i}. \tag{6.18}$$

which is derived in Theorem 6.5.4 under the assumption $|\phi| < 1$.

In Appendix 6.5, we find that first and second order stationary moments of h_t exist if ϕ and $|1 - p_{10} - p_{01}| < 1$.

For further details on the asymptotic second order stationarity and for the strictly stationarity see Appendix A. We refer also to Francq and Roussignol [26] and Francq and Zakoian [27] for a general discussion of the stationarity conditions for switching non linear AR and switching ARMA models.

6.3 Particle Filters

In Chapter 2 we reviewed the principal simulation based filtering techniques in a Bayesian perspective. We defined the single-move and the multi-move Gibbs samplers, the adaptive importance sampling and finally introduced particle filters. Furthermore we applied them to some simple stochastic volatility models in order to filter the hidden log-volatility. In this section we deal with particle filters, focusing on the problems of parameter estimation and hidden state filtering. We recall that particle filters are sequential Monte Carlo algorithms. They reveal quite useful for dynamic models, like $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$, which have elements of non-linearity and non-Gaussianity and provide a significant advantage over traditional filtering techniques. In particular, in many real situations data are processed on-line. When a new observation arrives, the estimate of the states and of the parameters has to be updated. Recursive techniques, like sequential Monte Carlo filters, are well appreciated. Moreover simulation based filtering allows to evaluate the likelihood function of complex dynamic models and allows also to perform model diagnostics.

In the following we focus on *Particle filters*, also referred in the literature as *Bootstrap filters*, *Interacting particle filters*, *Condensation algorithms*, *Monte Carlo filters* and on the joint estimation of the states and the parameters of the dynamic model.

We state a quite general formulation of the filtering problem in a Bayesian perspective, which does not usually admit an analytical solution. Denote by $\{\mathbf{x}_t;\, t \in \mathbb{N}\}$, $\mathbf{x}_t \in \mathcal{X}$, the hidden states of the system, by $\{\mathbf{y}_t;\, t \in \mathbb{N}_0\}$, $\mathbf{y}_t \in \mathcal{Y}$ the observable variable and by $\theta \in \Theta$ the parameters of the densities. In this section we suppose that the parameters are known. The Bayesian state space representation of a nonlinear, non-Gaussian dynamic model, is given by an *initial distribution* $p(\mathbf{x}_0)$, a *measurement density* $p(\mathbf{y}_t|\mathbf{x}_t)$ and a *transition density* $p(\mathbf{x}_t|\mathbf{x}_{t-1};\theta)$. Moreover, we assume that the Bayesian dynamic model

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1};\theta) \tag{6.19}$$

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t;\theta) \tag{6.20}$$

$$x_0 \sim p(\mathbf{x}_0;\theta) \quad \text{with } t = 1,\ldots,T. \tag{6.21}$$

is Markovian, that is the transition density depends on the past, only through the

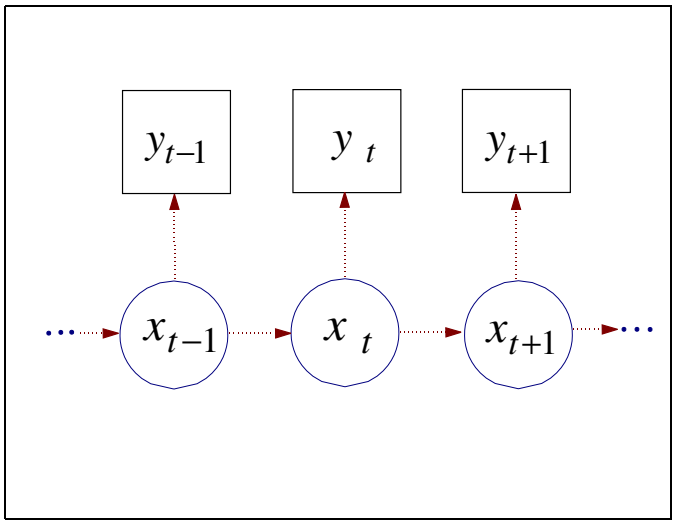

Figure 6.4: Causality structure of a Markovian dynamic model with hidden states. A box around the variable indicates the variable is known, while a circle indicates a hidden variable.

last value of the hidden state. The measurement density is a function of the current value of the hidden state. Fig. 6.4 shows the causality structure of the Bayesian dynamic model given in equations (6.19), (6.20) and (6.21). Note that models $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ do exhibit this structure.

When processing data on-line, at each time t, two quantities of interest are the estimate of the current hidden state of the system and the prediction on the state of the system at time $t+1$. In order to predict the future value of the state of the system, given the information available at time t, we use the Chapman-Kolmogorov equation, which characterises the hidden state evolution and gives us the following *prediction density*

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t};\theta) = \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{1:t};\theta)p(\mathbf{x}_t|\mathbf{y}_{1:t};\theta)d\mathbf{x}_t = \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_t;\theta)p(\mathbf{x}_t|\mathbf{y}_{1:t};\theta)d\mathbf{x}_t. \tag{6.22}$$

As the new observation $\mathbf{y}_{t+1}$ becomes available, it is possible using Bayes' theorem to update the prediction density and to filter the current state of the system. The *filtering density* is

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1};\theta) = \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\mathbf{y}_{1:t};\theta)p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} \tag{6.23}$$

where the marginal density at denominator is obtained as follows

$$p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta) = \int p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\mathbf{y}_{1:t};\theta)p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t};\theta)d\mathbf{x}_t. \tag{6.24}$$

Moreover the assumption of Markovian dynamic of the hidden states allows to obtain a recursive relation, which is useful when solving a filtering problem and sequentially processing data at the same time

$$p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta) = p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)\frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t;\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} \tag{6.25}$$

(for a proof see Appendix B). In the following we introduce some basic particle filter algorithms, with a particular attention to the auxiliary particle filter. Moreover we treat the problem of the joint estimation of the hidden states and of the parameters of the model.

6.3.1 State Filtering

Assume the parameters θ of the dynamic model given in equations (6.19), (6.20) and (6.21) are known. Different versions of the particle filter exist in the literature and different simulation approaches like rejection sampling, MCMC and importance sampling, can be used for the construction of a particle filter. To introduce particle filters, we will apply importance sampling reasoning to the smoothing problem.

Define the *smoothing density*, $p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta)$, as the density function of the hidden states of the system from 0 up to time $(t+1)$, conditionally on the observed variables from 1 up to time $(t+1)$. In order to solve the smoothing problem, the first basic idea used in particle filters, is to approximate the smoothing density

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta) \approx \sum_{i=1}^{N} w_t^i \delta_{\{\mathbf{x}_{0:t}^i\}}(d\mathbf{x}_{0:t}) \tag{6.26}$$

by simulating $\{\mathbf{x}_{0:t}^i\}_{i=1}^N$ from a proposal distribution $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t},\theta)$ and by correcting the weights of the resulting empirical density. The correction step comes from an importance sampling argument, thus the *unnormalized particles weights*[1] are defined

[1]Note that importance sampling requires to know the importance and the target distributions up to a proportionality constant, thus the unnormalized weights may not sum to one. However

as follows

$$w_t^i \triangleq \frac{p(\mathbf{x}_{0:t}^i|\mathbf{y}_{1:t};\theta)}{q(\mathbf{x}_{0:t}^i|\mathbf{y}_{1:t};\theta)}. \tag{6.28}$$

The second key idea used in particle filters, which makes them particulary appealing for on-line applications, consists in finding a recursive relation for the weights updating. Assume that the dynamic model of interest is the one described by equations (6.19), (6.20) and (6.21) and choose the importance density to factorise as follows: $q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta) = q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)q(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t+1};\theta)$, then the weights can be rewritten in a recursive form

$$w_{t+1}^i = w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i;\theta)p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i;\theta)}{q(\mathbf{x}_{t+1}^i|\mathbf{x}_{t+1}^i,\mathbf{y}_{t+1};\theta)} \tag{6.29}$$

(see Appendix B for a proof). For a discussion on the optimal choice of the importance density see Doucet [21]. However note that the most convenient choice of the importance density is the prior density: $q(\mathbf{x}_{t+1}^i|\mathbf{x}_{t+1}^i,\mathbf{y}_{t+1};\theta) = p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i;\theta)$. The resulting recursive relation then reduces to

$$w_{t+1}^i = w_t^i\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i;\theta). \tag{6.30}$$

At each time step t we may be interested not only in solving a smoothing problem, but also in filtering the current hidden state. In that case we search how to approximate the filtering density (Eq. (6.23)). Using the same importance sampling argument, the generic particle filter approximates equation (6.22) as follows

$$\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t};\theta) = \sum_{i=1}^{N} p(\mathbf{x}_{t+1}|\mathbf{x}_t^i;\theta)w_t^i\delta_{\{\mathbf{x}_t^i\}}(d\mathbf{x}_t) \tag{6.31}$$

which is called *empirical prediction density*. Thus by propagating the weights through the Chapman-Kolmogorov equation, it is possible to obtain the *empirical*

normalized importance sampling weights can be easily obtained as follows

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_{j=1}^N w_t^j} \quad i = 1,\ldots,N \text{ and } t = 1,\ldots,T. \tag{6.27}$$

The normalization procedure causes the loss of the unbiasness property.

filtering density, which approximates the filtering density given in Eq. (6.23)

$$\hat{p}(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1};\theta) = \sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t^i;\theta)w_t^i\delta_{\{\mathbf{x}_t^i\}}(d\mathbf{x}_t). \tag{6.32}$$

Assume now that the quantity $\mathbb{E}(f(\mathbf{x}_{t+1})|\mathbf{y}_{1:t+1})$ is of interest. It can be evaluated numerically by a Monte Carlo sample $\{\mathbf{x}_{t+1}^i, w_{t+1}^i\}_{i=1}^N$, simulated from the empirical filtering distribution (6.32) through importance sampling

$$\mathbb{E}(f(\mathbf{x}_{t+1})|\mathbf{y}_{1:t+1}) \approx \frac{\sum_{i=1}^N f(x_{t+1}^i)w_{t+1}^i}{\sum_{i=1}^N w_{t+1}^i}. \tag{6.33}$$

The generic particle filter developed through previous equations is called *Sequential Importance Sampling* (SIS) (see Appendix C for a pseudo-code representation). See also Doucet, Freitas and Gordon [22] for an updated review on the particle filter techniques and on the main convergence results for this kind of algorithms. It is well known in the literature (see for example Arulampalam, Maskell, Gordon and Clapp [4]), that basic SIS algorithms have a degeneracy problem. After some iterations the empirical distribution degenerates into a single particle, because the variance of the importance weights is non-decreasing over time (see Doucet [21]). In order to solve the degeneracy problem, the *Sampling Importance Resampling* (SIR) algorithm has been introduced by Gordon, Salmond and Smith [33]. This algorithm (see Appendix C for a pseudo-code representation) belongs to a wider class of bootstrap filters, which use a re-sampling step to generate a new set of particles with uniform weights. This step introduces diversity in particle set, avoiding degeneracy. Note however that the basic SIR algorithm produces a progressive impoverishment of the information contained in the particle set, because of the resampling step and of the fact that particles does not change over filter iterations. Many solutions have been proposed in literature. We recall here the *Regularised Particle Filter* proposed by Musso, Oudjane and LeGland [50], which is based on a discretisation of the continuous state space. Moreover Gilks and Berzuini [8] propose the SIR-Move algorithm, which moves particles after the re-sampling step. Thus particles value changes and impoverishment is partially avoided. Finally Pitt and Shephard [52] introduce the *Auxiliary Particle Filter* (APF) and applied it to a Gaussian ARCH-type stochastic volatility model. They find the filter works

well, although it is highly sensible to outliers. In the following we focus on the APF algorithm.

In order to avoid re-sampling, APF algorithm uses an auxiliary variable to select most representative particles and to mutate them through a simulation step. Then weights of the regenerated particles are updated through an importance sampling argument. In this way particles with low probability do not survive to the selection and the information contained in particles set is not wasted. In particular the auxiliary variable is a random particle index, which is used in the selection step to sample new particles. The random index is simulated from a distribution which contains and resumes the information on previous particle set. This feature is due to the use of μ_t^i in the measurement density. Note that the empirical filtering density given in Eq. (6.32) is a mixture of distributions, which can be reparameterised by introducing the allocation variable $i \in \{1, \ldots, N\}$. The joint distribution of the hidden state and the index i is

$$\begin{aligned} p(\mathbf{x}_{t+1}, i|\mathbf{y}_{1:t+1};\theta) &= \frac{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \mathbf{x}_{t+1}, i)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} p(\mathbf{x}_{t+1}, i|\mathbf{y}_{1:t};\theta) = \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} p(\mathbf{x}_{t+1}|i, \mathbf{y}_{1:t};\theta) p(i|\mathbf{y}_{1:t};\theta) = \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^i;\theta) w_t^i. \end{aligned} \tag{6.34}$$

The basic idea of the APF is to refresh the particle set while reducing the loss of information due to this operation. Thus the algorithm generates a new set of particles by jointly simulating the particle index i (*selection step*) and the selected particle value $\mathbf{x}_{t+1}$ (*mutation step*) from the reparameterised empirical filtering density and according to the following importance density

$$\begin{aligned} q(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1};\theta) &= q(\mathbf{x}_{t+1}^j|\mathbf{y}_{1:t+1};\theta) q(i^j|\mathbf{y}_{1:t+1};\theta) \\ &= p(\mathbf{x}_{t+1}^j|\mathbf{x}^{i^j};\theta)(p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j};\theta) w_t^{i^j}) \end{aligned} \tag{6.35}$$

for $j = 1, \ldots, N$. By following the usual importance sampling argument, the updating relation for the particle weights is

$$
\begin{aligned}
w_{t+1}^{j} &\triangleq \frac{p(\mathbf{x}_{t+1}^{j}, i^{j}|\mathbf{y}_{1:t+1};\theta)}{q(\mathbf{x}_{t+1}^{j}, i^{j}|\mathbf{y}_{1:t+1};\theta)} \\
&= \frac{p(\mathbf{x}_{t+1}^{j}|\mathbf{x}^{i^{j}};\theta)p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^{j};\theta)w_{t}^{i^{j}}}{p(\mathbf{x}_{t+1}^{j}|\mathbf{x}^{i^{j}};\theta)p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^{j}};\theta)w_{t}^{i^{j}}} \\
&= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^{j};\theta)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^{j}};\theta)}
\end{aligned}
\tag{6.36}
$$

In many applications of the particle filter techniques, parameters are treated as known and MCMC parameter estimates are used instead of the true values. MCMC is typically a off-line approach, it does not allow to sequentially update parameter estimates as new observations arrive. Moreover, when applied sequentially, MCMC estimation method is more time consuming than particle filter algorithms. Thus in the next section we will consider the filtering problem in presence of unknown static parameters, in a Bayesian perspective.

6.3.2 Parameter Estimation

When processing sequentially data, both the problems of hidden state filtering and of the parameters estimation arise. In engineering, a common way to solve this problem is to treat parameters as hidden states of the system. Berzuini et al. [9] develop this approach in a Bayesian framework. Thus standard particle filtering techniques apply here to estimate the joint posterior density $p(\mathbf{x}_{0:t}, \theta|\mathbf{y}_{1:t})$. Approximated posterior $p(\theta|\mathbf{y}_{0:t})$ is then obtained by marginalisation.

Observe that the parameters are fixed over time, thus particles relative to the parameter posterior distribution do not change, while the particles approximating hidden states are allowed to vary over filter iterations. As pointed out by Storvik [60], the degeneracy of the parameters weights produces a negative effect on the whole posterior distribution, which degenerates to a Dirac mass. Different solutions to the degeneracy problem have been proposed in the literature. For example Kitagawa [43] explicitly assumes an artificial evolution of the parameters, which are still considered as hidden states of the dynamic model. The assumption of time varying parameters introduces diversity in particles set avoiding the degeneracy problem, but produces higher variability in parameter estimates. Liu and West [46] use a kernel density estimation of the parameter posterior distribution as importance density to refresh

the particle set. This method produces slowly time-varying parameters and thus adds noise to the parameter estimates. In order to reduce the effect of the artificial variability, the authors adopt a kernel shrinkage technique.

An alternative approach can be founded in Storvik [60], which proposes a quite general particle filter for joint estimation of hidden states and non-dynamic parameters. The algorithm requires to know a set of sufficient statistics for the posterior distribution. Note however that the existence of sufficient statistic for the parameter θ is not necessary in principle, because the posterior distribution of the parameters $p(\theta|\mathbf{x}_{0:t}, \mathbf{y}_{0:t})$ can be always evaluated at each time step. A sequential algorithm, called practical filter, is proposed by Polson, Stroud and Müller [53]. The parameter and state joint filtering distribution is represented as a mixture of fixed lag-filtering distributions. They simulate from the joint filtering distribution by simulating sequentially from the parameter posterior and from the fixed-lag smoothing distribution. The method is particularly useful when a set of sufficient statistic for the posterior is known. A comparison (see Polson, Stroud and Müller [54]) with Storvik's [60] algorithm proves the superiority of the practical filter when apply to the basic continuous SV model.

Sequential methods, alternative to particle filters, can be founded in Andrieu and Doucet [3], who propose online Expectation-Maximization type algorithms, which do not degenerate, but require the knowledge of the hidden Markov process ergodic distribution and of a set of sufficient statistics for the posterior distribution.

In the following we refer to the algorithm due to Liu and West [46] and to the works of Lopes [47] and Lopes and Marigno [48], for some applications of the particle filter algorithms to MSSV models.

The problem of the joint estimation of parameters and states of a dynamic system can be stated in a Bayesian framework as follows. Define a Bayesian dynamic model with unknown parameters

$$(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}, \theta) \tag{6.37}$$

$$(\mathbf{y}_t|\mathbf{x}_t) \sim p(\mathbf{y}_t|\mathbf{x}_t, \theta) \tag{6.38}$$

$$x_0 \sim p(\mathbf{x}_0|\theta) \tag{6.39}$$

$$\theta \sim p(\theta)\,, \quad \text{with } t = 1, \ldots, T. \tag{6.40}$$

Note that unknown parameters are treated as random quantities, thus we denote the conditional density by $p(\cdot \mid \cdot, \theta)$, and assume a prior distribution $p(\theta)$. The state and parameters joint posterior distribution associated to this model is (*smoothing problem*)

$$\begin{aligned} p(\mathbf{x}_{0:t+1}, \theta | \mathbf{y}_{1:t+1}) &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{0:t+1}, \mathbf{y}_{1:t}, \theta) p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t}, \mathbf{y}_{1:t}, \theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} p(\mathbf{x}_{0:t}, \theta | \mathbf{y}_{1:t}) \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta) p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta) p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} p(\theta | \mathbf{x}_{0:t}, \mathbf{y}_{1:t}) \end{aligned} \tag{6.41}$$

The posterior distribution is written as the product of two components. The first is the filtering distribution and the second is the full posterior distribution of the parameters given hidden states and the observations. The completed posterior of the parameters is proportional to a function which can always be written in a recursive form

$$p(\theta | \mathbf{x}_{0:t}, \mathbf{y}_{1:t}) \propto p(\theta) p(x_0|\theta) \prod_{k=1}^{t} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \theta) p(\mathbf{y}_k|\mathbf{x}_k, \theta) \tag{6.42}$$

that can be evaluated in the simulated hidden states as a by product of the particle filter algorithm.

In the same way as for the smoothing problem, the joint filtering density of the current state $\mathbf{x}_t$ and of the parameter θ can be written as the product of two quantities (*filtering problem*)

$$\begin{aligned} p(\mathbf{x}_{t+1}, \theta | \mathbf{y}_{1:t+1}) &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \mathbf{y}_{1:t}, \theta) p(\mathbf{x}_{t+1}, \theta | \mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta) p(\mathbf{x}_{t+1}|\theta, \mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} p(\theta | \mathbf{y}_{1:t}) \end{aligned} \tag{6.43}$$

The filtering problem can thus be treated conditionally to the parameters value. It is possible for example to use the Kalman Filter or the HMM filtering algorithms to filter the states and the particle filter to estimate the parameters (see for example Chopin [14]). In MSSV model both the Kalman Filter and the HMM can not be used, thus Monte Carlo filter must be used for the joint estimation of parameters and states of the dynamic system. However, in a full simulation based approach, treating the parameters as fixed causes the degeneracy of the filter. To solve this problem

Liu and West [46] propose to approximate the posterior distribution $p(\theta|\mathbf{y}_{1:t})$ with a particle set $\{x_t^i, \theta_t^i, w_t^i\}$ and to reconstruct the parameter posterior distribution at time $(t+1)$ through a Gaussian kernel density estimation

$$
\begin{aligned}
& p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t+1}) \propto \\
& \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) p(\mathbf{x}_{t+1}|\theta_{t+1}, \mathbf{y}_{1:t}) p(\theta_{t+1}|\mathbf{y}_{1:t}) = \\
& = p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) \int_{\mathcal{X}} \int_{\Theta} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta_{t+1}) p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta_t) p(\theta_t|\mathbf{y}_{1:t}) \delta_{\theta_t}(\theta_{t+1}) d\theta_t d\mathbf{x}_t \approx \\
& \overset{Particle}{\approx} \sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta_{t+1}) \delta_{\theta_t}(\theta_{t+1}) w_t^i \delta_{\{(\mathbf{x}_t^i, \theta_t^i)\}}(d\mathbf{x}_t, d\theta_t) \approx \\
& \overset{Kernel}{\approx} \sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta_{t+1}) w_t^i \mathcal{N}(\theta_{t+1}|m_t^i, b\, V_t)
\end{aligned}
\tag{6.44}
$$

In this context, index t for parameters means that they are updated sequentially. Note that after particle approximation, another approximation has been introduced. The kernel reconstruction of the posterior, implies the substitution of the parameter transition density, $\delta_{\theta_t^i}(\theta_{t+1})$, by a Gaussian transition density $\mathcal{N}(\theta_{t+1}|m_t^i, b\, V_t)$. After the kernel reconstruction of the posterior density, a new set of particles can be generated by applying the APF algorithm to the states and to the parameters using the kernel posterior density estimate as parameters importance density. The reconstruction of the posterior distribution through Gaussian kernel density estimation is a technique introduced by West [64], [65] in order to obtain an Adaptive Importance Sampling algorithm. The use of an adapting importance function is particulary useful in the dynamic models, where the probability density function of the system can change over time.

Note that the posterior distribution is a mixture of distributions, that can be reparameterised, using an allocation variable i to indicate the mixture component

$$
p(\mathbf{x}_t, \theta_t, i) = p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta_{t+1}) w_t^i \mathcal{N}(\theta_{t+1}|m_t^i, b\, V_t) \tag{6.45}
$$

The main idea of APF applies here and the particle selection step is obtained by sampling the mixture index i together with states $\mathbf{x}_{t+1}$ and parameters θ_{t+1}. Sampling from the joint density (6.45) is obtained through importance sampling

with proposal density

$$q(\mathbf{x}_{t+1}, \theta_{t+1}, i|\mathbf{y}_{1:t+1}) = p(\mathbf{x}_{t+1}|\theta_{t+1}, \mathbf{x}_t^i)\mathcal{N}(\theta_{t+1}|m_t^i, b^2\, V_t)q(i|\mathbf{y}_{1:t+1}) \tag{6.46}$$

where the instrumental density, used to sample the random index, is $q(i|\mathbf{y}_{1:t+1}) = p(\mathbf{y}_{t+1}|\mu_{t+1}^i, m_t^i)w_t^i$. From previous assumptions on the proposal distribution, the weights updating equation is

$$\begin{aligned} w_{t+1}^j &\propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta_{t+1}^j)p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{ij}, \theta_{t+1}^j)\mathcal{N}(\theta_{t+1}^j|m_t^{ij}, b^2\, V_t)w_t^{ij}}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{ij}, m_t^{ij})p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{ij}, \theta_{t+1}^j)\mathcal{N}(\theta_{t+1}^j|m_t^{ij}, b^2\, V_t)w_t^{ij}} = \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta_{t+1}^j)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{ij}, m_t^{ij})}. \end{aligned} \tag{6.47}$$

with $j = 1, \ldots, N$. The algorithm avoids degeneracy by introducing diversity in particles. It is known that diversity produces the impoverishment of the information contained in particles. Thus Liu and West [46] propose a kernel shrinkage technique in order to reduce the effect of the artificial variability. The kernel density at time $t+1$ depends on the density at time t through the constraint on the conditional variance: $\mathbb{V}_t(\theta_{t+1}) = \mathbb{V}_t(\theta_t) \triangleq V_t$. It results that each component of the kernel density estimation of the posterior distribution is not centered on the particles,θ_t^i , but on the linear combination between particles and the empirical average of the particles value at the previous step

$$m_t^i = a\theta_t^i + (1-a)\bar{\theta}_t \tag{6.48}$$

In Appendix D, we give a proof of the kernel shrinkage relations given in equation (6.48), using standard theorems on the conditional normal distribution.
The resulting APF algorithm for states and parameters estimation is

Algorithm 6.3.1. *(see Liu and West [46])*
Given the initial set of particles $\{\mathbf{x}_t^j, \theta_t^j, w_t^j\}_{j=1}^N$, *for* $j = 1, \ldots, N$

1. *Calculate* $\mu_{t+1}^j = \mathbb{E}(\mathbf{x}_{t+1}|\mathbf{x}_t^j, \theta_t^j)$ *and* $m_t^j = a\theta_t^j + (1-a)\bar{\theta}_t$
2. *Simulate* $i^j \sim q(i|\mathbf{y}_{1:t+1}) \propto w_t^i\, p(\mathbf{y}_{t+1}|\mu_{t+1}^i, m_t^i)$ *with* $i \in \{1, \ldots, N\}$
3. *Simulate* $\theta_{t+1}^j \sim p(\theta_{t+1}|\theta_t^{ij}) = \mathcal{N}(\theta_{t+1}; m_t^{ij}, (1-a^2)V_t)$

4. *Simulate* $\mathbf{x}_{t+1}^j \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{i^j}, \theta_{t+1}^j)$

5. *Update particles weights:* $w_{t+1}^j \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta_{t+1}^j)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, m_t^{i^j})}$.

In Appendix D we give a proof of the weights updating relation. Although this filtering approach does not explicitly assume that parameters vary over time, the dynamic nature of the parameters results implicitly from the structure of the filtering algorithm. It is possible to show (see Appendix D), that the proposed filtering approach assumes time varying parameters with a Gaussian transition density. Note however that the particle filter algorithm uses an approximation of the parameter posterior distribution and maintains this approximation both in the importance density and also in the weight updating relation. In principle an exact weight updating relation must be determined and the approximation errors must be taken into account, before they accumulate and produce poor parameter estimates. Thus a weight correction step would be needed, which can be considered a variant of the Rao-Blackwellization argument (Casella and Robert [11]). In particular, consider the true parameter posterior distribution and look at the kernel density approximation as a way to obtain an adapting importance function, then the exact weights updating can be determined as follows

$$w_{t+1}^j \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta_{t+1}^j) p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{i^j}, \theta_{t+1}^j) w_t^{i^j} p(\theta_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, m_t^{i^j}) p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{i^j}, \theta_{t+1}^j) \mathcal{N}(\theta_{t+1}^j|m_t^{i^j}, b^2 V_t) w_t^{i^j}}. \tag{6.49}$$

where the parameter posterior distribution is known from relation (6.42) and can be approximated through particle filter

$$\begin{aligned} p(\theta_{t+1}|\mathbf{y}_{1:t}) &= \int_{\bigotimes_{k=1}^{t} \mathcal{X}_k} p(\theta|\mathbf{x}_{0:t}, \mathbf{y}_{1:t}) d\mathbf{x}_{0:t} \\ &\approx \sum_{i=1}^{N} \pi(\theta)\pi(x_0) \prod_{k=1}^{t} p(\mathbf{y}_k|\mathbf{x}_k^i, \theta) p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \theta) w_k^i \end{aligned} \tag{6.50}$$

This approximated weight updating is computationally unfeasible because of the high number of times $(t \times N)$ the transition and the measurement equations must be evaluated.

6.3.3 A Particle Filter Algorithm for MSSV Models

The general algorithm exhibited in the previous section applies both to the Gaussian model $\mathcal{M}_1$ and to heavy tail models $\mathcal{M}_2$ and $\mathcal{M}_3$. Lopes [47] gives a version of the algorithm for the gaussian model $\mathcal{M}_1$. In the following we exhibit the algorithm for a Student-t model. Remember that $\theta = (\alpha_1, \alpha_2, \phi, \nu, p_{12}, p_{22}, \sigma^2)$, then APF is

Algorithm 6.3.2. *(APF for Student-t MSSV model)*
Given an initial set of particles $\{\mathbf{x}_t^i, \theta_t^i, w_t^i\}_{i=1}^N$*:*

1. *Compute* $V_t = \sum_{j=1}^N (\theta_t^j - \bar{\theta}_t)(\theta_t^j - \bar{\theta}_t)' w_t^j$ *and* $\bar{\theta}_t = \sum_{j=1}^N \theta_t^j w_t^j$
2. *For* $j = 1, \ldots, N$*, update the following variables:*
   (a) $\tilde{s}_{t+1}^j = \underset{l \in 1,\ldots,k}{arg\,max}\, \mathbb{P}(s_{t+1} = l | s_t = s_t^j)$
   (b) $\mu_{t+1}^j = \alpha_{\tilde{s}_{t+1}^j}^j + \phi_t^j h_t^j$
   (c) $m_t^j = a\theta_t^j + (1-a)\bar{\theta}_t$
3. *For* $j = 1, \ldots, N$*:*
   (a) *Simulate* $k^j \in \{1, \ldots, N\}$ *with* $\mathbb{P}(k^j = l) \propto p(\mathbf{y}_{t+1} | \mu_{t+1}^l, m_t^l) w_t^l$
   (b) *Simulate* θ_{t+1}^j *from* $\mathcal{N}(m_t^{k^j}, b^2 V_t)$
   (c) *Simulate* $s_{t+1}^j \in \{1, \ldots, k\}$ *from* $\mathbb{P}(s_{t+1}^j = j | s_t^{k^j})$
   (d) *Simulate* h_{t+1}^j *from* $\mathcal{N}(\alpha_{s_{t+1}^j}^j + \phi_{t+1}^j h_t^j, (\sigma^2)_{t+1}^j)$
4. *Update weights* $w_{t+1}^j \propto p(\mathbf{y}_{t+1} | h_{t+1}^j, \nu_{t+1}^j) / p(\mathbf{y}_{t+1} | \mu_{t+1}^{k^j}, \nu_{t+1}^{k^j})$

Note that the model $\mathcal{M}_2$ is more difficult to estimate because the degrees of freedom ν determine the tail heaviness of the observable process. This element makes the weight updating relation more sensitive to the evolution of the parameters.
For the model $\mathcal{M}_3$ we propose the following adaptation of the algorithm of Liu and West [46]. In order to obtain an integral representation of the α-stable density, we introduce an auxiliary (or completing) variable $\mathbf{z}_t$. Then we suggest to approximate the integral by simulating $\mathbf{z}_t$ from its conditional distribution.

Algorithm 6.3.3. - *(APF for stable MSSV model)*
Given an initial set of particles $\{\mathbf{x}_t^i, \theta_t^i, w_t^i\}_{i=1}^N$*:*

1. *Compute* $V_t = \sum_{j=1}^{N}(\theta_t^j - \bar{\theta}_t)(\theta_t^j - \bar{\theta}_t)' w_t^j$ *and* $\bar{\theta}_t = \sum_{j=1}^{N} \theta_t^j w_t^j$

2. *For* $j = 1, \ldots, N$, *update the following variables:*

   (a) $\tilde{s}_{t+1}^j = \underset{l \in 1,\ldots,k}{arg\,max}\, \mathbb{P}(s_{t+1} = l | s_t = s_t^j)$

   (b) $\mu_{t+1}^j = \alpha_{\tilde{s}_{t+1}^j}^j + \phi_t^j h_t^j$

   and $m_t^j = a\theta_t^j + (1-a)\bar{\theta}_t$

3. *For* $j = 1, \ldots, N$*:*

   (a) *Simulate* $k^j \in \{1, \ldots, N\}$ *with* $\mathbb{P}(k^j = l) \propto \sum_{r=1}^{R} p(\mathbf{y}_{t+1}, z_t^{l,r} | \mu_{t+1}^l, m_t^l) w_t^l$

   (b) *Simulate* θ_{t+1}^j *from* $\mathcal{N}(m_t^{k^j}, b^2 V_t)$

   (c) *Simulate* $s_{t+1}^j \in \{1, \ldots, k\}$ *with* $\mathbb{P}(s_{t+1}^j = j | s_t^{k^j})$

   (d) *Simulate* h_{t+1}^j *from* $\mathcal{N}(\alpha_{s_{t+1}^j}^j + \phi_{t+1}^j h_t^j, (\sigma^2)_{t+1}^j)$

   (e) *Simulate* $z_{t+1}^{j,r}$ *from* $f(z | \mathbf{y}_{t+1}, h_{t+1}^j, m_t^j)$ *for* $r = 1, \ldots, R$

4. *Update weights* $w_{t+1}^j \propto \frac{\sum_{r=1}^{R} p(\mathbf{y}_{t+1}, z_{t+1}^{j,r} | h_{t+1}^j, \theta_{t+1}^j)}{\sum_{r=1}^{R} p(\mathbf{y}_{t+1}, z_{t+1}^{kj,r} | h_{t+1}^{kj}, \mu_{t+1}^{kj})}$

Note however that the numerical approximation of the stable density introduces further errors in the algorithm and the parameter estimation becomes difficult.

6.3.4 Convergence of the Particle Filter Algorithms

If we assume the parameter vector is a stochastic process with a Markovian transition kernel, then the particle filters developed for joint state filtering and parameter estimation converge a.s.. In fact, the dynamic models and particle filters studied in previous sections, satisfy the assumptions required for the a.s. convergence of the empirical posterior density to the true posterior

$$\hat{p}(x_{0:t}, \theta_{0:t} | y_{1:t}) \overset{a.s.}{\to} p(x_{0:t}, \theta_{0:t} | y_{1:t}) \tag{6.51}$$

The necessary assumptions for the a.s. convergence of quite general sequential Monte Carlo algorithms are in Crisan and Doucet [16]. The proofs of these convergence results are based on the rigourous mathematical analysis of the convergence of empirical densities, produced by Crisan [15]. See also Crisan and Doucet [17] for a useful survey on the convergence results on particle filters.

6.4 Simulation Study

In the following we verify the efficiency of the Auxiliary Particle Filter algorithm exhibited in the previous section through some applications on synthetic data. In the Tab. 6.1 we show the effect of the number of particles on the parameter estimates. An higher number of particles improves the precision of the estimates, overall for the parameters α_1, α_2 and ϕ. For all the experiments we use as prior a multivariate Gaussian distribution centered near the true parameters value. We try other initial values and find that for quite all starting value the APF estimates are close to the true parameters value. The result is not robust with respect all parameter setting. In that case the choice of the parameter δ related to the kernel shrinkage becomes important. There is a tradeoff between the high level of artificial noise (controlled by the parameter δ) which allows to explore the parameters space and the efficiency of the parameter estimates.

Table 6.1: Gaussian model $\mathcal{M}_1$. APF parameter estimates for an increasing number, M, of particles. Estimates on T=1,000 observations.

θ	True	$\hat{\theta}_{APF}$			
		M=100	M=250	M=1000	M=5000
α_1	-2.5	-3.271	-2.395	-2.242	-2.133
α_2	-1.0	-0.745	-1.011	-0.914	-0.923
ϕ	0.5	0.373	0.614	0.573	0.524
σ^2	0.1	0.006	0.255	0.376	0.354
p_{12}	0.010	0.197	0.249	0.014	0.126
p_{22}	0.975	0.832	0.862	0.974	0.877

In Appendix E we represent the result for the Gaussian model $\mathcal{M}_1$ on a sample of $T = 1,000$ observations and with a $M = 5,000$ constant size particle set. In Fig. 6.9 we exhibit the updating of the parameter estimate. In the same figure we report the quantiles at 2.75% and 97.5% of the parameter posterior distribution. Quantiles do not diverge and in some cases approach progressively to the parameter estimate indicating good estimates convergence over the filtering iterations. Results on the time evolution of the approximated posterior distribution for the parameters are represented in Fig. 6.10 to Fig. 6.15. The filtered hidden states are represented in Fig. 6.8.

The absence of degeneracy has been verified by estimating both the *survival rate* (see Fig. 6.6) and the *effective sample size* indicator (see Fig. 6.7). Survival rate measures the fraction of particles survived to the selection step with respect to the total number of particles in the set (see Appendix E). The survival rate reveals particle degeneracy when exhibiting a persistent high number of dead particles from a generation to the subsequent one.
The Fig. 6.6 exhibits the evolution over time of the survival rate of a set of $N = 5,000$ particles. Although for some filter iterations the rate falls under the 30% level, it does not remain persistently under that level. We can conclude that the filter does not show degeneracy problems.

The effective sample size is a function of the variance of the particle weight (see Appendix E). This degeneracy measure is less than or equal to N. It is equal to N when the importance function is exactly equal to the filtering density and tends to zero when the variance of the importance weights tends to infinity, this is when particle filter degenerates. Fig. 6.7 shows the estimated ESS relative to the particle filter applied to the gaussian model. Observe that the effective sample size varies over time, but it never stabilizes at zero. Thus we conclude again in favour of a non-degeneracy of our particle filter for the Gaussian model.

We apply particle filter to estimate the Student-t model $\mathcal{M}_2$. Estimation results for an increasing number of particles are represented in Tab. 6.2.

Table 6.2: Student-t model $\mathcal{M}_2$. APF parameter estimates for an increasing number, M, of particles. Estimates on T=1,000 observations.

θ	True	$\hat{\theta}_{APF}$			
		M=100	M=250	M=1000	M=5000
ν	8	7.563	8.851	9.534	7.927
α_1	-2.5	-1.437	-2.133	-2.051	-2.236
α_2	-1.0	-0.599	-0.662	-0.577	-0.973
ϕ	0.5	0.714	0.685	0.709	0.603
σ^2	0.1	0.080	0.039	0.038	0.083
p_{12}	0.010	0.107	0.024	0.210	0.101
p_{22}	0.975	0.890	0.881	0.877	0.890

We conclude that the algorithm need of an higher number of particles to produce better parameter estimates. Moreover the results obtained for both the Gaussian and

the Student-t models need further evaluation studies. In particular the sensitivity of the parameter estimates to the value of the transition probabilities p_{11} and p_{00} need to be studied.

6.5 Conclusion

Following some suggestions present in the literature on the SV models, in this work we develop heavy tail Markov switching stochastic volatility models. We discuss stationary conditions of MSSV models and in order to make inference we follow a recent literature on the simulation based approach. In particular we focus on the parameters and states sequential learning problem. We assume time varying parameters and apply the auxiliary particle filter algorithm. We test the efficiency of such algorithm on Student-t MSSV model, through some simulation studies.

Appendix A - Stationarity Conditions

In this appendix we derive some stationarity conditions for the stochastic volatility process with jumps. In particular we assume that the log-volatility depends on a Markov jump process $\{s_t\}_{t\in\mathbb{N}}$, which takes value in the finite countable state space $E = \{0, \dots, L\}$. Moreover define the following MSSV model

$$y_t = e^{h_t/2}\varepsilon_t \tag{6.52}$$

$$h_t = \alpha_{s_t} + \phi h_{t-1} + \sigma_\eta \eta_t \tag{6.53}$$

$$\alpha_{s_t} = \alpha + \beta s_t \tag{6.54}$$

$$\mathbb{P}(s_t = i | s_{t-1} = j) = p_{ij} \tag{6.55}$$

with $p_{ij} \leq 0,\ \forall\, i, j \in E$ and $\sum_{l=1}^{L} p_{il} \leq 1$. In the following we assume that $E = \{0, 1\}$ and the initial state s_0 of the process has probability measure μ_0. The transition probabilities $\{p_{ij}\}_{i,j\in E}$ define the following transition matrix:

$$\mathbb{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \tag{6.56}$$

Note that through the transition matrix and the initial probability measure, the Markov jump process is well defined due to the following theorem.

Theorem 6.5.1. *(Existence and uniqueness of the jump Markov process)*
Given the transition matrix $\mathbb{P}$ and the initial probability measure μ_0, there exists a unique probability measure $\mathbf{P}$ defined on the canonical space:

$$\Omega = E^{\mathbb{N}},\ \omega = (\omega_t)_{t\geq 0},\ s_t(\omega) = \omega_t \tag{6.57}$$
$$\mathcal{F}_t = \sigma(s_k,\ k \leq t),\ \mathcal{F} = \sigma(s_k,\ k \geq 0)$$

such that the process $((F), (\mathcal{F}_t)_{t\geq 0}, (s_t)_{t\geq 0}, \mathbf{P})$ verifies:

$$\mathbf{P}(s_0 = a_0, \dots, s_t = a_t) = \mu_0(a_0)\mathbb{P}(a_0, a_1)\dots\mathbb{P}(a_{t-1}, a_t). \tag{6.58}$$

Such a process is the Markov process defined by the initial measure μ_0 and by the transition matrix $\mathbb{P}$.

Proof. For a proof see for example Baldi et al.[5]).

□

A sufficient condition for the stationarity of the observable process with stochastic volatility is the stationarity of the hidden stochastic log-volatility process.

Theorem 6.5.2. *(Second order stationarity conditions)*
Given the MSSV process defined in equations (6.52)-(6.55), if the innovation process ε_t and the hidden process h_t are first order stationary then y_t is second order stationary.

Proof. By Jensen's inequality and under the independence assumption between h_t and ε_t:

$$\ln(\mathbb{E}_\pi(y_t^2)) \leq \mathbb{E}_\pi(\ln(y_t^2)) = \mathbb{E}_\pi(\ln(e^{h_t})) + \mathbb{E}_\pi(\varepsilon_t^2)$$

and if $\mathbb{E}_\pi(h_t^2)$ and $\mathbb{E}_\pi(\varepsilon_t^2)$ are finite when $t \to \infty$, then the previous quantity is finite. We conclude that $\mathbb{E}_\pi(y_t^2)$ is finite when $t \to \infty$.

□

The stationarity analysis is based on some basic properties of transition matrix of the jump Markov process and on more general conditions for the existence and the uniqueness of the ergodic distribution of the Markov chain $\{h_t, s_t\}_{t\in\mathbb{N}}$. In particular we are interested in the evaluation of the n-times composition of the transition matrix and in the ergodic probability of the Markov process $\{s_t\}_{t\in\mathbb{N}}$. In the next sections we find also first and second order moments of the hidden log-volatility process with switching and show the existence and uniqueness of the ergodic distribution.

Appendix A.1 - Transition Matrix Properties

Theorem 6.5.3. *(Transition matrix composition)*
Given the transition matrix $\mathbb{P}$*, the n-time composition is denoted by* $\mathbb{P}^n = \overbrace{\mathbb{P} \circ \mathbb{P} \circ \ldots \circ \mathbb{P}}^{n-times}$ *and is defined by the following equation:*

$$\mathbb{P}^n = \frac{1}{p_{01} + p_{11}} \begin{pmatrix} p_{10} & p_{01} \\ p_{10} & p_{01} \end{pmatrix} + \frac{(1 - p_{01} - p_{10})^n}{p_{10} + p_{01}} \begin{pmatrix} p_{01} & -p_{01} \\ -p_{10} & p_{10} \end{pmatrix} \tag{6.59}$$

Proof. In order to calculate the n-times composition of the transition matrix, we find the eigenvalues and the eigenvectors of $\mathbb{P}$. The eigenvalues are the roots of the characteristic polynomial associated to the transition matrix. The correspondent characteristic equation is:

$$|\lambda\, I - \mathbb{P}| = \begin{vmatrix} \lambda - p_{00} & p_{01} \\ p_{10} & \lambda - p_{11} \end{vmatrix} = 0 \quad \Rightarrow \quad (\lambda - p_{11})(\lambda - p_{00}) - p_{01}p_{10} = 0. \tag{6.60}$$

The characteristic equation is of second order and has the following solutions:

$$\lambda_1 = 1, \quad \lambda_2 = 1 - p_{01} - p_{10}. \tag{6.61}$$

To find the eigenvectors, consider that one eigenvector is the unit vector: $e_1 = (1, 1)$. The second eigenvector e_2 is defined by:

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (1 - p_{01} - p_{10}) \begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \quad x\, p_{10} + y\, p_{01} = 0 \tag{6.62}$$

which is satisfied by the vector $e_2 = (-p_{01}, p_{10})$. Through the eigenvalues and the eigenvector the transition matrix can be diagonalised: $\mathbb{P} = R\Lambda R^{-1}$, where $R = (e_1\, e_2)$ is the eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2\}$ is a diagonal matrix of eigenvalues. Thus the n-times composition of the transition matrix can be written as: $\mathbb{P}^n = R\Lambda^n R^{-1}$ and more explicitly:

$$
\begin{aligned}
\mathbb{P}^n &= R\Lambda^n R^{-1} = \\
&= \frac{1}{p_{01}+p_{10}} \begin{pmatrix} 1 & -p_{01} \\ 1 & p_{10} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1-p_{10}-p_{01})^n \end{pmatrix} \begin{pmatrix} p_{10} & p_{01} \\ -1 & 1 \end{pmatrix} = \\
&= \frac{1}{p_{01}+p_{10}} \begin{pmatrix} 1 & -p_{01}(1-p_{10}-p_{01})^n \\ 1 & p_{10}(1-p_{10}-p_{01})^n \end{pmatrix} \begin{pmatrix} p_{10} & p_{01} \\ -1 & 1 \end{pmatrix} = \\
&= \frac{1}{p_{01}+p_{10}} \begin{pmatrix} p_{10}+p_{01}(1-p_{10}-p_{01})^n & p_{01}-p_{01}(1-p_{10}-p_{01})^n \\ p_{10}-p_{10}(1-p_{10}-p_{01})^n & p_{01}+p_{10}(1-p_{10}-p_{01})^n \end{pmatrix} = \\
&= \frac{1}{p_{01}+p_{10}} \begin{pmatrix} p_{10} & p_{01} \\ p_{10} & p_{01}) \end{pmatrix} + \frac{(1-p_{10}-p_{01})^n}{p_{01}+p_{10}} \begin{pmatrix} p_{01} & -p_{01} \\ -p_{10} & p_{10} \end{pmatrix}
\end{aligned}
\tag{6.63}
$$

and this concludes the proof.

□

Another quantity which is useful in stationarity analysis is the ergodic distribution $\pi^T = (\pi_0\, \pi_1)$ of the jump Markov process. The stationary distribution is defined by the stationarity vectorial equation:

$$
\pi\mathbb{P} = \pi \quad \Rightarrow \quad \begin{pmatrix} \pi_0 \\ \pi_1 \end{pmatrix} \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} \pi_0 \\ \pi_1 \end{pmatrix} \tag{6.64}
$$

and by the constraint $\pi_0 + \pi_1 = 1$. The resulting stationary probability is:

$$
\begin{aligned}
\pi_0 &= \frac{p_{10}}{p_{01}+p_{10}} \\
\pi_1 &= \frac{p_{01}}{p_{01}+p_{10}}.
\end{aligned}
\tag{6.65}
$$

Appendix A.2 - First order moment

Observe that the autoregressive structure of the log-volatility process, see the equation (6.53), makes it dependent on the past history of the Markov jump process. This feature becomes evident after some recursive substitutions

$$\begin{aligned} h_t &= \alpha + \beta s_t + \phi h_{t-1} + \sigma_\eta \eta_t = \\ &= \alpha + \beta s_t + \phi(\alpha + \beta s_{t-1} + \phi h_{t-2} + \sigma_\eta \eta_{t-1}) + \sigma_\eta \eta_t = \\ &= \dots \\ &= \alpha \sum_{i=0}^{t-1} \phi^i + \beta \sum_{i=0}^{t-1} \phi^i s_{t-i} + \sigma_\eta \sum_{i=0}^{t-1} \phi^i \eta_{t-i} + \phi^t h_0. \end{aligned} \tag{6.66}$$

The system of stochastic difference equations (6.53), (6.54) and (6.55) admits an ergodic solution. In particular it is possible to find the ergodic solution for the process h_t.

Theorem 6.5.4. *(Ergodic solution)*
Assume that $h_0 = 0$ and $|\phi| < 1$, then the system of equations (6.53), (6.54) and (6.55), has the following ergodic solution h_t

$$h_t = \alpha \sum_{i=0}^{+\infty} \phi^i + \beta \sum_{i=0}^{+\infty} \phi^i s_{t-i} + \sigma_\eta \sum_{i=0}^{+\infty} \phi^i \eta_{t-i}. \tag{6.67}$$

Proof. Consider the process h_t

$$h_t = \alpha \sum_{i=0}^{+\infty} \phi^i + \beta \sum_{i=0}^{+\infty} \phi^i s_{t-i} + \sigma_\eta \sum_{i=0}^{+\infty} \phi^i \eta_{t-i}. \tag{6.68}$$

and suppose it is a solution of the system (6.53)-(6.55), then we show that it is still a solution of the system at time $t+1$

$$\begin{aligned} h_{t+1} &= \alpha + \beta y_t + \phi \left[\frac{\alpha}{1-\phi} + \beta \sum_{i=0}^{+\infty} \phi^i s_{t-1-i} + \sigma_\eta \sum_{i=0}^{+\infty} \phi^i \eta_{t-1-i} \right] + \sigma_\eta \eta_t = \\ &= \frac{-\phi\alpha + \alpha + \phi\alpha}{1-\phi} + \beta s_t + \beta \sum_{i=1}^{+\infty} \phi^i s_{t-i-1} + \beta s_{t-1} + \sigma_\eta \sum_{i=1}^{+\infty} \phi^i \eta_{t-1-i} + \sigma_\eta \eta_t = \\ &= \frac{\alpha}{1-\phi} + \beta s_t + \phi\beta \sum_{i=0}^{+\infty} \phi^i s_{t-i-1} + \phi\sigma_\eta \sum_{i=0}^{+\infty} \phi^i \eta_{t-i-1} + \sigma_\eta \eta_t = \\ &= \frac{\alpha}{1-\phi} + \beta \sum_{i=0}^{+\infty} \phi^i s_{t-i} + \sigma_\eta \sum_{i=0}^{+\infty} \phi^i \eta_{t-i} \stackrel{\Delta}{=} h_t \end{aligned} \tag{6.69}$$

□

We evaluate the asymptotic stationarity of the ergodic solution by calculating the moments of the process and by taking. Take the expectation of the process defined in (6.66) with respect to the ergodic probability π and consider the limit when $t \to +\infty$

$$\begin{aligned}\lim_{t\to+\infty} \mathbb{E}_\pi(h_t) &= \lim_{t\to+\infty} \left(\alpha \sum_{i=0}^{t-1} \phi^i + \beta \sum_{i=0}^{t-1} \phi^i \mathbb{E}_\pi(s_{t-i}) + \phi^t h_0 \right) = \\ &= \lim_{t\to+\infty} \left(\alpha \sum_{i=0}^{t-1} \phi^i + \beta \sum_{i=0}^{t-1} \phi^i \frac{p_{01}}{p_{01}+p_{10}} + \phi^t h_0 \right) = \\ &= \frac{\alpha}{1-\phi} + \frac{\beta}{1-\phi} \frac{p_{01}}{p_{01}+p_{10}}.\end{aligned} \tag{6.70}$$

where the expected value of the jump process is calculated with respect to the ergodic probability given in (6.65) $\mathbb{E}_\pi(s_{t-i}) = 0\,\pi_0 + 1\,\pi_1 = p_{01}/(p_{01}+p_{10})$.

Appendix A.3 - Second order moment

In order to evaluate the second order asymptotic stationarity of the log-volatility process, consider the variance of the process under the ergodic probability and take the limit when $t \to +\infty$

$$\begin{aligned}&\lim_{t\to+\infty} \mathbb{V}_\pi(h_t) = \\ &= \lim_{t\to+\infty} \left(\mathbb{V}_\pi(\beta \sum_{i=0}^{t-1} \phi^i s_{t-i}) + \sigma_\eta \sum_{i=0}^{t-1} \phi^{2i} \right) = \\ &= \lim_{t\to+\infty} \left(\beta^2 \sum_{i=0}^{t-1} \phi^{2i} \mathbb{V}_\pi(s_{t-i}) + 2\beta^2 \sum_{i<j} \phi^i \phi^j \mathbb{C}ov(s_{t-i}, s_{t-j}) + \sigma_\eta \sum_{i=0}^{t-1} \phi^{2i} \right)\end{aligned} \tag{6.71}$$

Under the assumption that $|\phi| < 1$, the first and third terms of the sum have finite limits and reduce respectively to

$$\lim_{t\to+\infty} \left(\beta^2 \sum_{i=0}^{t-1} \phi^{2i} \mathbb{V}_\pi(s_{t-i}) \right) = \frac{\beta^2}{1-\phi^2} \frac{p_{01}}{p_{01}+p_{10}} (1 - \frac{p_{01}}{p_{01}+p_{10}}) \tag{6.72}$$

and

$$\lim_{t\to+\infty} \sigma_\eta^2 \sum_{i=0}^{t-1} \phi^{2i} = \sigma_\eta^2 \frac{1}{1-\phi^2} \tag{6.73}$$

The covariance term becomes

$$
\begin{aligned}
& \lim_{t\to+\infty} 2\beta^2 \sum_{i<j} \phi^i \phi^j \mathbb{C}ov(s_{t-i}, s_{t-j}) = \\
= & \lim_{t\to+\infty} 2\beta^2 \sum_{i<j} \phi^i \phi^j \frac{p_{01}p_{10}}{(p_{10}+p_{01})^2}(1-p_{01}-p_{10})^{(j-i)} = \\
= & \lim_{t\to+\infty} \left(2\beta^2 \sum_{i=0}^{t-1} \sum_{j=i+1}^{t-1} \phi^i \phi^j \frac{p_{01}p_{10}}{(p_{10}+p_{01})^2}(1-p_{01}-p_{10})^{(j-i)} \right) = \\
= & \lim_{t\to+\infty} \left(2\beta^2 \frac{p_{01}p_{10}}{(p_{10}+p_{01})^2} \sum_{i=0}^{t-1} \sum_{j=1}^{t-1-i} \phi^{2i}\phi^j (1-p_{01}-p_{10})^j \right) = \\
= & \lim_{t\to+\infty} \left(2\beta^2 \frac{p_{01}p_{10}}{(p_{10}+p_{01})^2} \sum_{i=0}^{t-1} \phi^{2i} \frac{1-[\phi(1-p_{01}-p_{10})]^{t-1-i}}{1-\phi(1-p_{01}-p_{10})} \phi(1-p_{01}-p_{10}) \right) = \\
= & \lim_{t\to+\infty} \left(\frac{2\beta^2 p_{01}p_{10}}{(p_{10}+p_{01})^2} \frac{\phi(1-p_{01}-p_{10})}{1-\phi(1-p_{01}-p_{10})} \left\{ \sum_{i=0}^{t-1} \phi^{2i} - \sum_{i=0}^{t-1} \phi^{2i} [\phi(1-p_{01}-p_{10})]^{t-i-1} \right\} \right) = \\
= & \lim_{t\to+\infty} \left(\frac{2\beta^2 p_{01}p_{10}}{(p_{10}+p_{01})^2} \frac{\phi(1-p_{01}-p_{10})}{1-\phi(1-p_{01}-p_{10})} \left\{ \frac{1-\phi^{2t}}{1-\phi^2} - \frac{[\phi(1-p_{01}-p_{10})]^t - \phi^{2t}}{\phi(1-p_{01}-p_{10})-\phi^2} \right\} \right) = \\
= & \frac{2\beta^2}{1-\phi^2} \frac{p_{01}p_{10}}{(p_{10}+p_{01})^2} \frac{\phi(1-p_{01}-p_{10})}{1-\phi(1-p_{01}-p_{10})}.
\end{aligned}
\tag{6.74}
$$

The last equation has been obtained under the following stationarity conditions: $|\phi| < 1$ and $|\phi(1-p_{01}-p_{10})| < 1$. The first condition is required for the stationarity of the variance term. The second condition is satisfied due to the existence of the ergodic probability of the jump process. Note that the auto-covariance of the jump Markov process has been calculated through the equation (6.59)

$$
\begin{aligned}
\mathbb{C}ov_\pi(s_{t-i}, s_{t-j}) &= \mathbb{E}_\pi(s_{t-i}s_{t-j}) - \mathbb{E}_\pi(s_{t-i})\mathbb{E}_\pi(s_{t-i}) = \\
&= \mathbb{P}_\pi(s_{t-i}=1)\mathbb{P}_\pi(s_{t-j}=1) - \mathbb{E}_\pi(s_{t-i})\mathbb{E}_\pi(s_{t-i}) = \\
&= \mathbb{P}_\pi(s_{t-i}=1)\mathbb{P}^{j-i}(1,1) - (\frac{p_{01}}{p_{01}+p_{10}})^2 = \\
&= \frac{p_{01}}{p_{01}+p_{10}}(p_{01}+p_{10}(1-p_{01}-p_{10})^{j-i}) - (\frac{p_{01}}{p_{01}+p_{10}})^2 = \\
&= \frac{p_{01}p_{10}}{p_{01}+p_{10}}(1-p_{01}-p_{10})^{j-i}.
\end{aligned}
\tag{6.75}
$$

with $i \leq j$.

Finally we check the stationarity of the autocovariance function of the process. Assume that $\tau \leq t-1$, then

$$
\begin{aligned}
&\mathbb{C}ov_\pi(h_t, h_{t+\tau}) = \\
&= \mathbb{C}ov_\pi(\beta \sum_{i=0}^{t-1} \phi^i s_{t-i} + \sigma_\eta \sum_{i=0}^{t-1} \phi^i \eta_{t-i}, \beta \sum_{i=0}^{t+\tau-1} \phi^i s_{t+\tau-i} + \sigma_\eta \sum_{i=0}^{t+\tau-1} \phi^i \eta_{t+\tau-i}) = \quad (6.76)\\
&= \mathbb{C}ov_\pi(\beta \sum_{i=0}^{t-1} \phi^i s_{t-i}, \beta \sum_{i=0}^{t+\tau-1} \phi^i s_{t+\tau-i}) + \mathbb{C}ov_\pi(\sigma_\eta \sum_{i=0}^{t-1} \phi^i \eta_{t-i}, \sigma_\eta \sum_{i=0}^{t+\tau-1} \phi^i \eta_{t+\tau-i}) = \\
&= \mathbb{C}ov_\pi(\beta \sum_{i=0}^{t-1} \phi^i s_{t-i}, \beta \sum_{i=0}^{t+\tau-1} \phi^i s_{t+\tau-i}) + \sigma_\eta^2 \phi^\tau \frac{1-\phi^{2t}}{1-\phi^2} = \\
&= \beta^2 \sum_{i=0}^{t-1} \sum_{j=0}^{t+\tau-1} \phi^i \phi^j \mathbb{C}ov_\pi(s_{t-i}, s_{t+\tau-j}) + \sigma_\eta^2 \phi^2 \frac{1-\phi^{2t}}{1-\phi^2} = \quad (6.77)\\
&= \beta^2 \sum_{i=0}^{t-1} \sum_{j=\tau}^{t-1} \phi^i \phi^{j-\tau} \mathbb{C}ov_\pi(s_{t-i}, s_{t-j}) + \sigma_\eta^2 \phi^2 \frac{1-\phi^{2t}}{1-\phi^2} = \\
&= \beta^2 \sum_{i=0}^{t-1} \sum_{j=\tau}^{t-1} \phi^i \phi^{j-\tau} \frac{p_{01}p_{10}}{p_{01}+p_{10}} (1-p_{01}-p_{10})^{|j-i|} + \sigma_\eta^2 \phi^2 \frac{1-\phi^{2t}}{1-\phi^2}
\end{aligned}
$$

Previous quantity depends on t, thus we process is not second order stationary. Moreover the limit when $t \to +\infty$ is finite and depends only on τ, under the assumption that $|\phi(1-p_{01}-p_{10})| < 1$

$$
\begin{aligned}
&\frac{p_{01}p_{10}}{p_{01}+p_{10}} \beta^2 \sum_{i=0}^{+\infty} \sum_{j=\tau}^{+\infty} \phi^i \phi^{j-\tau} (1-p_{01}-p_{10})^{|j-i|} + \frac{\sigma_\eta^2 \phi^2}{1-\phi^2} = \quad (6.78)\\
&= \frac{p_{01}p_{10}}{p_{01}+p_{10}} \beta^2 \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \phi^i \phi^j (1-p_{01}-p_{10})^{|j+\tau-i|} + \frac{\sigma_\eta^2 \phi^2}{1-\phi^2} \leq \\
&\leq \frac{p_{01}p_{10}}{p_{01}+p_{10}} \beta^2 \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \phi^i \phi^j (1-p_{01}-p_{10})^{|j+\tau|+|i|} + \frac{\sigma_\eta^2 \phi^2}{1-\phi^2} = \\
&= \frac{p_{01}p_{10}}{p_{01}+p_{10}} \beta^2 \frac{(1-p_{01}-p_{10})^\tau}{(1-\phi(1-p_{01}-p_{10}))^2} + \frac{\sigma_\eta^2 \phi^2}{1-\phi^2} < +\infty
\end{aligned}
$$

It is possible to prove that the covariance is finite also in the case $\tau > t-1$.

After previous considerations, we conclude that the jump log-volatility process is asymptotically stationary of second order.

Appendix A.4 - Ergodic Distribution

In this part of the appendix we show that the second order stationarity conditions obtained in previous sections are necessary conditions for the existence and uniqueness of the ergodic distribution of the hidden Markov process $\{h_t, s_t\}_{t\in\mathbb{N}}$. On the stationarity conditions of a Markov switching functional autoregressive process, the only available results are due to Francq and Roussognol [26]. Francq and Zakoian [27] analyse stationarity conditions of a Markov-switching multivariate autoregressive moving average process.

In the following we will refer mainly to the work of Francq and Roussignol [26]. Introduce the following multivariate functional autoregressive process with values in $\mathbb{R}^S$

$$h_t = F(h_{t-1}, s_t, \theta) + G(\eta_t, s_t, \theta) \quad \forall t \geq 1 \tag{6.79}$$

where $\{\eta_t\}_{t\in\mathbb{N}}$ is a sequence of i.i.d. random processes, $\theta \in \Theta$ the parameters of the model and $\{s_t\}_{t\in\mathbb{N}}$ a discrete Markov chain independent of $\{\eta_t\}_{t\in\mathbb{N}}$, with values in the finite state space $E = \{1, 2, \ldots, L\}$ and with stationary transition probabilities $\mathbb{P}(s_t = j | s_t = i) = p_{ij}$. Then the following theorem holds.

Theorem 6.5.5. *(Existence and uniqueness of the ergodic probability of $\{x_t, s_t\}_{t\in\mathbb{N}}$) Suppose the following conditions*

(i) The Markov chain $\{s_t\}_t$ is irreducible and aperiodic;

(ii) For all $i \in E$ the random vector $G(\eta_t, i)$ has density $f_i(\cdot)$ with respect to the Lebesgue measure of $\mathbb{R}^S$ and $\mathbb{E}(||G(\eta_t, i)||) < \infty$ where $||\cdot||$ is the Euclidean norm;

(iii) There exist $a_1, a_2, \ldots, a_L$ such that $\forall\,(x, y) \in \mathbb{R}^S$, $||F(x, i) - F(y, i)|| \leq$

$a_i||x-y||$ and such that the matrix

$$Q = \begin{pmatrix} p_{1,1}a_1 & p_{2,1}a_1 & \cdots & p_{L,1}a_1 \\ p_{1,2}a_2 & p_{2,1}a_2 & \cdots & p_{L,2}a_2 \\ \vdots & \vdots & \vdots & \vdots \\ p_{1,L}a_L & p_{2,L}a_L & \cdots & p_{L,L}a_L \end{pmatrix} \tag{6.80}$$

has spectral radius strictly less than 1;

are satisfied. Then The Markov chain defined by Eq. (6.79) admits a unique invariant probability μ. The second marginal of μ is equal to the invariant probability of $\{s_t\}_{t\in\mathbb{N}}$. A stationarity Markov process $\{h_t, s_t\}_{t\in\mathbb{N}}$ satisfying (6.79) with μ as initial distribution is an aperiodic ergodic Harris process. Moreover, for all process $\{h_t, s_t\}_{t\in\mathbb{N}}$ satisfying (6.79) and all μ-integrable function g from $\mathbb{R}^S \times E$ to $\mathbb{R}$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_k, s_k) = \mu(g) \quad a.s. \tag{6.81}$$

Proof. For a proof see Francq and Roussignol [26]

□

The theorem applies to the hidden log-volatility process. In particular the assumption *(ii)* is satisfied because the random variable $G(\eta_t, s)$ has normal density with mean zero and finite variance σ_η.
The third assumption is also satisfied because

$$||F(x,s) - F(y,s)|| = ||\alpha + \beta s + \phi y - (\alpha + \beta s + \phi x)|| = |\phi|\,||y - x|| \tag{6.82}$$

thus $a_1 = a_2 = |\phi|$ and the spectral radius of

$$Q = \begin{pmatrix} p_{0,0}|\phi| & p_{1,0}|\phi| \\ p_{0,1}|\phi| & p_{1,0}|\phi| \end{pmatrix} \tag{6.83}$$

is $\lambda_1 = |\phi|$ and $\lambda_2 = |\phi||(1 - p_{01} - p_{10})|$. The assumption *(iii)* requires that $|\phi| < 1$ and $|(1 - p_{01} - p_{10})| < 1$. These conditions are satisfied if we require the second order stationarity of the process $\{s_t\}$, (see Appendix A.3).

Appendix B - Particles Weights when Parameters are Known

Recursive formulas are quite useful when processing data sequentially. In this appendix we give a proof of two recursive relations which are important for sequential Monte Carlo algorithms. We assume that parameters are known and denote with $p(\cdot\,|\,\cdot\,;\theta)$ a generic parametric density function with known parameters.

The recursive filtering formula given in equation (6.25) is obtained by applying Bayes' theorem and using the Markov property of the dynamic system (6.19), (6.20) and (6.21).

Proof. (*Recursive filtering relation*)

Consider the filtering density

$$\begin{aligned} p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta) &= \frac{p(\mathbf{x}_{0:t+1},\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} = \\ &= p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)\frac{p(\mathbf{x}_{t+1},\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\mathbf{x}_{0:t};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} = \\ &= p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)\frac{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\mathbf{x}_{0:t+1};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)}p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t};\theta) = \\ &= p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)\frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_{t};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} \end{aligned}$$

□

In the following we derive the recursive relations given in equations (6.29) and (6.30), which is used to update sequentially the weights of the particles.

Proof. (*Recursive weight updating relation*)

Starting from the definition of importance weights

$$
\begin{aligned}
w_{t+1} &\stackrel{\Delta}{=} \frac{p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta)} = \\
&\stackrel{Bayes}{=} \frac{p(\mathbf{x}_{0:t+1},\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta)p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} = \\
&= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)p(\mathbf{x}_{t+1},\mathbf{y}_{t+1}|\mathbf{y}_{1:t},\mathbf{x}_{0:t};\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta)p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} = \\
&= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta)} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{0:t+1},\mathbf{y}_{1:t};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t};\theta) = \quad (6.84) \\
&\stackrel{Markov}{=} \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)}{q(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1};\theta)} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_t;\theta) = \\
&= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t};\theta)} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t;\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)q(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t+1};\theta)} = \\
&= w_t \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t;\theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t};\theta)q(\mathbf{x}_{t+1}|\mathbf{x}_{0:t},\mathbf{y}_{1:t+1};\theta)}.
\end{aligned}
$$

Thus particle weights updating recursive relation is

$$
w_{t+1} \propto w_t \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t;\theta)}{q(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{t+1};\theta)}. \tag{6.85}
$$

Moreover, if we assume that the importance density is the prior on the hidden states, then $q(\mathbf{x}_{t+1}|\mathbf{x}_t,\mathbf{y}_{t+1};\theta) = p(\mathbf{x}_{t+1}|\mathbf{x}_t;\theta)$, equation (6.85) simplifies to

$$
w_{t+1} \propto w_t\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1};\theta). \tag{6.86}
$$

□

Appendix C - Particle Filter Algorithms

In the following we give a pseudo-code representation of two basic particle filter algorithms, i.e. the SIS, SIR and refer to Chapter 2 for further details on such algorithms and for the presentation of APF algorithm.

Algorithm 6.5.1. - *SIS Particle Filter*
Given the initial set of particles $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$*:*

1. *Simulate* $\mathbf{x}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}; \theta)$
2. *Update the weights:* $w_{t+1}^i \propto w_t^i \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i;\theta)\, p(\mathbf{x}_{t+1}|\mathbf{x}_t^i;\theta)}{q(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\mathbf{y}_{t+1};\theta)}$

Note that if the instrumental distribution $q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}; \theta)$ is set equal to $p(\mathbf{x}_{t+1}|\mathbf{x}_t^i; \theta)$ then the weights updating relation becomes: $w_{t+1}^i \propto w_t^i\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i; \theta)$. This is the natural choice, because this density represents a sort of prior at time t for the state x_{t+1}. However, as underlined in Pitt and Shephard [52] this strategy is sensitive to outliers. Crisan and Doucet [16] analyse the choice of that importance distributions, which satisfy the assumptions necessary to the a.s. convergence of the sequential Monte Carlo algorithm.

Algorithm 6.5.2. - *SIR Particle Filter*
Given the initial set of particles $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$*:*

1. *Simulate* $\mathbf{x}_{t+1}^i \sim q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}; \theta)$
2. *Update the weights:* $\tilde{w}_{t+1}^i \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i; \theta)$
3. *Normalize the weights:* $\bar{w}_{t+1}^i = \tilde{w}_{t+1}^i \,(\sum_{j=1}^N \tilde{w}_{t+1}^j)^{-1}$, *for* $i = 1, \ldots, N$.
4. *Simulate* $\{x_{t+1}^i\}_{i=1}^N$ *from the empirical density* $\{\mathbf{x}_t^i, \bar{w}_t^i\}_{i=1}^N$
5. *Assign* $w_{t+1}^i = 1/N$, *for* $i = 1, \ldots, N$.

Note that in the SIR particle filter, we assumed: $q(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}; \theta) = p(\mathbf{x}_{t+1}|\mathbf{x}_t^i; \theta)$. Moreover, due to the resampling step, the weights are uniformly distributed over the particle set: $w_t^i = 1/N$, thus the weights updating relation becomes: $\tilde{w}_{t+1}^i \propto w_t^i\, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i; \theta) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i; \theta)$.

Appendix D - Joint Estimation of States and Parameters

We show some analytical aspects of the joint estimation of the parameters and the states of the Bayesian dynamic model given in equations (6.37), (6.38), (6.39) and (6.40). Introduce the following notation for the conditional moments $\mathbb{V}_t(\cdot) = Var(\cdot|\mathbf{y}_{1:t})$, $\mathbb{C}_t(\cdot,\cdot) = Cov(\cdot,\cdot|\mathbf{y}_{1:t})$ and $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathbf{y}_{1:t})$. Denote with I the identity matrix.

Appendix D.1 - Parameters Transition Density

Assume that parameters evolve over time

$$\theta_{t+1} = \theta_t + \xi_{t+1}, \quad with \quad \xi_{t+1} \sim N(0, W_{t+1}). \tag{6.87}$$

Note that the noise component produces artificial variability in the posterior distribution of the parameters. In order to reduce the variability Liu and West [46] suggest to impose the following constraint on the variance-covariance matrix of the parameter $\mathbb{V}_t(\theta_{t+1}) = \mathbb{V}_t(\theta_t) = V_t$. It follows that

$$\begin{aligned} &\mathbb{V}_t(\theta_{t+1}) = \mathbb{V}_t(\theta_t) + \mathbb{V}_t(\xi_{t+1}) + 2\mathbb{C}_t(\xi_{t+1}, \theta_t) \Leftrightarrow \\ &\mathbb{C}_t(\xi_{t+1}, \theta_t) = -\frac{\mathbb{V}_t(\xi_{t+1})}{2} = -\frac{W_{t+1}}{2}, \end{aligned} \tag{6.88}$$

In order to control the transition of the parameters between time t and $(t+1)$ they use a technique of shrinkage between gaussian kernels. The resulting parameters *transition density* is a Gaussian. The shrinkage technique has already been used by West [65] in order to reconstruct the posterior distribution in an adaptive importance sampling scheme.

In the following we prove the result given in Eq. (6.48).

Proof. The joint density of θ_{t+1} and θ_t is a Gaussian density, characterised by the following moments

$$\mathbb{E}_t(\theta_t) \triangleq \bar{\theta}_t \tag{6.89}$$

$$\mathbb{E}_t(\theta_{t+1}) = \mathbb{E}_t(\xi_{t+1}) + \mathbb{E}_t(\theta_t) = \mathbb{E}_t(\theta_t) \triangleq \bar{\theta}_t \tag{6.90}$$

$$\mathbb{V}_t(\theta_t) = \mathbb{V}_t(\theta_{t+1}) = V_t \tag{6.91}$$

$$\mathbb{C}_t(\theta_{t+1}, \theta_t) = \mathbb{C}_t(\theta_t + \xi_{t+1}, \theta_t) = V_t + \mathbb{C}_t(\xi_{t+1}, \theta_t) = V_t - \frac{W_{t+1}}{2} \tag{6.92}$$

and by straightforward calculations, the distribution of θ_{t+1} conditional to θ_t is Gaussian, with following conditional mean and variance

$$\begin{aligned} \mathbb{E}_t(\theta_{t+1}|\theta_t) &= \bar{\theta}_t + (V_t - \frac{W_{t+1}}{2})V_t^{-1}(\theta_t - \bar{\theta}_t) \\ &= \bar{\theta}_t + (I - \frac{W_{t+1}}{2}V_t^{-1})\theta_t - (I - \frac{W_{t+1}}{2}V_t^{-1})\bar{\theta}_t \\ &= A_{t+1}\theta_t + (I - A_{t+1})\bar{\theta}_t \end{aligned} \tag{6.93}$$

where $A_{t+1} = (I - \frac{W_{t+1}}{2}V_t^{-1})$.

$$\begin{aligned} \mathbb{V}_t(\theta_{t+1}|\theta_t) &= \mathbb{V}_{\theta_t}(\theta_{t+1}) - \mathbb{C}_t(\theta_{t+1}, \theta_t)\mathbb{V}_t^{-1}(\theta_t)\mathbb{C}_t(\theta_{t+1}, \theta_t) \\ &= V_t - (V_t - \frac{W_{t+1}}{2})V_t^{-1}(V_t - \frac{W_{t+1}}{2}) \\ &= V_t - (I - \frac{W_{t+1}}{2}V_t^{-1})(I - \frac{W_{t+1}}{2}V_t^{-1})V_t \\ &= (I - A_{t+1}^2)V_t. \end{aligned} \tag{6.94}$$

Conclude that

$$p(\theta_{t+1}|\theta_t) = N(A_{t+1}\theta_t + (I - A_{t+1})\bar{\theta}_t, (I - A_{t+1}^2)V_t) \tag{6.95}$$

□

In order to simplify the estimation problem Liu and West [46] assume that the variance-covariance matrix of the noise is proportional to V_t and to a discount factor δ

$$W_{t+1} = V_t(\frac{1}{\delta} - 1) \tag{6.96}$$

Thus previous quantities become: $A_{t+1} = I\,\frac{3\delta-1}{2\delta}$, $\mathbb{V}_t(\theta_{t+1}|\theta_t) = (1 - (\frac{3\delta-1}{2\delta})^2)$ and $\mathbb{E}_t(\theta_{t+1}|\theta_t) = \frac{3\delta-1}{2\delta}\theta_t + (\frac{1-\delta}{2\delta})\bar{\theta}_t$. Denote $a = \frac{3\delta-1}{2\delta}$, then the distribution in equation (6.95) simplifies to:

$$p(\theta_{t+1}|\theta_t) = N(\theta_{t+1}; a\theta_t + (1-a)\bar{\theta}_t, (1-a^2)V_t) = N(\theta_{t+1}; m_t, (1-a^2)V_t) \quad (6.97)$$

Appendix D.2 - SIS Algorithm for State and Parameter Joint Estimation

The *sequential importance sampling* (SIS) particle filter is the starting point to understand and to develop other particle filters like the *auxiliary particle filter*. Thus, in the following we exhibit a basic SIS algorithm for the joint estimation of states $\{\mathbf{x}_t, t \in \mathbb{N}\}$, $x_t \in \mathcal{X}$ and parameters $\{\theta_t, t \in \mathbb{N}\}$, $\theta_t \in \Theta$. In the Bayesian model, given in equations (6.37)-(6.40), the parameters are fixed over time, but for estimation purposes we let parameters vary over time. In particular the proof in this appendix is based on the hypothesis that parameters evolution is described by a Gaussian random walk: $\theta_{t+1} = \theta_t + \epsilon_t$. We use the Liu and West's kernel shrinkage technique in order to reduce the effects on the parameters estimates of the artificial diversity introduced in the particle filter. We show also why the algorithm of Liu and West [46] can be view as a reinterpretation of a dynamic model with time varying parameters. In the following we give the pseudo-code representation of the algorithm and the proof of the weights updating relation.

Algorithm 6.5.3. - ***SIS for state and parameter estimation***
Given the initial set of particles $\{\mathbf{x}_t^i, \theta_t^i, w_t^i\}_{i=1}^N$, *for* $i = 1, \ldots, N$

1. *Simulate* $\theta_{t+1}^i \sim p(\theta_{t+1}|\theta_t^i)$

2. *Simulate* $\mathbf{x}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta_{t+1}^i)$

3. *Update the weights:* $w_{t+1}^i \propto w_t^i \, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta_{t+1}^i)$

Proof. (Recursive Weights Updating Relation)
Consider the joint posterior density of the parameters and the hidden states

$$\begin{aligned} p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t+1}) &\overset{Bayes}{=} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}, \mathbf{y}_{1:t}) p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} = \\ &\overset{Markov}{=} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} \propto \\ &\propto \quad p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t}) = \end{aligned} \tag{6.98}$$

$$\overset{Kolmog.}{=} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})\int_{\mathcal{X}}\int_{\Theta} p(\mathbf{x}_{t+1},\theta_{t+1}|\mathbf{x}_t,\theta_t,\mathbf{y}_{1:t})p(\mathbf{x}_t,\theta_t|\mathbf{y}_{1:t})d\theta_t d\mathbf{x}_t =$$

$$= p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})\int_{\mathcal{X}}\int_{\Theta} p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta_{t+1},\theta_t,\mathbf{y}_{1:t})p(\theta_{t+1}|\mathbf{x}_t,\theta_t,\mathbf{y}_{1:t})p(d\mathbf{x}_t,d\theta_t|\mathbf{y}_{1:t}) =$$

$$= p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})\int_{\mathcal{X}}\int_{\Theta} \overbrace{p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta_{t+1})p(\theta_{t+1}|\theta_t)}^{Joint\ transition\ density} p(\mathbf{x}_t,\theta_t|\mathbf{y}_{1:t})d\theta_t d\mathbf{x}_t$$

Observe that the joint transition density is expressed as the product of the state transition density, conditional to the parameters, and the parameters transition density. At time t the parameters transition density can be chosen to be a normal distribution centered on the previous value of the particle: θ_t^i, but this choice produces higher variability in parameter estimates. In order to solve the problem, Liu and West [46] use a Gaussian kernel shrinkage technique, which leads to more stable estimate. The resulting transition density is the Gaussian distribution in equations (6.95) and (6.97), with mean and variance estimated on the simulated posterior distribution.

Assume to have, at time t, a set of particles $\{\mathbf{x}_t^i,\theta_t^i,w_t^i\}_{i=1}^N$, which approximates the prior distribution $p(\mathbf{x}_t,\theta_t|\mathbf{y}_{1:t})$. The resulting empirical distribution is

$$p(\mathbf{x}_t,\theta_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N} w_t^i \delta_{\{\mathbf{x}_t^i,\theta_t^i\}}(d\mathbf{x}_t,d\theta_t) \tag{6.99}$$

and the last equation in (6.98) can be approximated as follows

$$\sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})p(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\theta_{t+1})p(\theta_{t+1}|\theta_t^i)w_t^i\delta_{\{\mathbf{x}_t^i,\theta_t^i\}}(d\mathbf{x}_t,d\theta_t) \tag{6.100}$$

In SIS particle filter, the new set of particles $\{\mathbf{x}_{t+1}^i,\theta_{t+1}^i,w_{t+1}^i\}_{i=1}^N$ is generated by simulating each pair $\{\mathbf{x}_{t+1}^i,\theta_{t+1}^i\}$ from the instrumental density $q(\mathbf{x}_{t+1},\theta_{t+1}|\mathbf{y}_{1:t+1})$. The weights updating equation is determined by an importance sampling argument. Choose the instrumental density to be the product of the priors of θ_{t+1} and $\mathbf{x}_{t+1}$:

$$q(\mathbf{x}_{t+1},\theta_{t+1}|\mathbf{y}_{1:t+1}) = p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta_{t+1})p(\theta_{t+1}|\theta_t) \tag{6.101}$$

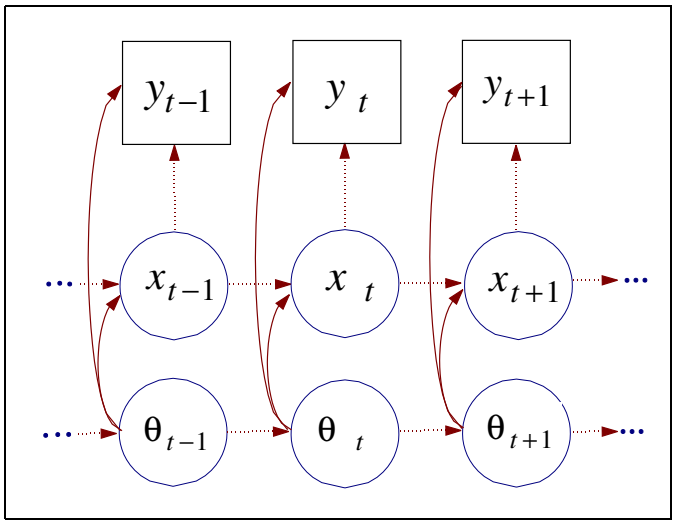

Figure 6.5: Causality structure of a Markovian dynamic model with hidden states and time varying parameter. A box around the variable indicates the variable is known, while a circle indicates a hidden variable.

then the weights updating equation is given by the following correction step

$$\begin{aligned} w_{t+1}^i &\propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta_{t+1}^i) p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \theta_{t+1}^i) p(\theta_{t+1}^i|\theta_t^i) w_t^i}{p(\mathbf{x}_{t+1}^i|\mathbf{x}_t^i, \theta_{t+1}^i) p(\theta_{t+1}^i|\theta_t^i)} = \\ &= w_t^i \, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i, \theta_{t+1}^i). \end{aligned} \tag{6.102}$$

□

Appendix D.3 - APF Algorithm for States and Parameters Joint Estimation

Auxiliary Particle Filter can be derived from the basic SIS algorithm, exhibited in the previous appendix. APF uses the auxiliary variable j to select randomly particles and to mutate selected particles. The auxiliary variable is simulated from a distribution, which summarizes and conserves the information contained in previous particle set. This feature is obtained also by using the variable μ_t. In this way the re-sampling step does not cause the impoverishment of the information contained in the actual particle set.

Algorithm 6.5.4. - *APF for states and parameters estimation (see Liu and West [46])*
Given the initial set of particles $\{\mathbf{x}_t^j, \theta_t^j, w_t^j\}_{j=1}^N$, *for* $j = 1, \dots, N$

1. *Calculate* $\mu_{t+1}^j = \mathbb{E}(\mathbf{x}_{t+1}|\mathbf{x}_t^j, \theta_t^j)$ *and* $m_t^j = a\theta_t^j + (1-a)\bar{\theta}_t$
2. *Simulate* $i^j \sim q(i|\mathbf{y}_{1:t+1}) \propto w_t^i\, p(\mathbf{y}_{t+1}|\mu_{t+1}^i, m_t^i)$ *with* $i \in \{1, \dots, N\}$
3. *Simulate* $\theta_{t+1}^j \sim p(\theta_{t+1}|\theta_t^{i^j}) = \mathcal{N}(\theta_{t+1}; m_t^{i^j}, (1-a^2)V_t)$
4. *Simulate* $\mathbf{x}_{t+1}^j \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{i^j}, \theta_{t+1}^j)$
5. *Update particles weights:* $w_{t+1}^j \propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta_{t+1}^j)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, m_t^{i^j})}$.

Proof. (Recursive Weights Updating Relation)
Consider the *filtering density* or joint posterior density for the parameters and the states

$$
\begin{aligned}
& p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t+1}) = \\
& \overset{Bayes}{=} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}, \mathbf{y}_{1:t}) p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} = \\
& \overset{Markov}{=} \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1}) p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})} \propto
\end{aligned}
\tag{6.103}
$$

$$
\begin{aligned}
&\propto \quad p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})p(\mathbf{x}_{t+1},\theta_{t+1}|\mathbf{y}_{1:t}) = \\
&\overset{Kolmog.}{=} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})\int_{\mathcal{X}}\int_{\Theta} p(\mathbf{x}_{t+1},\theta_{t+1}|\mathbf{x}_t,\theta_t,\mathbf{y}_{1:t})p(\mathbf{x}_t,\theta_t|\mathbf{y}_{1:t})d\theta_t d\mathbf{x}_t = \\
&= \quad p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})\int_{\mathcal{X}}\int_{\Theta} p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta_{t+1})p(\theta_{t+1}|\theta_t)p(\mathbf{x}_t,\theta_t|\mathbf{y}_{1:t})d\theta_t d\mathbf{x}_t = \\
&= \quad p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})\int_{\mathcal{X}}\int_{\Theta} \overbrace{p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta_{t+1})p(\theta_{t+1}|\theta_t)}^{Joint\ transition\ density} p(\mathbf{x}_t,\theta_t|\mathbf{y}_{1:t})d\theta_t d\mathbf{x}_t.
\end{aligned}
$$

Observe that the joint transition density is decomposed in the product of the state transition density conditional to the parameters and the parameters transition density. Liu and West [46] use a Gaussian kernel shrinkage technique, which provides more stable estimates. The resulting parameter transition density is the Gaussian distribution exhibited in equations (6.95) and (6.97).
Assume to have, at time t, a set of particles $\{\mathbf{x}_t^j, \theta_t^j, w_t^j\}_{j=1}^N$, which approximates the prior distribution $p(\mathbf{x}_t, \theta_t|\mathbf{y}_{1:t})$. The resulting empirical distribution is:

$$
p(\mathbf{x}_t, \theta_t|\mathbf{y}_{1:t}) \approx \sum_{j=1}^{N} w_t^j \delta_{\{\mathbf{x}_t^j, \theta_t^j\}}(d\mathbf{x}_t, d\theta_t) \tag{6.104}
$$

and the last equation in (6.103) can be approximated as follows

$$
\sum_{j=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})p(\mathbf{x}_{t+1}|\mathbf{x}_t^j,\theta_{t+1})p(\theta_{t+1}|\theta_t^j)w_t^j \delta_{\{\mathbf{x}_t^j,\theta_t^j\}}(d\mathbf{x}_t, d\theta_t) \tag{6.105}
$$

Note that the previous density is a mixture of distribution and in APF particle filter, it is reparameterised through the allocation variable i as follows: $p(\mathbf{x}_t, \theta_t, i) = p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_{t+1})p(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\theta_{t+1})p(\theta_{t+1}|\theta_t^i)w_t^i$. The index i represents the auxiliary variable and is sampled together with the new set of particles, according to the instrumental probability: $q(i|\mathbf{y}_{1:t+1}) = p(\mathbf{y}_{t+1}|\mu_{t+1}^i, m_t^i)w_t^i$, where μ_{t+1}^i is a variable which resumes the information contained in the particle set $\{\mathbf{x}_t^j, \theta_t^j, w_t^j\}_{j=1}^N$, and $m_t = \mathbb{E}_t(\theta_{t+1}|\theta_t)$ is the mean of the parameters transition density. Given the index i, the new set of particles $\{\mathbf{x}_{t+1}^j, \theta_{t+1}^j, w_{t+1}^j\}_{j=1}^N$ is generated by simulating $(\mathbf{x}_{t+1}^j, \theta_{t+1}^j)$ from the instrumental density $q(\mathbf{x}_{t+1}, \theta_{t+1}|i, \mathbf{y}_{1:t+1})$. The weights updating equation

is determined by an importance sampling argument. Choose the conditional instrumental density to be the product of the priors of θ_{t+1} and $\mathbf{x}_{t+1}$, given i, with $i = 1, \ldots, N$

$$q(\mathbf{x}_{t+1}, \theta_{t+1}|i, \mathbf{y}_{1:t+1}) = p(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{x}_t^i, \theta_t^i) = p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta_{t+1})p(\theta_{t+1}|\theta_t^i) \tag{6.106}$$

then the weights updating equation is given by the following correction step

$$\begin{aligned} w_{t+1}^j &\propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta_{t+1}^j)p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{ij}, \theta_{t+1}^j)p(\theta_{t+1}^j|\theta_t^{ij})w_t^{ij}}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{ij}, m_t^{ij})p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{ij}, \theta_{t+1}^j)p(\theta_{t+1}^j|\theta_t^{ij})w_t^{ij}} = \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta_{t+1}^j)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{ij}, m_t^{ij})}. \end{aligned} \tag{6.107}$$

with $j = 1, \ldots, N$.

□

Appendix E - Simulation Results

In the appendix we analyse the results of particle filter techniques when applied to the Gaussian MSSV model $\mathcal{M}_1$ and to the heavy tails models $\mathcal{M}_2$ and $\mathcal{M}_3$. We show how to detect the degeneracy problem and to check for the convergence of the parameters estimation.

Appendix E.1 - Gaussian MSSV model

The results of the particle filtering algorithm consist in the parameter estimates and in the log-volatility and Markov-switching state filtering. Fig. 6.9 shows parameter estimates for the Gaussian model $\mathcal{M}_1$ on a sample of $T = 1,000$ observations and with a $M = 5,000$ constant size particle set. In the same figure we report the quantiles at 2.75% and 97.5% of the parameter posterior distribution. Quantiles do not diverge and in some cases approach progressively to the parameter estimate indicating good estimates convergence over the filtering iterations.
To validate the estimation results a more accurate analysis of the particle filter outputs must be performed. It is well known that degeneracy is a very common problem in particle filtering, thus we monitor the time evolution of the particles set.

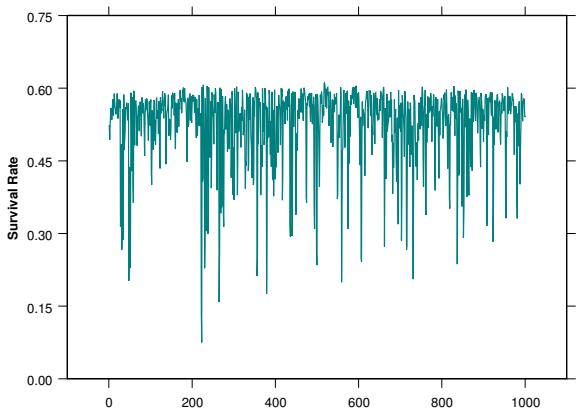

Figure 6.6: Survival rate of the particle set at each time step.

In particular we employ the *Survival Rate*, which evaluates the number of particles survived to the selection step with respect to the total number of particles in the

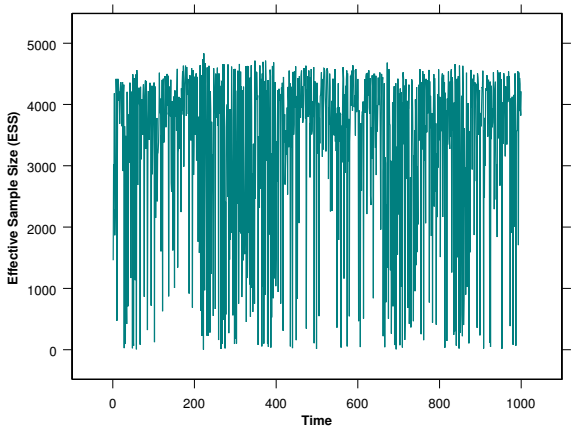

Figure 6.7: Time evolution of the estimated Effective Sample Size.

set. The survival rate reveals particles set degeneracy when exhibiting a persistent high number of dead particles from a generation to the subsequent one. Survival rate has been calculated as follow

$$SR_t = \frac{N - \sum_{i=1}^{N} \mathbb{I}_{\{0\}}(Card(I_{i,t}))}{N} \tag{6.108}$$

where $I_{i,t} = \{j \in \{1, \ldots, N\} | i_t^j = i\}$ is the set of all random index values, which are selecting, at time t, the $i - th$ particle. Note that if at time t the particle k does not survive to the selection step then the set $I_{k,t}$ becomes empty. Fig.6.6 exhibits the evolution over time of the survival rate of a set of $N = 5,000$ particles. Although for some filter iterations the rate falls under the 30% level, it does not remain persistently under that level. We can conclude that the filter does not show degeneracy problems.

In order to complete the degeneracy analysis we evaluate the *Effective Sample Size* indicator. This degeneracy measure has been introduced by Liu and Chen [45] and for the general dynamic system of equations (6.19)-(6.21) is defined as

$$ESS_t \triangleq \frac{N}{1 + Var(w_k^{*i})} \tag{6.109}$$

where the weights $w_k^{*i} = p(x_k^i | y_{1:k}; \theta) / q(x_k^i | x_{k-1}^i; \theta)$ cannot be calculated explicitly.

Thus the following estimator has been used

$$\hat{ESS}_t = \frac{1}{\sum_{i=1}^{N} (\tilde{w}_t^i)^2} \tag{6.110}$$

where the weights $\tilde{w}_t^i$ have been defined in equation (6.27). Observe that this degeneracy measure is less than or equal to N. It is equal to N when the importance function is exactly equal to the filtering density and tends to zero when the variance of the importance weights tends to infinity, this is when particle filter degenerates. Fig. 6.7 shows the estimated ESS relative to the particle filter applied to the gaussian model. Observe that the effective sample size varies over time, but it never stabilizes at zero. Thus we conclude again in favour of a non-degeneracy of our particle filter for the gaussian model.

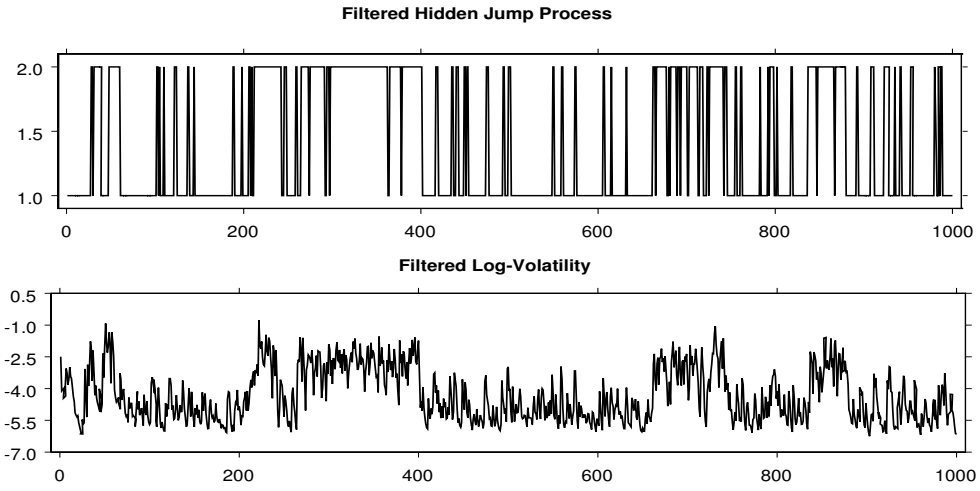

Figure 6.8: Filtered log-volatility and Markov switching states.

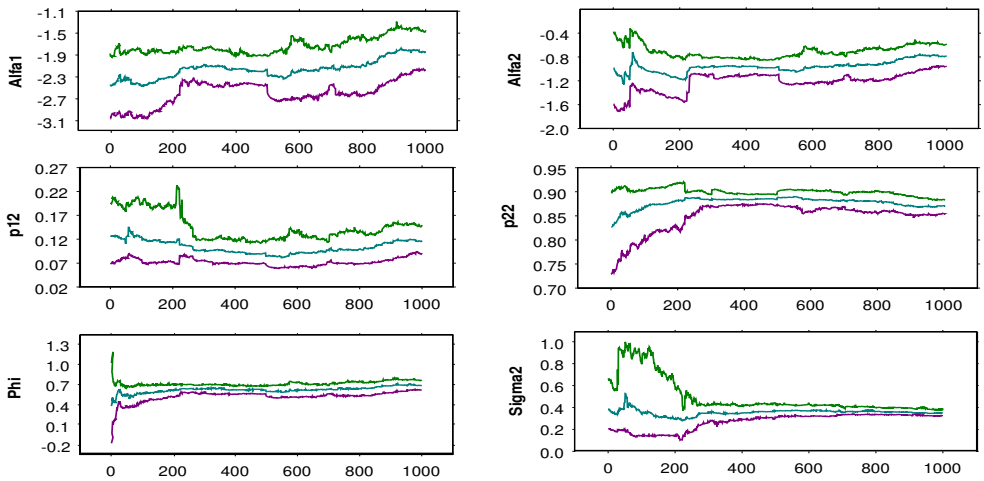

Figure 6.9: Parameter estimates for the Gaussian MSSV model. Graphs exhibit also the posterior mean and quantiles at 0.275 and 0.975 level.

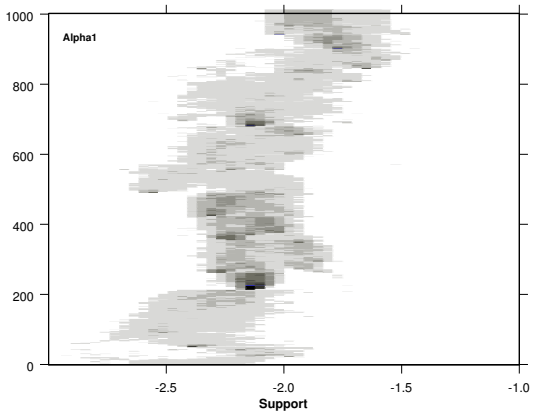

Figure 6.10: Time evolution of the α_1's density function for the Gaussian MSSV model.

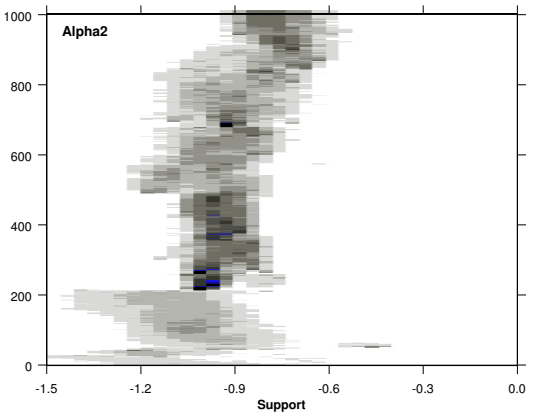

Figure 6.11: Time evolution of the α_2's density function for the Gaussian MSSV model.

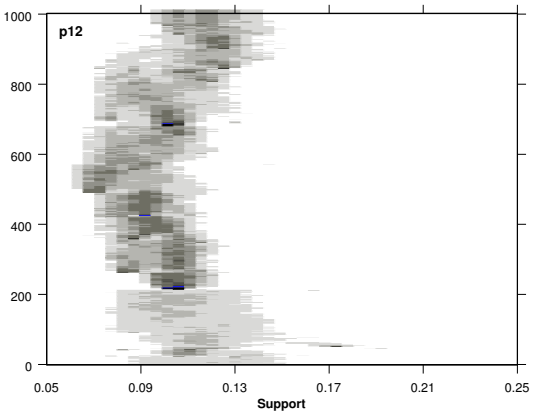

Figure 6.12: Time evolution of the p_{12}'s density function for the Gaussian MSSV model.

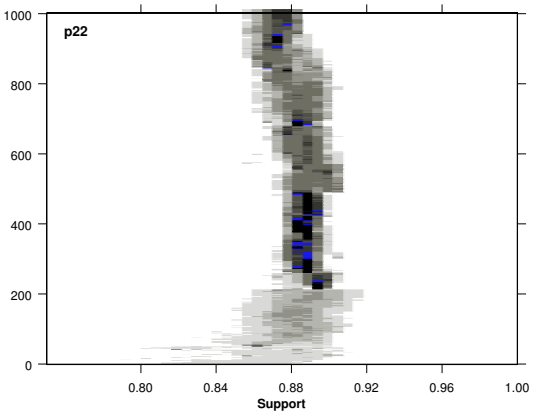

Figure 6.13: Time evolution of the p_{22}'s density function for the Gaussian MSSV model.

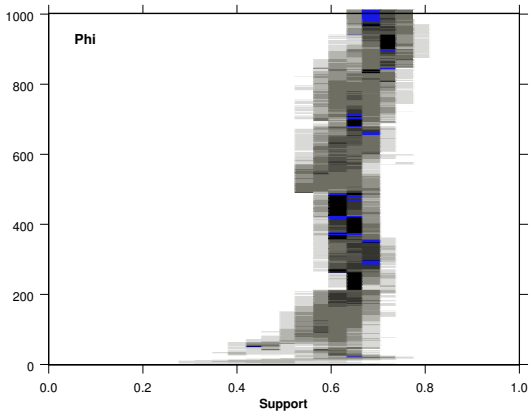

Figure 6.14: Time evolution of the ϕ's density function for the Gaussian MSSV model.

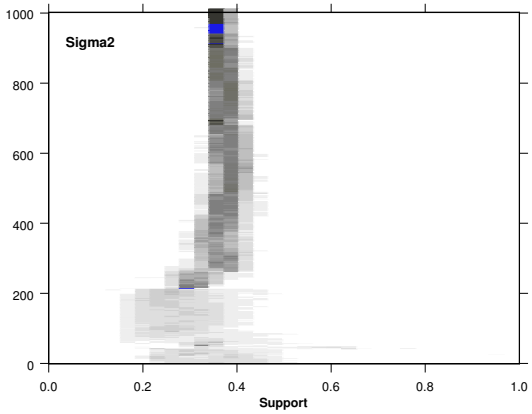

Figure 6.15: Time evolution of the σ^2's density function for the Gaussian MSSV model.

Bibliography

[1] Andersen T.G. (1994), Stochastic Autoregressive Volatility: A Framework for Volatility Modelling, *Mathematical Finance*, 4, pp. 75-102.

[2] Andersen T.G. and Sørensen B. (1996), GMM Estimation of a Stochastic Volatility Models: A Monte Carlo Study, *Journal of Business & Economic Statistics*, 14, pp. 328-352.

[3] Andrieu C. and Doucet A. (2002), Online Expectation-Maximization Type Algorihtms for Parameter Estimation in General State Space Models, *Working paper*, Dep. Of Mathematics, Bristol University.

[4] Arulampalam S., Maskell S., Gordon N. and Clapp T. (2001), A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking, *Technical Report*, QinetiQ Ltd., DSTO, Cambridge.

[5] Baldi P., Mazliak L. and Priouret P. (2000), *Martingales et chaînes de Markov* 2^{nd} *Ed.*, Hermann Éditeurs, Paris.

[6] Barndorff-Nielsen O.E. and Shephard N. (2000), Modelling by Lévy Processes for Financial Econometrics, *Working paper*, n. W03, University of Oxford.

[7] Barndorff-Nielsen O.E. and Shephard N. (1999), Non-Gaussian OU based models and some of their uses in Financial economics, *Working paper*, n. W09, University of Oxford.

[8] Berzuini C., Gilks W.R. (2001), Following a moving average-Monte Carlo inference for dynamic Bayesian models, *J.R. Statist. Soc. B*, vol. 63, pp.127-146.

[9] Berzuini C., Best N.G, Gilks W.R. and Larizza C., (1997), Dynamic conditional independence models and Markov chain Monte Carlo Methods, *Journal of the American Statistical Association*, Vol. 92, pp. 1403-1441.

[10] Buckle D.J. (1995), Bayesian inference for stable distribution, *Journal of American Statistical Association*, Vol. 90, pp. 605-613.

[11] Casella G. and Robert C.P., (1996), Rao-Blackwellization of Sampling Schemes, *Biometrika*, 83, 81-94.

[12] Clark P.K. (1973), A Subordinated Stochastic Process Model With Finite Variances for Speculative Prices, *Econometrica*, 41, pp. 135-156.

[13] Chib S., Nardari F. and Shephard N. (2002), Markov chains Monte Carlo methods for stochastic volatility models, *Journal of Econometrics*, 108(2002), pp. 281-316.

[14] Chopin N. (2001), Sequential inference and state number determination for discrete state-space models through particle filtering, *Working paper*, CREST 2001-34.

[15] Crisan D. (2001), Particle filters - A theoretical perspective, in *Sequential Monte Carlo Methods in Practice*, eds Doucet A., Freitas J.G. and Gordon J., (2001), Springer Verlag, New York.

[16] Crisan D. and Doucet A. (2000), Convergence of sequential Monte Carlo methods, *Technical Report 381*, CUED-F-INFENG.

[17] Crisan D. and Doucet A. (2002), A survey of convergence results on particle filtering for practitioners, *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 736–746.

[18] Danielsson J. (1994), Stochastic volatility in asset prices: estimation with simulated maximum likelihood, *Journal of Econometrics*, 61, pp. 375-400.

[19] Danielsson J. and Richard J. (1993), Accelerated Gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics*, 8, pp. 153-173.

[20] De Jong P. and Shephard N. (1995), The Simulation Smoother for Time Series Models, *Biometrika*, Vol. 82, Issue 2, pp. 339-350.

[21] Doucet A., Godsill S. and Andrieu C. (2000), On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*,Vol. 10, pp. 197-208.

[22] Doucet A., Freitas J.G. and Gordon J., *Sequential Monte Carlo Methods in Practice*, Springer Verlag, New York.

[23] Durbin J. and Koopman S. (1997), Monte Carlo maximum likelihood estimation for non Gaussian state space models, *Biometrika*, 84, pp.669-684.

[24] Eraker B. (2001), MCMC Analysis of Diffusion Models with application to Finance, *Journal of Business & Economic Statisics*, April 2001, Vol. 19, No.2, pp. 177-191.

[25] Fridman M. and Harris L. (1998), A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models, *Journal of Business & Economic Statistics*, Vol. 16, n. 3, pp. 284-291.

[26] Francq C., Roussignol M. (1998), Ergodicity of autoregressive process with Markov-switching and consistency of the maximum-likelihood estimator, *Statistics*, 32, pp. 151-173.

[27] Francq C., Zakoian J.M. (2001), Stationarity of multivariate Markov-switching ARMA models, *Journal of Econometrics*, 102, pp. 339-364.

[28] Gallant A.R. and Tauchen G. (1996), Which Moments to Match, *Econometric Theory*, 12, pp. 657-681.

[29] Gallant A.R., Hsieh D. and Tauchen G. (1997), Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics*, 81, pp.159-192.

[30] Geyer C. (1994), On the convergence of the Monte Carlo maximum likelihood calculations, *Journal of the Royal Statistical Society*, B 65, pp. 261-274.

[31] Geyer C. (1996), Estimation and Optimization of functions, in *Markov Chain Monte Carlo in Practice*, Gilks W., Richardson S. and Spiegelhalter D. eds., Chapman and Hall, London, pp. 241-258.

[32] Godsill S. (1999), MCMC and EM-based methods for inference in heavy-tailed processes with α-stable innovations, *Working paper*, Signal Processing Group, Dep. of Engineering, University of Cambridge.

[33] Gordon N., Salmond D. and Smith A.F.M. (1993), Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation, *IEE Proceedings-F*, 1993, Vol. 140, pp. 107-113.

[34] Gouriéroux C., Monfort A. and Renault E. (1993), Indirect Inference, *Journal of Applied Econometrics*, 8, pp. 85-118.

[35] Hamilton J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57/2, pp. 357-384.

[36] Harrison J. and West M., (1997), *Bayesian Forecasting and Dynamic Models* 2nd Ed., Springer Verlag, New York.

[37] Harvey A. and Shephard N. (1996), Estimation of an asymmetric model of asset prices, *Journal of Business & Economic Statistics*, 14/4, pp. 429-434.

[38] Harvey A., Ruiz E. and Shephard N. (1994), Multivariate Stochastic Variance Models, *The Review of Economic Studies*, Vol. 61, Issue 2, pp. 247-264.

[39] Hull J. and White A. (1987),The pricing of options on assets with stochastic volatilities, *Journal of Finance*, 42, pp. 281-300.

[40] Jacquier E., Polson N. G. and Rossi P. E. (1994), Bayesian Analysis of Stochastic Volatility Models, *Journal of Business & Economic Statistics*, Vol. 12, n.4, pp. 69-87.

[41] Johannes M.S. and Polson N. (2002), MCMC methods for Financial Econometrics, in *Handbook of Financial Econometrics, forthcomming*, May 2002, http://www-1.gsb.columbia.edu/faculty/mjohannes/.

[42] Johannes M.S., Polson N. and Sroud J. (2002), Nonlinear Filtering of Stochastic Differential Equations with Jumps, March 2002, http://www-1.gsb.columbia.edu/faculty/mjohannes/.

[43] Kitagawa G. (1998), A self-organizing state space model, *Journal of Am. Statist. Ass.*, 93, pp. 1203-1215.

[44] Kim S., Shephard N. and Chib S. (1998), Stochastic volatility: likelihood inference and comparison with arch models, *Review of Economic Studies*, 65, pp. 361-393.

[45] Liu J.S. and Chen R. (1998), Sequential Monte Carlo Methods for Dynamical System., *Journal of the American Statistical Association*, 93, pp. 1032-1044.

[46] Liu J. and West M. (2001), Combined Parameter and State Estimation in Simulation-Based Filtering, in *Sequential Monte Carlo Methods in Practice*, eds Doucet A., Freitas J.G. and Gordon J., (2001), Springer Verlag, New York.

[47] Lopes H.F. (2001), Sequential analysis of stochastic volatility models: some econometric applications, *Working paper*, University of Rio de Janeiro, 2001.

[48] Lopes H.F. and Marigno C. (2001), A particle filter algorithm for the Markov switching stochastic volatility model, *Working paper*, University of Rio de Janeiro 2001.

[49] Melino A. and Turnbull S.M. (1990), Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, 45, pp. 239-265.

[50] Musso C, Oudjane N. and LeGland F. (2001), Improving Regularised Particle Filters, in *Sequential Monte Carlo in Practice, eds Doucet A., Freitas J.G. and Gordon J.*, (2001), Springer Verlag, New York.

[51] Nelson D.B. (1988), *Time Series Behavior of Stock Market Volatility and Returns*, unpublished PhD. dissertation, Massachusetts Institute of Technology, Economics Dept.

[52] Pitt M. and Shephard N. (1999), Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, Vol. 94(446), pp. 590-599.

[53] Polson N.G., Stroud J.R. and Müller P. (2003), Practical Filtering with sequential parameter learning, *Technical Report*,Graduate School of Business, University of Chicago.

[54] Polson N.G., Stroud J.R. and Müller P. (2003), Practical Filtering for Stochastic Volatility Models, *Working paper*, 2003.

[55] Samorodnitsky G. and Taqqu M.S. (1994), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinte Variance*, Chambridge University Press.

[56] Shephard N. (1993), Fitting nonlinear time series models, with applications to stochastic variance models, *Journal of Applied Econometrics*,8,pp.135-152.

[57] Shephard N. and Pitt M. K. (1997), Likelihood Analysis of Non-Gaussian Measurement Time Series, *Biometrika*, Vol. 84, Issue 3, pp. 653-667.

[58] So M.K.P., Lam K. and Li W.K. (1994), Multivariate Modelling of the Autoregressive Random Variance Process, *Journal of Time Series Analysis*, 18, pp. 429-446.

[59] So M.K.P., Lam K. and Li W.K. (1998), A stochastic volatility model with Markow switching, *Journal of Business & Economic Statistics*, 16, pp. 244-53.

[60] Storvik G. (2002), Particle filters for state space models with the presence of unknown static parameters, *IEEE Trans. on Signal Processing*, 50, pp. 281-289.

[61] Taylor S.J. (1982), Financial Returns Modelled by the Product of Two Stochastic Processes, a Study of Daily Sugar Prices 1961-79, in *Time Series Analysis: Theory and Practice 1* ed. O.D. Anderson, Amsterdam: North-Holland, pp. 203-226.

[62] Taylor S.J. (1986), *Modelling Financial Time Series*, New York, Wiley.

[63] Taylor S.J. (1994), Modelling Stochastic Volatility, *Mathematical Finance*, 4, pp.183-204.

[64] West M. (1992), Mixture models, Monte Carlo, Bayesian updating and dynamic models, *Computer Science and Statistics*, 24, pp. 325-333.

[65] West M. (1993), Approximating posterior distribution by mixtures, *Journal of Royal Statistical Society*, B, 55, pp. 409-442.

[66] Wiggins J. (1987), Options Values Under Stochastic Volatility: Theory and Empirical Evidence, *Journal of Financial Economics*, 19, pp. 351-372.